

Price Dynamics in Partially Segmented Markets *

Robin Greenwood
Harvard University and NBER

Samuel G. Hanson
Harvard University and NBER

Gordon Y. Liao
Harvard University

First draft: March 2015

Abstract

We develop a dynamic model of financial markets in which capital moves quickly between securities within a given asset class, but moves slowly across the markets for different asset classes. In our model, most investors specialize in a single asset class such as government bonds, corporate bonds, or equities. However, a smaller number of generalist investors have the ability to flexibly allocate capital across markets, albeit only gradually. As a result, short-run demand curves for each individual asset class are steeply downward-sloping and closely-related asset classes can be highly segmented in the short term. Over the long-run, capital flows more freely across the boundaries of asset classes, increasing cross-market integration and muting the price impact of supply shocks. Using this framework, we show how supply shocks in one asset class are transmitted over time to other asset classes and how specialist and generalist investors adjust their portfolios in response. While prices of a given asset class initially overreact to a supply shock in that market, under plausible conditions, prices underreact in related markets. We explore several applications, including the design and impact of central bank asset purchase programs, and the role of corporate issuance in promoting market integration.

*We thank Daniel Bergstresser, Jeremy Stein, Adi Sunderam, and seminar participants at Brandeis and the University of Texas for useful feedback. We thank David Biery for research assistance.

1 Introduction

How do large supply shocks in one asset market affect the pricing of securities in other related markets? For example, suppose that the Federal Reserve announces that it plans to quickly sell \$1 trillion of long-term U.S. Treasury bonds. How would such an announcement impact the yields on long-term U.S. Treasury bonds? Assuming that the shock leads yields on Treasury bonds to rise, how should we expect the yields on corporate bonds and mortgage-backed securities, which are also exposed to interest rate risk, to react? And, how quickly should this shock be reflected in these related asset markets?

If the markets for different asset classes are tightly integrated, then a shock that affects the pricing of a risk factor in one asset class will have a similar effect on other asset classes exposed to the same risk. When markets are more segmented, however, prices of risk in one market may be disconnected from those in other markets. Segmentation arises because institutional, informational, and behavioral frictions lead investors to specialize in a particular asset class or a narrow set of securities (Merton [1987], Grossman and Miller [1988], Shleifer and Vishny [1997]). Although specialization can facilitate arbitrage across securities within an asset class, it can impede arbitrage across asset classes. Following a large supply shock, specialists' limited willingness to trade across markets may lead the pricing of risk in one market to become temporarily disconnected from that in other related markets.

However, in real world financial markets, the degree of segmentation depends on time horizon. Over the long run the forces of arbitrage ensure that capital will flow from underpriced markets to overpriced markets. Nonetheless, the process of market integration can be slow, because investors with the flexibility to trade across asset classes do not do so immediately. For example, investment committees at pension funds and endowments—who have the flexibility to allocate capital across markets—typically reallocate capital annually. Duffie and Strulovici (2012) explain how differences in risk prices across markets provide incentives for capital reallocation and, thus, affect the speed of cross-market capital movement.

In this paper, we develop a dynamic model of financial markets in which capital moves quickly between securities within a given asset class, but more slowly between markets for different asset classes. Our main objective is to understand how supply and demand shocks in one market are reflected in prices and investor behavior in neighboring markets. We develop our model using a stylized depiction of risky fixed-income assets that trade in partially segmented markets. As we explain below, fixed income markets are a natural setting for understanding segmentation, because bond yields naturally encode information about future risk premia. And, from a practical perspective, the fault lines between different market segments within fixed-income markets are particularly stark when compared to, say, equity markets.

We consider a setting in which two long-term risky assets trade in partially segmented markets. Both assets are exposed to a common fundamental risk factor, making them partial substitutes. This means that, absent frictions, their prices would be tightly linked by cross-market arbitrage. To introduce market segmentation, we assume that there are two sets of risk-averse market specialists,

each of whom can flexibly trade one of the risky assets as well as a short-term risk-free asset. Specialists are unable to allocate capital across the two markets. However, markets are partially integrated by risk-averse generalist investors or “asset allocators,” who periodically reevaluate their portfolios and shift between the two risky assets. This setup is similar to Gromb and Vayanos (2002), except that the cross-market arbitrageurs are slow-moving as in Duffie (2010). Because of the gradual nature of cross-market arbitrage, markets are more integrated in the long run than the short run.

What happens in this setting when there is an unanticipated supply shock in one market? For concreteness, we will often think of the first asset as being long-term U.S. Treasuries and the second asset as long-term corporate bonds, both of which are heavily exposed to interest rate risk. So, to make the thought experiment concrete, suppose that the Federal Reserve announces that it will quickly sell \$1 trillion of long-term U.S. Treasury bonds, permanently expanding the amount of interest rate risk that investors need to bear in equilibrium.

How will the Treasury market react to this sudden expansion in supply? The risk premium on long-term Treasury bonds will rise, lifting their yields. Treasury market specialists will react immediately to the shock, absorbing the increased supply into their inventories. However, Treasury yields will overreact—the short-run price impact will exceed the long-run impact—because the amount of capital that can initially accommodate the shock is limited to specialists and a handful of generalists. Over the long-run, additional generalist investors will allocate more capital to the Treasury market, muting the price impact of the supply shock at longer horizons. Price dynamics of this sort are similar to those described in Duffie (2010).

Our more novel results pertain to the behavior of yields in the second related market, in this case, long-term corporate bonds. Although the Federal Reserve is not directly selling corporate bonds in our thought experiment, the corporate bond market will be indirectly affected as generalist investors increase their holdings of long-term Treasuries and reduce their holdings of long-term corporate bonds. These cross-market capital flows drive down the prices of corporate bonds and push up corporate bond yields. In this way, the trading of generalist asset allocators transmits supply shocks across markets, serving to increase market integration. While yields in the Treasury market initially overreact to this supply shock, under plausible parameter values, we show that corporate bond yields will underreact to the shock: the short-run price impact is less than the long-run impact. The overreaction of Treasury yields and the underreaction of corporate yields are driven by the fact that generalists only reallocate their portfolios slowly. As a result, it takes time for financial markets to fully digest the supply shock.

If there were no specialists and all investors were generalists, then the two markets would be fully integrated in the sense that exposures to common risk factors would have the same prices in the two markets. However, when markets are partially segmented, risk prices can differ across markets. This occurs because risks cannot be easily unbundled from assets and because markets receive periodic supply shocks, making cross-market arbitrage risky for generalists, as in Gromb and Vayanos (2002). For example, interest rate risk may not be priced identically in the corporate bond market as in the Treasury market. Following a supply shock, the premia associated with similar

risk exposures can differ significantly between the two asset markets. As asset allocators react to these pricing discrepancies, the difference in risk premia between the two markets will gradually narrow. However, the difference will not vanish in the long run because of the permanent risks associated with cross-market arbitrage.

The price dynamics in our model depend critically on the fractions of specialists in each market, the number of time periods it takes generalists to fully rebalance their portfolios, and the degree of substitutability between the two asset markets. The fraction of specialists and generalist investors play an especially important role. When there are no generalists, the markets are completely segmented and the supply shock is not transmitted across markets. When there are a small number of slow-moving generalists, the Treasury market overreacts while the corporate market under-reacts to the shock to Treasury supply. However, if there are many slow-moving generalists, markets are well-integrated and supply shocks can result in short-run overreaction in both markets.

We also use our model to explore the impact of anticipated supply shocks. For example, suppose that, instead of announcing that it will immediately sell \$1 trillion of long-term bonds, the Fed announces that these sales will take place in two years. How will the Treasury and corporate bond markets react to this announcement? Although yields in both markets react immediately to this announcement, both markets exhibit significant underreaction: the short-run price impact is significantly less than the long-run impact. Immediately upon the announcement, generalists begin to gradually adjust in the direction of the anticipated shock, buying Treasuries and selling corporate bonds. Treasury specialists provide temporary liquidity to these generalists by selling their Treasury holdings, planning to replenish their inventories once the supply shock actually lands in two years. The result is a protracted adjustment process starting from the announcement and continuing until well after the supply shock actually arrives. Longer delays between the announcement and the arrival of the supply shock result in a more gradual adjustment process.

In describing our model thus far, we have not distinguished between a risky “asset” and the “market” in which it trades. This distinction arises when we introduce multiple risky assets into each market that differ in their degree of exposure to common risk factors. For example, the “market” for U.S. Treasury securities contains bonds of many different maturities, which have different exposures to interest rate risk. Extending our framework to allow for multiple securities per asset market, we show that a particular conditional CAPM prices all assets in the first market and that a different conditional CAPM prices all assets in the second market. Critically, these two market-specific pricing models are linked over time by the cross-market arbitrage activities of slow-moving asset allocators. For example, the pricing of interest rate risk for 5-year Treasuries is always perfectly consistent with the pricing of interest rate risk for 10-year Treasuries. However, the pricing of interest rate risk in the overall Treasury market may differ somewhat from that in the corporate bond market. And these cross-market differences will be most pronounced following the arrival of major shocks that take time for slow-moving generalists to digest.

The question of how asset prices adjust across partially segmented markets is of enormous practical importance. Consider the recent large-scale purchases of long-term government bonds—i.e., quantitative easing policies—by central banks in the United States, United Kingdom, Japan, and

Europe. A key question about these policies is whether they impact the prices of financial assets outside of the market for government bonds. The favored methodology for answering this question has been to use event studies of short-run price changes immediately surrounding central bank policy announcements. A number of these studies have concluded that the effects of quantitative easing are most pronounced in the market in which the central bank is transacting, with only modest spillovers to other related markets (Woodford [2012] and Krishnamurthy and Vissing-Jorgensen [2013]). Others have suggested that at longer horizons, the spillovers are more significant. Maimsky (2014) suggests that if one expands the measurement window by a few days or weeks, however, the effects in other markets may be much larger. More generally, our model suggests that the short-run price impact of an announced supply change on different markets may not accurately reveal the long-run impact, which is often of greater interest to policymakers. We illustrate this idea more formally by analyzing the statistical power of short-run event studies within our model. We show that the horizon at which statistical power is maximized is often much shorter than the horizon at which the long-run price impact is achieved.

We also argue that our model is a useful representation of corporate arbitrage. It is increasingly recognized that corporations behave like arbitrageurs in their financing activities, issuing securities in markets in which prices are temporarily high (Baker and Wurgler [2000] and [2012]). More recent work has suggested that corporations may help integrate different asset markets through their financing activities (Greenwood, Hanson, and Stein [2010] and particularly Ma [2015]). Although research has emphasized the many advantages that corporations may have compared to professional arbitrageurs, one of the major disadvantages is that they tend to move quite slowly, much like the generalists in our model.

Our findings are closely tied to two strands of research in finance and economics. The idea that front-line arbitrageurs in financial markets are highly specialized traces back to Merton (1987) and Grossman and Miller (1988), and is a central tenet of the theory underlying limits-of-arbitrage (De Long et al [1990], Shleifer and Vishny [1997], and Gromb and Vayanos [2002]). More recently, a number of researchers have convincingly demonstrated downward-sloping demand curves for individual financial asset classes, which would be puzzling if markets were fully integrated (Gabaix, Krishnamurthy, and Vigneron [2007], Gârleanu, Pedersen, and Poteshman [2009], Greenwood and Vayanos [2014], and Hanson [2014]).

Second, our paper is related to research on “slow-moving capital.” Here, our model is most closely related to and builds on Duffie (2010). Duffie and Strulovici (2012) also present a model of the movement of capital across two partially segmented markets, but their focus is on the endogenous speed of capital mobility, which we take as exogenous. Our contribution here is to explicitly characterize the dynamics of prices across related asset markets and to describe the patterns of cross-market arbitrage in response to large supply or demand shocks.

2 Model

We develop the model in two steps. We first develop a benchmark model for pricing a long-term fixed-income asset that is exposed to both interest rate risk and default risk. Absent default risk, this model is a simplified, discrete-time version of the term structure models developed in Vayanos and Vila (2009) and Greenwood and Vayanos (2014) in which interest rate risk is priced by a set of specialized, risk-averse bond arbitrageurs.

In the second step, we introduce two fixed-income assets that are exposed to different risks and, therefore, are partial substitutes. Specifically, one asset is default-free while the other is exposed to default risk. We consider a market structure in which there are two separate groups of specialist arbitrageurs who trade these two assets. Building on Duffie (2010), we introduce a set of generalist investors who allocate capital between the two markets, but who do so slowly. These cross-market capital flows transmit supply shocks between markets and serve to integrate the markets over time. Despite the existence of these generalist investors, the two markets are not perfectly integrated even in the long run because cross-market arbitrage is risky. Using this model, we show how bond risk premia, bond yields, and the bond holdings of different investors adjust following shocks to asset supply.

We develop our model using fixed-income securities for two main reasons. First, fixed-income markets are a natural place to study segmentation because bond yields naturally encode information about future risk premia. For example, yield spreads between similar assets are an intuitive way to measure the degree of market integration. Second, in practice fixed-income markets can be highly specialized, despite the fact that fixed-income assets often share similar risk exposures. For example, the investors who specialize in high-grade corporate bonds are often distinct from those who specialize in trading U.S. Treasuries. Nevertheless, while we write the model in this way, none of our results are specific to fixed-income assets and we will consider applications in which the segmentation is between the fixed-income market more broadly and equity markets.

2.1 Single asset model

2.1.1 Defaultable perpetuities

Consider a perpetual, defaultable bond that promises to pay a coupon of C each period. Let $P_{L,t}$ denote the price of the bond at time t . The gross return on the bond from t to $t + 1$ is

$$1 + R_{L,t+1} = \frac{(1 - Z_{t+1})(P_{L,t+1} + C)}{P_{L,t}}, \quad (1)$$

where $0 \leq Z_{t+1} < 1$ is the default realization at time $t + 1$. If $Z_{t+1} \equiv 0$, the bond is default-free. If Z_{t+1} is stochastic, the bond is defaultable with high realizations of Z_{t+1} corresponding to larger default losses at time $t + 1$. This formulation of default risk follows Duffie and Singleton's (1999) "recovery of market value" assumption which has become standard in the credit risk literature.¹

¹Consider the return on a large portfolio of defaultable bonds. Suppose that fraction h_{t+1} of the bonds default at $t + 1$ and are worth $(1 - L_{t+1})(P_{L,t+1} + C)$ and that fraction $(1 - h_{t+1})$ of the bonds do not default and are worth

Using the Campbell-Shiller (1988) log-linear approximation and defining $\theta \equiv 1/(1+C) < 1$, the one-period log return on the defaultable perpetuity is

$$r_{L,t+1} \equiv \ln(1 + R_{t+1}^L) \approx \underbrace{\frac{1}{1-\theta}}_D y_{L,t} - \underbrace{\frac{\theta}{1-\theta}}_{D-1} y_{L,t+1} - z_{t+1}, \quad (2)$$

where $y_{L,t}$ is the log yield-to-maturity at time t and

$$D = \frac{1}{1-\theta} = \frac{C+1}{C} \quad (3)$$

is the Macaulay duration when the perpetuity is trading at par.²

2.1.2 Risk factors

Investors in defaultable long-term bonds are exposed to three different types of risk: interest rate risk, default risk, and supply risk. First, investors are exposed to *interest rate risk*. In our model, investors face an exogenous short-term interest rate that evolves randomly over time and will suffer a capital loss on the bond holdings if short-term rates rise unexpectedly. Second, investors face *default risk*: the future period-by-period default realization is unknown and evolves randomly over time. Finally, investors are exposed to *supply risk*: there are random supply shocks which impact the prices and yields on long-term bonds, holding fixed the expected future path of short-term interest rates and expected future defaults. Thus, using Campbell's (1991) terminology, interest rate risk and default risk are forms of fundamental "cash flow" risk, whereas supply risk is a form of "discount rate" risk.

We make the following concrete assumptions:

- **Short-term interest rates:** The log short-term riskless rate available to investors between time t and $t+1$, denoted r_t , is known at time t . We assume that r_t also follows an exogenous AR(1) process

$$r_{t+1} = \bar{r} + \rho_r (r_t - \bar{r}) + \varepsilon_{r,t+1}, \quad (4)$$

$(P_{L,t+1} + C)$. Then return on the credit portfolio is

$$1 + R_{L,t+1} = \frac{h_{t+1}(1 - L_{t+1})(P_{L,t+1} + C) + (1 - h_{t+1})(P_{L,t+1} + C)}{P_{L,t}} = \frac{(1 - h_{t+1}L_{t+1})(P_{L,t+1} + C)}{P_{L,t}}.$$

This maps into our framework with $Z_{t+1} = h_{t+1}L_{t+1}$ where h_{t+1} is the fraction of the surviving portfolio that defaults at $t+1$ and L_{t+1} is the loss-given-default as a fraction of market value.

²The Campbell-Shiller (1988) approximation of the log return is $r_{L,t+1} \approx \kappa + \theta p_{L,t+1} + (1-\theta)c - p_{L,t} - z_{t+1}$ where $\theta = 1/(1 + \exp(c - \bar{p}_L))$ and $\kappa = -\log(\theta) - (1-\theta)\log(\theta^{-1} - 1)$. Iterating forward, implies $p_{L,t} = (1-\theta)^{-1}\kappa + c - \sum_{i=0}^{\infty} \theta^i E_t[r_{L,t+i+1} + z_{t+i+1}]$. Applying this approximation to promised cashflows ($z_{t+i+1} \equiv 0$) and the yield-to-maturity, defined as the *constant return* that equates price and the discounted value of *promised* cashflows, we obtain $p_{L,t} = (1-\theta)^{-1}\kappa + c - (1-\theta)^{-1}y_{L,t}$. Equation (2) follows by substituting this expression for $p_{L,t}$ into the Campbell-Shiller return approximation. Assuming the steady-state price is par ($\bar{p}_L = 0$), we have $\theta = 1/(1+C)$. Thus, bond duration is $-\partial p_{L,t}/\partial y_{L,t} = (1-\theta)^{-1} = (1+C)/C$. Since $P_{L,t} = C/Y_{L,t}$ and $-\partial p_{L,t}/\partial y_{L,t} = -(\partial P_{L,t}/\partial Y_{L,t})((1+Y_{L,t})/P_{L,t}) = (Y_{L,t} + 1)/Y_{L,t}$ this corresponds to the Macaulay duration when the perpetuity is trading at par ($Y_{L,t} = C$).

where $Var_t[\varepsilon_{r,t+1}] = \sigma_r^2$. One can think of the short-term rate as being determined outside the model either by monetary policy or by a stochastic short-term storage technology that is available in perfectly elastic supply.

