Rate-Amplifying Investor Demand and the Excess Sensitivity of Long-Term Interest Rates*

SAMUEL G. HANSON
Harvard Business School

DAVID O. LUCCA
Federal Reserve Bank of New York

JONATHAN H. WRIGHT
Johns Hopkins University

December 22, 2020

Abstract

Long-term nominal interest rates are known to be highly sensitive to high-frequency (daily or monthly) movements in short-term rates. We find that, since 2000, this high-frequency sensitivity has grown even stronger in U.S. data. By contrast, the association between low-frequency changes (at 6- or 12-month horizons) in long- and short-term rates, which was also strong before 2000, has weakened substantially. This puzzling post-2000 pattern arises because increases in short rates temporarily raise the term premium component of long-term yields, leading long rates to *temporarily* overreact to changes in short rates. The frequency-dependent excess sensitivity of long-term rates that we observe in recent years is best understood using a model in which (i) declines in short rates trigger "rate-amplifying" shifts in investor demand for long-term bonds and (ii) the arbitrage response to these demand shifts is slow. We study, both theoretically and empirically, how such rate-amplifying demand can be traced to mortgage refinancing activity, investors who overextrapolate recent changes in short rates, and investors who "reach for yield" when short rates fall. We discuss the implications of our findings for the validity of event-study methodologies and for the transmission of monetary policy.

*Earlier versions of this paper circulated under the title "The Excess Sensitivity of Long-Term Rates: A Tale of Two Frequencies" and "Interest Rate Conundrums in the Twenty First Century" We thank John Campbell, Anna Cieslak, Gabriel Chodorow-Reich, Richard Crump, Thomas Eisenbach, Robin Greenwood, Andrei Shleifer, Eric Swanson, Jeremy Stein, and Adi Sunderam as well as seminar participants at the 2018 AEA meetings, Harvard University, the Federal Reserve Bank of New York, Johns Hopkins University, Northwestern Kellogg, the Society for Computational Economics 2017 International Conference, University of Georgia, and University of Southern California for helpful comments. Hanson gratefully acknowledges funding from the Division of Research at Harvard Business School. All errors are our sole responsibility. The views expressed here are the authors' and are not representative of the views of the Federal Reserve Bank of New York or of the Federal Reserve System. Emails: shanson@hbs.edu; david.lucca@ny.frb.org; wrightj@jhu.edu.

The sensitivity of long-term interest rates to movements in short-term rates is a central feature of the term structure and is thought to play a crucial role in the transmission of monetary policy. Short-term nominal interest rates are determined by current monetary policy and its near-term expected path. Shocks to monetary policy and the macroeconomy are generally thought to be short-lived, so long-term rates should not be highly sensitive to changes in short rates if the expectations hypothesis holds (Shiller, 1979). However, a large literature demonstrates that long-term nominal rates are far more sensitive to high-frequency changes in short rates than is predicted by this standard view (Shiller et al., 1983; Cochrane and Piazzesi, 2002; Gürkaynak et al., 2005; Giglio and Kelly, 2018), implying either the existence of surprisingly persistent shocks to short rates and/or "excess sensitivity" of long-term rates when judged relative to an expectations-hypothesis baseline. Despite its importance for monetary policy, the deeper forces underpinning this puzzling degree of high-frequency sensitivity—and the extent to which it has evolved over time—remain poorly understood.

This paper provides new evidence that, in the past two decades, the sensitivity of long-term rates has grown even stronger at high-frequencies, but has weakened substantially at lower frequencies. This puzzling post-2000 pattern arises because increases in short rates temporarily raise the term premium component of long-term bond yields, leading long rates to temporarily overreact to changes in short rates. We argue that this short-lived excess sensitivity arises because changes in short-term rates induce non-standard shifts in investor demand for long-term bonds—stemming from a combination of institutional and psychological factors—that amplify the impact of changes in short rates relative to the expectations-hypothesis. We study, both theoretically and empirically, how such "rate-amplifying" demand shocks can arise through three distinct channels: mortgage refinancing activity, investors who overextrapolate recent changes in short-term rates, and investors who "reach for yield" when short rates decline.

We begin by documenting an important and previously unrecognized fact about the term structure of nominal interest rates: since 2000, the sensitivity of long-term yields to changes in short-term rates has become highly frequency-dependent. Prior to 2000, the sensitivity of long yields to changes in short rates was similarly strong at high and low frequencies. However, since 2000, the association between high-frequency changes (at daily or 1-month horizons) in long- and short-term rates has strengthened even further. By contrast, the relationship between low-frequency changes (at 6- or 12-month horizons) has weakened substantially since 2000.[1] Concretely, between 1971 and 1999, a daily regression of changes in 10-year U.S. Treasury yields on changes in 1-year yields delivers a coefficient of 0.56; and the analogous regression using 12-month changes gives nearly the same coefficient. But, between 2000 and 2019, the coefficient from the daily regression jumps to 0.87, while the coefficient from the corresponding 12-month regression drops to just 0.23. Figure 1, which plots the sensitivity of 10-year yields to changes in 1-year yields as a function of horizon in both the pre-2000 and post-2000 samples, summarizes this key finding. This pattern is not specific to the U.S.: we find similar results for Canada, Germany, and the U.K.

What explains this puzzling post-2000 tendency of short- and long-term rates to move together at high frequencies but not at low frequencies? As a matter of statistical description, we show that this pattern arises because, all else equal, past increases in short rates predict a subsequent flattening of the yield curve—and subsequent declines in long-term yields and forward rates—in the post-2000 data. More

[1] We do not mean to argue that there was a *discrete* change in the underlying data-generating process around 2000. Instead, our reading of the evidence is that the underlying data-generating process has changed *gradually* over time.

formally, yield-curve dynamics have become "path-dependent" or "non-Markovian" since 2000. Loosely speaking, to form the best forecast of future bond yields and returns, it is not enough to know the current shape of the yield curve: one also needs to know how the yield curve has shifted in recent months.

These predictable reversals in long-term rates are linked to a new form of short-lived bond return predictability: since 2000, the expected returns on long-term bonds (in excess of those on short-term bonds) are *temporarily* elevated following increases in short rates. Thus, relative to an expectations-hypothesis baseline, long rates temporarily overreact to changes in short rates, exhibiting what Mankiw and Summers (1984) dubbed "excess sensitivity." For instance, in the post-2000 data, we estimate that 10-year yields rise by 66 basis points in response to a 100 bps monthly increase in 1-year yields. Over the next 6 months, 10-year yields are expected to fall by 36 bps, reversing over half of the initial response.

What deeper forces underpin the evolving sensitivity of long-term yields to movements in short rates? Gürkaynak et al. (2005) note that the strong sensitivity of long-term nominal rates could be consistent with the expectations hypothesis if one adopts the view that long-run inflation expectations are unanchored and are continuously being updated in light of incoming news—i.e., if one allows for highly persistent shocks to expected inflation. We argue that the narrative in Gürkaynak et al. (2005) is a good explanation for the high degree of sensitivity observed prior to 2000. Indeed, consistent with the expectations-hypothesis logic of their explanation, in the pre-2000 data, we find no evidence that the reaction of long yields to movements in short rates tends to reverse predictably.

However, in the post-2000 period, the strong high-frequency sensitivity of long-term nominal rates primarily reflects the sensitivity of long-term *real* rates to nominal short rates, rather than the sensitivity of break-even inflation (Beechey and Wright, 2009; Hanson and Stein, 2015; Abrahams et al., 2016). To the extent that one shares the widespread view that expected future real rates at distant horizons do not fluctuate meaningfully at high frequencies (see Gürkaynak et al., 2005), this makes it hard to square the strong high-frequency sensitivity of long rates since 2000 with expectations-hypothesis logic. To resolve this puzzle, Hanson and Stein (2015) argue that the strong post-2000 sensitivity works through the term premium component of long-term yields: shocks to short rates move term premia in the same direction. Consistent with this view, we find strong evidence that the reaction of long-term yields to movements in short rates tends to predictably reverse in the post-2000 data, giving rise to short-lived shifts in the expected returns to holding long-term bonds.

So, how can we best understand our key empirical finding that, in recent decades, the sensitivity of long-term yields to changes in short rates declines steeply in the horizon over which these changes are computed? Because this finding reflects a form of short-lived bond return predictability, the most natural explanations involve temporary supply-and-demand imbalances in less-than-perfect financial markets (Grossman and Miller (1988) and Duffie (2010)). Therefore, we develop a model that emphasizes the role of what we call "rate-amplifying" shocks to the supply-and-demand for long-term bonds.

In our model, risk-averse bond arbitrageurs can either invest in short- or long-term nominal bonds. While monetary policy pins down the interest rate on short-term nominal bonds, long-term bonds are available in a net supply that varies randomly over time. This net supply, which arbitrageurs must hold in equilibrium, equals the gross supply of long-term bonds net of the amount inelastically demanded by other, non-arbitrageur investors. To induce risk-averse arbitrageurs to absorb an increase in net supply of long-term bonds, the expected return on long-term bonds in excess of that on short-term bonds must rise as in Greenwood and Vayanos (2014) and Vayanos and Vila (2020), thereby lifting the term premium component of long-term bonds yields. The key friction in these models—which has proved invaluable in understanding the effect of Quantitative Easing policies—stems from the limited risk-bearing capacity of the specialized fixed-income arbitrageurs who must absorb shocks to the net supply of long-term bonds.

Consistent with Gurkaynak et al. (2005), we assume there was a large persistent component of short-term nominal rates before 2000, reflecting shocks to trend inflation as in Stock and Watson (2007). The existence of this highly persistent component in combination with expectations-hypothesis logic, explains the strong sensitivity of long rates at both high and low frequencies before 2000. In the post-2000 period, the volatility of this persistent component of short rates has dropped sharply. From an expectations-hypothesis perspective, this should have reduced the sensitivity of long rates at all frequencies. In the data, this occurs at low frequencies, but we see greater-than-ever sensitivity at high frequencies.

The key contribution of our model is to explain how such frequency-dependent sensitivity arises in the post-2000 data. Our explanation rests on two key ingredients: (i) "rate-amplifying" shifts in the supply or demand for long-term bonds that move term premia in the same direction as short-term rates and (ii) and slow-moving arbitrage capital. The first key ingredient is our assumption that shocks to the net supply of long-term bonds are *positively* correlated with shocks to short rates. This can either be because increases in short rates are associated with increases in the gross supply of long-term bonds or with non-standard reductions in the demand of other, non-arbitrageur investors for long-term bonds. This assumption implies that increases in short rates are associated with increases in the term premium component of long-term rates, generating "excess sensitivity" relative to the expectations hypothesis. This reduced-form assumption is consistent with several distinct amplification mechanisms that we detail below—each rooted in well-known institutional frictions and facets of investor psychology—that have arguably grown in importance recent decades.[2]

The second key ingredient is that arbitrage capital is slow-moving as in Duffie (2010). As a result, these rate-amplifying demand shocks encounter a short-run arbitrage demand curve that is steeper than the long-run arbitrage demand curve, generating a short-lived imbalance in the market for long-term bonds. This slow-moving capital dynamic implies that the shifts in term premia triggered by movements in short rates are transitory. As a result, the excess sensitivity of long rates is greatest when measured at high frequencies. Furthermore, we show that frequency-dependent excess sensitivity is most pronounced when the underlying rate-amplifying demand shocks are themselves short-lived in nature. In summary, the combination of rate-amplifying demand shocks and slow-moving arbitrage capital enables our model to match the frequency-dependent sensitivity of long rates observed since 2000.