- **Default losses:** We assume that the log default process $z_t = -\ln(1 - Z_t)$ follows

$$z_{t+1} = \bar{z} + \rho_z(z_t - \bar{z}) + \varepsilon_{z,t+1} \quad (5)$$

where $Var_t[\varepsilon_{z,t+1}] = \sigma_z^2$.

- **Supply:** We assume that the perpetuity is available in an exogenous, time-varying supply s_t . For tractability, we assume that supply follows an AR(1) process

$$s_{t+1} = \bar{s} + \rho_s(s_t - \bar{s}) + \varepsilon_{s,t+1}, \quad (6)$$

where $Var_t[\varepsilon_{s,t+1}] = \sigma_s^2$.

For simplicity, we will assume that $\varepsilon_{s,t+1}$, $\varepsilon_{r,t+1}$, and $\varepsilon_{z,t+1}$ are mutually orthogonal. However, it is straightforward to relax this assumption.

2.1.3 Specialist demand and market clearing

There is a unit mass of specialized bond arbitrageurs, each with risk tolerance τ . Specialist arbitrageurs can earn an uncertain future return of $r_{L,t+1}$ from t to $t+1$ by investing in the defaultable long-term bond. Alternatively, they can earn a certain return of r_t by investing at the short-term interest rate. Specialist arbitrageurs are concerned with their interim wealth.

Formally, we assume that at date t specialist arbitrageurs have mean-variance preferences over their wealth at $t+1$. This means that arbitrageurs choose their holdings of the perpetuity to solve

$$\max_{b_t} \left\{ b_t E_t[r_{L,t+1}] - (2\tau)^{-1} (b_t)^2 Var[r_{L,t+1}] \right\}, \quad (7)$$

where $rx_{L,t+1} \equiv r_{L,t+1} - r_t$ is the log excess returns on the defaultable long-term bond over the short-term interest rate between t and $t+1$. Thus, arbitrageur demand for the risky bond is

$$b_t^* = \tau \frac{E_t[rx_{L,t+1}]}{Var_t[rx_{L,t+1}]} \quad (8)$$

Equation (8) simply says that arbitrageurs borrow at the short-term rate and invest in risky long-term bonds when the expected return on perpetuities exceeds that the short rate ($E_t[rx_{L,t+1}] > 0$). Conversely, arbitrageurs sell short bonds and invest at the short rate when $E_t[rx_{L,t+1}] < 0$. And they respond more aggressively to these movements in risk premia when they are more risk tolerant and when the variance of excess bond returns is low.

Market clearing ($b_t^* = s_t$) implies that the bond risk premium, $E_t [rx_{L,t+1}]$ is given by

$$E_t [rx_{L,t+1}] = \tau^{-1} V_L^{(1)*} s_t, \quad (9)$$

where $V_L^{(1)*} = \text{Var}_t [(D-1)y_{L,t+1} + z_{t+1}]$ is the equilibrium variance of 1-period excess bond returns.

Thus, bond risk-premia are increasing in bond supply, s_t . When a positive supply shock hits the market, bond risk premia jump instantaneously. If the shock is almost permanent ($\rho_s \approx 1$), the impact on the risk premium will be long lived. If the shock is highly transient ($\rho_s \approx 0$), bond supply will rapidly revert to steady-state ($\bar{s}$) and risk premia will quickly to their steady-state level, $\tau^{-1} V_L^{(1)*} \bar{s}$.

2.1.4 Solution and equilibrium yields

To solve the model, we conjecture that equilibrium bond yields take the form

$$y_{L,t} = \alpha_0 + \alpha_r (r_t - \bar{r}) + \alpha_z (z_t - \bar{z}) + \alpha_s (s_t - \bar{s}). \quad (10)$$

Using this conjecture, in the Internet Appendix we show that the equilibrium variance of 1-period excess bond returns, $V_L^{(1)*}$, must satisfy the following quadratic equation

$$V_L^{(1)*} = \left(\frac{\theta}{1 - \rho_r \theta} \sigma_r \right)^2 + \left(\frac{1}{1 - \rho_z \theta} \sigma_z \right)^2 + \left(\tau^{-1} \frac{\theta}{1 - \rho_s \theta} \sigma_s \right)^2 (V_L^{(1)*})^2. \quad (11)$$

The total risk premium can be decomposed into compensation for bearing interest rate risk, compensation for bearing credit risk, and compensation for bearing supply risk:

$$\begin{aligned} E_t [rx_{L,t+1}] &= \overbrace{\tau^{-1} \left(\frac{\theta}{1 - \rho_r \theta} \sigma_r \right)^2 s_t}^{\text{Interest rate risk premium}} + \overbrace{\tau^{-1} \left(\frac{1}{1 - \rho_z \theta} \sigma_z \right)^2 s_t}^{\text{Credit risk premium}} \\ &\quad + \overbrace{\tau^{-1} \left(\tau^{-1} \frac{\theta}{1 - \rho_s \theta} \sigma_s \right)^2 (V_L^{(1)*})^2 s_t}^{\text{Supply risk premium}}. \end{aligned} \quad (12)$$

As in Greenwood and Vayanos (2014), when there is supply risk ($\sigma_s^2 > 0$) an equilibrium only exists if bond arbitrageurs are sufficiently risk tolerant.³ If arbitrageurs are sufficiently risk tolerant, there are two possible solutions to (11): one in which yields are highly sensitive to supply shocks and one in which yields are less sensitive to supply shocks. What is the intuition for the multiplicity of equilibria? If yields are highly sensitive to supply shocks, 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. Conversely, if yields

³If τ is too small and there are supply shocks ($\sigma_s^2 > 0$), no equilibrium exists because bonds become extremely risky for arbitrageurs and it is impossible to clear the market. See the Internet Appendix.

are less sensitive to supply shocks, then bonds become less risky for arbitrageurs and arbitrageurs are willingly absorb supply shocks even if they are only compensated by modest changes in yields.

Following Greenwood and Vayanos (2014), we focus on the well-behaved and economically relevant equilibrium in which yields are less sensitive to supply shocks, which corresponds to the smaller root of (11).⁴ It is then straightforward to show that $V_L^{(1)*}$ is increasing in σ_r^2 , σ_z^2 , σ_s^2 , ρ_r , ρ_z , ρ_s , and $D [= (1 - \theta)^{-1}]$ and decreasing in τ . Thus, for a given level of bond supply, the total risk premium is larger when short-term rates are more volatile, when there is greater uncertainty about future defaults, and when supply shocks are more volatile. Furthermore, the risk premium is larger when each of these three processes is more persistent. Finally, the risk premium is increasing in the duration of the perpetuity and is decreasing in arbitrageur risk tolerance.

Rewriting equation (2) as $y_{L,t} = E_t[(1 - \theta)(r_t + rx_{L,t+1} + z_{t+1}) + \theta y_{L,t+1}]$ and iterating forward, we see that the equilibrium yield on the defaultable perpetuity is a weighted average of expected future short rates, future default losses, and future future risk premia

$$y_{L,t} = (1 - \theta) \sum_{i=0}^{\infty} \theta^i E_t \left[\overbrace{r_{t+i}}^{\text{Short rate}} + \overbrace{z_{t+i+1}}^{\text{Default loss}} + \overbrace{\tau^{-1} V_L^{(1)*} s_{t+i}}^{\text{Risk premium}} \right]. \quad (13)$$

Because of the coupon-bearing nature of the long-term bond, equation (13) shows that expected short rates, default losses, and risk premia in the near future have a larger effect on bond yields than those in the distant future.⁵ Making use of the assumed AR(1) dynamics for r_t , z_t , and s_t , we can express the equilibrium yield as

$$y_{L,t} = \overbrace{\left[\bar{r} + \frac{1 - \theta}{1 - \rho_r \theta} (r_t - \bar{r}) \right]}^{\text{Expected future short rates}} + \overbrace{\left[\bar{z} + \frac{1 - \theta}{1 - \rho_z \theta} \rho_z (z_t - \bar{z}) \right]}^{\text{Expected future default losses}} + \overbrace{\left[\tau^{-1} V_L^{(1)*} \bar{s} + \tau^{-1} V_L^{(1)*} \frac{1 - \theta}{1 - \rho_s \theta} (s_t - \bar{s}) \right]}^{\text{Risk premium}}. \quad (14)$$

Equation (14) shows that the perpetuity yield is more sensitive to movements in short rates when the short-rate process is more persistent and when bond duration is shorter (i.e., $\partial^2 y_{L,t} / \partial r_t \partial \rho_r > 0$ and $\partial^2 y_{L,t} / \partial r_t \partial D < 0$). Similarly, the yield is more sensitive to movements in current default losses (z_t) when the default process is more persistent and when bond duration is shorter. Yields are more sensitive to bond supply when short-rates are more volatile or more persistent or when defaults are more volatile or more persistent. Finally, yields are also more sensitive to supply shocks when risk tolerance is low, supply shocks are more volatile, or supply shocks are more persistent.⁶

⁴ As $\sigma_s^2 \rightarrow 0$, this smaller root converges the solution for $V_L^{(1)*}$ when $\sigma_s^2 = 0$ (i.e., to $((\theta\sigma)_r / (1 - \rho_r\theta))^2 + (\sigma_z / (1 - \rho_z\theta))^2$) whereas the larger root diverges to infinity as $\sigma_s^2 \rightarrow 0$. All of the relevant comparative statics on $V_L^{(1)*}$ have the intuitive signs at the smaller root, but have the opposite signs at the larger root.

⁵ This is similar to Campbell and Shiller's (1988) analysis of stock prices and is unlike the case of zero-coupon bonds.

⁶ The sign of $\partial^2 y_{L,t} / \partial s_t \partial D$ is ambiguous since $\partial V_{rx}^* / \partial D > 0$, but $\partial[(1 - \theta) / (1 - \rho_s\theta)] / \partial D < 0$. This corresponds to the finding in Vayanos and Greenwood (2014) that, depending on the persistence of supply shocks, a current

2.2 Model with two partially segmented markets

2.2.1 Securities markets

There are two perpetual risky securities, A and B . A is default-free and exposed only to interest rate risk. Borrowing notation from above, security A pays a coupon of C_A each period, so the gross return on A is $1 + R_{A,t+1} = (P_{A,t+1} + C_A) / P_{A,t}$. The log excess return on the A perpetuity over the short-term interest rate from time t to $t + 1$ is

$$rx_{A,t+1} = \frac{1}{1 - \theta_A} y_{A,t} - \frac{\theta_A}{1 - \theta_A} y_{A,t+1} - r_t, \quad (15)$$

where $\theta_A = 1 / (1 + C_A)$.

The second security, B , is subject to default risk. Apart from default risk, the two securities would be perfect substitutes. Specifically, the B security carries a *promised* coupon payment of C_B each period. However, the gross return on the B security from time t to $t + 1$ is $1 + R_{B,t+1} = (1 - Z_{t+1})(P_{B,t+1} + C_B) / P_{B,t}$ where $0 \leq Z_{t+1} \leq 1$ is the default realization at time $t+1$. Therefore, the log excess return on B from time t to $t + 1$ is

$$rx_{B,t+1} = \frac{1}{1 - \theta_B} y_{B,t} - \frac{\theta_B}{1 - \theta_B} y_{B,t+1} - z_{t+1} - r_t, \quad (16)$$

where $\theta_B = 1 / (1 + C_B)$. The additional z_{t+1} term in equation (16) reflects the time $t + 1$ default realization that is specific to the B perpetuity. The variance of z_{t+1} determines, in part, the degree of substitutability between assets A and B .

We assume that the processes for the short rate r_t and for default losses z_t are as in equations (4) and (5) above. However, we assume that the two markets are subject to different supply shocks which also limits their substitutability for investors with shorter horizons. The net supply that investors must hold of asset A evolves according to

$$s_{A,t+1} = \bar{s}_A + \rho_{s_A} (s_{A,t} - \bar{s}_A) + \varepsilon_{s_A,t+1}, \quad (17)$$

where $Var_t [\varepsilon_{s_A,t+1}] = \sigma_{s_A}^2$. Similarly, the net supply that investors must hold of asset B evolves as

$$s_{B,t+1} = \bar{s}_B + \rho_{s_B} (s_{B,t} - \bar{s}_B) + \varepsilon_{s_B,t+1}, \quad (18)$$

where $Var_t [\varepsilon_{s_B,t+1}] = \sigma_{s_B}^2$. We continue to assume that $\varepsilon_{r,t+1}$, $\varepsilon_{z,t+1}$, $\varepsilon_{s_A,t+1}$, and $\varepsilon_{s_B,t+1}$ are mutually orthogonal.

increase in bond supply can have a greater impact on the yields of intermediate or long-dated bonds. Specifically, highly persistent supply shocks have the greatest impact on long-dated yields, while transitory supply shocks have the greatest impact on intermediate-dated yields.

2.2.2 Market participants

There are three types of investors, with identical risk tolerance τ . Investors are distinguished by their ability to transact in different markets and by the frequency with which they can rebalance their portfolios. Fast-moving A -specialists are free to adjust their holdings of the A asset and the riskless short-term rate each period; however, A -specialists cannot hold the B asset. A -specialists are present in mass p_A and we denote their demand for A by $b_{A,t}$. Analogously, fast-moving B -specialists can freely adjust their holdings of the B asset and the riskless short-term rate each period, but cannot hold the A asset. B -specialists are present in mass p_B and their demand for asset B is $b_{B,t}$.

The third group of investors is a set of slow-moving generalists who can adjust their holdings of A and B asset, as well as the riskless short-term rate, but can do so only every k periods. Fraction $1/k$ of generalists investors are active each period and can re-allocate their portfolios between the A and B assets. However, they must then maintain this same portfolio allocation for the next k periods. As in Duffie (2010), this is a reduced form way to model the real-world informational and agency frictions that limit the speed of capital flows across markets. Generalists are present in mass $1 - p_A - p_B$.

The market structure we have described here is a natural way to capture the industrial organization of real world capital markets. Due to agency and informational problems, savers are only willing to give delegated managers the discretion to adjust their portfolios quickly if the manager gives them an open-ended claim (e.g., Stein (2005)). As a result, fast-moving investors often have endogenously short horizons. Furthermore, in an effort to reduce discretion, savers may impose numerous governance procedures which limit the ability of generalists to respond rapidly to shocks. Similarly, most institutions, such as endowments, pensions, and insurance companies, that have greater long-run flexibility to re-allocate capital across asset classes are subject to governance mechanisms—themselves a response to informationally and agency frictions—that limit the speed of any such capital movement. In short, we believe that a model with fast-moving specialists and slow-moving generalists is a tractable, reduced-form way to capture real-world arbitrage frictions. We later discuss empirical support for these assumptions.

Fast-moving A -specialists and B -specialists have mean-variance preferences over 1-period portfolio returns. Thus, their demands are given by

$$b_{A,t}^* = \tau \frac{E_t[r x_{A,t+1}]}{Var_t[r x_{A,t+1}]}, \quad (19)$$

and

$$b_{B,t}^* = \tau \frac{E_t[r x_{B,t+1}]}{Var_t[r x_{B,t+1}]}. \quad (20)$$

Since they only rebalance their portfolios every k periods, slow-moving generalist investors have mean-variance preferences over their k -period *cumulative* portfolio excess return. Defining $r x_{A,t \rightarrow t+k} \equiv \sum_{i=1}^k r x_{A,t+i}$ and $r x_{B,t \rightarrow t+k} \equiv \sum_{i=1}^k r x_{B,t+i}$ as the cumulative k -period returns from

t to $t + k$ on A and B , the k -period portfolio excess return of generalists who are active at t is⁷

$$rx_{d_t, t \rightarrow t+k} = d_{A,t} \cdot rx_{A, t \rightarrow t+k} + d_{B,t} \cdot rx_{B, t \rightarrow t+k}. \quad (21)$$

Thus, generalist investors who are active at time t choose their holdings of asset A and B , denoted $d_{A,t}$ and $d_{B,t}$, to solve

$$\max_{d_{A,t}, d_{B,t}} \left\{ E_t[rx_{d_t, t \rightarrow t+k}] - (2\tau)^{-1} (Var_t[rx_{d_t, t \rightarrow t+k}]) \right\}. \quad (22)$$

This implies that

$$\begin{bmatrix} d_{A,t}^* \\ d_{B,t}^* \end{bmatrix} = \frac{\tau}{1 - R_{AB}^{2(k)}} \begin{bmatrix} \frac{E_t[rx_{A, t \rightarrow t+k}]}{Var_t[rx_{A, t \rightarrow t+k}]} - \beta_{B|A}^{(k)} \frac{E_t[rx_{B, t \rightarrow t+k}]}{Var_t[rx_{B, t \rightarrow t+k}]} \\ \frac{E_t[rx_{B, t \rightarrow t+k}]}{Var_t[rx_{B, t \rightarrow t+k}]} - \beta_{A|B}^{(k)} \frac{E_t[rx_{A, t \rightarrow t+k}]}{Var_t[rx_{A, t \rightarrow t+k}]} \end{bmatrix}, \quad (23)$$

where, for example, $\beta_{B|A}^{(k)}$ is the coefficient from a linear regression of $rx_{B, t \rightarrow t+k}$ and $rx_{A, t \rightarrow t+k}$ and $R_{AB}^{2(k)}$ is the goodness of fit from this regression.⁸

Equation (23) says that, all else equal, generalist investors allocate more capital to market A when asset A becomes more attractive from a narrow risk-reward standpoint (i.e., $d_{A,t}^*$ is increasing in $E_t[\sum_{i=1}^k rx_{A, t+i}]/Var_t[rx_{A, t \rightarrow t+k}]$). Further, assuming B and A co-move positively ($\beta_{B|A}^{(k)} > 0$), generalists allocate less capital to market A when asset B becomes more attractive from a risk-reward standpoint (i.e., $d_{A,t}^*$ is decreasing in $E_t[\sum_{i=1}^k rx_{B, t+i}]/Var_t[rx_{B, t \rightarrow t+k}]$). In this way, the response of generalist investors transmits supply shocks in the B market to the A market, promoting cross-market integration over time. Cross-market capital flows become more responsive to differences in risk-reward between markets when the two assets are closer substitutes (i.e., when $R_{AB}^{2(k)}$ is higher). In the limit as the two assets become perfect substitutes, $R_{AB}^{2(k)}$ approaches 1 and generalist investors become extremely aggressive in exploiting any cross-market pricing differences.