What are the key source of rate-amplifying demand that together explain why long-term yields temporarily overreact to changes in short rates in the recent U.S. data? We explore three rate-amplification channels that may help explain why increases in short rates to trigger temporary supply-and-demand balances in the market for long-term bonds: (i) shifts in the effective gross supply of long-term bonds due to mortgage refinancing waves (Hanson, 2014; Malkhozov et al., 2016), (ii) shifts in the demand for long-term bonds from biased investors who overextrapolate recent changes in short rates (Giglio and Kelly, 2018; D'Arienzo, 2020), and (iii) shifts in the demand from investors who "reach for yield" when short

[2]Where it creates no confusion, we will simply refer to this collection of mechanisms as "rate-amplifying demand shocks" even though it is most natural to think of some of these specific mechanism as operating on the supply-side of the market.

rates fall (Hanson and Stein, 2015). For each channel, we first show how it can be used to microfound rate-amplifying shocks to the net supply of long-term bonds similar to those we previously assumed in reduced-form. Next, we discuss why the strength of each channel may have grown in recent decades: the key underlying trend here is the increasing financialization of interest-rate risk. Finally, by looking at the relationship between bond yields and different financial quantities, we empirically assess the extent to which each channel contributes to the frequency-dependent sensitivity of long-term rates we observe since 2000. Given the difficulties inherent in precise attribution, we believe that the primary contribution of these empirical exercises is to rule in the general class of rate-amplifying mechanisms we emphasize. That said, we find evidence that mortgage refinancing and investor overextrapolation both help explain why long yields rates have temporarily overreacted to short rates since 2000 in the U.S. By contrast, we find less evidence that reaching-for-yield plays an important role in driving our key empirical findings.

This paper contributes to the vast literature demonstrating that, contrary to the expectations hypothesis, the expected excess returns on long-term bonds vary meaningfully over time. Seminal contributions here include Fama and Bliss (1987), Campbell and Shiller (1991), and Cochrane and Piazzesi (2005). In contrast to most of the existing literature on bond return predictability—which focuses on business-cycle frequency variation in expected bond excess returns, our empirical findings point to a new, short-lived form of bond return predictability that has emerged in recent decades.

Our findings have important implications for how economists should interpret event-study evidence based on high-frequency changes in long-term bond yields. Macroeconomic news—including news about monetary policy—comes out in a lumpy manner, and the short-run change in long-term yields around news announcements is often used as a measure of the expected longer-run impact of news shocks. Nakamura and Steinsson (2018) is a prominent recent example of this increasingly popular approach to identification in macroeconomics.[3] However, if, as we show, some of the impact of a news shock on long-term yields tends to wear off quickly over time, then a shock's short- and long-run impact will be quite different. And, the event-study approach will necessarily capture only the short-run impact. For instance, it is common for news announcements to cause large jumps in 10-year forward rates, but we show that a large portion of these jumps are due to *transient* shifts in term premia. As a result, event-study methodologies are likely to provide biased estimated of the longer-run impact of news on long-term yields.

Our results also have implications for monetary policy transmission. In the textbook New Keynesian view (Gali, 2008), the central bank adjusts short-term nominal rates. This affects long-term rates via the expectations hypothesis, which in turn influences aggregate demand. Stein (2013) points out that the excess sensitivity of long-term yields—whereby shocks to short rates move term premia in the same direction—should strengthen the effects of monetary policy relative to the textbook view. Stein (2013) refers to this as the "recruitment" channel of monetary transmission. We find that the behavior of interest rates does not conform to the textbook New Keynesian view in which term premia are constant. Nonetheless, our findings suggest that the recruitment channel may not be as strong as Stein (2013) speculates since a portion of the resulting shifts in term premia are transitory and, thus, likely to have only modest effects on aggregate demand. We do not argue that there is no recruitment channel, just that it is smaller than one might conclude based on the high-frequency response of term premia to policy

[3]Earlier papers examining the high-frequency response of long-term rates to news about monetary policy include Evans and Marshall (1998), Kuttner (2001), and Cochrane and Piazzesi (2002).

shocks documented in Hanson and Stein (2015), Gertler and Karadi (2015), and Gilchrist et al. (2015).

The plan for the paper is as follows. In Section 1, we document our key stylized facts about the changing high- and low-frequency sensitivity of long-term interest rates. In Section 2, we show that past increases in short rates predict a future reversals in long-term yields in the post-2000 data, reflecting a new form of bond return predictability. Section 3 develops the economic modelling framework that we use to interpret our findings. We build on this framework in Section 4 where we explicitly model three specific rate-amplification mechanisms—mortgage refinancing waves, investor overextrapolation, and investor reaching-for-yield—and then assess empirically the extent to which each mechanism helps explain our key findings. Section 5 discusses the implications of our findings for event-study identification strategies that exploit high-frequency movements in long-term yields, the transmission of monetary policy, bond market "conundrums," and affine term structure models. Section 6 concludes.

1 The sensitivity of long-term rates to short-term rates

This section presents our main finding. Between 1971 and 2000, the sensitivity of long-term rates to changes in short-term rates was similarly strong at both high- and low-frequencies. Since 2000, the association between high-frequency changes in short- and long-term interest rates has grown even stronger. By contrast, the association between low-frequency changes in short- and long-term interest rates has weakened substantially. As a result, the sensitivity of long-term rates has become surprisingly frequency-dependent since 2000. We first document these basic facts for the U.S. We then contrast the patterns we see in the post-2000 data with those observed in the U.S. prior to the 1970s. Finally, we show that the sensitivity of long-term rates has evolved in a similar fashion in Canada, Germany, and the U.K.

Baseline findings for the U.S. We begin by regressing changes in 10-year Treasury yields or forward rates on changes in 1-year nominal Treasury yields. Specifically, we estimate regressions of the form:

$$y_{t+h}^{(10)} - y_t^{(10)} = \alpha_h + \beta_h (y_{t+h}^{(1)} - y_t^{(1)}) + \varepsilon_{t,t+h} \tag{1.1}$$

and

$$f_{t+h}^{(10)} - f_t^{(10)} = \alpha_h + \beta_h (y_{t+h}^{(1)} - y_t^{(1)}) + \varepsilon_{t,t+h}, \tag{1.2}$$

where $y_t^{(n)}$ is the continuously compounded n-year zero-coupon yield in period t and $f_t^{(n)}$ is the n-year-ahead instantaneous forward rate. To do so, we obtain historical data on the nominal and real U.S. Treasury yield curve from Gürkaynak et al. (2007) and Gürkaynak et al. (2010). We decompose nominal yields into real yields and inflation compensation, defined as the difference between nominal and real yields derived from Treasury Inflation-Protected Securities (TIPS). Our sample begins in August 1971, which is when reliable data on 10-year nominal yields first become available and ends in December 2019. For real yields and inflation compensation, we only study the post-2000 sample, since data on TIPS are not available until 1999. All data are measured as of the end of the relevant period—e.g., the last trading day of each month.

In standard monetary economics models, the central bank sets overnight nominal interest rates, and other interest rates are influenced by the expected path of overnight rates. A large literature argues that central banks in the U.S. and abroad have increasingly relied on communication—implicit or explicit signaling about the future path of overnight rates—as an active policy instrument (Gurkaynak et al., 2005; Lucca and Trebbi, 2009). To capture news about the near-term path of monetary policy that would not impact the current overnight rate, we take the short rate to be the 1-year nominal Treasury rate which follows approaches in the recent literature (Campbell et al., 2012; Gertler and Karadi, 2015; Gilchrist et al., 2015; Hanson and Stein, 2015).

Panel A in Table 1 reports estimated coefficients β_h in equation (1.1) for zero-coupon nominal yields, real yields, and inflation compensation using daily data and end-of-month data with $h = 1, 3, 6, 12$ months—i.e., for daily, monthly, quarterly, semi-annual, and annual changes in yields.[4] The results are shown for the pre-2000 and post-2000 samples separately. We base this sample split on a number of break-date tests that we will discuss shortly. Figure 1 plots the estimated coefficients β_h in equation (1.1) for nominal yields versus monthly horizon h for the pre-2000 and post-2000 samples.

Since we use overlapping h-month changes in equation (1.1) when $h > 1$, we report Newey and West (1987) standard errors using a lag truncation parameter of $\lceil 1.5 \times h \rceil$; when $h = 1$, we report heteroskedasticity-robust standard errors. To address the tendency for statistical tests based on Newey and West (1987) standard errors to over-reject in finite samples, we compute p-values using the asymptotic theory of Kiefer and Vogelsang (2005) which gives more conservative p-values and has better finite-sample properties than traditional Gaussian asymptotic theory.

At a daily frequency, the regression coefficients have risen significantly between the pre-2000 and post-2000 samples. Specifically, the daily coefficient for 10-year yields in Panel A has risen from $\beta_{day} = 0.56$ in the pre-2000 sample to $\beta_{day} = 0.87$ in the post-2000 sample and this increase is highly statistically significant (p-val< 0.001). At the same time, Panel A shows the coefficients at lower frequencies are much smaller after 2000. For example, the coefficient for $h = 12$-month changes in 10-year yields is $\beta_{12} = 0.56$ before 2000 but only $\beta_{12} = 0.23$ in the post-2000 sample and this difference is statistically significant (p-val < 0.001).

Combining these two observations, Figure 1 shows our main finding: in the post-2000 sample, the coefficient β_h is a steeply declining function of the horizon h over which yield changes are calculated. By contrast, β_h is a relatively constant function of horizon h in the pre-2000 sample. In other words, Table 1 and Figure 1 show that, prior to 2000, there was a strong tendency for short- and long-term rates to rise and fall together at both high- and low-frequencies. While the high-frequency relationship has grown even stronger since 2000, the low-frequency relationship has weakened significantly. Furthermore, Table 1 shows that the majority of the decline in β_h as a function of h during the post-2000 sample is due to the real component of long-term yields.

This is a surprising result: one would not expect β_h to vary strongly with monthly horizon h as in the post-2000 data. In a standard term-structure models with a single factor, we have $y_t^{(10)} = \alpha + \beta \cdot y_t^{(1)}$ for some $\beta \in (0, 1)$, implying that $\beta_h = \beta$ for all h, regardless of whether or not the expectations hypothesis holds. More generally, even accounting for multiple risk factors, term premia only fluctuate at business-cycle frequencies in conventional asset-pricing models, implying that β_h should be quite stable across monthly horizons h. And, as detailed in Section 3 below, if there are both persistent and transient shocks to short rates, the expectations hypothesis implies that β_h should be slightly increasing in h as it was

[4] Bond maturities are in years and time periods are in months, except when we estimate regressions at a daily frequency.

in the pre-2000 data. Thus, our finding that β_h is a steeply decreasing function of horizon h since 2000 suggests that term structure dynamics have shifted in an important way.

Panel B of Table 1 reports the corresponding β_h coefficients in equation (1.2) using changes in instantaneous forwards as the dependent variable. Like 10-year yields, the sensitivity of 10-year forward rates to changes in short-term rates has risen at high frequencies, but has declined markedly at low frequencies. Specifically, the coefficient for daily changes in 10-year forward rates rose from $\beta_{day} = 0.39$ in the pre-2000 sample to $\beta_{day} = 0.49$ in the post-2000 sample. By contrast, the coefficient at a 12-month horizon fell from $\beta_{12} = 0.39$ in the pre-2000 sample to $\beta_{12} = -0.13$ after 2000.

We use two approaches to date the timing of the break and both approaches suggest that there was a break around 2000. First, we estimate equations (1.1) and (1.2) using 10-year rolling windows. The estimated coefficients for $h = 12$-month changes are shown in Figure 2 for 10-year yields and forwards. These β_{12} coefficients decline substantially in more recent windows. The second approach is to test for a structural break in equations (1.1) and (1.2) for $h = 12$-month changes, allowing for an unknown break date. We use the test of Andrews (1993) who conducts a Chow (1960) test at all possible break dates, and then takes the maximum of the Wald test statistics. Figure 3 plots the Wald test statistic for each possible break date in equations (1.1) and (1.2) along with the Cho and Vogelsang (2017) critical values for a null of no structural break. The strongest evidence for a break is in 1999 or 2000 in both equations (1.1) and (1.2) and the break is highly statistically significant.

To clarify, we do not intend to argue that there was a *discrete* change in the underlying data-generating process in 2000. Instead, consistent with the rolling-window regressions shown in Figure 2, our reading of the data is that the underlying data-generating process has changed gradually over time—a gradual change which then becomes discernible when we compare the behavior of yields in across different samples. Nonetheless, throughout the remainder of the paper, we will adopt the heuristic of simply splitting the data into two samples: pre- and post-2000.

Robustness. In the Internet Appendix, we conduct a variety of robustness checks on our key findings. First, we show that very similar results obtain when we use long-term private yields as the dependent variable in equation (1.1). Specifically, we examine long-term corporate bond yields with Moody's ratings of Aaa and Baa, the 10-year swap yield, and the yield on Fannie Mae mortgage-backed-securities. For of all these long-term yields, the sensitivity to changes in 1-year Treasury rates was similar irrespective of frequency before 2000. After 2000, the sensitivity at high frequencies increases while the sensitivity at low frequencies declines significantly.

Second, we obtain similar results using different proxies for the short-term rate—i.e., using changes in 3-month, 6-month, or 2-year Treasury yields—as the independent variable in equation (1.1).

Third, one might be concerned about our use of overlapping changes in equations (1.1) and (1.2) when $h > 1$. Instead of computing Newey and West (1987) standard errors with a lag truncation parameter of $\lceil 1.5 \times h \rceil$, we find that one would draw almost identical inferences using Hansen and Hodrick (1980) standard errors with a lag truncation parameter equal to h. Going further, we show the estimates and our inferences are similar if we simply use non-overlapping h-month changes.

Finally, one might wonder if our dating of this break is driven by distortions stemming from the 2009–2015 period when overnight nominal rates were stuck at the zero lower bound in the U.S. Our use of

1-year rates as the independent variables in equations (1.1) and (1.2) limits any potential distortions since 1-year nominal yields continued to fluctuate from 2009 to 2015 (Swanson and Williams, 2014). Indeed, even if we end our sample period in 2007 or 2008, we still detect a break around 2000. For instance, if the post-2000 sample ends in December 2008, we find a daily $\beta_{day} = 0.77$ and a yearly $\beta_{12} = 0.20$, which are essentially indistinguishable from the numbers in Table 1.

U.S. evidence prior to the Great Inflation. Since data on the Treasury term structure is far more limited prior to the 1970s (Gürkaynak et al., 2007), our baseline findings use data beginning in 1971. Nevertheless, it is useful to examine the sensitivity of long yields to short rates before the Great Inflation, which ran from the late-1960s to the mid-1980s. Specifically, one plausible explanation for the strong sensitivity of long-term nominal yields during the 1971-1999 sample is that this was a period when long-run inflation expectations became unanchored and were continuously being revised in response to news (Gürkaynak et al., 2005). Since long-run inflation expectations have become firmly moored in recent decades, it is useful to compare the patterns we see since 2000 to the those observed prior to the Great Inflation—another period when inflation expectations were more firmly anchored. In the Internet Appendix, we thus examine the sensitivity of long-term Treasury yields to changes in short-term yields from 1953 (when the relevant data become available) to 1968 (when long-run inflation expectations began to drift up). Consistent with the view that long-run inflation expectations were better-anchored prior to the Great Inflation, the 1953-1968 coefficients are lower than those in the 1971-2000 sample. However, while the level of the β_h coefficients is lower in the 1953-1968 sample, we do not see the strong dependence on horizon h that is so evident in the post-2000 data. In summary, while the unanchoring and then reanchoring of long-run inflation expectations may help explain shifts in the *level* of β_h over time, the strongly *frequency-dependent* sensitivity of long-term rates that we see since 2000 appears to be something new under the sun.

International evidence. Our focus is on the U.S., but it is useful to consider whether these same patterns are also observed in other large, highly-developed economies. In the Internet Appendix, we report estimates of equation (1.1) for the U.K., Germany, and Canada for both pre-2000 and post-2000 samples. Our data for Canada, Germany, and the U.K. begin in 1986, 1972, and 1985, respectively. We find similar patterns for Canada, Germany, and the U.K. Specifically, for all three countries, β_h is strongly decreasing in h in the post-2000 data, but not in the pre-2000 data.

2 Yield-curve dynamics and bond return predictability

In this section, we first pinpoint the term-structure dynamics that account for the stronger high-frequency sensitivity and weaker low-frequency sensitivity of long rates to short rates that we see in the post-2000 data. Specifically, we demonstrate that this frequency-dependent sensitivity of long-term rates arises because, all else equal, past increases in short rates predict a subsequent flattening of the yield curve—and a subsequent decline in long-term yields and forwards—in the post-2000 data. Statistically, this means that post-2000 yield curve dynamics are “path-dependent” or “non-Markovian”: it is not enough to know the current shape of the yield curve. Instead, to form the best forecast of future bond yields and returns, one also needs to know how the yield curve has shifted in recent months.

Second, we show that these non-Markovian dynamics are themselves a reflection of a new short-lived form of bond return predictability. Specifically, since 2000, the expected returns on long-term bonds over those on short-term bonds are temporarily elevated following past increases in short rates. Thus, relative to an expectations-hypothesis baseline, long-term yields exhibit excess sensitivity at high frequencies and *temporarily overreact* to changes in short rates.

2.1 Non-Markovian yield-curve dynamics

In this subsection, we show that strong horizon-dependence of β_h in the post-2000 period arises because yield curve dynamics have become non-Markovian: to form the best forecast of future yields, one needs to know the current shape of the yield curve *and* how short rates have recently changed.

Predicting level and slope. When examining term structure dynamics, it is useful and customary to study the dynamics of yield-curve factors, especially level and slope factors (Litterman and Scheinkman, 1991). We define the level factor as the 1-year yield ($L_t \equiv y_t^{(1)}$) and the slope factor as the 10-year yield less the 1-year yield ($S_t \equiv y_t^{(10)} - y_t^{(1)}$)—a.k.a., the "term spread."[5] Most term structure models are Markovian with respect to current yield curve factors, meaning that the conditional mean of future yields depends only on today's yield-curve factors. However, our key finding—the post-2000 horizon-dependence of the relationship between long- and short-term yields—suggests that it may be useful to include lagged factors when forecasting yields. This idea has proven useful in several other contexts, including in Cochrane and Piazzesi (2005), Duffee (2013), and Feunou and Fontaine (2014). Specifically, we consider the following system of predictive monthly regressions:

$$L_{t+1} = \delta_{0L} + \delta_{1L} L_t + \delta_{2L} S_t + \delta_{3L}(L_t - L_{t-6}) + \delta_{4L}(S_t - S_{t-6}) + \varepsilon_{L,t+1} \tag{2.1a}$$

$$S_{t+1} = \delta_{0S} + \delta_{1S} L_t + \delta_{2S} S_t + \delta_{3S}(L_t - L_{t-6}) + \delta_{4S}(S_t - S_{t-6}) + \varepsilon_{S,t+1}. \tag{2.1b}$$

These regressions include level and slope as well as their changes over the prior six months, which is a simple way of allowing for longer lags without estimating too many free parameters.

Table 2 reports estimates of equations (2.1a) and (2.1b) for both the pre-2000 and post-2000 samples. We include specifications omitting all lagged changes (imposing $\delta_3 = \delta_4 = 0$), omitting lagged changes in slope (imposing $\delta_4 = 0$), and including all predictors. Based on the Akaike information criterion (AIC) or Bayesian information criterion (BIC), the model in column (1) with no lagged changes is chosen in the pre-2000 sample, while the model in column (5) with lagged changes in level is selected in the post-2000 sample. As shown in the bottom panel, in the post-2000 sample, the lagged change in level is a highly significant negative predictor of the future slope—i.e., increases in the level of yields predict subsequent yield-curve flattening. For example, as shown in column (5), a 100 basis point increase in the level over the prior 6-months is associated with a 11 basis *per-month* decline in slope in the post-2000 sample (p-val < 0.001). By contrast, as shown in column (2), the coefficient on $L_t - L_{t-6}$ in the pre-2000 sample is

[5] The level and slope factors are sometimes defined as the first two principal components of a set of yields. For simplicity, we have defined the level and slope factors using fixed maturities on the yield curve. This choice makes little difference and we find similar results if we examine the first two principal components.

zero. And, we can easily reject the hypothesis that the coefficients on $L_t - L_{t-6}$ in the pre- and post-2000 samples are equal (p-val < 0.001).[6]

The model in equations (2.1a) and (2.1b) can match the puzzling post-2000 horizon-dependent behavior of β_h that we documented above. This model can be written as a restricted vector autoregression (VAR) in $\mathbf{y}_t = (L_t, S_t)'$ of the form: $\mathbf{y}_{t+1} = \boldsymbol{\mu} + \mathbf{A}_1\mathbf{y}_t + \mathbf{A}_2\mathbf{y}_{t-6} + \boldsymbol{\varepsilon}_{t+1}$. Let $\Gamma_{ij}(h)$ denote the ijth element of the autocovariance of $\mathbf{y}_t$ at a lag of h months—i.e., the ijth element of $\mathbf{\Gamma}(h) = E[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_{t-h} - E[\mathbf{y}_{t-h}])']$. Given the estimated parameters from equations (2.1a) and (2.1b), we can work out $\Gamma_{ij}(h)$ to obtain the VAR-implied values of β_h in equation (1.1):

$$\beta_h = \frac{Var(L_t - L_{t-h}) + Cov(S_t - S_{t-h}, L_t - L_{t-h})}{Var(L_t - L_{t-h})} = 1 + \frac{2\Gamma_{12}(0) - \Gamma_{12}(h) - \Gamma_{12}(-h)}{2(\Gamma_{11}(0) - \Gamma_{11}(h))}. \tag{2.2}$$

In the pre-2000 sample, Table 1 reported estimates of $\beta_1 = 0.46$ and $\beta_{12} = 0.56$. In the post-2000 sample, the estimates are $\beta_1 = 0.66$ and $\beta_{12} = 0.23$. Table 2 reports the VAR-implied values of β_1 and β_{12} from equation (2.2). In the pre-2000 data, all of the VAR models can roughly match both β_1 and β_{12}. In the post-2000 sample, all models can match β_1, but only the models that include lagged changes in level—i.e., models that allow for non-Markovian dynamics—can match the sharp decline in β_{12}. Specifically, if the post-2000 VAR does not include lagged changes ($\delta_3 = \delta_4 = 0$) as in column (4), the VAR-implied values of β_{12} would be 0.59 and would be nowhere near what we observe in the data.

Predictable reversals in long-term rates. We next show that these non-Markovian dynamics imply that, in the post-2000 data, there are predictable reversals in long-term rates following past increases in short-term rates—i.e., long-term rates *temporarily* overreact to changes in short-terms rates. To see this explicitly, in Table 3 we estimate specifications that are reminiscent of the Jorda (2005) "local projection" approach to estimating impulse-response functions. Specifically, we predict the future changes in 10-year yields and forwards from month t to $t+h$ using the current level (L_t) and slope (S_t) of the yield curve as well as the prior month's change in level ($L_t - L_{t-1}$) and slope ($S_t - S_{t-1}$):

$$z_{t+h} - z_t = \delta_0^{(h)} + \delta_1^{(h)} L_t + \delta_2^{(h)} S_t + \delta_3^{(h)} (L_t - L_{t-1}) + \delta_4^{(h)} (S_t - S_{t-1}) + \varepsilon_{t \to t+h}. \tag{2.3}$$

Table 3 reports estimates of equation (2.3) for $z_t = y_t^{(10)}$ and $f_t^{(10)}$ in the pre- and post-2000 samples for $h =$ 3-, 6-, 9-, and 12- month changes. In Figure 4, we plot the coefficients $\delta_3^{(h)}$ on $L_t - L_{t-1}$ for $h = 1, 2, ..., 12$, tracing out the expected future change in z_t from month t to $t+h$ in response to an unexpected change in the level of rates between $t-1$ and t.[7]

Figure 4 plots the coefficients $\delta_3^{(h)}$ on $L_t - L_{t-1}$ versus monthly horizon h for both 10-year yields and 10-year forward rates. Figure 4 shows that, in the post-2000 data, there are predictable reversals in both 10-year yields and forwards following a past increase in short-term rates. However, there is no such reversal in the pre-2000 data. For instance, for 10-year yields, Table 3 reports that $\delta_3^{(6)} = -0.36$

[6] In unreported results, we also find that past changes in the level factor are associated with declines in the slope factor over the following month when the change in the level is computed over the prior 3 or 12 months.