As we increase their investment horizon k , generalists worry less about their near-term wealth and, thus, worry less about bearing temporary discount rate risk. Said differently, given that supply shocks and the associated return predictability are mean reverting, $Var_t[rx_{A, t \rightarrow t+k}]/k$ and $Var_t[rx_{B, t \rightarrow t+k}]/k$ will be decreasing in k . Thus, relative to short-horizon investors, long-

⁷Formally, this means we assume that slow-moving generalists re-invest all capital initially allocated to the A market (B market) in the A market (B market) over their k -period investment horizon. Also, note that our implicit log-linearization of the portfolio return omits the second-order Jensen's inequality adjustments familiar from Campbell and Viceira (1999, 2001, and 2006). This omission does not alter the core economic intuition of the model and avoids cluttering our equations. Furthermore, in the case of low-volatility fixed-income instruments, most of these adjustment are quantitatively small.

⁸We obtain similar results if we alter equation (22) to reflect the fact that the cumulative return from rolling over an investment at the short-rate for k periods, $\sum_{i=0}^{k-1} r_{t+i}$, is unknown at time t . As in Campbell and Viceira (2001), this adds an I-CAPM-like hedging motive for holding long-duration assets that have high excess returns when $\sum_{i=0}^{k-1} r_{t+i}$ turns out to be lower than expected. Formally, this means that generalists solve $\max_{d_{A,t}, d_{B,t}} \left\{ E_t[r_{P, t \rightarrow t+k}] - \frac{1}{2\tau} (Var_t[r_{P, t \rightarrow t+k}]) \right\}$ where $r_{P, t \rightarrow t+k} = (\sum_{i=0}^{k-1} r_{t+i}) + d_{A,t} \cdot (\sum_{i=1}^k rx_{A, t+i}) + d_{B,t} \cdot (\sum_{i=1}^k rx_{B, t+i})$. The solution takes the same form as (23), replacing $E_t[\sum_{i=1}^k rx_{A, t+i}]$ with $E_t[\sum_{i=1}^k rx_{A, t+i}] - \tau^{-1} Cov_t[\sum_{i=1}^k rx_{A, t+i}, \sum_{i=0}^{k-1} r_{t+i}]$ and similarly for asset B .

horizon investors care more about permanent fundamental risk (shocks to r_t and z_t) and less about transitory supply risk (discount rate risk) as in Campbell and Vuolteenaho (2004). By contrast, short-horizon investors care equally about cash flow and discount rate risk. Thus, as we vary the mix between long-horizon generalists and short-horizon specialists for a fixed k , or as we increase k for a given mix of specialist and generalist investors, we alter the steady-state pricing of cash flow risk as opposed to discount rate risk. In this way, there is a connection with Merton (1973), Campbell (1993), Campbell and Viceira, Campbell and Vuolteenaho (2004), and Hanson, Shleifer, Stein, and Vishny (2014).

2.2.3 Equilibrium yields

In market A , there is a mass p_A of fast-moving specialists, each with demand $b_{A,t}^*$, and a mass $(1 - p_A - p_B)k^{-1}$ of active slow-moving generalists, each with demand $d_{A,t}^*$. These investors must accommodate the active supply, which is the total supply of $s_{A,t}$ less any supply held off the market by inactive generalist investors, $(1 - p_A - p_B)k^{-1} \sum_{j=1}^{k-1} d_{A,t-j}^*$. Thus, the market-clearing condition for asset A is

$$\begin{array}{c} \text{Specialist} \\ \text{demand} \\ \underbrace{p_A b_{A,t}} \end{array} + \underbrace{\begin{array}{c} \text{Active generalist} \\ \text{demand} \\ (1 - p_A - p_B)k^{-1} d_{A,t}^* \end{array}} = \underbrace{\begin{array}{c} \text{Total bond} \\ \text{supply} \\ s_{A,t} \end{array}} - \underbrace{\begin{array}{c} \text{Inactive generalist} \\ \text{holdings} \\ (1 - p_A - p_B)(k^{-1} \sum_{i=1}^{k-1} d_{A,t-i}^*) \end{array}}. \quad (24)$$

The market-clearing condition for asset B is analogous.

We conjecture that equilibrium yields and generalist demands are linear functions of a state vector, $\mathbf{x}_t$, that includes the steady-state deviations of the short-term interest rate, the default (cash-flow) realization, the supply of asset A , the supply of asset B , inactive generalist holdings of asset A , and inactive generalist holdings of asset B . Formally, we conjecture that long-term yields in market A and B are linear in the state vector

$$y_{A,t} = \alpha_{A0} + \boldsymbol{\alpha}'_{A1} \mathbf{x}_t, \quad (25)$$

$$y_{B,t} = \alpha_{B0} + \boldsymbol{\alpha}'_{B1} \mathbf{x}_t, \quad (26)$$

and that the demands of slow-moving generalists are

$$d_{A,t} = \delta_{A0} + \boldsymbol{\delta}'_{A1} \mathbf{x}_t, \quad (27)$$

$$d_{B,t} = \delta_{B0} + \boldsymbol{\delta}'_{B1} \mathbf{x}_t, \quad (28)$$

where the $2(1+k) \times 1$ state vector, $\mathbf{x}_t$, is given by

$$\mathbf{x}_t = [r_t - \bar{r}, z_t - \bar{z}, s_{A,t} - \bar{s}_A, s_{B,t} - \bar{s}_B, d_{A,t-1} - \delta_{A0}, \dots, d_{A,t-(k-1)} - \delta_{A0}, d_{B,t-1} - \delta_{B0}, \dots, d_{B,t-(k-1)} - \delta_{B0}]'. \quad (29)$$

These assumptions imply that the state vector follows an AR(1) process

$$\mathbf{x}_{t+1} = \Gamma \mathbf{x}_t + \boldsymbol{\epsilon}_{t+1}. \quad (30)$$

This means that the transition matrix Γ depends on generalist demands so we can write $\Gamma = \Gamma(\boldsymbol{\delta}_{1A}, \boldsymbol{\delta}_{1B})$.

As we shown in the Internet Appendix, equilibrium yields take same basic form as (14) with only specialist investors. For market A , the yield is

$$\begin{aligned} y_{A,t} = & \overbrace{\left\{ \bar{r} + \left(\frac{1 - \theta_A}{1 - \rho_r \theta_A} \right) (r_t - \bar{r}) \right\}}^{\text{Expected future short rates}} \\ & + \overbrace{\left[(p_A \tau)^{-1} V_A^{(1)*} (\bar{s}_A - (1 - p_A - p_B) \delta_{A0}) \right]}^{\text{Unconditional term premia}} \\ & + \overbrace{\left[(p_A \tau)^{-1} V_A^{(1)*} \left(\begin{array}{c} \frac{1 - \theta_A}{1 - \theta_A \rho_{s_A}} (s_{A,t} - \bar{s}_A) \\ - (1 - \theta_A) (1 - p_A - p_B) k^{-1} \sum_{i=0}^{\infty} \theta_A^i E_t [\sum_{j=0}^{k-1} (d_{A,t+i-j}^* - \delta_{A0})] \end{array} \right) \right]}^{\text{Conditional term premia}}. \end{aligned} \quad (31)$$

The yield for market B asset has an extra term relating to expected future default losses, but is otherwise similar

$$\begin{aligned} y_{B,t} = & \overbrace{\left\{ \bar{r} + \left(\frac{1 - \theta_B}{1 - \rho_r \theta_B} \right) (r_t - \bar{r}) \right\}}^{\text{Expected future short rates}} + \overbrace{\left\{ \bar{z} + \frac{1 - \theta}{1 - \rho_z \theta} \rho_z (z_t - \bar{z}) \right\}}^{\text{Expected future default losses}} \\ & + \overbrace{\left[(p_B \tau)^{-1} V_B^{(1)*} (\bar{s}_B - (1 - p_A - p_B) \delta_{B0}) \right]}^{\text{Unconditional term/credit premia}} \\ & + \overbrace{\left[(p_B \tau)^{-1} V_B^{(1)*} \left(\begin{array}{c} \frac{1 - \theta_B}{1 - \theta_B \rho_{s_B}} (s_{B,t} - \bar{s}_B) \\ - (1 - \theta_B) (1 - p_A - p_B) k^{-1} \sum_{i=0}^{\infty} \theta_B^i E_t [\sum_{j=0}^{k-1} (d_{B,t+i-j}^* - \delta_{B0})] \end{array} \right) \right]}^{\text{Conditional term/credit premia}} \end{aligned} \quad (32)$$

Although equations (31) and (32) show that yields in markets A and B take a similar algebraic form, the risk premia in the two markets will not be the same because of the different risks that market specialists must bear in equilibrium.

2.2.4 Technical note: Equilibrium multiplicity and equilibrium selection

As explained further in the Internet Appendix, solving the model involves finding a solution to a system of $8k$ polynomial equations in $8k$ unknowns. Specifically, we need to determine the way that equilibrium yields and active generalist demand in markets A and B respond to shifts in asset supply in A and B : this generates 8 unknowns. We also need to determine how equilibrium yields and active generalist demand in A and B respond to the holdings of inactive generalists: this

generates $8(k - 1)$ unknowns.

As in the single-asset case, in the presence of supply shocks, a solution only exists if investors are sufficiently risk tolerant (i.e., for τ sufficiently large). And, there can be a multiplicity of equilibrium solutions. However, as above, there is a unique solution that has well-behaved limiting behavior. And, as we show below, this solution delivers comparative statics that accord with common sense.

What is the economic intuition for the multiplicity of equilibria? In our two-asset model with slow-moving investors, there are three separate forces that give rise to equilibrium multiplicity:

1. Since specialists have short-horizons, a steeply-downward sloping demand curve creates a self-fulfilling form of discount rate risk for specialists, just as in the single-asset model. However, the relevant and well-behaved solution features a smaller equilibrium response of A yields to A supply shocks, and similarly for asset B . As above, the solutions featuring a larger response to supply shocks explode in the limiting case where supply risk vanishes.
2. Although cross-market generalists have longer horizons, they are also concerned about the supply or discount rate risk associated with cross-market arbitrage. As a result, the degree of equilibrium segmentation between the A and B can be self-fulfilling. If generalists behave as if markets A and B are highly segmented, then cross-market arbitrage becomes very risky, so generalists do not aggressively integrate markets in response to supply shocks and segmentation persists in equilibrium. Specifically, yields in A are not highly responsive to B supply shocks and may even *decline* in response to an increase in the supply of B , and vice versa. Conversely, if generalists behave as if markets are highly integrated, then cross-market arbitrage becomes less risky, so generalists aggressively integrate markets in response to supply shocks and integration persists in equilibrium: yields in A are highly responsive to B supply shocks and vice versa. However, the relevant and well-behaved solution always features more aggressive cross-market arbitrage and, thus, tighter cross-market integration. The other solutions with weak cross-market arbitrage explode in the limiting case where supply risk vanishes. Specifically, in order to induce generalists to absorb a B supply shock, the yields in A *drop massively* in response to a tiny rise in the supply of B .
3. The final source of multiplicity stems from the way that active generalists and, therefore, bond yields react to the holdings of inactive generalists. In the unique, well-behaved equilibrium, active generalists reduce their holdings less than one-for-one in response to abnormally large holdings of inactive generalists. As a result, large holdings of inactive generalists reduce equilibrium yields. However, there are also equilibria in which active generalists “over-react” to the holdings of inactive generalists, reducing their holding more than one-for-one. This can lead to situations where large holdings of inactive generalists, which reduce the current supply that needs to be absorbed by specialists and active generalists, actually *raises* equilibrium yields. This equilibrium behaves oddly in the limit where the number of generalists grows vanishingly small, with a tiny number of active generalists taking extremely large bets.

2.3 Defining market integration

What do we mean by “market integration”? We define markets as being *integrated in the short-run* when *conditional risk premia* are the same in both markets at each date. For example, the pricing of interest rate risk is conditionally integrated across markets if, at each date, the expected return per unit of exposure to short-rate shocks is the same in markets A and B . Similarly, we will say that markets are *integrated in the long-run* when average, or *unconditional risk premia* are the same in both markets. Unconditional integration is therefore a weaker form of market integration than conditional integration.

Market integration has nothing to do with the speed by which *fundamental cash flow news* is reflected in securities prices. In our model, fundamental cash flow news is reflected instantaneously in both markets. All of the action in our model concerns the way that nonfundamental supply or discount rate news is reflected in different markets over time. To see this, consider the terms in curly brackets in equations (31) and (32) above. Both risky security A and risky security B share exposure to news about changes in future short rates, but these are reflected identically in their yields.

In our model, the degree of market integration depends on who can bear risk at different horizons and is driven by two parameters: $(1 - p_A - p_B)$ and k . The first parameter, $(1 - p_A - p_B)$, is the population share of generalists. This parameter determines the degree of long-run integration between markets. For instance, if $(1 - p_A - p_B) \approx 1$, markets will be well integrated in the long-run even even if k is large. The second parameter, k , indexes the speed with which generalist capital can flow between markets. Thus, k determines the degree of short-run integration. Markets are perfectly segmented if $(1 - p_A - p_B) = 0$ or $k \rightarrow \infty$. If either of these conditions holds, the two markets operate independently of each other.

Formally, collect all of the j -period returns in a vector $\mathbf{rx}_{t \rightarrow t+j}$ and the asset supplies in a vector $\mathbf{s}_t$. Markets are *integrated in the short-run* if

$$\begin{aligned} E_t [\mathbf{rx}_{t \rightarrow t+j}] &= \tau^{-1} \text{Var} [\mathbf{rx}_{t \rightarrow t+j}] \mathbf{s}_t \\ &= \beta_t [\mathbf{rx}_{t \rightarrow t+j}, rx_{M_t, t \rightarrow t+j}] E_t [rx_{M_t, t \rightarrow t+j}] \end{aligned} \quad (33)$$

for some j where $rx_{M_t, t \rightarrow t+j} = \mathbf{s}'_t \mathbf{rx}_{t \rightarrow t+j}$, $E_t [rx_{M_t, t \rightarrow t+j}] = \mathbf{s}'_t E_t [\mathbf{rx}_{t \rightarrow t+j}]$, and $\beta_t [\mathbf{rx}_{t \rightarrow t+j}, rx_{M_t, t \rightarrow t+j}] = \text{Var} [\mathbf{rx}_{t \rightarrow t+j}] \mathbf{s}_t / (\mathbf{s}'_t \text{Var} [\mathbf{rx}_{t \rightarrow t+j}] \mathbf{s}_t)$. In other words, markets are integrated in the short-run if, at each date, a conditional-CAPM based on the current market portfolio ($rx_{M_t, t \rightarrow t+j} = \mathbf{s}'_t \mathbf{rx}_{t \rightarrow t+j}$) prices both the A and B assets. In our model, markets are integrated in the short-run if and only if $(1 - p_A - p_B) = 1$ and $k = 1$. In this case, the model collapses to a multi-asset analog of Greenwood and Vayanos (2014).

Similarly, markets are *integrated in the long-run* if

$$\begin{aligned} E [\mathbf{rx}_{t \rightarrow t+k}] &= \tau^{-1} \text{Var} [\mathbf{rx}_{t \rightarrow t+k}] E [\mathbf{s}_t] \\ &= \beta [\mathbf{rx}_{t \rightarrow t+j}, rx_{\overline{M}, t \rightarrow t+j}] E [rx_{\overline{M}, t \rightarrow t+j}], \end{aligned} \quad (34)$$

where $E[rx_{\overline{M},t \rightarrow t+j}] = E[s'_t] E[\mathbf{r}\mathbf{x}_{t \rightarrow t+j}]$ and $\beta[\mathbf{r}\mathbf{x}_{t \rightarrow t+j}, rx_{\overline{M},t \rightarrow t+j}] = Var[\mathbf{r}\mathbf{x}_{t \rightarrow t+j}] E[s_t] / (E[s'_t] Var[\mathbf{r}\mathbf{x}_{t \rightarrow t+j}] E[s_t])$. In other words, markets are integrated in the short-run if the same unconditional-CAPM based on the average market portfolio ($rx_{\overline{M},t \rightarrow t+j} = E[s'_t] \mathbf{r}\mathbf{x}_{t \rightarrow t+j}$) prices both the A and B assets on average. In our model, markets are integrated in the long-run if and only if $(1 - p_A - p_B) = 1$, irrespective of the value of k .

Economically, the reason markets are not integrated is because cross-market arbitrage is risky for generalists as in Gromb and Vayanos (2002). Specifically, unless $(1 - p_A - p_B) = 1$ and $k = 1$, short-run integration fails because generalists demand compensation for the unique risks associated with their short-run cross-market arbitrage activities. Similarly, unless $(1 - p_A - p_B) = 1$, long-run integration fails because generalists are engaged in a risky cross-market arbitrage trade even in the long run and must be compensated for its risks. Thus, in the general case where $(1 - p_A - p_B) < 1$ and $k > 1$, both short-run and long-run integration fail.

How should one think about the relevant values for $(1 - p_A - p_B)$ and k empirically? Clearly, the relevant values of $(1 - p_A - p_B)$ and k depend heavily on the two markets being considered. For instance, U.S. Treasury and U.S. Agency bonds are often overseen by the same portfolio manager within a large institution. As a result, we would expect $(1 - p_A - p_B)$ to be near 1 and k to be low, so the two markets would be tightly integrated even in the short-run: Treasury supply shocks would be rapidly transmitted to Agency debt markets and vice versa. However, in other cases, such as U.S. Treasury bonds and corporate bonds or the fixed-income market and the equity market, it is natural to think that $(1 - p_A - p_B)$ is well below 1 and that $k > 1$. Although these two different asset classes are often held by the same generalists—e.g., pension funds or endowments, most of these investors are quite slow to reallocate capital.