[7] The inclusion of L_t and S_t as controls means that $\delta_3^{(h)}$ and $\delta_4^{(h)}$ capture the response of z_t to an unexpected changes in the level and slope between $t-1$ and t. The estimated $\delta_3^{(h)}$ coeffiecients are similar if we omit $S_t - S_{t-1}$ from the regression. The estimated $\delta_3^{(h)}$ coeffiecients are also quite similar, albeit with slightly larger standard errors, if we omit all controls.

(p-val = 0.07) after h = 6-months in the post-2000 data. (The difference between $\delta_3^{(6)}$ in the pre- and post-2000 data is significant with a p-value of 0.04.) Table 1 showed that, since 2000, a 100 bps increase in short-rates in month t is associated with a 66 bps contemporaneous rise of long-term yields. Thus, Table 3 suggests that 36 bps—or more than half—of this initial response is expected to reverse within 6 months. As in Table 1, the post-2000 reversion in 10-year forwards is even larger in magnitude and is statistically stronger. For 10-year forwards, we have $\delta_3^{(6)} = -0.52$ (p-val < 0.01) and the difference between $\delta_3^{(6)}$ in the pre- and post-2000 data is highly significant (p-val < 0.01). In summary, Table 3 and Figure 4 show that long-term rates appear to temporarily overreact to changes in short-term rates in the post-2000 data, but there was no such tendency before 2000.

To better understand these results, we decompose 10-year yields into the sum of a level component and a slope component as in Table 2—i.e., $y_t^{(10)} = L_t + S_t$—and plot the coefficients $\delta_3^{(h)}$ versus h for both level ($z_t = L_t$) and slope ($z_t = S_t$). Consistent with Table 2, Table 3 shows that the predictable reversals in long-term yields reflects the juxtaposition of two opposing forces in the post-2000. First, past increases in short-term rates predict subsequent increases in short-term rates in the post-2000 data, perhaps owing to the Fed's growing desire to gradually adjust short rates (Stein and Sunderam, 2018). However, past increase in short rates strongly predict a subsequent flattening of the yield curve since 2000. Since the latter effect outweighs the former, we see predictable reversals in long-term yields post-2000.

2.2 Predicting bond returns

In this subsection, we recast our main finding—the fact that, in recent years, β_h is so large at high frequencies and then declines rapidly as a function of horizon h—as a result about bond return predictability. Specifically, we show that this result arises because past increases in the level of rates lead to *temporary* rise in the expected return on long-term bonds relative to those on short-term bonds—i.e., a temporary rise in bond term premia. Thus, our findings reflect a new form of bond return predictability.

Results for 10-year bonds. The k-month log excess return on n-year bonds over the riskless return on k-month bills, $(k/12)\, y_t^{(k/12)}$, is:

$$rx_{t\to t+k}^{(n)} \equiv (k/12)\,(y_t^{(n)} - y_t^{(k/12)}) - (n - k/12)(y_{t+k}^{(n-k/12)} - y_t^{(n)}). \tag{2.4}$$

We first forecast the k-month excess return on n = 10-year zero-coupon bonds using level, slope, and the 6-month past changes in these two yield-curve factors:

$$rx_{t\to t+k}^{(10)} = \delta_0 + \delta_1 L_t + \delta_2 S_t + \delta_3(L_t - L_{t-6}) + \delta_4(S_t - S_{t-6}) + \varepsilon_{t\to t+k}. \tag{2.5}$$

In Table 4, we report the results from estimating these predictive regressions for k = 1, 3, and 6-month returns. Panel A reports the results for the pre-2000 sample and Panel B shows the post-2000 results.[8,9]

[8] The yield on k-month Treasury bills, $y_t^{(k/12)}$, is from the yield curve estimates in Gürkaynak et al. (2007). However, this curve is based on coupon securities with at least three months to maturity and does not fit the very short end of the curve well in the pre-2000 data. Therefore, we take the 1-month bill yield from Ken French's website for the pre-2000 sample.

[9] We obtain broadly similar results in Table 4 if we forecast returns using 3- or 12-month past changes in level and slope. And, the return predictability associated with past changes in level remains similar if, instead of controlling for level and slope, we control for the first five forward rates as in Cochrane and Piazzesi (2005).

In the post-2000 data, Table 4 shows the past change in the level of rates is a robust predictor of the excess returns on long-term bonds. However, there is no such predictability in the pre-2000 data. For instance, in column (6) of Panel B, we see that, all else equal, a 100 bps increase in short-term rates over the prior 6 months is associated with a $\delta_3 = 166$ bps (p-val < 0.01) increase in expected 3-month bond returns and the difference between δ_3 in the pre- and post-2000 data is statistically significant (p-val < 0.01). In untabulated results, we find that the post-2000 return predictability associated with past increases in the level of rates is short-lived and generally dissipates after $k = 6$ months. In other words, past increases in the level of rates lead to a temporary increase in the risk premia on long-term bonds.[10]

To draw out the connection to the predictable curve flattening discussed above, we show that these results for 10-year returns are related to predictability of the returns on what we refer to as "level-mimicking" and "slope-mimicking" portfolios. Specifically, we follow Joslin et al. (2014) and construct bond portfolios that locally mimic changes in the level and slope factors. Consider a factor-mimicking portfolio that places weight w_n on zero-coupon bonds with n years to maturity. The k-month excess return on this portfolio from t to $t+k$ is $rx^P_{t\to t+k} = (\sum_n w_n \cdot rx^{(n)}_{t\to t+k})/\,|\sum_n w_n|$. The level-mimicking portfolio has a weight of -1 on 1-year bonds and no weight on any other bonds. For small k, we have $rx^{(10)}_{t\to t+k} \approx -10 \cdot (\Delta_k L_{t+k} + \Delta_k S_{t+k})$ and $rx^{(1)}_{t\to t+k} \approx -1 \cdot \Delta_k L_{t+k}$. Thus, the level- mimicking portfolio has a k-month excess return of $rx^{LEVEL}_{t\to t+k} = -1 \cdot rx^{(1)}_{t\to t+k} \approx \Delta_k L_{t+k}$. The slope-mimicking portfolio has a weight of 1 on 1-year bonds and of -0.1 on 10-year bonds, so $rx^{SLOPE}_{t\to t+k} = (1 \cdot rx^{(1)}_{t\to t+k} - 0.1 \cdot rx^{(10)}_{t\to t+k})/0.9 \approx \Delta_k S_{t+k}/0.9$. Finally, we note that the excess returns on 10-year bonds are just a linear combination of the excess returns on the level- and slope-mimicking portfolios: $rx^{(10)}_{t\to t+k} = -9 \cdot rx^{SLOPE}_{t\to t+k} - 10 \cdot rx^{LEVEL}_{t\to t+k}$.

In the two bottom blocks of Table 4, we estimate equation (2.5) using $rx^{LEVEL}_{t\to t+k}$ and $rx^{SLOPE}_{t\to t+k}$ as the dependent variable. In the post-2000 sample, the excess returns on the slope-mimicking portfolio depend negatively on $L_t - L_{t-6}$, but the excess returns on the level-mimicking portfolio depends positively on $L_t - L_{t-6}$.[11] While the two effects partially offset when predicting 10-year excess returns, the net effect is positive and statistically significant in the post-2000 data. Furthermore, the results in Table 4 where we forecast $rx^{LEVEL}_{t\to t+k}$ and $rx^{SLOPE}_{t\to t+k}$ are entirely consistent with those in Table 2.[12]

In summary, we find that, since 2000, term premia on long-term bonds are *temporarily* elevated following past increases in short rates. This implies that, relative to an expectations-hypothesis baseline, long-term rates *temporarily overreact* to movements in short rates, exhibiting what Mankiw and Summers (1984) called "excess sensitivity" at high frequencies.

Results for other bond maturities. In the Internet Appendix, we examine the predictability for bond maturities other than $n = 10$ years. If, as we argue, past increases in short rates temporarily raise the compensation that investors require for bearing interest-rate risk, this should have a larger impact on the expected returns of long-term bonds than intermediate bonds. However, such a short-lived increase in the compensation for bearing interest rate risk should have relatively constant or even a hump-shaped

[10] Consistent with the vast literature on lower-frequency movements in bond risk premia initiated by Fama and Bliss (1987) and Campbell and Shiller (1991), we find $\delta_2 > 0$—i.e., expected bond returns are high when the yield curve is steep.

[11] The latter fact is consistent with Piazzesi et al. (2015) and Cieslak (2018), who account for it either with expectational errors or time-varying risk premia. Brooks et al. (2017) also show that the federal funds rate displays short-term momentum.

[12] In the Internet Appendix, we follow Joslin et al. (2014) and control for macroeconomic variables in equation (2.5). Even after controlling for macroeconomic variables, we find that past changes in level continue to have significant incremental predictive power for the future excess returns on 10-year bonds and the returns on the slope-mimicking portfolio.

effect on the yield and forward curves. The intuition is that the impact on bond yields equals the effect on a bond's average expected returns over its lifetime. As a result, a *temporary* rise in the compensation for bearing interest rate risk can have a greater impact intermediate-term yields than on long-term yields. Indeed, this is precisely what we find in the post-2000 data.

2.3 Interpreting the evidence

Before developing our economic modelling framework in the next section, we pause here to interpret our results. Specifically, our findings point towards the view that, in recent years, the term premium on long-term bonds is increasing in the recent *change* in short-term rates, all else equal. This simple non-Markovian assumption can match the facts that, in the post-2000 data, (i) the sensitivity of long-term yields β_h in equation (1.1), declines with horizon h and (ii) that, controlling for current yield curve factors, past changes in short-term interest rates predict future yield-curve flattening, future declines in long-term rates, and high future excess returns on long-term bonds.

To develop these ideas, we shift notation slightly. Rather than identifying specific maturities, we now refer to the long-term yield as y_t and the short rate as i_t . We split the long-term yield into an expectations-hypothesis component, eh_t, that reflects expected future short-term rates and a term premium component, tp_t, that reflects expected future bond risk premia: $y_t = eh_t + tp_t$. Thus, by definition, β_h—the total sensitivity of long-term yields at horizon h—is the sum of the expectations-hypothesis, β_h^{eh} and the term premium, β_h^{tp}, components:

$$\overbrace{\frac{Cov\left[y_{t+h}-y_t, i_{t+h}-i_t\right]}{Var\left[i_{t+h}-i_t\right]}}^{\beta_h} = \overbrace{\frac{Cov\left[eh_{t+h}-eh_t, i_{t+h}-i_t\right]}{Var\left[i_{t+h}-i_t\right]}}^{\beta_h^{eh}} + \overbrace{\frac{Cov\left[tp_{t+h}-tp_t, i_{t+h}-i_t\right]}{Var\left[i_{t+h}-i_t\right]}}^{\beta_h^{tp}}. \tag{2.6}$$

First, consider the expectations-hypothesis piece, β_h^{eh}. For now, assume the short-rate follows a univariate AR(1) process, implying $eh_t = \alpha^{eh} + \beta^{eh} \cdot i_t$ and $\beta_h^{eh} = \beta^{eh}$ for all h. Next, consider the term premium piece. In conventional asset-pricing theories, term premia only vary at business-cycle frequencies, so one would not expect β_h^{tp} to vary strongly with monthly horizon h. Thus, conventional theories suggest that $tp_t \approx \alpha^{tp} + \beta^{tp} \cdot i_t$, implying that $\beta_h^{tp} = \beta^{tp}$ and $\beta_h = \left(\beta^{eh} + \beta^{tp}\right)$ for all h. In other words, it is difficult for conventional theories to match the strong horizon-dependence of β_h that is so evident in the post-2000 data.