3 Dynamics of Securities Prices in the Short- and Long-Run

Following large supply shocks, how do prices adjust across securities markets? Here we use our model to explore asset price dynamics following shocks. We are particularly interested in understanding how these dynamics depend on the model's underlying parameters, especially $(1 - p_A - p_B)$ and k .

Below in Table 1 we list the benchmark set of parameter values that we use in these exercises. Parameters are calibrated under the assumption that market A is the U.S. Treasury market and that market B is the investment grade corporate bond market. For simplicity, we calibrate the model so that one period corresponds to one year. The total average supply of assets in each market is normalized to be one unit. Further details on the calibration are provided in Internet Appendix B.

We begin our analysis by choosing $k = 4$ years and $p_A = p_B = 0.45$, but later show comparative statics for these parameter choices. Based on these values, our simulations assume that most of the capital in each market is operated by specialists, with 10% being controlled by flexible generalist investors, one-fourth of whom re-allocate their portfolio each year. Our choice of $k = 4$ is arbitrary, but we think of this as capturing the empirically relevant case of pension funds reallocating their asset exposures on an annual or biennial basis. Below we discuss how results depend on all these

parameters.

3.1 Unanticipated supply shocks

We first consider the impact of a unanticipated supply shock that increases the supply of asset A (Treasuries) by 50% in period 10. To make the intuition as stark as possible, we focus on the case of a near-permanent supply shock.⁹ Specifically, Figure 1 illustrates the evolution of expected annual returns and bond yields in market A (Treasuries) and market B (corporate bonds). Figure 2 shows how specialists and generalist investors adjust their holdings in response to the shock. Finally, Figure 3 plots the yield spread between the B asset (corporate bonds) and the A asset (Treasuries).

Prior to the supply shock in period 10, Figure 1 shows that the risk premium in market B (corporate bonds) is 0.58% per annum versus a risk premium of 0.46% in market A (Treasuries). The additional risk premium of 0.12% obtains because market B (corporate bonds) are subject to default risk, which exposes corporate bond holders to an additional source of cash flow risk and amplifies the supply risk facing corporate bond holders. The yield in market B is 5.18% per annum versus a yield of 4.46% in market A . The steady-state yield in market A equals the average short-term riskfree rate of $\bar{r} = 4.00\%$ plus the steady-state risk premia of 0.46%. The 0.72% steady-state yield spread between the B and A markets equals the difference in steady-state risk premia of 0.12% plus the market B 's expected default losses of $\bar{z} = 0.60\%$ per annum.

When the supply shock hits the A (Treasury) market in period 10, expected returns and yields in both markets react immediately. Panel A in Figure 1 shows that expected returns in market A overreact and reach a peak of 0.68% before ultimately falling back to a long-run level of 0.62%. In contrast, expected returns in market B under-react, rising slowly from 0.58% to a new long-run level of 0.65%. The overreaction of expected returns for asset A is illustrative of a general property of models that feature slow-moving capital, namely, the relative *steepness* of short-run demand curves and relative *flatness* of long-run demand curves.

Recall that $k = 4$ in this example, so by period 13 all generalist investors have re-allocated their portfolios in response to the supply shock at period 10. However, the gradual adjustment of generalists gives rise to modest echo effects after period 13, generating a series of damping oscillations that converge to the new long-run equilibrium. As in Duffie (2010), these oscillations arise because generalists who reallocate soon after the supply shock hits take large opportunistic positions that temporarily reduce the supply of A (and increase the supply of B) that needs to be actively absorbed in later periods.

In market A , conditional risk premia are the sum of risk premia related to interest rate risk, the supply of A , and the supply of B . Changes in total risk premia are driven primarily by the pricing of interest rate risk. Conditional risk premia in market B can be similarly decomposed into its components, which also include a premium for cash flow risk. Following the supply shock, the

⁹Because we set $\rho_{s_A} = 0.999$ in our benchmark calibration, the supply shock has a half-life of 693 years and is nearly permanent.

premium associated with interest rate risk differs significantly between the two asset markets. As asset allocators react to this pricing discrepancy, the difference in interest rate risk premia between the two markets gradually narrows. However, the difference does not vanish in the long run because of the permanent risks associated with cross-market arbitrage. Furthermore, because markets are partially segmented, large supply shocks can have surprising effects on seemingly unrelated risk premia in our model. Specifically, because it triggers significant cross-market capital flows, the shock to the supply of asset A (Treasuries) actually raises the risk premium that corporate bond investors earn for bearing default risk, even though Treasury bonds themselves have no exposure to default risk.

Figure 1.B shows the reactions of bond yields in both markets. The overreaction of the A market and the underreaction of the B market is more muted in yield space than in risk premium space. This is natural since bond yields reflect weighted averages of future bond risk premia.¹⁰ Market A yield overreacts by 12% of the total long-run impact and market B yield under-reacts by 18% of the total long-run impact.

Why does market A overreact while market B underreacts? Figure 2.A shows how the positions of different market participants evolve over time. Following the initial supply shock in market A , both specialist demand in A ($b_{A,t}$) and active generalist demand in A ($d_{A,t}$) spike upwards. As a partial hedge against their increased holdings of A , active generalists reduce their holdings in market B ($d_{B,t}$). Specifically, as shown in equation (23), this reduction in active generalists' B holdings is motivated by a need to reduce the common short-rate risk (and supply risk) across their holdings in both markets. To fill the void left by the generalists, specialists in market B must hold more of the B asset ($b_{B,t}$). As time passes and more generalists reallocate their portfolios in response to the supply shock, the active demands for A ($b_{A,t}$ and $d_{A,t}$) decline slowly towards their new long-run levels.

In our model, the dynamics of bond risk premia are tied to the dynamics of the “active supply” of A and B that must be absorbed by active market participants each period. By active supply we mean the total supply less the assets that are being held off the market by inactive generalists, corresponding to the right-hand side of equation (24). The evolution of the active supplies is shown in Figure 2.B. The initial supply shock to A in period 10 immediately increases the active supply in A but has no immediate effect on the active supply of B . This is because slow-moving generalists have yet to reduce their holdings in market B . Over the ensuing periods, generalists gradually increase their holdings of A and reduce their holdings of B . Therefore, the active supply in A gradually declines while the active supply in B gradually rises. The dynamics of active supply mirror those for bond risk premia shown in Figure 1.A.

We can assess the evolution of market segmentation over time by examining the dynamics of yields spreads. Figure 3 shows that $y_B - y_A$, the yield spread between the two markets, compresses due to the increased supply of asset A . The compression of yield spread is driven by the convergence in risk premia between the two markets. Because yield A overreacts and yield B underreacts, the

¹⁰Specifically, generalizing (14) we have $y_{A,t} = (1 - \theta_A) \sum_{i=0}^{\infty} \theta_A^i E_t[r_{t+i} + \tau^{-1} V_A^{(1)} b_{A,t+i}^*]$ and $y_{B,t} = (1 - \theta_B) \sum_{i=0}^{\infty} \theta_B^i E_t[r_{t+i} + z_{t+i+1} + \tau^{-1} V_B^{(1)} b_{B,t+i}^*]$.

yield spread overreacts even more (38%) than the A market yield.

3.2 Comparative statics

We next perform a number of comparative statics exercises. The critical parameters governing price dynamics in our model are investor risk tolerance τ , the mix between specialist and generalist investors p_A and p_B , the rebalancing frequency of generalist investors k , and the fundamental cash-flow risk of a cross-market arbitrage position σ_z^2 .

In Table 2, we show how the price impact of the same supply shock studied above—an unanticipated shock that increases the supply in market A by 50%—varies as we change these key parameters of our model one at a time. (All other parameters are held constant at the values listed in Table 1.) For a given set of model parameters, we summarize the impact of the supply shock on both the A and B markets by listing the yields and expected annual returns in (i) the period before the shock arrives (labeled as “pre-shock”), (ii) the period when the shock arrives (labeled as “short-run”), and (iii) in $2k$ periods after the shock arrives (labeled as “long-run”).

We define the degree to which bond yields over- or underreact as the difference between the short-run change and the long-run change, expressed as a percentage of the long-run change¹¹

$$\%Over-Reaction(y) \equiv \frac{(y_t - y_{t-1}) - (y_{t+2k} - y_{t-1})}{(y_{t+2k} - y_{t-1})}.$$

Our measure of over-reaction for risk premia, $\%Over-Reaction(E[rx])$, is defined analogously.¹²

According to this definition, in our benchmark set of parameters, yields in market A overreact by approximately 12%, while yields in market B underreact by 18%.

The second row in Table 2 shows that, if market participants are more risk tolerant, this reduces the price impact of the supply shock on both market A and market B . Changing investor risk tolerance has a very similar impact on the short- and long-run response of yields to shocks. Thus, the degree of overreaction or underreaction in each market, when expressed in percentage terms, is unchanged.

We next change the mix between generalist and specialist investors. p_A and p_B indicate the relative fraction of specialists in market A and B , respectively. In row 3, we set $p_A = p_B = 0.5$ so there are no generalists and the two markets are completely segmented: a supply shock in the market A is not transmitted to the market B and vice versa.

In contrast, in the case of many generalists and few specialists, the markets are well integrated, so that both the A market and the B market overreact to a supply shock that directly hits only

¹¹Since our supply shock is not quite permanent, we subtract off the constant $(1 - \rho_{s_A}^{2k})/\rho_{s_A}^{2k}$ from $\%Over-Reaction$ to ensure that our measure is zero the case of perfectly conditionally-integrated ($1 - p_A - p_B = k = 1$) or perfectly segmented markets ($1 - p_A - p_B = 0$) in which there is no “over-reaction” but only “reaction.” This is because in these limiting cases we have $[(y_t - y_{t-1}) - (y_{t+2k} - y_{t-1})] / [y_{t+2k} - y_{t-1}] = [\alpha_{s_A} s_{A,t} - \alpha_{s_A} s_{A,t} \rho_{s_A}^{2k}] / \alpha_{s_A} s_{A,t} \rho_{s_A}^{2k} = (1 - \rho_{s_A}^{2k})/\rho_{s_A}^{2k}$. For $k = 4$ and $\rho_{s_A} = 0.999$, this constant is 0.8%.

¹²Our measure of overreaction compares the short-run and long-run price impact, holding fixed the exogenous model parameters such as $(1 - p_A - p_B)$ and k . Of course, it is also interesting to compare the price impact of supply shocks as we vary these parameters values. Table 2 performs both types of exercises.

the A markets. In this case, the two markets behave as essentially one, and the result is similar to that in Duffie (2010).

We also change the mix between market A specialists and market B specialists, holding fixed the overall mix between generalists and specialists. Table 2 shows that if we hold the total number of specialists the same at $p_A + p_B = 0.8$, then as we increase the proportion of specialists in B and decrease the proportion of specialists in A , we get more over-reaction in A . The B market is only modestly affected by this change in the relative fraction of specialists, because the supply shock is primarily being absorbed by generalists anyway.

k is the number of periods it takes for generalists to fully reallocate their portfolios. When $k = 2$, half the generalists reallocate their portfolio each period, and the other half reallocate in the next period. As k increases, there is more over-reaction in market A , and more under-reaction in market B . Note that k also affects the long-run unconditional risk premium. If $\sigma_{s_A}^2, \sigma_{s_B}^2 > 0$, then the steady-state fraction of the both assets held by generalists rises and the steady-state risk premium declines as we increase k . This is because, on average, longer horizon investors have a comparative advantage at bearing transitory supply shocks (discount rate shocks).¹³ In summary, as we increase k , *supply shocks* have a large impact on conditional risk premia, but *supply risk* has a smaller effect on unconditional risk premia.

Finally, we ask whether the supply shock scenario we have considered has symmetric effects if it is delivered in market B , the market in which the risky asset also has exposure to cash flow risk. The bottom two lines of the table show that the answer is yes: the price impact on market A when the supply shock hits market B is exactly the same as the price impact on market B in our base case where the supply shock hits A . The table shows that the degree of market integration depend on the amount of default risk in market B . As we increase σ_z^2 , we have less long-run and short-run integration between the two markets, because A and B are less substitutable. If B -specific risk is large, then the market with the shock will have a larger peak (because there is less ability to integrate between the markets). However, there is not much effect on under-reaction in the market that does not receive the shock (i.e., under-reaction is consistent across scenarios).

As we reduce the amount default risk in market B , shocks to the supply of B have a large price impact on market A because the two assets are closer substitutes and cross-market arbitrage is less risky. In the limit when $\sigma_z^2 = 0$, the markets would be perfect substitutes and thus perfectly integrated: conditional risk premia would be identical in the two markets.

3.3 Anticipated supply shocks

We next study the dynamics following the *announcement* of a large *future change* in asset supply. As an example of a large pre-announced supply change, consider the Large Scale Asset Purchase programs (often called “Quantitative Easing”) initiated by the Federal Reserve between 2008 and

¹³Formally, let $E[rx_{t+1}^A] = (p_A \tau)^{-1} V_A^{(1)*} (\bar{s}^A - (1 - p_A - p_B) \delta_{A0})$ denote the unconditional risk premium. If $\sigma_{s_A}^2, \sigma_{s_B}^2 > 0$ and $(1 - p_A - p_B) > 0$, $E[rx_{t+1}^A]$ is decreasing in k and δ_{A0} is increasing in k . However, when $\sigma_{s_A}^2, \sigma_{s_B}^2 = 0$, $E[rx_{t+1}^A]$ and δ_{A0} are independent of k .

2013. The Fed's initial announcement of long-term bond purchases occurred on November 25, 2008, but asset purchases did not begin until January 2009 and continued in the months thereafter.

To mimic the announcement of a future increase in the supply of asset A , we assume that $s_{A,t}$ jumps up at some time t and we simultaneously increase the demands of inactive generalist investors for A such that the *active supply* of asset A does not change at time t . This means that, unlike the case of an unanticipated supply shock, it would be possible to clear the market at time t without any increase in the holdings of A specialists or active generalists. Thus, the only reason that prices change when a future supply change is announced is because the announcement leads active long-horizon generalists to opportunistically increase their current holdings of A .

Formally, letting $\varepsilon_t[X_t] = E_t[X_t] - E_{t-1}[X_t]$ denote the time t innovation to some random process X_t , an anticipated supply shock is defined so that the innovation to the right-hand of equation (24) is zero

$$\varepsilon_t[s_{A,t}] - (1 - p_A - p_B) k^{-1} \sum_{j=1}^{k-1} \varepsilon_t[d_{A,t-j}] = 0. \quad (35)$$

Furthermore, if we vary $\{\varepsilon_t[d_{A,t-j}]\}_{j=1}^{k-1}$ holding fixed $\sum_{j=1}^{k-1} \varepsilon_t[d_{A,t-j}] = (k/(1 - p_A - p_B)) \varepsilon_t[s_{A,t}]$, we can alter the announced *timing* of a supply change holding fixed the announced size of the *cumulative* change. As we discuss shortly, exercises of this sort can be used to evaluate different asset purchase strategies for a central bank seeking to impact long-term interest rates or for share repurchasing firm seeking to give boost to its stock price.

We begin with the simpler case of the pre-announcement of a one-time, near-permanent jump in the supply of asset A . Specifically, Figure 4 shows the dynamics of risk premia and bond yields when market participants learn in period 5 that the supply of asset A will increase in period 8. Panel A of Figure 4 shows that annual risk premia in market A actually *decline* slightly after the announcement in period 5 and before the supply rises in period 8. And risk premia in market A still jump up when bond supply actually jumps to its new level in period 8. The risk premium in market B , which is only indirectly affected due to cross-market arbitrage by generalists, rises gradually over time. What drives these dynamics? Upon the announcement, generalists begin to gradually adjust in the direction of the anticipated shock, buying asset A and selling asset B . Risk premia decline in market A because A specialists provide temporary liquidity to generalists, planning to replenish their inventories once the supply shock actually lands.

Panel B of Figure 4 shows the evolution of bond yields in response to the anticipated supply shock. Yields in *both* market A and market B rise *gradually* to their new steady-state levels. The yields in market A underreact by 51% and the yields in market B underreact by 46%. Why do A yields rise gradually when the supply shock is pre-announced but over-reacted in Figure 1 when the same shock was unanticipated? First, risk premia in the near future have a larger effect on bond yields than those in the distant future. Second, risk-premia in the A market are only expected to rise significantly once the supply actually increases at time 8. In combination, these two facts imply that yields in market A must rise gradually over time.

Figure 5 shows how the positions of market participants evolve over time in response to this

announced supply increase. Panel A shows that active generalist opportunistically increase the holdings of A (d_A) and decrease their holdings of B (d_B) when the shock is announced at time 5.¹⁴ The gradual build up of generalist demand in A is responsible for the slight decline in annual A market risk premia from periods 5 to 7 before the upward jump in period 8. Similarly, the gradual reduction of generalists' demand for B results in a slow rise in B market risk premia. In contrast to the generalists, the fast specialist demand in market A (b_A) decreases initially then increases. This is because specialists can adjust quickly, and thus they have the ability to front-run the anticipated change in supply – specialists reduce their portfolio holdings of A just before the positive supply shock and increase holding of A immediately after the shock.

Figure 6 compares the dynamics of prices in the case of anticipated versus unanticipated increase in supply. In both cases, the supply of asset A increases by 50% at time 8. In the anticipated case, this supply increase is announced at time 5, while in the unanticipated case the shock is not pre-announced and a surprise at time 8. Whether or not a shock is anticipated, the *long-run* impact on yields and risk premia is the same. However, the short-run effect can be quite different. When the shock is a surprise, yields and risk premia in the A market overreact more strongly. Pre-announcing the supply shock mobilizes slow-moving generalists before the supply of A actually rises at time 8. This early mobilization reduces the active supply of asset A that must be absorbed when the shock lands at time 8, damping the overreaction of market A . While introducing a delay between announcement and the supply increase reduces overreaction, this also lengthens the amount of time it takes for the full impact of the supply increase to be reflected in prices. This tension may be relevant for central bankers designing asset purchase programs: allowing for a long period of time between a purchase announcement and implementation results in less profits for market specialists and, therefore, less short-term cost of policy implementation. Of course, pre-announcement necessarily delays the desired impact on asset prices.

We can also use the model to describe more complex paths of policy announcements. Consider a closer representation of central bank asset purchase and sale. An announcement is made regarding a policy that is to be implemented over multiple future periods, analogous to the Federal Reserve's announcement that it would purchase \$40 billion of MBS each month in QE3. In Figure 7, asset sales are announced in period 5 and carried out from period 6 to 8. Panel A shows the paths of risk premia rising gradually over the periods of policy implementation. Panel B shows the paths of yields also rising gradually. In summary, we observe substantial under-reactions in returns and yields in both markets at the time of the policy announcement in period 5. An event study during the time of the initial announcement would capture some of the market reaction, but it would far understate the long-run impact.

¹⁴Why do generalists buy A bonds *in advance* of the supply increase? Generalists have long-horizons ($k = 4$ in this example), but can only adjust their portfolios slowly. Although generalists expect A bond yields to rise over the next 4 periods and, therefore, expect to suffer a capital loss on A bonds, they expect this capital loss to be more than offset by an increase in income from holding A bonds. As a result, the expected cumulative 4-period excess return from holding A bonds rises when a future supply increase is announced at t . Formally, since $E_t[r_{x_{A,t \rightarrow t+4}}] = \{\sum_{i=0}^3 E_t[y_{A,t+i} - r_{t+i}]\} - \{(\theta_A / (1 - \theta_A)) E_t[y_{A,t+4} - y_{A,t}]\}$, this means that change in the first term in curly braces outweighs the second term in curly braces. Using equation (23), the significant rise in $E_t[r_{x_{A,t \rightarrow t+4}}]$ then translates in an increase in $d_{A,t}$ and decline in $d_{B,t}$.

In Figure 8, we show the paths of generalists and specialist demands in response to a series of anticipated supply shocks. Generalists and specialists both increase their demands for A in order to accomodate the overall increase in supply. Generalists also reduce their demand for B to partially hedge against their purchase of A . However, this market integration occurs at a slow pace since only one-tenth of the generalists can reallocate between the two markets each period. Lastly, specialists in B increase their demand (b_B) to fill in for the generalists that have left market B for market A .

4 Model Extensions

4.1 Multiple securities per market

We now explore how the main results of our model carry over to a more complex setting in which there are multiple risky securities trading in each market.

4.1.1 Markets and securities

There are N bonds in market A , denoted $A_1, A_2, \dots, A_N$. As above, we assume that the A -market bonds are default-free and are only exposed to interest rate risk.¹⁵ However, the bonds have different durations. Specifically, security A_n has duration D_{A_n} . Following the same logic as in equation (2), the log excess return on bond A_n over the short-term interest rate from time t to $t + 1$ is

$$rx_{A_n,t+1} = \frac{\overbrace{1}^{D_{A_n}}}{1 - \theta_{A_n}} y_{A_n,t} - \frac{\overbrace{\theta_{A_n}}^{D_{A_n}-1}}{1 - \theta_{A_n}} y_{A_n,t+1} - r_t, \quad (36)$$

where $\theta_{A_n} = 1 - 1/D_{A_n}$.

We also assume that there are N defaultable bonds in market, denoted $B, B_1, B_2, \dots, B_N$. We assume that the return on security B_n from time t to $t + 1$ is

$$1 + R_{B_n,t+1} = (1 - Z_{t+1})^{\psi_n} (1 - U_{n,t+1}) \frac{(\delta_{B_n} P_{B_n,t+1} + C_{B_n})}{P_{B_n,t}}, \quad (37)$$

where Z_{t+1} is a default process common to all bonds in the B market, ψ_n is the exposure of

¹⁵Technically, we model all of these bonds as perpetuities. In order to have perpetuities with different durations, we need to introduce a set of “geometrically decaying perpetuities.” Specifically, consider a perpetuity that promises to pay a decaying stream $C, \delta C, \delta^2 C, \delta^3 C, \dots$ where $(1 - \delta) \in [0, 1]$ denotes the geometric decay rate. Thus, $\delta = 0$ corresponds to 1-period debt and $\delta = 1$ is a consol bond with a constant promised coupon. Assuming a yield of $Y_{L,t}$, the price of this security is $P_{L,t} = \sum_{j=1}^{\infty} (1 + Y_{L,t})^{-j} \delta^{j-1} C = C / (1 + Y_{L,t} - \delta)$ which implies a Macaulay duration of $-\partial P_{L,t} / \partial Y_{L,t} = (1 + Y_{L,t}) / (1 + Y_{L,t} - \delta)$. Suppose the perpetuity’s price is $\bar{P}_L = 1$ and its yield is $\bar{Y}_L$ in the steady-state. This implies a coupon of $C = 1 - \delta + \bar{Y}_L$ and a steady-state duration of $-\partial P_{L,t} / \partial Y_{L,t} = (C + \delta) / C$ which is increasing in δ .

Using the same steps as above, the log return on the decaying perpetuity from t to $t + 1$ is approximately $r_{L,t+1} = \log[(\delta P_{L,t+1} + C) / P_{L,t}] \approx (1 - \theta)^{-1} y_{L,t} - \theta (1 - \theta)^{-1} y_{L,t+1}$ where $\theta = \delta / (\delta + \exp(c - \bar{p}_L))$. Since the steady-state price is par ($\bar{p}_L = \log(1) = 0$), we have $\theta = \delta / (\delta + C)$. Thus, bond duration is $-\partial P_{L,t} / \partial Y_{L,t} = (1 - \theta)^{-1} = (\delta + C) / C$ which corresponds to the Macaulay duration when the perpetuity is trading at par.

Thus, we assume that security A_n has a geometric decay rate if δ_{A_n} and a coupon of $C_{A_n} = 1 - \delta_{A_n} + \bar{Y}_{A_n}$, implying a duration of $D_{A_n} = (1 - \theta_{A_n})^{-1} = (1 + \bar{Y}_{A_n}) / (1 + \bar{Y}_{A_n} - \delta_{A_n})$.

perpetuity B_n to this systematic default factor, and $U_{n,t+1}$ is an idiosyncratic default process that is specific to security B_n . Therefore, the log excess return on bond B_n from time t to $t+1$ is

$$rx_{B_n,t+1} = \frac{1}{1 - \theta_{B_n}} y_{B_n,t} - \frac{\theta_{B_n}}{1 - \theta_{B_n}} y_{B_n,t+1} - \psi_n z_{t+1} - u_{n,t+1} - r_t, \quad (38)$$

where $\theta_{B_n} = 1 - 1/D_{B_n}$ and $u_{n,t+1} = -\ln(1 - U_{n,t+1})$.

We assume that the processes for the short rate r_t and for the common default process z_t are as in equations (4) and (5) above. We assume that idiosyncratic default process for security B_n follows

$$u_{n,t+1} = \bar{u}_n + \rho_{u_n} \cdot (u_{n,t} - \bar{u}_n) + \varepsilon_{u_n,t+1}. \quad (39)$$

We assume the net supplies that investors must hold in the A assets are

$$\mathbf{q}_{A,t} = \mathbf{q}_{A,0} + \mathbf{q}_{A,1} \cdot s_{A,t} \quad (40)$$

where $s_{A,t}$ follows

$$s_{A,t+1} = \rho_{s_A} s_{A,t} + \varepsilon_{s_A,t+1}. \quad (41)$$

Similarly, the net supplies that investors must hold in the B assets are

$$\mathbf{q}_{B,t} = \mathbf{q}_{B,0} + \mathbf{q}_{B,1} \cdot s_{B,t},$$

where $s_{B,t}$ follows

$$s_{B,t+1} = \rho_{s_B} s_{B,t} + \varepsilon_{s_B,t+1}. \quad (42)$$

Finally, we assume that $\varepsilon_{r,t+1}$, $\varepsilon_{z,t+1}$, $\varepsilon_{s_A,t+1}$, $\varepsilon_{s_B,t+1}$, $\varepsilon_{u_1,t+1}$, $\varepsilon_{u_2,t+1}$, ..., $\varepsilon_{u_3,t+1}$ are mutually orthogonal.

4.1.2 Market participants

As above, there are three-types of investors, each with risk tolerance τ . A -specialists are present in mass p_B , B -specialists are present in mass p_B , and generalists are present in mass $(1 - p_A - p_B)$.

Fast-moving A -specialists are free to adjust their holdings of all securities in the A market (and the riskless short-term rate) each period, but cannot hold the B assets. Let $b_{A_n,t}$ denote the demand of A specialists for asset A_n and let $\mathbf{b}_{A,t}$ denote the $N \times 1$ vector of their holdings of each of the N assets in market A . Collecting the excess returns on these N securities in a vector, the excess return on A -specialists portfolio is thus $rx_{A,t+1} = (\mathbf{b}_{A,t})' \mathbf{r}\mathbf{x}_{A,t+1}$.

Fast-moving A -specialists have mean-variance preferences over 1-period portfolio returns and solve

$$\begin{aligned} & \max_{\mathbf{b}_{A,t}} \left\{ E_t [rx_{A,t+1}] - (2\tau)^{-1} \text{Var}_t [rx_{A,t+1}] \right\} \\ &= \max_{\mathbf{b}_{A,t}} \left\{ \mathbf{b}_{A,t}' E_t [\mathbf{r}\mathbf{x}_{A,t+1}] - (2\tau)^{-1} \mathbf{b}_{A,t}' \text{Var}_t [\mathbf{r}\mathbf{x}_{A,t+1}] \mathbf{b}_{A,t} \right\}. \end{aligned}$$

Thus, the demands of A -specialists are given by

$$\mathbf{b}_{A,t}^* = \tau (\text{Var}_t [\mathbf{r}\mathbf{x}_{A,t+1}])^{-1} E_t [\mathbf{r}\mathbf{x}_{A,t+1}].$$

Since this implies $\text{Var}_t [rx_{A,t+1}] = \tau \cdot E_t [rx_{A,t+1}]$, we have

$$E_t [\mathbf{r}\mathbf{x}_{A,t+1}] = \tau^{-1} \text{Var}_t [\mathbf{r}\mathbf{x}_{A,t+1}] \mathbf{b}_{A,t}^* = \beta_t [\mathbf{r}\mathbf{x}_{A,t+1}, rx_{A,t+1}] \cdot E_t [rx_{A,t+1}].$$

Thus, the 1-period returns on all A -market securities will be priced by a local conditional-CAPM that is specific to the A -market—i.e., where the relevant “market portfolio” is the time t portfolio of A -market specialists, $rx_{A,t+1} = (\mathbf{b}_{A,t}^*)' \mathbf{r}\mathbf{x}_{A,t+1}$.

Obviously, a completely symmetric analysis hold for the two B assets. Thus, we will have two partially linked conditional-CAPMs: one for the two A assets and another for the two B assets. The key question is how these two conditional-CAPMs will be linked together in equilibrium by the cross-market arbitrage activities of generalist. As above, slow-moving generalists are present in mass $1 - p_A - p_B$. Fraction $1/k$ of generalists investors are active each period and choose the portfolios of securities from the A and B markets that they will hold over the following k periods.

4.1.3 The risk of cross-market arbitrage

With multiple assets, the key question concerns the risks that generalists face when they undertake cross-market arbitrage. Note that assets in market A are exposed to three common shocks: shocks to r_{t+1} , shocks to $s_{A,t+1}$, and shocks to $s_{B,t+1}$. Assets in market B are exposed to these three common shocks as well as common shocks to z_{t+1} . In addition, asset B_n is potentially exposed to idiosyncratic shocks to $u_{n,t+1}$.

An interesting complication arises if generalists are able to freely choose their holdings of the N securities in market A and the N securities in market B . In this case, it may be possible to use A assets to construct a “factor-mimicking portfolio” that is only exposed to shocks to r_{t+1} and to construct a similar factor-mimicking portfolio using only B assets. If this is possible then, unless the risks associated with shocks to r_{t+1} are being priced the same in the A and B markets at each date, generalists will have a riskless arbitrage opportunity.

Roughly speaking, it will be *possible* to construct (nearly-perfectly) factor-mimicking portfolios if both A and B markets contain many redundant securities. Even if it is *possible* to construct factor-mimicking portfolios in both A and B markets, this may not be feasible for slow-moving generalists. For instance, generalists may lack the expertise to construct their complicated (long-short) mimicking portfolios or may face institutional frictions that make this infeasible.

We can distinguish between at least four cases:

1. **Case 1:** It is possible to construct factor-mimicking portfolios in both markets A and B for each of the common risk factors.
2. **Case 2:** It is not possible to construct factor-mimicking portfolios in this way.

3. **Case 3:** It is possible to construct factor-mimicking portfolio in this way, but generalists are not capable of doing so: generalists function of coarse asset-allocators as opposed to granular cross-market arbitrageurs.

We discuss each of these three cases in greater detail.

Case 1: It is possible to construct factor-mimicking portfolios and generalists can do so Suppose generalists can freely choose positions in all $2N$ securities. Also, suppose that the B assets are not exposed to idiosyncratic shocks.

In this case, if $N \geq 4$ and all securities are non-redundant (technically, their factor loadings must linearly independent), then it will be possible to construct factor-mimicking portfolios for shocks to r_{t+1} , shocks to $s_{A,t+1}$, and shocks to $s_{B,t+1}$ using only A assets and using only B assets.¹⁶ If this is the case, then active generalists will work to perfectly integrate factor pricing between the A and B markets in the short-run. Of course, because generalists are slow-moving, the risk factor prices that prevail in both the A and B markets will be subject to slow-moving capital effects, e.g., risk factor prices will over-react to shocks to the supply of that risk factor. However, because cross asset-class arbitrage is riskless, the same risk factors prevail in all asset markets.

Even if the B market assets are subject to idiosyncratic default shocks, this outcome will obtain in the limit where we hold constant the total supply of A and B assets but allow $N \rightarrow \infty$. In this case, investors portfolios will be arbitrarily granular, implying that it will be easy for investors to diversify away these idiosyncratic shocks.

Thus, generalists can freely choose positions in all $2N$ securities, Case 1 will be a good approximation in the case when N is large relative to the number of common risk factors in the A and B markets.

Case 2: It is not possible to construct factor-mimicking portfolios Next suppose generalists can freely choose positions in all $2N$ securities and that B assets are exposed to idiosyncratic shocks. However, suppose that N is not large, so the existence of these idiosyncratic shocks mean that it is not possible to construct accurate factor-mimicking portfolios using the B market assets. Specifically, the N securities in market B are exposed to $N + 4$ risk factors—4 that are common and N that are asset-specific. A factor-mimicking portfolio is a set of N unknown positions in the B assets that must satisfy $N + 4$ linear equations. In general, there is no such solution. And, if N is small and idiosyncratic volatility is large, then any portfolio will do a poor job of mimicking the each common factor.¹⁷ As a result, cross-market arbitrage will remain risky for generalists, so cross-market integration will be imperfect, both in the short-run (conditionally) as well as in the long-run (conditionally).

¹⁶If $N = 4$, there will be a unique set of factor-mimicking portfolios using B market securities. If $N > 4$, there will be multiple possible factor-mimicking portfolios.

¹⁷By contrast, as noted above, as N grows large a law of large numbers effective reduces the number of equations to 4.

This case might be empirically relevant, if the idiosyncratic risks are not security specific, but are risk factors that are shared by a large subset of securities in a market (e.g., industry default factors).

Case 3: It is possible to construct factor-mimicking portfolios, but face limitations that prevent them from doing so Next suppose that the B assets are either not exposed to idiosyncratic shocks or that N is large, so it is possible to construct accurate factor-mimicking portfolios. However, suppose that generalists are not cannot freely choose positions in all $2N$ securities. For instance, generalists may lack the expertise to construct their complicated (long-short) mimicking portfolios or may face institutional frictions that make this infeasible. In this telling, generalists function coarse asset-allocators and not granular cross-market arbitraguers. They can overweight Treasuries relative to corporate bonds, might even be able to independently vary the duration of both portfolios as well as the credit risk exposure of the corporate portfolios. Nonetheless, the degrees of freedom in within market asset allocation are less than the number of common risk factors. In this case, the markets for A and B will again be imperfectly integrated in both the short- and long-run, because cross-market arbitrage is risky for generalists with granular portfolios of this sort.

To give a few concrete examples, we might first suppose generalists can only move baseline $\mathbf{s}_{A,0}$ and $\mathbf{s}_{B,0}$ up and down so that

$$\begin{aligned}\mathbf{d}_{A,t} &= \mathbf{s}_{A,0} \cdot d_{A,t} \\ \mathbf{d}_{B,t} &= \mathbf{s}_{B,0} \cdot d_{B,t}.\end{aligned}$$

In this case, generalists only have one-degree of freedom: how much to allocate to a baseline portfolio in each market and cannot allocate at all within markets. A less extreme example would be to give generalists two degrees of freedom in each market. For instance, we might assume that

$$\begin{aligned}\mathbf{d}_{A,t} &= \mathbf{s}_{A,0} \cdot d_{A-0,t} + \mathbf{s}_{A,1} \cdot d_{A-1,t} \\ \mathbf{d}_{B,t} &= \mathbf{s}_{B,0} \cdot d_{B-0,t} + \mathbf{s}_{B,1} \cdot d_{B-1,t}.\end{aligned}$$

The bottom line here is that cross-market arbitrage remains risky for coarse asset-allocators such as real-world pension funds, endowments, and macro hedge funds. As a result, cross-asset class market integration is likely to be imperfect in the real world.

4.2 Endogenizing the Amount of Specialization

We could endogenize p_A and p_B , for instance, by allowing them to adjust so that the portfolios of A -specialists, B -specialists, and generalists have the same Sharpe ratios in the model's steady state.¹⁸

¹⁸To compare the multiperiod generalist returns with 1-period specialist returns, the Sharpe ratio of generalists would have to be divided by $\sqrt{k}$ to make it comparable with the 1-period Sharpe ratio of specialists.