To generate horizon-dependent sensitivity, consider, instead, the following non-Markovian assumption:

$$tp_t = \alpha^{tp} + \beta^{tp} \cdot i_t + \delta^{tp} \cdot \left(i_t - i_{t-1}\right), \tag{2.7}$$

where $\delta^{tp} > 0$. This assumption implies that term premia depend on the current level of short-term rates and the recent *change* in short rates. Under this assumption, one can show that:

$$\beta_h = \beta^{eh} + \beta^{tp} + \delta^{tp} \cdot \left(1 - \gamma_h\right) \text{ where } \gamma_h \equiv \frac{Cov\left[i_{t+h-1} - i_{t-1}, i_{t+h} - i_t\right]}{Var\left[i_{t+h} - i_t\right]}. \tag{2.8}$$

The key is then to note that γ_h—the coefficient from a regression of $(i_{t+h-1} - i_{t-1})$ on $(i_{t+h} - i_t)$—is an increasing function of h. When $\delta^{tp} > 0$, this in turn explains why β_h^{tp} is decreasing in h.[13] Furthermore, when $\delta^{tp} > 0$, controlling for current level of short-term rates, past changes in short-rates predict future yield curve flattening, declines in long-term yields, and high excess returns on long-term bonds.[14]

In summary, our empirical findings can all be seen as consequences of the fact that, in recent decades, the term premia on long-term bonds is increasing in the recent change in short-term rates. Put differently, our findings reflect a new form of short-lived bond return predictability.

3 A model of temporary bond market overreaction

How can we best understand our key empirical finding that, in recent decades, the sensitivity of long-term yields to changes in short rates is steeply declining in the horizon over which these changes are computed? Because this finding reflects a form of short-lived return predictability, the most natural class of explanations involves temporary supply-and-demand imbalances in less-than-perfect financial markets (Grossman and Miller (1988) and Duffie (2010)). Therefore, we develop a model that emphasizes the role of what we call "rate-amplifying" shocks to the supply-and-demand for long-term bonds. Naturally, these rate-amplifying supply-and-demand shocks—which may either stem from institutional factors or investor psychology—will lead long yields to *temporarily* over-react if they trigger a short-lived supply-and-demand imbalance in the market for long bonds. In Section 4, we will build on this framework and explicitly model three rate-amplification mechanisms—mortgage refinancing waves, investor overextrapolation, and investor reaching-for-yield—and then assess empirically the extent to which each mechanism helps explain our key finding. The model in this section emphasizes the common underlying structure and shared asset-pricing implications of these rate-amplification mechanisms. By contrast, Section 4 emphasizes the idea that different amplification mechanisms have implications for different financial quantities.

3.1 Model setting

Overview In our model, time is discrete and infinite. Risk-averse arbitrageurs can either hold risky long-term nominal bonds or riskless short-term nominal bonds. The nominal interest rate on short-term bonds follows an exogenous stochastic process. Long-term bonds are available in a given net supply that must be absorbed by the arbitrageurs in our model. Since arbitrageurs' risk-bearing capacity is limited, shifts in the net supply of long-term bonds impact the term premium component of long-term yields as in Greenwood and Vayanos (2014) and Vayanos and Vila (2020).

The first key assumption is that there are rate-amplifying supply-and-demand shocks: shocks to the net supply of long-term bonds are *positively* correlated with shocks to short rates. This can either be because increases in short rates are associated with increases in the gross supply of long-term bonds or with reductions in the demands of other, non-arbitrageur investors. These rate-amplifying supply-and-demand shocks mean that arbitrageurs must increase their exposure to long-term bonds when short

[13] For instance, if i_t follows an AR(1) of the form $i_{t+1} - \bar{\imath} = \rho_i (i_t - \bar{\imath}) + \varepsilon_{i,t+1}$, then $\gamma_h = \left(2\rho_i - \rho_i^{h-1} - \rho_i^{h+1}\right) / \left(2 - 2\rho_i^h\right)$. We have $\gamma_1 = -(1-\rho_i)^2 / (2 - 2\rho_i) < 0$ and $\lim_{h\to\infty} \gamma_h = \rho_i > 0$. And, treating γ_h as continous in h, we have $\partial\gamma_h/\partial h > 0$.

[14] Our findings do not directly speak to the sign and magnitude of β^{tp}. The fact that $\beta_h - \beta_1 = -\delta^{tp} \cdot (\gamma_h - \gamma_1)$ declines rapidly with h in the post-2000 data, tells us about δ^{tp}—i.e., the way that term premia depend on past *changes* in short rates. By contrast, this finding tells us little about the low-frequency relationship between term premia and the *level* of short rates—i.e., about β^{tp}.

rates rise. To induce arbitrageurs to absorb these net supply shocks, the term premium component of long-term yields must increase when short rates rise, generating "excess sensitivity" of long-term yields when judged relative to the expectations hypothesis.

The second key assumption, following Duffie (2010), is that arbitrage capital is slow-moving: these net supply shocks walk down a short-run demand curve that is steeper than the long-run demand curve.[15] This slow-moving capital dynamic implies that an increase in short-term interest rates leads to a *temporary* supply-and-demand imbalance in the market for long-term bonds and, thus, a short-lived increase in bond risk premia. As a result, the excess sensitivity of long-term yields is greatest when measured at short horizons. Furthermore, we show that this frequency-dependent excess sensitivity is most pronounced when the underlying rate-amplifying net supply shocks are themselves transitory in nature.

The model can match our key finding in Section 1—β_h has fallen for large h and risen for small h post-2000—if (i) shocks to short-term nominal rates have become less persistent and (ii) the kinds of rate-amplification mechanisms we emphasize have grown in importance. We argue that (i) is justified by the strong evidence that shocks to the persistent component of inflation have become less volatile since the mid-1990s (Stock and Watson, 2007). Similarly, we argue that (ii) is justified since these rate amplification mechanisms appear to have become more powerful over time.

Short- and long-term nominal bonds. At time t, investors learn that short-term bonds will earn a riskless log return of i_t in nominal terms between time t and $t+1$. Short-term nominal bonds are available in perfectly elastic supply at this interest rate. One can think of the short-term nominal interest rate as being determined outside the model by monetary policy.

Long-term nominal bonds are available in a given net supply s_t that must be absorbed by the arbitrageurs in our model. The long-term nominal bond is a perpetuity that pays a coupon of $K > 0$ each period. Let P_t denote the price of this bond at time t, so the return on long-term bonds from t to $t+1$ is $1 + R_{t+1} = (P_{t+1} + K)/P_t$. To generate a tractable linear model, we use the well-known Campbell and Shiller (1988) log-linear approximation to the return on this perpetuity. Defining $\phi \equiv 1/(1+K) \in (0,1)$, the log excess return on long-term bonds over short-term bonds from t to $t+1$ is approximately:

$$rx_{t+1} \equiv \ln(1 + R_{t+1}) - i_t \approx \frac{1}{1-\phi} y_t - \frac{\phi}{1-\phi} y_{t+1} - i_t, \tag{3.1}$$

where y_t is the log yield-to-maturity on long-term bonds at time t and $D = 1/(1-\phi)$ is the bond's duration. Iterating equation (3.1) forward and taking expectations, the yield on long-term bonds is:

$$y_t = (1-\phi) \sum\nolimits_{j=0}^{\infty} \phi^j E_t \left[i_{t+j} + rx_{t+j+1} \right]. \tag{3.2}$$

The long-term yield is the sum of (i) an expectations hypothesis piece $eh_t = (1-\phi)\sum_{j=0}^{\infty} \phi^j E_t [i_{t+j}]$ that reflects expected future short rates and (ii) a term premium piece $tp_t = (1-\phi)\sum_{j=0}^{\infty} \phi^j E_t [rx_{t+j+1}]$ that reflects expected future excess returns on long-term bonds over short-term bonds.

[15] In this way, our model is a cousin of the model in Greenwood et al. (2018), who incorporate slow-moving capital effects into a model of the term structure.

Arbitrageurs. There are two groups of risk-averse arbitrageurs in the model, each with identical risk tolerance τ, who differ solely in the frequency with which they can rebalance their bond portfolios.

The first group of arbitrageurs are "fast-moving" and are free to adjust their holdings of long-term and short-term bonds each period. Fast-moving arbitrageurs are present in mass q and we denote their demand for long-term bonds at time t by b_t. Fast-moving arbitrageurs have mean-variance preferences over 1-period portfolio log returns. Thus, their demand for long-term bonds at time t is:

$$b_t = \tau \frac{E_t\left[rx_{t+1}\right]}{Var_t\left[rx_{t+1}\right]}. \tag{3.3}$$

The second group of arbitrageurs are "slow-moving" and can only rebalance their holdings of long-term and short-term bonds every $k \geq 2$ periods. Slow-moving arbitrageurs are present in mass $1 - q$. A fraction $1/k$ of slow-moving arbitrageurs are active each period and can rebalance their portfolios, but then cannot trade again for the next k periods. As in Duffie (2010), this is a reduced-form way to model the forces—whether due to institutional frictions or limited attention—that may limit the speed of arbitrage capital flows. Since they only rebalance every k periods, slow-moving arbitrageurs have mean-variance preferences over their k-period *cumulative* portfolio excess return. Thus, the demand for long-term bonds from the subset of slow-moving arbitrageurs who are active at time t is:

$$d_t = \tau \frac{E_t[\sum_{j=1}^{k} rx_{t+j}]}{Var_t[\sum_{j=1}^{k} rx_{t+j}]}. \tag{3.4}$$

Risk factors. Holders of long-term bonds face two different types of risk. First, they are exposed to *short rate risk.* they will suffer a capital loss on their long-term bond holdings if short-term rates unexpectedly rise. Second, they are exposed to *supply risk*: there are shocks to the net supply of long-term bonds that impact the term premium component of long-term bond yields. We make the following assumptions about the evolution of these two risk factors.

Short-term nominal interest rates. The short-term nominal interest rate is the sum of a highly persistent component $i_{P,t}$ and a more transient component $i_{T,t}$:

$$i_t = i_{P,t} + i_{T,t}. \tag{3.5}$$

A natural interpretation is that the persistent component reflects long-run inflation expectations and the transient component reflects cyclical variation in short-term real rates and expected inflation. The persistent component $i_{P,t}$ follows an exogenous AR(1) process:

$$i_{P,t+1} = \bar{\imath} + \rho_P \left(i_{P,t} - \bar{\imath}\right) + \varepsilon_{P,t+1}, \tag{3.6}$$

where $0 < \rho_P < 1$ and $Var_t\left[\varepsilon_{P,t+1}\right] = \sigma_P^2$. The transient component $i_{T,t}$ also follows an exogenous AR(1):

$$i_{T,t+1} = \rho_T i_{T,t} + \varepsilon_{T,t+1}, \tag{3.7}$$

where $0 < \rho_T \leq \rho_P < 1$ and $Var_t\left[\varepsilon_{T,t+1}\right] = \sigma_T^2$.

If $\rho_T < \rho_P$ and σ_P is large relative to σ_T, then short-term nominal rates will be highly persistent. As a result, long-term nominal rates will be highly sensitive to movements in short-term nominal rates due to standard expectations-hypothesis logic. Indeed, a large value of σ_P is a good explanation for the high sensitivity of long-term rates observed in the 1970s, 1980s, and the 1990s when long-run inflation expectations were not well-anchored (Gürkaynak et al., 2005). However, long-run inflation expectations have become firmly anchored in recent decades and there is strong evidence that shocks to the persistent component of nominal inflation have become far less volatile since the mid-1990s.