In the simple two-asset version model consider above, one equilibrium strategy mix is $p_A^* = p_B^* = 0$. The reason is obvious: since there is no benefit from specialization in this set up, investors will always maximize their Sharpe ratios by being a well-diversified generalists. Specifically, if we consider an price equilibrium where $p_A^* = p_B^* = 0$, then prices must be such that the market portfolio held by generalists has the highest Sharpe ratio of all possible portfolios, including those of A or B specialists. As a result, $p_A^* = p_B^* = 0$ will be an equilibrium in portfolio strategies.

However, this need not be the case in the multi-asset extension considered above if we generalists function as coarse cross-market asset-allocations as opposed to highly granular cross-market arbitrageurs. In this more realistic case, there will be different benefits to specialization—i.e., the ability to do highly granular cross-security arbitrage within a given market—and to generalization—i.e., obtaining a more diversified exposure to common supply shocks. As a result, this set up will deliver interior values of p_A^* and p_B^* .

5 Discussion and Applications

5.1 Event studies and assessment of Quantitative Easing programs

In response to a rapidly evolving financial crisis and worldwide recession, in late 2008 and early 2009, central banks around the world announced their intention to aggressively purchase government bonds and other securities. The Fed announced on November 25, 2008 its intention to buy \$100 billion in GSE debt and \$500 billion in MBS, and followed up two months later with a plan to purchase Treasuries. The Bank of England followed in quick succession, announcing 50 billion pounds of purchases in January 2009, and in March 2009 expanding the purchase program to include government bonds. The Bank of Japan, already engaged in an asset purchase program since earlier in the decade, announced in December 2008 that the size of monthly purchases of JGB securities was to increase. At the time of writing, these central banks preside over large portfolios of securities: by December 2013, the Fed’s open market portfolio included \$3.7 trillion of securities.

A crucial question in assessing the effectiveness of these programs is whether they impacted securities prices beyond government bonds. Suppose, for example, that the impact of the asset purchase programs was limited to the markets in which the purchases were being made (Treasury bonds and MBS), perhaps because these markets are highly segmented from other securities. Such a finding should dampen any enthusiasm for these programs, and cast doubt that asset purchases could affect lending in the broader economy.

Our model proposes a natural framework for understanding how these asset purchase programs should spill across different securities markets. According to our model, the largest short-run effects of these programs should be in the securities being purchased. In the long run, however, changes in risk premia induced in the primary market should spill over to secondary markets. Differences between the short-run and long-run should reflect the degree to which the programs were anticipated, the length of time between the announcement date and implementation, and the effective degree of segmentation between securities markets.

Most empirical studies of the asset purchase programs have focused on the very short-run impact on prices or yields, measured using event studies.¹⁹ In one of the first of such studies, Gagnon, Raskin, Remache, and Sack (2011) report interest rate changes around a set of Fed announcement days between November 2008 and January 2010. Cumulating over all 1-day LSAP announcements, they show a decline of yields of 62 basis points for the 10-year US Treasury Bond, 123 basis points for agency MBS, and 74 basis points for the Baa corporate bond index. Krishnamurthy and Vissing-Jorgensen (2011) extend this analysis to QE2 and also discuss the impact on other assets, including low grade corporate bonds. After controlling for other factors, Krishnamurthy and Vissing-Jorgensen conclude that the effects of asset purchases were most pronounced among the assets being purchased (MBS and Treasuries in the case of QE1 and Treasuries in QE2), effectively suggesting a high degree of segmentation between different fixed income markets. Neeley (2013) shows that the average cumulative change in yields in 10-year bond yields in Australia, Canada, Germany, Japan, and the UK to major Fed LSAP announcements was 53 basis points, compared to the 107 basis points change in the yield of the US 10-year bond. Neeley finds no significant impact on major stock market indexes outside the US. In summary, at short horizons, there is modest evidence (and some disagreement among researchers) that asset purchases impacted other fixed income markets, and almost no evidence that they had any impact on the equity markets. Several other studies have confirmed these basic conclusions.

At the same time, some researchers have recognized that short horizon announcement returns may not capture the full impact of the asset purchase programs. In their empirical assessment of the Bank of England’s quantitative easing program, Joyce, Lasoasa, Stevens and Tong (2010) suggest that it may have impacted corporate bonds and equities, although “these effects might be expected to take time to feed through, as it will take time for investors and asset managers to rebalance their portfolios.” They further suggest that the impact on QE may subsequently be reflected in corporate issuance. Fratzcher, Lo Duca, and Straub (2013) suggest that the Fed’s QE programs triggered portfolio flows that ultimately impacted emerging market asset prices and exchange rates. Studying QE programs in Mamaysky (2014) suggests that QE might ultimately spill into the asset markets through portfolio allocation, but notes that “it is unlikely that such portfolio flows can take place quickly.”

Approaches to measuring the long-run effects of QE vary across researchers. Joyce, Lasoasa, Stevens and Tong report the cumulative change in asset prices for the longer period between March 4 2009 and May 31 2010 in addition to the more standard analysis of 1-day announcement returns. For example, they show that corporate bond yields fall by a cumulative 70 basis points around asset purchase announcements, but by 400 basis points over the longer period. Mamaysky takes a more tailored approach to each asset market: he chooses an announcement window that maximizes the statistical power of the measured return.²⁰ Doing this, he shows that the impact of QE on equity markets and high yield bond markets is much larger than what one would measure using a

¹⁹The event study methodology was originally developed in the 1970s to tackle questions of informational efficiency of stock prices.

²⁰For example, in U.S. stocks, he shows that the highest power is at 15 days, and he documents.

shorter window. But even this approach surely understates the full long-run effect, because, as we have noted, the full impact of supply shocks can easily occur over a window over which the power to detect them is low.

Our model clarifies the broader issue at stake: event studies are a useful methodology for detecting short-run changes in prices, but lack statistical power at longer horizons. But if markets are segmented, the true long run effects may take time to materialize. More formally, consider the effects of σ_z on the detection of price spillover in QE event studies. When σ_z is large relative to σ_r , our model suggests that the econometricians may have little power to detect this gradual shift in excess returns: there is simply too much cash-flow news noise in market B to reliably detect cross-market spillovers using a handful of QE announcement events. The statistical power would increase with the number of events, but power is still decreasing in σ_z . More generally, our framework cautions against drawing negative conclusions about price spillovers based on a handful of events for some pair of assets.²¹

Figure 9 shows the detectability of yield reactions to an unanticipated supply shock in market A . In an environment with low short-rate volatility and low supply shock volatility, a supply shock can have statistically significant impact in the first market but not in the second market. The parametrization of Figure 9 correspond to the low volatility environment during QE when the short-term rate is pinned down at zero and supply risk is low. During this period, risk-taking by speculators was low (τ is small) and mortgage convexity was not an important contributor of supply shocks (as all the pre-payment options were either already exercised or deep in-the-money). Supply shocks induced by the usual hedging demand from insurance companies and pension funds were also lower than normal due to low rate volatility. In such an environment, event studies would have detected significant impact in one market, e.g. the MBS market, and little spillover to the second market, e.g. the treasury market. Even though the impact in the second market exists, the confidence interval is much wider and thus it would not be detectable by conventional event studies.

In summary, our model suggests that researchers must be cautious in their use of event studies in assessing the impact of large supply and demand shocks on market prices. Measuring long-run impact of supply shocks across markets is inherently difficult because the full economic impact may occur over such a long time that it is swamped by noise.

²¹Studying this case allows us to think about the Great Rotation: the claim by practitioners that QE has trigger a slow rotation of capital from bonds to stocks. Since $(1 - p_A - p_B)$ and k are both high, our model predicts that the spillover from bonds (A) to stocks (B) in required return space will be gradual. Furthermore, when σ_z^2 is large, our model suggests that the econometricians may have little power to detect this gradual shift in excess returns: there is simply too much cash-flow news noise in stocks to reliably detect cross-market spillovers using a handful of QE announcement events. The statistical power would increase with the number of events, but power is still decreasing in σ_z^2 . More generally, our framework cautions against drawing negative conclusions about price spillovers based on a handful of events for some pair of assets.

5.2 Corporate Arbitrage

In our model, generalist “asset allocators” play a critical role in integrating prices across markets. Earlier we suggested that these investors were best thought of as pension funds or endowments, who adjust asset allocation at an annual frequency. But an equally valid interpretation of the generalists in our model would be corporations who continually access the capital markets for new capital. Corporations have flexibility over which securities to issue, and can thus execute a form of arbitrage between otherwise segmented securities markets.

Consider a firm with financing needs that has the choice of issuing in the equity or debt markets. Suppose that debt markets have just received a positive demand shock. For example, perhaps bond market mutual funds experienced outflows, causing an increase in prices and a reduction in yields in corporate bond markets. How should a firm with access to debt markets respond? The typical firm may have preferences over its target capital structure, but attractive conditions in the debt markets encourage the firm to issue a greater share of its financing in debt than it ordinarily would.

A growing literature in behavioral corporate finance has suggested that corporations have advantages vis a vis professional arbitrageurs in conducting arbitrage. For example, Ma (2015) points out that corporations ultimately redeem their securities for cash flows generated by the firm. As a result, they are less subject to subsequent changes in the market prices of their securities. In contrast, professional arbitrageurs may be reluctant to aggressively buy an underpriced security, for example, because in the event that the security becomes more underpriced in the future, the arbitrageur may experience outflows that ultimately prevent the arbitrageur from capitalizing on the mispricing.

Is it reasonable to think of corporations as slow-moving generalists in the sense implied by our model? Several recent papers in behavioral corporate finance speak to the slow-moving nature of corporate arbitrage. Greenwood, Hanson, and Stein (2010) suggest, for example, that corporations adjust the maturity of their debt issuance in response to government supply shocks, a phenomenon they dub “gap filling.” They show that when corporate issuance is measured annually, there is only a modest correlation between government supply shocks and the maturity of corporate issuance. However, at horizons of two years or longer, or when the data are measured in levels rather than issuance, the evidence is much stronger. Graham, Leary, and Roberts (2014) show similar results in their study of corporate leverage between 1920 and 2010. Corporate leverage tends to rise when government debt falls, suggesting that corporations respond to supply shocks. But this result is stronger when the authors study the level of corporate leverage, rather than annual changes. Ma (2015) studies cross-market arbitrage by corporations directly. She shows that firms actively trade across both the equity and debt markets in an attempt to arbitrage the market for their own securities. But her evidence of cross market timing, similar to other studies, is annual.

6 Conclusion

Arbitrage is a highly specialized activity. Specialized arbitrageurs ensure that options on IBM stock are priced consistently. Different arbitrageurs ensure that two similar maturity Treasury bonds trade at the same yield. Yet another set of traders ensure that the price of the S&P 500 futures contract does not deviate significantly from the basket of underlying stocks. Even outside of pure arbitrage, the business of money management is highly segmented across markets. The vast majority of mutual funds offer by Fidelity and Vanguard, for example, two of the world's largest asset managers, are in vehicles that specialize in either fixed-income or equity markets.

While specialization brings many benefits, the boundaries of securities markets are tested when there are large shocks to the demand for or supply of an entire asset class. In this paper, we developed a model to describe securities prices when shocks must draw in arbitrageurs from other related asset markets. We use the model to study the process by which capital flows across markets, and how quickly and by what magnitude prices adjust in different markets. Even when a large amount of capital is mobile in the long run, different asset markets need not be fully integrated because market segmentation creates risks for arbitrageurs.

Our model is useful for measuring the degree of effective segmentation between asset markets populated by groups of different specialists. In this paper, we have taken the existence of these market specialists as given. But what determines the degree of asset class specialization in the first place? Why are some asset classes dominated by specialists while others are widely held in the portfolios of generalists? And what determines the boundaries of specialists' expertise and, hence, the fault lines of different asset classes? Answer these question remains an important task for future research.

References

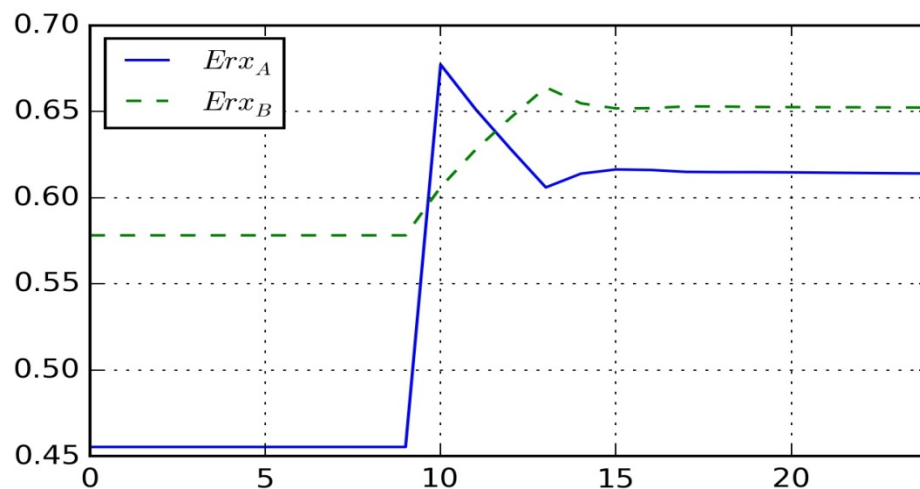
- Bauer, M. and C. Neely, 2013, “International Channels of the Fed’s Unconventional Monetary Policy,” *Federal Reserve Bank of St. Louis* Working Paper 2012-028B.
- Bernanke, B. and K. Kuttner, 2005, “What Explains the Stock Market’s Reaction to Federal Reserve policy?”, *The Journal of Finance*, 60 (3), 1221–1257.
- Campbell, John Y, 1993, “Intertemporal Asset Pricing without Consumption Data,” *American Economic Review* 83, 487-512.
- Campbell, J. Y. and R. J. Shiller, 1988, “Stock Prices, Earnings, and Expected Dividends,” *Journal of Finance* 43, 661-76.
- Campbell, J. Y. and T. Vuolteenaho, 2004, "Bad Beta, Good Beta," *American Economic Review* 94, 1249-1275.
- Chen, H., V. Curdia, and A. Ferrero, 2011, “The Macroeconomic Effects of Large-scale Asset Purchase Programs,” *Federal Reserve Bank of New York* Staff Report No. 527. (Also see <http://www.newyorkfed.org/aboutthefed/staffreviewMay2012.pdf>.)
- D’Amico, S., W. English, D. Lopez-Salido, and E. Nelson, 2012, “The Federal Reserve’s Large-scale Asset Purchase Programs: Rationale and Effects,” Federal Reserve Board.
- 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*, forthcoming.
- DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, 1990a, “Noise Trader Risk in Financial Markets,” *Journal of Political Economy* 98, 703-738.
- Duffie, D., and B. Strulovici, “Capital Mobility and Asset Pricing”, *Econometrica* 2012, Volume 80:2469-2509.
- Duffie, D., “Asset Price Dynamics with Slow-Moving Capital”, *Journal of Finance* 2010, Volume 65: 1238-1268.
- Duffie, D., and K. Singleton, 1999, "Modeling Term Structures of Defaultable Bonds" , *Review of Financial Studies* 1999, Volume 12: 687-720.
- Cochrane, J. H., 2011, “Presidential Address: Discount Rates,” *Journal of Finance* 66, 1047-1108.
- Fawley, B. and C. Neely, 2013, “Four Stories of Quantitative Easing,” Federal Reserve Bank of St. Louis Review, 95 (1), 51–88.
- Fuhrer, J. and G. Olivei, 2011, “The Estimated Macroeconomic Effects of the Federal Reserve’s Large-scale Treasury Purchase Program,” FRBB Public Policy Briefs No. 11-2.
- Gagnon, J., M. Raskin, J. Remache, and B. Sack, 2011a, “The Financial Market Effects of the Federal Reserve’s Large-scale Asset Purchases,” *International Journal of Central Banking* 7 , 3–43.
- Gagnon, J., M. Raskin, J. Remache, and B. Sack, 2011b, “Large-scale Asset Purchases by the Federal Reserve: Did they Work?”, FRBNY Economic Policy Review 17, 41–59.

- Glick, R. and S. Leduc, 2013, "Unconventional Monetary Policy and the Dollar," FRBSF Economic Letter 2013-09.
- Greenwood, R. and D. Vayanos, 2014, "Bond Supply and Excess Bond Returns," *Review of Financial Studies* 27, 663-713.
- Greenwood, R., S.G. Hanson and J.C. Stein, 2010, "A Gap-Filling Theory of Corporate Debt Maturity" *Journal of Finance*.
- Gromb, D. and D. Vayanos, 2010, "Limits of Arbitrage: The state of the Theory," *Annual Review of Financial Economics*, 2, 251-275
- Gromb, Denis and Dimitri Vayanos, 2015, The Dynamics of Financially Constrained Arbitrage, Working paper.
- Grossman, Sanford J & Miller, Merton H, 1988. "Liquidity and Market Structure," *Journal of Finance* 43, 617-37.
- Hamilton, J. and J. 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. G., A.Shleifer, J.C. Stein, and R.W. Vishny, 2015,"Banks as Patient Fixed-Income Investors." forthcoming *Journal of Financial Economics*.
- Hooper, P., T. Slok, and M. Luzzetti, 2013, "Impact of Fed QE on Global Markets," *Deutsche Bank Global Economics Perspectives*, May, 23, 2013.
- Froot, K.A., and P. G. J. O'Connell, 1999 "The Pricing of U.S. Catastrophe Reinsurance," NBER Chapters, in: The Financing of Catastrophe Risk, pages 195-232 National Bureau of Economic Research, Inc.
- Krishnamurthy, A., O. Vigneron and X. Gabaix, "Limits of Arbitrage: Theory and Evidence from the Mortgage-Backed Securities Market", *Journal of Finance*, 62(2), April 2007, p. 557-595.
- Hanson, Samuel G., 2014, "Mortgage Convexity," *Journal of Financial Economics*, forthcoming.
- 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*, Fall 2011, 215-265.
- Krishnamurthy, A. and A. Vissing-Jorgensen, 2012, "The aggregate demand for treasury debt," *Journal of Political Economy*, 120, 233-267.
- Ma, Y., 2015, "Non-financial Firms as Arbitrageurs in Their Own Securities", Working paper.
- Mamaysky, Harry, 2014, "The Time Horizon of Price Responses to Quantitative Easing", Working Paper.
- Merton, Robert C, 1987. "A Simple Model of Capital Market Equilibrium with Incomplete Information", *Journal of Finance* 42, 483-510.
- Modigliani, F. and R. Sutch, 1966, "Innovations in Interest Rate Policy," *American Economic Review* 56, 178-197.

- Modigliani, F. and R. Sutch, 1967, “Debt Management and the Term Structure of Interest Rates: An Empirical Analysis of Recent Experience,” *Journal of Political Economy* 75 , 569–589.
- Neely, C., 2012, “The Large-scale Asset Purchases had Large International Effects,” *Federal Reserve Bank of Saint Louis Working Paper*.
- Pedersen, L.H., M. Mitchell , and T. Pulvino, 2007. “Slow Moving Capital,” *American Economic Review* 97, 215-220.
- Rosa, C., 2012, “How ‘Unconventional’ are Large-scale Asset Purchases: The Impact of Monetary Policy on Asset Prices,” FRBNY Staff Report No. 560.
- Shleifer, A., and R. W. Vishny, 1997. “The Limits of Arbitrage,” *Journal of Finance* 52, 35-55.
- Stehn, S. and N. Weisberger, 2013, “The Equity Market Response to Fed Policy,” US Economics Analyst, Goldman Sachs, Issue 13/20.
- 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*, Spring 2011, 151–188.
- Williams, J., 2012, “The Federal Reserve’s Unconventional Policies,” FRBSF Economic Letter, 2012-34.
- Vayanos, D. and J. Vila, 2009, “A Preferred-Habitat Model of the Term Structure of Interest Rates,” NBER Working Paper No. 15487.

Figure 1: Price-impact of an unanticipated supply shock. This figure shows the impact on annual bond risk premia and bond yields in markets A and B of an unexpected shock that increases the supply of asset A by 50% in period 10. Panel A shows the evolution of annual bond risk premia in market A , $E_t[rx_{A,t+1}]$, and market B , $E_t[rx_{B,t+1}]$, over time. Panel B shows the evolution of bond yields in market A , $y_{A,t}$, and market B , $y_{B,t}$, over time.

Panel A: Annual bond risk premia



Panel B: Bond yields

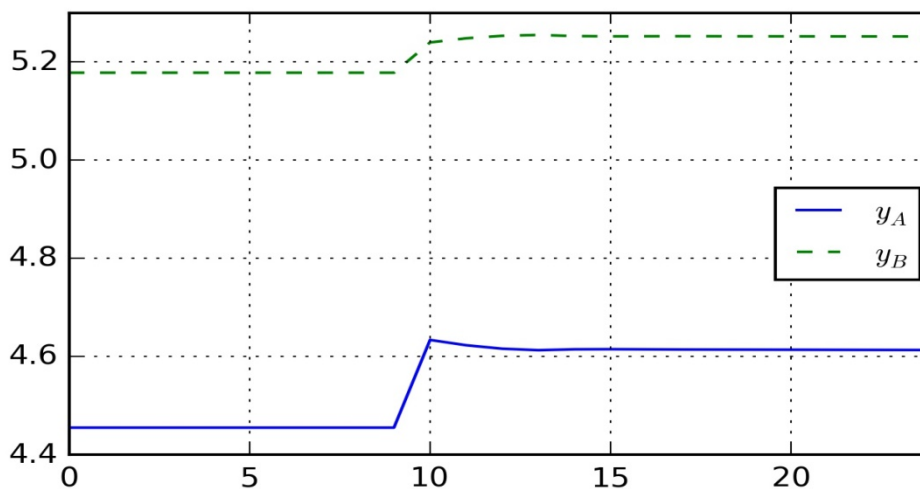
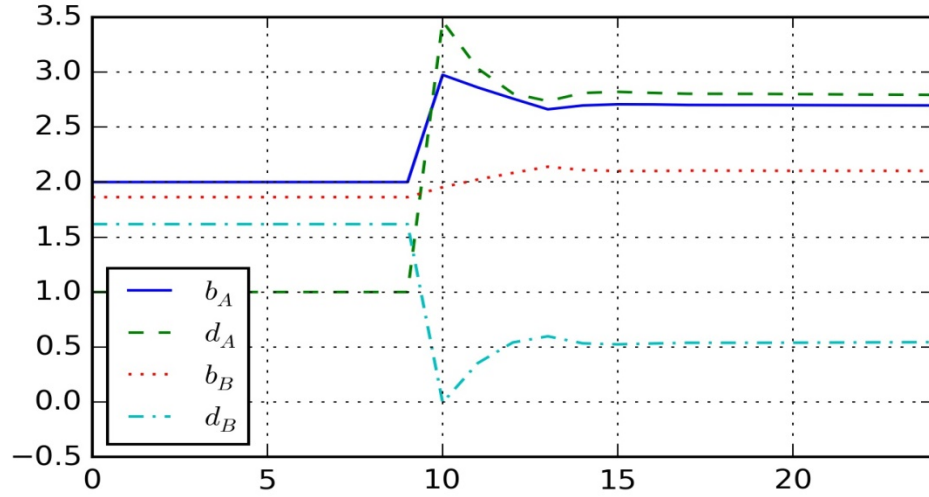


Figure 2: Portfolio adjustments in response to an unanticipated supply shock. This figure shows the impact on investor positions and active asset supplies of an unexpected shock that increases the supply of asset A by 50% in period 10. Panel A shows the evolution of specialists holdings in markets A and B ($b_{A,t}$ and $b_{B,t}$) as well as the positions of active generalists ($d_{A,t}$ and $d_{B,t}$). Panel B shows the evolution of the “active supplies” of assets A and B . The active supply of A is $s_{A,t} - (1 - p_A - p_B)k^{-1} \sum_{i=1}^{k-1} d_{A,t-i}$ and the active supply of B is defined analogously.

Panel A: Specialist holdings and positions of active generalists



Panel B: Active asset supply

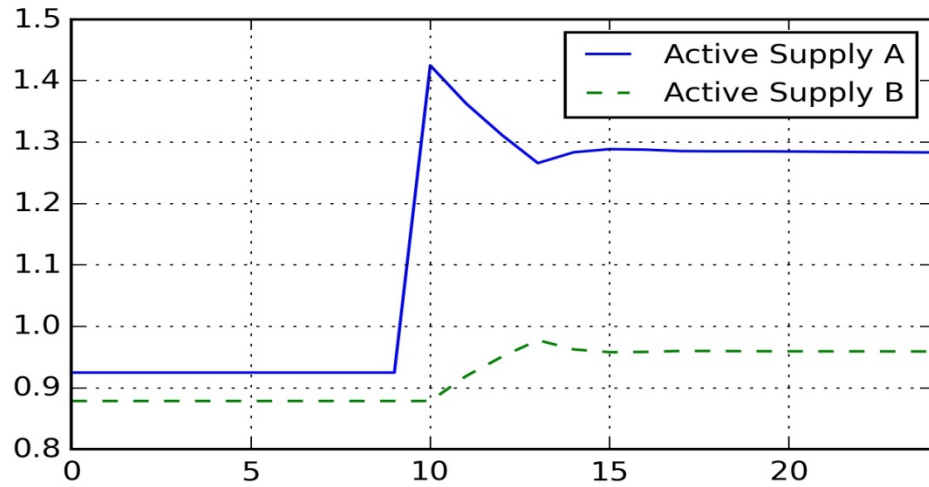


Figure 3: Yield spread impact of an unanticipated supply shock. This figure shows the impact on the yield spread between asset B and asset A , $y_{B,t} - y_{A,t}$, of an unexpected shock that increases the supply of asset A by 50% in period 10.

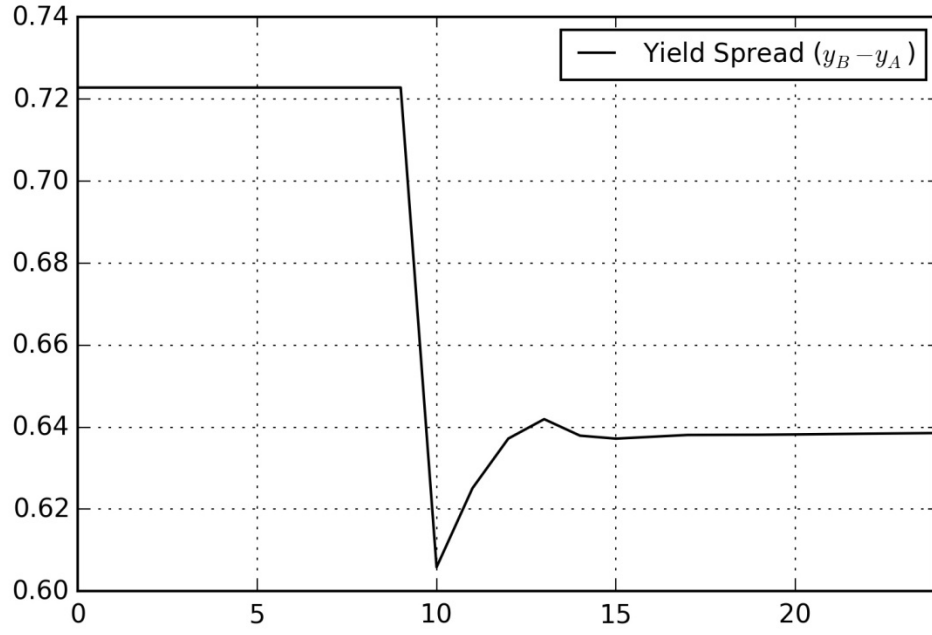
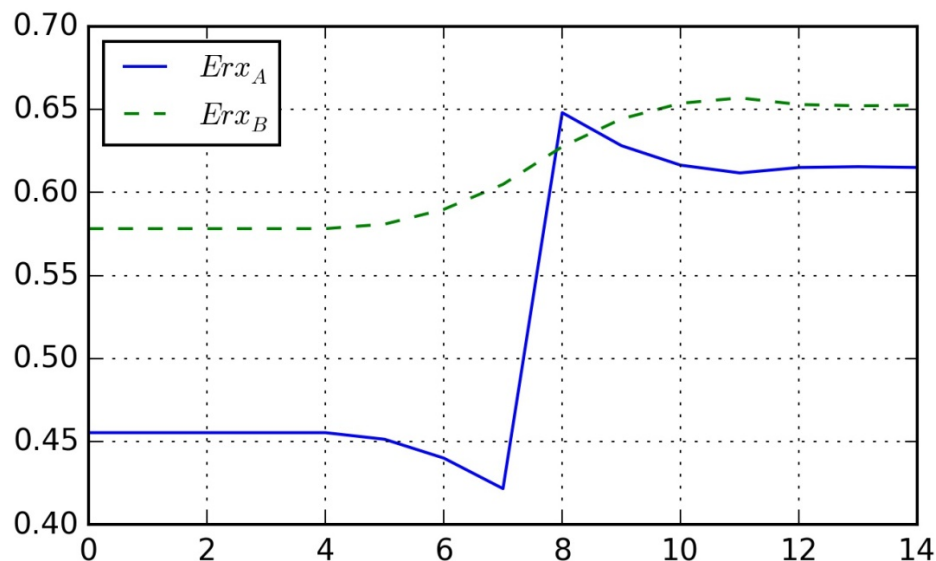


Figure 4: Price-impact of an anticipated supply shock. This figure shows the impact on annual bond risk premia and bond yields in markets A and B of an anticipated supply shock: there is an announcement at time 5 that the supply of asset A will increase by 50% at time 8. Panel A shows the evolution of annual bond risk premia in market A , $E_t[rx_{A,t+1}]$, and market B , $E_t[rx_{B,t+1}]$, over time. Panel B shows the evolution of bond yields in market A , $y_{A,t}$ and market B , $y_{B,t}$ over time.

Panel A: Annual bond risk premia



Panel B: Bond yields

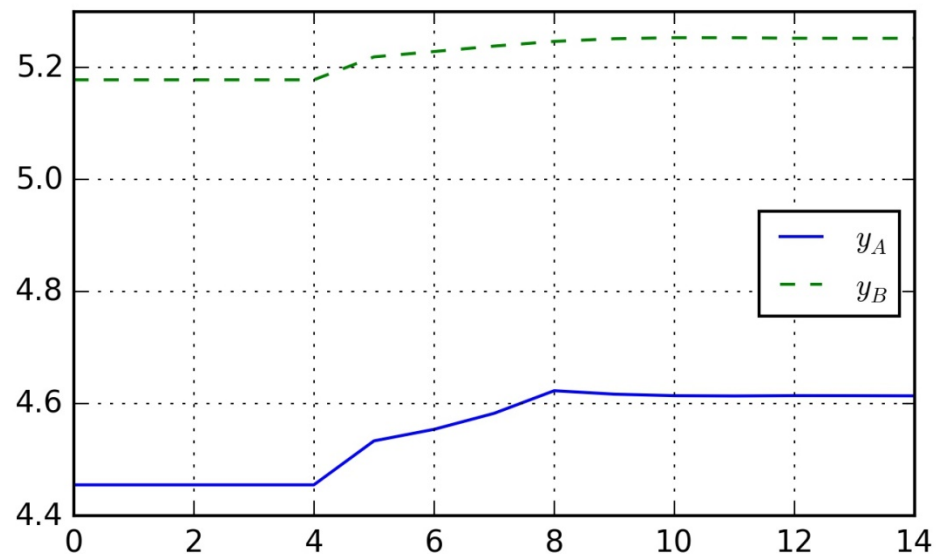
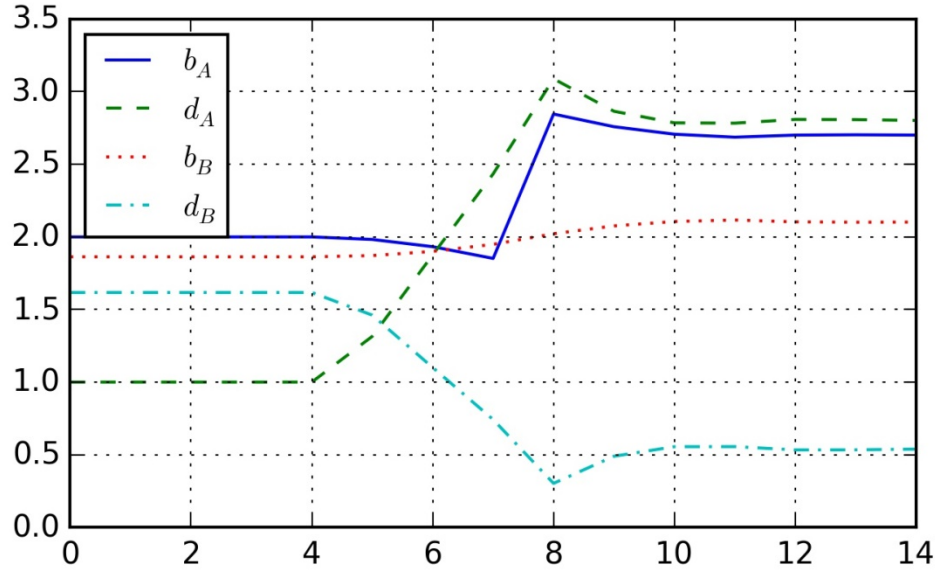


Figure 5: Portfolio adjustments in response to an anticipated supply shock. This figure shows the impact on investor positions and active asset supplies of an anticipated supply shock: there is an announcement at time 5 that the supply of asset A will increase by 50% at time 8. Panel A shows the evolution of specialists holdings in markets A and B ($b_{A,t}$ and $b_{B,t}$) as well as the positions of active generalists ($d_{A,t}$ and $d_{B,t}$). Panel B shows the evolution of the “active supplies” of assets A and B . The active supply of A is $s_{A,t} - (1 - p_A - p_B)k^{-1} \sum_{i=1}^{k-1} d_{A,t-i}$ and the active supply of B is defined analogously.

Panel A: Specialist holdings and positions of active generalists



Panel B: Active asset supply

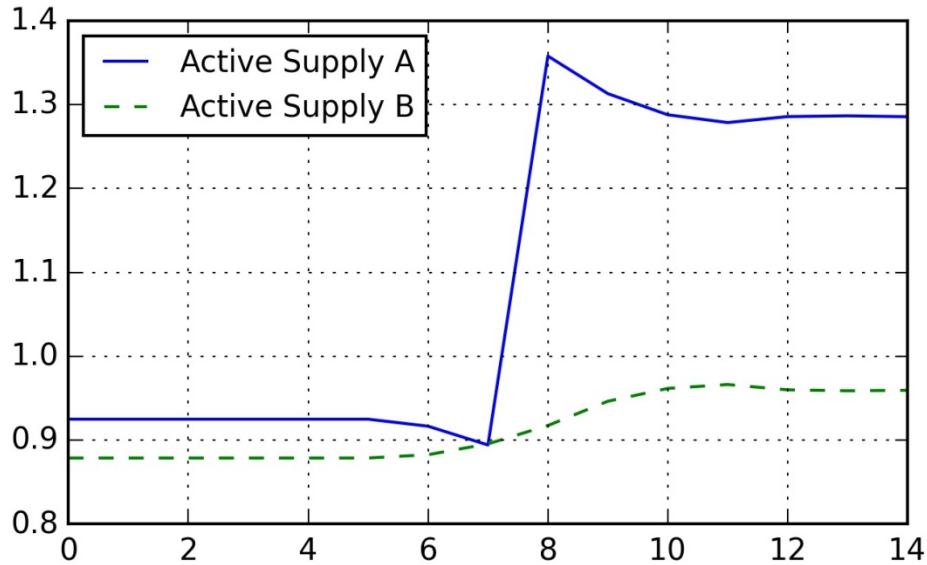
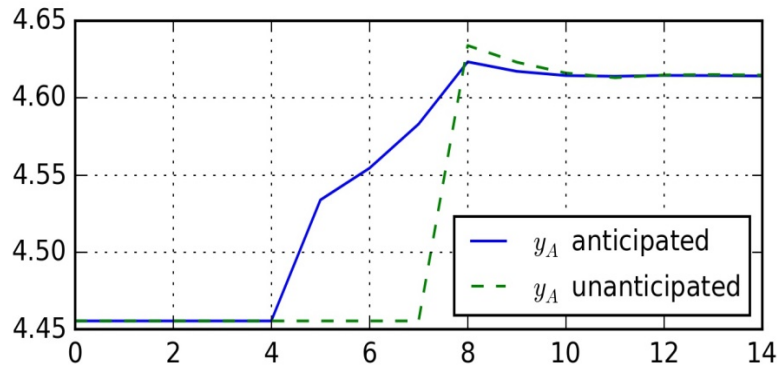
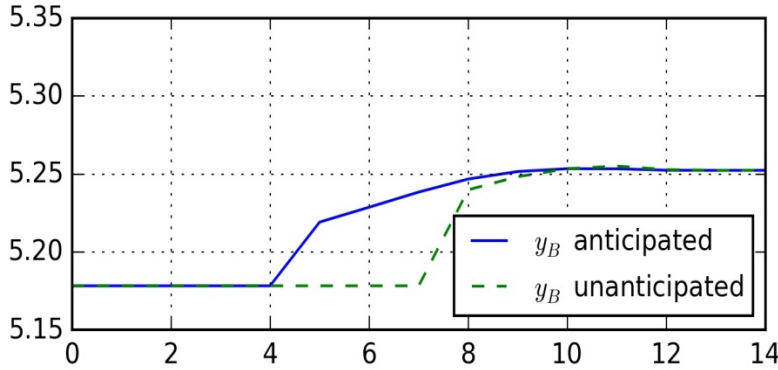


Figure 6: Comparison of anticipated and unanticipated supply shocks. This figure compares the price impact of anticipated and unanticipated supply shocks. The yield dynamics are the same as depicted in Figure 1.B and Figure 4.B and are presented here for easy of comparison. In both cases, the supply of A asset increases by 50%. The solid lines show the yield dynamics when the supply shock is pre-announced in period 5 and arrives in period 8. The dashed lines show the yield dynamics when the supply shock unexpectedly arrives in period 5 without any prior announcement. Panel A shows yields in market A , Panel B shows yields in market B , and Panel C shows the yields spread, $y_B - y_A$.