Rate-amplifying shocks to the supply and demand for long-term bonds. Long-term nominal bonds are available in an exogenous, time-varying net supply s_t that must be held in equilibrium by fast-moving and slow-moving arbitrageurs. This net supply equals the gross supply of long-term bonds minus the demand from other, non-arbitrageur investors outside the model who have inelastic demands.

We assume that s_t follows an AR(1) process:

$$s_{t+1} = \overline{s} + \rho_s (s_t - \overline{s}) + C\varepsilon_{P,t+1} + C\varepsilon_{T,t+1} + \varepsilon_{s,t+1}, \tag{3.8}$$

where $0 < \rho_s \leq \rho_T < 1$, $C \geq 0$, and $Var_t[\varepsilon_{s,t+1}] = \sigma_s^2$.

When $C > 0$, there are *rate-amplifying* net supply shocks—shocks to short rates are positively associated with shocks to the net supply of long-term bonds—and C parameterizes the strength of these rate-amplification mechanisms. This specification for net bond supply in equation (3.8) is a reduced-form way of capturing several different rate-amplifying supply-and-demand mechanisms. In Section 4, we will explore three different rate-amplification mechanisms: (i) temporary mortgage refinancing waves that are triggered by declines in interest rates, (ii) investors who tend to overextrapolate recent changes in short-term interest rates, and (iii) investors who "reach for yield" by buying more long-term bonds when short-term interest rates are low. Rate-amplifying net supply shocks can arise either because increases in short rates are associated with increases in the gross supply of long-term bonds (as in the mortgage refinancing channel) or because they are associated with reductions in the demands of other, non-arbitrageur investors (as in the investor overextrapolation and reaching-for-yield channels).

By contrast, the $\varepsilon_{s,t+1}$ shocks in equation (3.8) capture other forces that are unrelated to short rates which also impact the net supply of long-term bonds. While the model can be solved for any arbitrary correlation structure between the $\varepsilon_{P,t+1}$, $\varepsilon_{T,t+1}$, and $\varepsilon_{s,t+1}$ shocks, we assume, for simplicity, that these three underlying shocks are mutually orthogonal.

The difference between the persistence of these rate-amplifying net supply shocks and the persistence of the underlying shocks to short-term rates plays an important role in our model's ability to generate excess sensitivity that is most pronounced at high frequencies. To see why, note that equation (3.8) implies that the net supply of long-term bond is given by

$$\begin{aligned} s_t &= \overline{s} + C[(i_{P,t} - \overline{\imath}) - (\rho_P - \rho_s) \textstyle\sum_{j=0}^{\infty} \rho_s^j (i_{P,t-j-1} - \overline{\imath})] \\ &\quad + C[i_{T,t} - (\rho_T - \rho_s) \textstyle\sum_{j=0}^{\infty} \rho_s^j i_{T,t-j-1}] + [\textstyle\sum_{j=0}^{\infty} \rho_s^j \varepsilon_{s,t-j}]. \end{aligned} \tag{3.9}$$

When $\rho_s < \rho_T$, the rate-amplifying net supply shocks are less persistent than the underlying short rate shocks. As a result, net bond supply is increasing in the differences between the current level of each component of the short rate and a geometric moving-average of past values of that component. Thus, when $\rho_s < \rho_T$, net bond supply s_t will be high when short rates have recently risen—i.e., there will be short-lived supply-and-demand imbalances in the market for long-term bonds. By contrast, if $\rho_s = \rho_T = \rho_P$, net supply will be just as persistent as the underlying shocks to short-term rates. In this case, $s_t = \bar{s} + C(i_t - \bar{\imath}) + [\sum_{j=0}^{\infty} \rho_s^j \varepsilon_{s,t-j}]$ and only the current *level* of short rates—as opposed to recent changes in short rates—impacts net bond supply.

3.2 Equilibrium yields

At time t, there is a mass q of fast-moving arbitrageurs, each with demand b_t, and a mass $(1-q)\,k^{-1}$ of active slow-moving arbitrageurs who rebalance their portfolios, each with demand d_t. These arbitrageurs must accommodate the *active net supply*, which is the total net supply s_t of long-term bonds less any supply held off the market by inactive slow-moving arbitrageurs who do not rebalance the portfolios, $(1-q)\,k^{-1}\sum_{j=1}^{k-1} d_{t-j}$. Thus, the market-clearing condition for long-term bonds at time t is:

$$\overbrace{qb_t}^{\text{Fast demand}} + \overbrace{(1-q)k^{-1}d_t}^{\text{Active slow demand}} = \overbrace{s_t}^{\text{Total net supply}} - \overbrace{(1-q)(k^{-1}\textstyle\sum_{j=1}^{k-1} d_{t-j})}^{\text{Inactive slow holdings}}. \tag{3.10}$$

We conjecture that equilibrium yields y_t and the demands of active slow-moving arbitrageurs d_t are linear functions of a state vector, $\mathbf{x}_t$, that includes the steady-state deviations of both components of short-term nominal interest rates, the net supply of long-term bonds, and holdings of bonds by inactive slow-moving arbitrageurs. Formally, we conjecture that the yield on long-term bonds is $y_t = \alpha_0 + \boldsymbol{\alpha}_1'\mathbf{x}_t$ and that slow-moving arbitrageurs' demand for long-term bonds is $d_t = \delta_0 + \boldsymbol{\delta}_1'\mathbf{x}_t$, where the $(k+2)\times 1$ dimensional state vector, $\mathbf{x}_t$, is given by $\mathbf{x}_t = [i_{P,t} - \bar{\imath},\ i_{T,t},\ s_t - \bar{s}, d_{t-1} - \delta_0, \cdots, d_{t-(k-1)} - \delta_0]'$. These assumptions imply that the state vector follows a VAR(1) process $\mathbf{x}_{t+1} = \boldsymbol{\Gamma}\mathbf{x}_t + \boldsymbol{\epsilon}_{t+1}$, where $\boldsymbol{\Gamma}$ depends on the parameters $\boldsymbol{\delta}_1$ governing slow-moving arbitrageurs' demand.

In the Internet Appendix, we show how to solve for equilibrium yields in this setting. A rational expectations equilibrium of our model is a fixed point of a specific operator involving the "price-impact" coefficients, $(\boldsymbol{\alpha}_1')$, which show how the state variables impact bond yields, and the "demand-impact" coefficients, $(\boldsymbol{\delta}_1')$, which show how these variables impact the demand of active slow-moving investors. Specifically, let $\boldsymbol{\omega} = (\boldsymbol{\alpha}_1', \boldsymbol{\delta}_1')'$ and consider the operator $\mathbf{f}(\boldsymbol{\omega}_0)$ which gives (i) the price-impact coefficients that will clear the market for long-term bonds and (ii) the demand-impact coefficients consistent with optimization on the part of active slow-moving investors when agents conjecture that $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ at all future dates. A rational expectations equilibrium of our model is a fixed point $\boldsymbol{\omega}^* = \mathbf{f}(\boldsymbol{\omega}^*)$. Solving the model involves numerically finding a solution to a system of $2k$ non-linear equations in $2k$ unknowns.

An equilibrium solution only exists if arbitrageurs are sufficiently risk tolerant (i.e., for τ sufficiently large). When an equilibrium exists, there can be multiple equilibria. Equilibrium non-existence and multiplicity of this sort are common in overlapping-generations, rational-expectations models such as ours where risk-averse arbitrageurs with finite investment horizons trade an infinitely-lived asset that is subject to supply shocks. Different equilibria correspond to different self-fulfilling beliefs that arbitrageurs can hold about the price-impact of supply shocks and, hence, the risks of holding long-term bonds. However, we always find a unique equilibrium that is stable in the sense that equilibrium is robust to a small perturbation in arbitrageurs' beliefs regarding the equilibrium that will prevail in the future. Consistent with the "correspondence principle" of Samuelson (1947), this unique stable equilibrium has comparative statics that accord with standard economic intuition. We focus on this unique stable equilibrium in our analysis. See Greenwood et al. (2018) for an extensive discussion of these issues.

3.3 The sensitivity of long-term yields

We now explain the factors that shape the sensitivity of long-term rates in our model and how this sensitivity depends on horizon. Consider the model-implied counterpart of the empirical regression coefficient in equation (1.1). In the model, the coefficient β_h from a regression of $y_{t+h} - y_t$ on $i_{t+h} - i_t$ is:

$$\beta_h = \frac{Cov\left[y_{t+h} - y_t, i_{t+h} - i_t\right]}{Var\left[i_{t+h} - i_t\right]} = \frac{\boldsymbol{\alpha}_1'(2\mathbf{V} - \boldsymbol{\Gamma}^h\mathbf{V} - \mathbf{V}(\boldsymbol{\Gamma}')^h)\mathbf{e}}{\mathbf{e}'(2\mathbf{V} - \boldsymbol{\Gamma}^h\mathbf{V} - \mathbf{V}(\boldsymbol{\Gamma}')^h)\mathbf{e}}, \tag{3.11}$$

where $\mathbf{V} = Var\left[\mathbf{x}_t\right]$ denotes the variance of the state vector $\mathbf{x}_t$ and $\mathbf{e}$ denotes the $(k+2) \times 1$ vector with ones in the first and second positions and zeros elsewhere.[16]

We can then establish the following result:

Proposition 1. *The dependence of the model-implied coefficient β_h defined in (3.11) on the time-horizon h is determined by (i) the persistence ρ_x of the three shocks $x \in \{s, T, P\}$, (ii) the volatilities of the two short-rate shocks σ_T, and σ_P, (iii) the strength of the rate-amplification mechanisms C, and (iv) the degree to which arbitrage capital is slow moving q.*

1. *When $C = 0$, shifts in the net supply of long-term bonds are unrelated to changes in short-term rates. As a result, changes in term premia are unrelated to shifts in short-term rates and long-term yields do not exhibit excess sensitivity relative to the expectations hypothesis. Furthermore,*
   - *(a) if $\rho_T = \rho_P$, then β_h is independent of h. Also, β_h is independent of σ_T, and σ_P in this case.*
   - *(b) if $\rho_T < \rho_P$, then β_h is a mildly* increasing *function of h. Also, in this case, the level of β_h falls with σ_T and rises with σ_P for all h.*
2. *When $C > 0$, shifts in the net supply of long-term bonds are positively correlated with changes in short-term rates. As a result, changes in term premia are positively correlated with changes in short-term rates and long-term yields exhibit excess sensitivity. Furthermore,*
   - *(a) if $\rho_s = \rho_T = \rho_P$, and all capital is fast-moving ($q = 1$), then β_h is independent of h;*
   - *(b) if $\rho_s \leq \rho_T = \rho_P$ and either (i) supply shocks are transient ($\rho_s < \rho_T$) or (ii) capital is slow-moving ($q < 1$), then β_h is decreasing in h;*
   - *(c) if $\rho_s \leq \rho_T < \rho_P$, then β_h can be non-monotonic in h.*

Proof. See the Internet Appendix for all proofs. □

When there are no rate-amplifying net supply shocks ($C = 0$), long-term interest rates are not excessively sensitive to short-term rates relative to the expectations hypothesis. To the extent that short-term interest rates contain both a transient and a more persistent component ($\rho_T < \rho_P$), one

[16] To derive this expression, note that $y_{t+h} - y_t = \boldsymbol{\alpha}_1'\left(\mathbf{x}_{t+h} - \mathbf{x}_t\right)$ and $i_{t+h} - i_t = \mathbf{e}'\left(\mathbf{x}_{t+h} - \mathbf{x}_t\right)$. Since the state-vector $\mathbf{x}_t$ follows a VAR(1) process $\mathbf{x}_{t+1} = \boldsymbol{\Gamma}\mathbf{x}_t + \boldsymbol{\epsilon}_{t+1}$ with $\Sigma = Var\left[\boldsymbol{\epsilon}_{t+1}\right]$, we have $vec(\mathbf{V}) = \left(\mathbf{I} - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}\right)^{-1} vec(\boldsymbol{\Sigma})$. Noting that $Cov\left[\mathbf{x}_{t+j}, \mathbf{x}_t'\right] = \boldsymbol{\Gamma}^j\mathbf{V}$ and $Cov[\mathbf{x}_t, \mathbf{x}_{t+j}'] = \mathbf{V}\left(\boldsymbol{\Gamma}'\right)^j$, we have $Var\left[\mathbf{x}_{t+h} - \mathbf{x}_t\right] = 2\mathbf{V} - \boldsymbol{\Gamma}^h\mathbf{V} - \mathbf{V}(\boldsymbol{\Gamma}')^h$ and the result follows.

should actually expect the β_h coefficients to increase in horizon h when $C = 0$. This effect arises since movements in the persistent short rate component are associated with larger movements in long-term yields by standard expectations-hypothesis logic and because the persistent component dominates changes in short rates at longer horizons. Furthermore, when $\rho_T < \rho_P$, the level of β_h for any horizon h depends on σ_T and σ_P. For instance, an increase in σ_P raises the fraction of total short-rate variation at all horizons that is due to the persistent component. Since shocks to the persistent component of short rates have larger impact on long-term yields, an increase in σ_P raises β_h at all horizons.