Panel A: Yield of bond A



Panel B: Yield of bond B



Panel C: Yield spread between bonds B and A

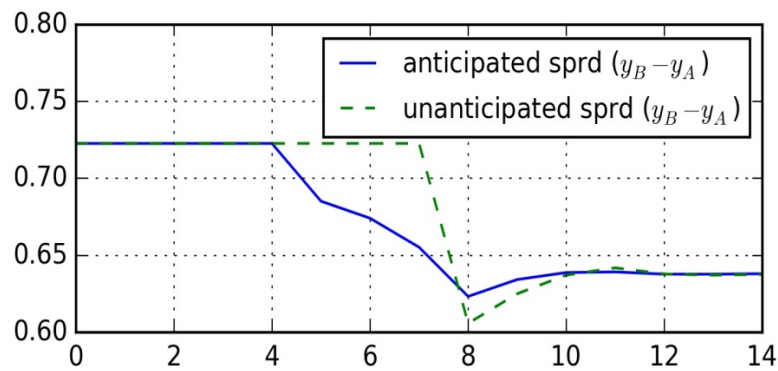
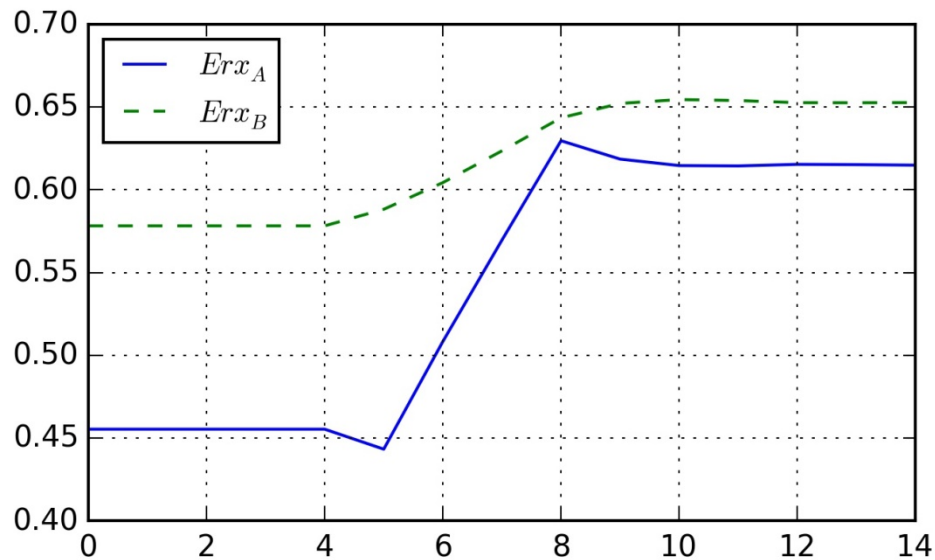


Figure 7: Price-impact of an anticipated gradual supply shock. This figure shows the impact on annual bond risk premia and bond yields in markets A and B of an anticipated gradual supply shock: there is an announcement at time 5 that the supply of asset A will increase by 50% with the increase spread out equally between time 6 and time 8. Panel A shows the evolution of annual bond risk premia in market A , $E_t[rx_{A,t+1}]$, and market B , $E_t[rx_{B,t+1}]$, over time. Panel B shows the evolution of bond yields in market A , $y_{A,t}$ and market B , $y_{B,t}$ over time.

Panel A: Annual bond risk premia



Panel B: Bond yields

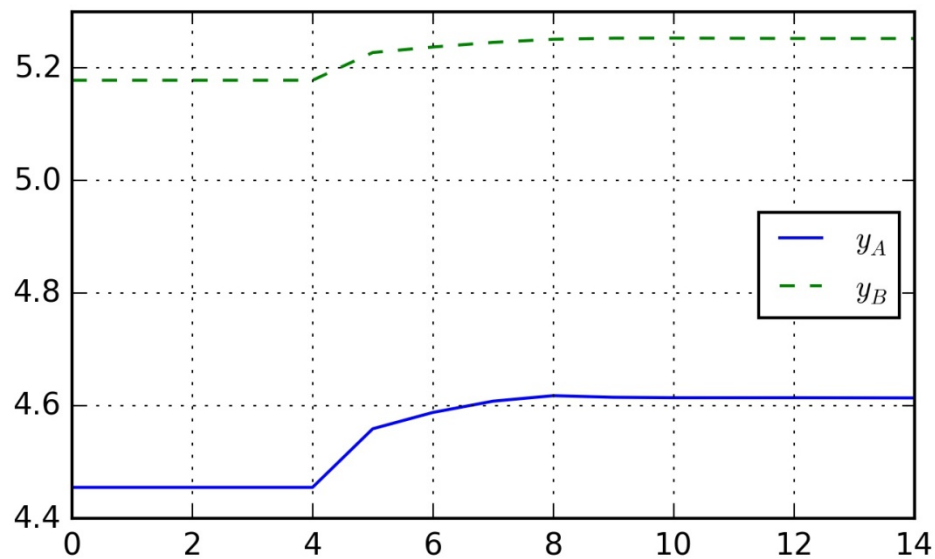
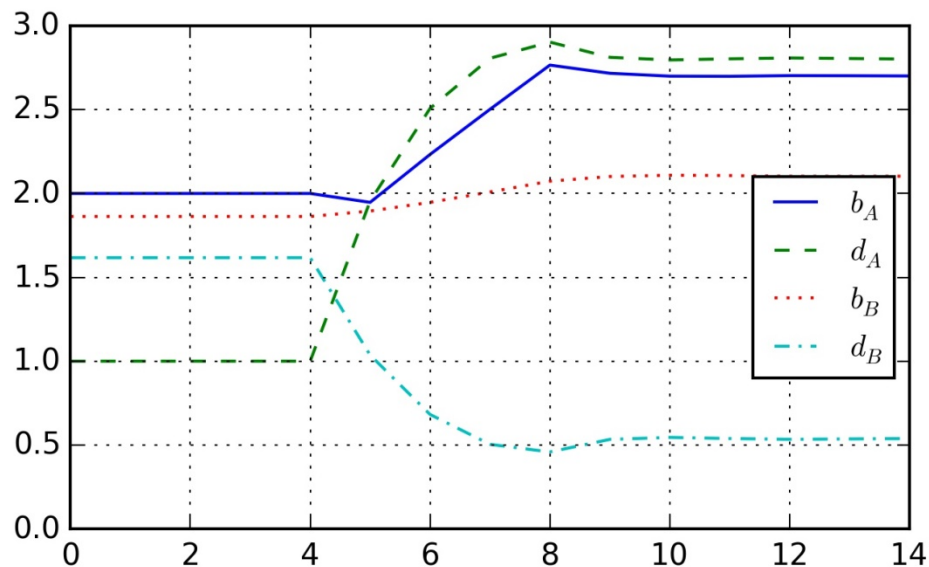


Figure 8: Portfolio adjustments in response to an anticipated gradual supply shock. This figure shows the impact on investor positions and active asset supplies of an anticipated gradual supply shock: there is an announcement at time 5 that the supply of asset A will increase by 50% with the increase spread out equally between time 6 and time 8. Panel A shows the evolution of specialists holdings in markets A and B ($b_{A,t}$ and $b_{B,t}$) as well as the positions of active generalists ($d_{A,t}$ and $d_{B,t}$). Panel B shows the evolution of the “active supplies” of assets A and B . The active supply of A is $s_{A,t} - (1 - p_A - p_B)k^{-1} \sum_{i=1}^{k-1} d_{A,t-i}$ and the active supply of B is defined analogously. Panel A: Specialist holdings and positions of active generalists



Panel B: Active asset supply

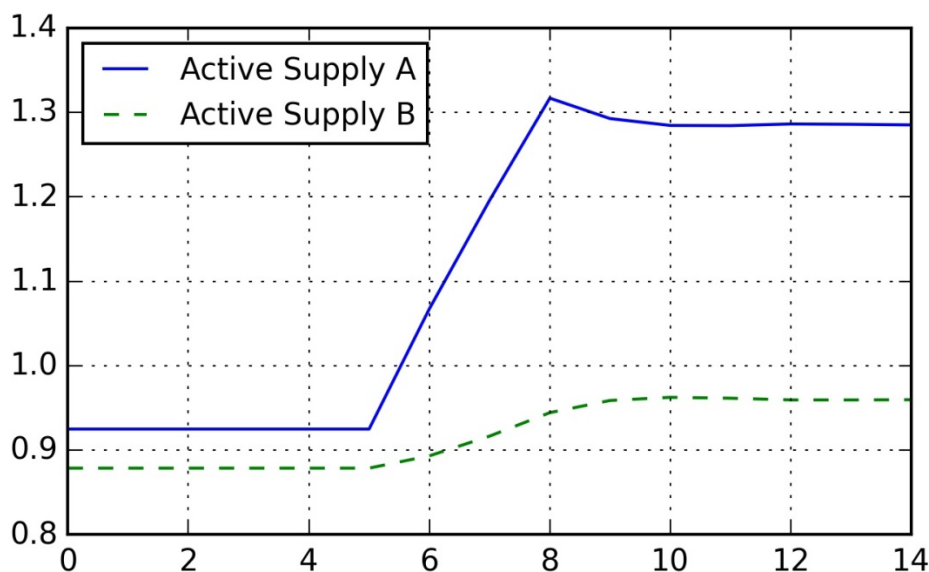


Figure 9: Event studies confidence interval. The yields of markets A and B and the 95% confidence intervals of the respective market yield are shown. An unexpected shock that increases the supply of asset A by 50% is delivered in period 10. The following parameters are used: $\tau = 0.5$, $\sigma_{sA} = \sigma_{sB} = 0$, $\sigma_r = 1\%$, $\rho_r = 0.95$, $\sigma_z = 0.2\%$, $\bar{z} = 0.2\%$. All other parameters are the same as the values listed in Table 1.

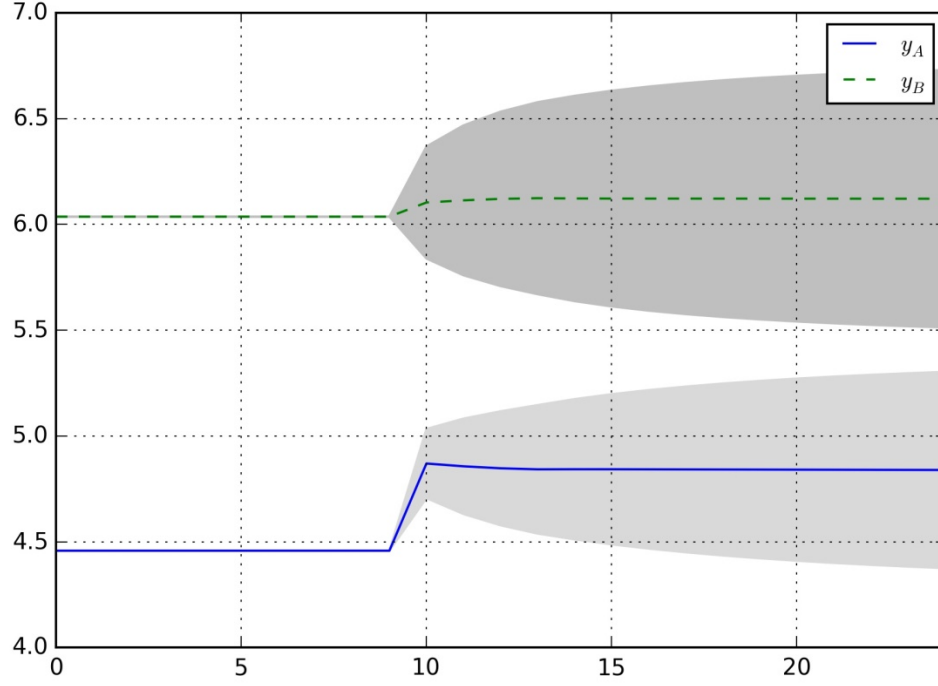


Table 1: Benchmark model parameters: This table presents the benchmark model parameters that we used throughout our numerical exercises. Parameters are calibrated under the assumption that market A is the U.S. Treasury market and that market B is the investment grade corporate bond market. We calibrate the model so that one period corresponds to one year. The total average supply of assets in each market is normalized to be one unit. Further details on the calibration are provided in Appendix B.

Parameter	Description	Value
p_A, p_B	Percentage of investors that are specialists in A and B	45%
k	Number of periods between generalist portfolio rebalancing	4
$\bar{r}$	Average short-term riskless rate	4%
σ_r	Volatility of annual shocks to short-term riskless rate	13%
ρ_r	Annual persistence of short-term riskless rate	0.85
$\bar{z}$	Expected default losses per annum	0.60%
σ_z	Volatility of annual shocks to default losses	0.60%
ρ_z	Annual persistence of default losses	0.85
$\bar{s}_A, \bar{s}_B$	Average asset supplies	1
$\sigma_{s_A}, \sigma_{s_B}$	Volatility of annual supply shocks	0.60%
ρ_{s_A}, ρ_{s_B}	Annual persistence of supply shocks	0.999
D_A, D_B	Macaulay duration in years (implies $\theta_A = \theta_B = 0.8$)	5 years
τ	Investor risk tolerance	50

Table 2: Comparative statics: This table shows how the price impact of the same supply shock---an unanticipated shock that doubles the supply in market A ---varies as we change key model parameters model one at a time. All other parameters are held constant at the values listed in Table 1. For a given set of model parameters, we summarize the impact of the supply shock on both the A and B markets by listing the yields and expected annual returns in (i) the period before the shock arrives (labeled as "pre-shock"), (ii) the period when the shock arrives (labeled as "short-run"), and (iii) in 2k periods after the shock arrives (labeled as "long-run"). Finally, we report the degree to which bond yields over- or underreact as the difference between the short-run change and the long-run change, expressed as a percentage of the long-run change

$$\%Over-Reaction(y) = \frac{(y_t - y_{t-1}) - (y_{t+2k} - y_{t-1})}{(y_{t+2k} - y_{t-1})}.$$

Our measure of over-reaction for risk premia is defined analogously.

		Market A								Market B							
		Risk premia, $E_t[rx_{A,t+1}]$				Yields, $y_{A,t}$				Risk premia, $E_t[rx_{B,t+1}]$				Yields, $y_{B,t}$			
		Pre shock	Short run	Long run	Over react	Pre shock	Short run	Long run	Over react	Pre shock	Short run	Long run	Over react	Pre shock	Short run	Long run	Over react
Supply shock hits market A																	
Base case		0.46	0.22	0.16	38%	4.46	0.18	0.16	12%	0.58	0.03	0.07	-63%	5.18	0.06	0.07	-18%
More risk tolerant	$\tau = 60$	0.37	0.18	0.13	38%	4.37	0.15	0.13	12%	0.47	0.02	0.06	-63%	5.07	0.05	0.06	-18%
No Generalists	$p_A = p_B = 0.5$	0.47	0.23	0.23	0%	4.47	0.23	0.23	0%	0.66	0.00	0.00	NA	5.26	0.00	0.00	NA
More Generalists	$p_A = p_B = 0.24$	0.40	0.30	0.12	143%	4.40	0.17	0.12	38%	0.49	0.11	0.11	1%	5.09	0.12	0.11	8%
More B specialists	$p_A = 0.3, p_B = 0.6$	0.55	0.33	0.20	62%	4.55	0.24	0.20	18%	0.51	0.03	0.08	-62%	5.11	0.06	0.08	-17%
Fast-adjusting Generalists	$k = 2$	0.46	0.20	0.16	22%	4.46	0.17	0.16	4%	0.58	0.05	0.07	-36%	5.18	0.07	0.07	-6%
Slow-adjusting Generalists	$k = 6$	0.45	0.23	0.16	45%	4.45	0.19	0.16	18%	0.58	0.02	0.08	-75%	5.18	0.05	0.07	-28%
Supply shock hits market B																	
Base case		0.46	0.03	0.07	-63%	4.46	0.06	0.07	-18%	0.58	0.30	0.22	39%	5.18	0.24	0.22	12%
More default risk	$\sigma_z = 1.2\%$	0.48	0.02	0.06	-69%	4.48	0.04	0.06	-21%	1.14	0.65	0.51	27%	5.74	0.56	0.51	9%