The existence of rate-amplifying demand shocks ($C > 0$) naturally generates excess sensitivity relative to the expectations hypothesis. However, Part 2.(a) of Proposition 1 shows that rate-amplifying demand ($C > 0$) need not generate *horizon-dependent* excess sensitivity—i.e., *temporary* overreaction of long-term yields when judged relative to the expectations hypothesis. To generate horizon-dependent excess sensitivity (β_h coefficients that decline with h), Part 2.(b) clarifies that either (i) the rate-amplifying demand shocks must be less persistent than the underlying short-rate shocks ($\rho_s < \rho_T$) or (ii) these rate-amplifying demand shocks must be met by a slow-moving arbitrage response.[17] Under either of these conditions, shifts in short-term rates give rise to a short-lived supply-and-demand imbalance in the market for long-term bonds, leading long-term yields to *temporarily* overreact to short rates. In practice, we suspect that *both* transitory rate-amplifying demand shocks and slow-moving capital play a role in explaining why β_h declines steeply with h in the recent data. Furthermore, these two mechanisms reinforce one another: it is easier to quantitatively match the steep decline in β_h as a function of h in calibrations that feature both of these elements.

3.4 Model calibration

Our main empirical findings are that β_h has declined in the post-2000 sample at low frequencies (high h) but has risen at high frequencies (low h), leading β_h to decline steeply with horizon h in the post-2000 data. Guided by Proposition 1, we now discuss how to best understand the changing sensitivity of long-term interest rates. We assume that $\rho_s < \rho_T < \rho_P$ throughout and focus on the role of changes in σ_P (the volatility of persistent short rate shocks) and C (the strength of any rate-amplifying mechanisms).

We argue that our framework can match these surprising patterns if two underlying parameters shifted from the 1971-1999 period to the post-2000 period:

1. σ_P **has fallen**: Shocks to the persistent component of short-term nominal rates have become less volatile in the post-2000 period.

2. C **has risen**: The rate-amplifying supply-and-demand mechanisms that we emphasize have become more important in recent decades.

[17] Technically, when capital is slow-moving ($q < 1$) and $\rho_s \leq \rho_T = \rho_P$, β_h is only guaranteed to be *locally* decreasing in h for $h \leq k$—i.e., for horizons shorter than over which all slow-moving arbitrageurs will have rebalanced their portfolios. Specifically, while we always have $\beta_h < \beta_{h-1}$ for $h \leq k$, we can have $\beta_h > \beta_{h-1}$ for $h > k$. However, even when there are local non-monoticities, β_h is globally decreasing in the sense that $\lim_{h\to\infty} \beta_h < \beta_1$. What explains the potential for these local non-monoticities? As in Duffie (2010), the gradual adjustment of slow-moving arbitrageurs can gives rise to modest echo effects for $h > k$, generating a series of damping oscillations that converge to $\lim_{h\to\infty} \beta_h$. These oscillations arise because the slow-moving arbitrageurs who reallocate soon after a supply shock lands take large opportunistic positions. These positions temporarily reduce the active supply of long-term bonds and need to be re-absorbed in later periods.

Illustrative calibration. We consider an illustrative calibration of the model in which each period is a month. We assume the following parameters were the same in the 1971-1999 and post-2000 periods:[18]

- **Persistence:** $\rho_P = 0.995$, $\rho_T = 0.96$, and $\rho_s = 0.80$. We assume that the rate-amplifying net supply shocks are less persistent than the underlying short rate shocks. These parameters imply that shocks to the persistent short rate component have a half-life of 11.5 years, shocks to the transient component have a half-life of 1.4 years, and shocks to the net supply of long-term bond have a half-life of 3 months.

- **Slow-moving capital:** $q = 30\%$ and $k = 12$. Thus, $1 - q = 70\%$ of the arbitrageurs are slow-moving and only rebalance their bond portfolios every 12 months. These assumptions capture the idea that many large institutional investors only rebalance their portfolios annually.

- **Volatility of the transient component of short rates:** $\sigma_T^2 = 0.15\%$.

- **No independent net supply shocks:** $\sigma_s^2 = 0$. Thus, the net supply shocks induced by shocks to short rates are the only reason term premia vary. Given our focus, this assumption is without loss of generality.

- **Other parameters:** $\tau = 0.5$ and $\phi = 119/120$, so the duration of the perpetuity is $D = 1/(1-\phi) = 120$ months—i.e., 10 years.

We assume that two model parameters, C and σ_P, changed between the 1971-1999 and the post-2000 periods. For the pre-2000 period, we assume:

- **A large persistent component of short rates:** $\sigma_P^2 = 0.15\%$. The implied standard deviation of the short rate is 4.12% which compares with a volatility of 1-year yields of 2.63% during the 1971-1999 sample.

- **No rate-amplifying net supply shocks:** $C = 0$. We assume there were no rate-amplifying net supply shocks in the 1971-1999 period.

By contrast, for the post-2000 period, we assume:

- **A small persistent component of short rates:** $\sigma_P^2 = 0.012\%$. The implied standard deviation of the short rate is 1.77% which is similar to the post-2000 volatility of 1-year yields of 1.85%.

- **Net supply shocks induced by short rate shocks:** $C = 0.55$. Thus, we assume a meaningful increase in the strength of rate-amplifying supply-and-demand mechanisms.

Model-implied regression coefficients β_h. The first graph in Figure 5 plots the model-implied regression coefficients β_h from equation (3.11) against the monthly horizon h for the pre- and post-2000 calibrations. In the pre-2000 calibration where σ_P is large and $C = 0$, β_h is high and is largely independent of horizon h. In fact, β_h rises gradually with h in the pre-2000 calibration—as it does in the pre-2000

[18]The pre-Great Inflation data discussed in Section 1 are consistent with a regime where C was also negligible and σ_P was large, although not as large as during the 1971-1999 period.

data—because the more persistent component of short rates dominates when changes are computed at longer horizons. By contrast, in the post-2000 calibration where σ_P is smaller and C is large, β_h declines steeply with h. And, since σ_P is lower, β_h eventually reaches a lower level for large h.

β_h declines with h in the post-2000 calibration because short rate shocks give rise to transient rate-amplifying shocks to the net supply of long-term bonds ($C > 0$ and $\rho_s < \rho_T$) that encounter a short-run demand curve that is steeper than the long-run demand curve due to slow-moving capital ($q < 1$ and $k > 1$), triggering short-lived market imbalances. However, the second graph in Figure 5 shows that β_h only declines moderately with h in our post-2000 calibration if we drop the assumption that arbitrage capital is slow-moving. Thus, from a quantitative perspective, transient rate-amplifying net supply shocks and slow-moving capital are both helpful for matching the fact that β_h declines so steeply with h in the post-2000 data.

Figure 6 shows the model-implied impulse response functions in the post-2000 calibration following a 100 bp shock to short rates that lands in month $t = 12$. (We assume there is a 50 bp shock to both the persistent and transient components of the short rate.) The long-term yield is the sum of an expectations-hypothesis component and a term premium component: $y_t = eh_t + tp_t$. Thus, the term spread is $y_t - i_t = (eh_t - i_t) + tp_t$. The figure shows impulse responses for short-term rates (i_t), long-term yields (y_t), the term spread ($y_t - i_t$), and the term premium (tp_t).

The initial shock to short rates leads to a rise in term premia. Thus, relative to the expectations-hypothesis, long-term rates are excessively sensitive to short rates. However, the rise in term premia wears off quickly, explaining our key finding that β_h declines sharply with horizon h. Nonetheless, the impulse to short rates causes the yield curve to flatten on impact as in the data. This is because $(eh_t - i_t)$ falls on impact and this flattening due to the expectations hypothesis outweighs the steepening due to the rise in term premia. However, the initial rise in short rates predicts additional yield curve flattening—and predictable reversals in long-term yields—over the following months.

Matching related findings. In addition to matching the fact that the β_h coefficients decline steeply with h in the post-2000 period, the model can also match the related empirical findings documented above. First, the model is consistent with our return forecasting evidence: in the post-2000 calibration, bond risk premia $E_t[rx_{t+1}]$ will be elevated when short-term rates have recently risen. To see this, note that risk premia are:

$$E_t[rx_{t+1}] = \tau^{-1}V^{(1)} \cdot b_t = (\tau q)^{-1} V^{(1)} \cdot (s_t - (1-q)k^{-1} \textstyle\sum_{j=0}^{k-1} d_{t-j}). \tag{3.12}$$

The idea is that, when $C > 0$, fast-moving arbitrageurs will be bearing greater interest-rate risk when short-rates have recently risen—i.e., $b_t = q^{-1}(s_t - (1-q)k^{-1} \sum_{j=0}^{k-1} d_{t-j})$ will be higher—and they will require compensation for bearing this extra risk. Again, there are two reasons why increases in short rates lead to increases in b_t and, hence, $E_t[rx_{t+1}]$. First, even if there are no slow-moving arbitrageurs ($q = 1$), when supply shocks are less persistent than short rates (i.e., $\rho_s < \rho_T$), equation (3.9) shows that supply s_t will be high when short rates have recently risen. Second, even if supply shocks are as persistent as short-rate shocks ($\rho_s = \rho_T = \rho_P$), when there is slow-moving capital, b_t will be high when short rates have recently risen since some slow-moving arbitrageurs will not have rebalanced their portfolios in response to the related supply shock.

Second, let $L_t = i_t$ and $S_t = y_t - i_t$ denote the model-implied level and slope factors. If we estimate equation (2.1b) in data simulated from the model, we find that past increases in the level of rates predict a flattening of the yield curve in the post-2000 calibration but not in the pre-2000 calibration. In the post-2000 calibration, past increases in the level of rates are associated with a higher current risk premium on long-term bonds. Since the risk premium is $E_t[rx_{t+1}] = S_t - \phi(1-\phi)^{-1}(E_t[\Delta S_{t+1}] + E_t[\Delta L_{t+1}])$, all else equal, $E_t[\Delta S_{t+1}]$ is lower when short rates have recently risen. Thus, the model generates the non-Markovian dynamics emphasized in Section 2.1.

4 Rate-amplification mechanisms

What are the key source of rate-amplifying demand that together explain why long yields temporarily overreact to changes in short rates in the recent U.S. data? In this section, we explore three rate-amplification mechanisms that may help explain why increases in short rates to trigger temporary supply-and-demand balances in the market for long-term bonds: (i) the mortgage refinancing channel, (ii) the investor overextrapolation channel, and (iii) the reaching-for-yield channel.

For each channel, we first show how it can be used to microfound rate-amplifying shocks to the net supply of long-term bonds similar to those we introduced in reduced-form in Section 3 and we then embed each channel our general modelling framework. (The Appendix provides additional details and illustrative calibrations of these three microfounded models).

Next, we discuss why the strength of each channel may have increased in recent decades. The key underlying trend here is the increasing financialization of interest-rate risk—e.g., the growth of mortgage securitization or of open-ended bond funds.[19]

Finally, by examining the relationship between bond yields and different financial quantities, we assess empirically the extent to which each channel contributes to the frequency-dependent sensitivity of long-term rates we observe after 2000 in the U.S. Of course, the precise mix of rate-amplifying mechanisms may vary somewhat over time and across geographic regions (e.g., the U.S. versus the Eurozone). So, given the difficulties inherent in precise attribution, we believe that the primary contribution of these empirical exercises is to rule in the general class of rate-amplifying mechanisms we emphasize. With those caveats in mind, we find evidence that mortgage refinancing and investor overextrapolation both help explain why long yields rates have temporarily overreacted to short rates since 2000 in the U.S. By contrast, we find less evidence that reaching-for-yield plays an important role in driving the short-lived overreaction of long-term U.S. yields.

4.1 Mortgage refinancing

Negative shocks to short-term rates trigger mortgage refinancing waves in the U.S. that lead to temporary declines in the duration of outstanding fixed-rate mortgages and, thus, reductions in the effective gross supply of long-term bonds (Hanson, 2014; Malkhozov et al., 2016). Because these induced supply shifts are large relative to bond investors' risk appetites, declines in short rates are associated with temporary

[19] We do not mean to argue that there was a *discrete* change in the underlying structure of fixed-income markets around 2000. Instead, our argument is that the sorts of rate-amplifying supply-and-demand mechanisms that we emphasize have gradually become more important, leading to a gradual change in the data-generating process for bond yields that becomes discernible when we compare the pre-2000 and post-2000 samples.

declines in bond term premia. As we detail below, there is evidence that strength of this channel has grown in recent decades in the U.S. because mortgage-backed securities became a larger share of the bond market (Hanson, 2014). To be sure, the mortgage refinancing channel is only relevant in countries such as the U.S. where fixed-rate mortgages with an embedded prepayment option are an important source of mortgage financing. However, Domanski et al. (2017) and Shin (2017) point to a similar rate-amplification mechanism—one that may be more important in the Eurozone—stemming from the desire of insurers and pensions to dynamically match the duration of their assets and liabilities.[20]

Modeling the mechanism. Most fixed-rate, residential mortgages in the U.S. give the borrower the option to prepay at any time without a penalty. When long-term rates fall, the option to prepay and refinance older, higher-coupon mortgages at current mortgage rates becomes more attractive to individual borrowers. Since refinancing entails large financial and nonfinancial costs, households only gradually exercise their prepayment options following a decline in mortgage rates, leading the effective maturity or "duration" of outstanding mortgages—i.e., the sensitivity of mortgage prices to changes in interest rates—to decline when long-term rates fall. What is more, the expected amount of aggregate mortgage refinancing activity varies significantly over time: depending on the past path of mortgage rates, there are times when many households move from being far from refinancing to being close to refinancing and vice versa. The resulting shifts in expected refinancing activity trigger large changes to the total quantity of interest rate risk that must be borne by bond market investors, leading to transient, but sizable fluctuations in bond term premia (Hanson, 2014; Malkhozov et al., 2016).

This mortgage refinancing channel can be used to micro-found a specification for the net supply of long-term bonds that is similar to equation (3.9)—i.e., one that depends on the difference between current interest rates and a moving-average of past rates. Following (Malkhozov et al., 2016), we assume that (i) there is a constant quantity M of outstanding fixed-rate mortgages with an embedded prepayment option; (ii) the primary mortgage rate, y_t^M, equals the long-term yield, y_t, plus a constant spread; (iii) the average coupon on outstanding mortgages evolves according to $c_{t+1}^M - c_t^M = -\eta \cdot (c_t^M - y_t^M)$, where $(c_t^M - y_t^M)$ is the "refinancing incentive" at time t and $\eta \in [0, 1]$ is the sensitivity of c_{t+1}^M to the refinancing incentive at t; (iv) the average "duration" or effective maturity of outstanding mortgages is $DUR_t^M = \overline{DUR}^M - N \cdot \left(c_t^M - y_t^M\right)$, where $\overline{DUR}^M > 0$ and $N > 0$ is the "negative convexity" of the average mortgage; and (v) the effective gross supply of long-bonds at time t is $s_t = M \cdot DUR_t^M$.

Each of these assumptions captures a well-known and reliable empirical regularity about the U.S. mortgage market. In particular, assumption (iii) captures the fact that, when the refinancing incentive $(c_t^M - y_t^M)$ is higher, more households refinance their existing high-coupon mortgages at time t, leading the average coupon to fall from t to $t + 1$. Assumption (iv) captures the fact that, when the refinancing incentive is higher, more households are expected to refinance their existing mortgages in the near future, implying that the average outstanding mortgage behaves more like a short-term bond. In other words,

[20] Domanski et al. (2017) and Shin (2017) argue that the convexity of insurers' and pensions' liabilities is greater than the convexity of their assets. Thus, as interest rates decline, the duration of their liabilities increases more than the duration of their assets, and insurers and pensions increase their demand for long-term bonds to match asset and liability duration. Holding fixed the gross supply of long-term bonds, this means that the net supply of long-term bonds that must be held by arbitrageurs is lower when short rates are low. This mechanism is arguably quite important in European bond markets where insurers and pensions play an especially important role. And this dynamic may have grown in recent years as regulators have pushed insurers to more prudently manage their interest rate exposures.

mortgage prices become less sensitive to long-term yields when the refinancing incentive is high, so bond investors must bear less interest rate risk when the refinancing incentive is high. Together these assumptions imply that the effective gross supply of long-term bonds at time t is:

$$s_t = M \cdot \overline{DUR}^M + MN \cdot (y_t - \eta \sum_{j=0}^{\infty} (1-\eta)^j \, y_{t-1-j}). \tag{4.1}$$

Thus, the mortgage refinancing channel implies that bond investors must bear greater interest rate risk when the long-term yields are high relative to their backward-looking, geometric average—i.e., when long-term rates have recently risen. And, the strength of this channel is given by the product MN.

There have been two structural shifts in the U.S. bond market in recent decades that are relevant for the strength of the refinancing channel. First, mortgage-backed securities (MBS) have become a larger share of the U.S. bond market over time. In the language of the model, this means that M has risen. For instance, from 1976 to 1999, MBS accounted for 21% of the value of the Bloomberg-Barclays Aggregate Index on average, a proxy for the broad U.S. investment-grade bond market. From 2000 to 2019, the corresponding figure was 33%. As a result, movements in the duration of the outstanding mortgages (DUR_t^M) now generate far larger shifts in the effective supply of long-term bonds when judged relative to the overall U.S. bond market. Second, due a secular decline in refinancing costs and frictions, mortgage refinancing has become more interest-rate elastic over time (Bennett and Peristiani, 2001; Fuster et al., 2019). More elastic mortgage refinancing corresponds to a rise in both N and η. As a result, the association between DUR_t^M and recent changes in long-term rates has grown stronger. Together these changes suggest that the strength of the mortgage refinancing channel may have grown in recent decades.[21]

To solve our model of mortgage refinancing, we substitute the expression for supply in equation (4.1) into the market-clearing condition in (3.10) from Section 3, thereby allowing for the possibility that the arbitrage response to mortgage refinancing waves is slow-moving. Specifically, we continue to assume that fraction q of bond investors are fast-moving with demands given by equation (3.3) and fraction $(1-q)$ are slow-moving and only rebalance their portfolios every $k \geq 2$ periods with demands given by (3.4). We can then establish the following proposition:

Proposition 2. ***Mortgage refinancing model.*** *Consider the mortgage refinancing model and for simplicity suppose $\rho_T = \rho_P$. When $MN > 0$, long-term yields are excessively sensitive to short rates when judged relative to the expectations hypotheisis. When $MN > 0$ and $\eta = 0$, this excess sensitivity is only horizon-dependent—i.e., the model-implied regression coefficient β_h in equation (3.11) only declines with horizon h—when arbitrage capital is slow moving ($q < 1$). By contrast, when $MN > 0$ and $\eta > 0$, β_h declines with horizon h even if all arbitrage capital is fast-moving ($q = 1$).*

When $\eta > 0$, shocks to short rates trigger shifts in effective bond supply that are less persistent than the underlying short rate shocks, giving rise to horizon-dependent excess sensitivity even in the absence of slow-moving capital. However, we are best able to quantitatively match the post-2000 behavior of the

[21] The rise in N means that a given change in the refinancing incentive $(c_t^M - y_t^M)$ now has a larger impact on mortgage duration DUR_t^M. Specifically, in our pre-2000 sample, the coefficient from a regression of DUR_t^M on $(y_t^M - c_t^M)$ is $N = 0.19$. In the post-2000 sample, the corresponding coefficient is $N = 0.91$ and the difference is statistically significant (p-value $<$ 0.001). The increase in η means that c_t^M has become a faster-moving average of past long-term rates, implying that the refinancing incentive more closely tracks recent changes in long rates. Specifically, in the pre-2000 sample, the coefficient from a regression of $(c_{t+1}^M - c_t^M)$ on $(y_t^M - c_t^M)$ is $\eta = 0.013$. In the post-2000 sample, the corresponding coefficient is $\eta = 0.022$ and the difference is statistically significant (p-value $= 0.011$).

β_h coefficients using calibrations of this model in which (i) MN has risen substantially from the pre-2000 level and (ii) the resulting rate-amplifying supply shocks are met by a slow-moving arbitrage response.

In addition to explaining the post-2000 behavior of the β_h coefficients, our model of mortgage refinancing predicts that (1) mortgage duration DUR_t^M is high when interest rates have recently risen, (2) the level of mortgage duration positively predicts future excess returns on long-term bonds (i.e., $E_t[rx_{t+1}]$ is high when DUR_t^M is high), and (3) the level of mortgage duration predicts subsequent yield-curve flattening (i.e., $E_t[\Delta S_{t+1}]$ is low when DUR_t^M is high).

Evidence from mortgage-related quantities. To empirically assess the contribution of the mortgage refinancing channel to our findings, we use two proxies for the impact of mortgage refinancing on the effective supply long-term bonds:

- $y_t^M - c_t^M$ is the mortgage refinancing *disincentive* in month t. Here y_t^M is the average primary rate for 30-year, fixed-rate mortgages from Freddie Mac's Primary Mortgage Market Survey and c_t^M is the average outstanding coupon of MBS in the Bloomberg-Barclays U.S. MBS index. The index covers pass-through MBS backed by conventional fixed-rate mortgages that are guaranteed by Ginnie Mae, Fannie Mae, and Freddie Mac. This refinancing disincentive measure, which is associated with a higher duration on outstanding mortgages, is available beginning in Jan-1976.
- DUR_t^M is the duration-to-worst of the Bloomberg-Barclays U.S. MBS index in month t, a measure of the sensitivity of MBS prices to changes in long-term yields. This MBS duration measure is available on a monthly basis beginning in Jan-1976.[22]

The correlation between $y_t^M - c_t^M$ and DUR_t^M is 0.55 from 1976 to 1999 and 0.66 in the post-2000 sample.

Using each of these proxies (X_t) for mortgage duration, we first estimate

$$X_t = \gamma_0 + \gamma_1 L_t + \gamma_2 S_t + \gamma_3 (L_t - L_{t-6}) + \gamma_4 (S_t - S_{t-6}) + \varepsilon_t^{MBS}, \tag{4.2}$$

for the pre- and post-2000 samples. We are mainly interested in the coefficient on the 6-month change in level ($L_t - L_{t-6}$), which tells us how MBS duration responds to recent changes in the level of short-term rates. Second, we estimate

$$rx_{t\to t+3}^Z = \delta_0 + \delta_1 L_t + \delta_2 S_t + \delta(L_t - L_{t-6}) + \delta_4 (S_t - S_{t-6}) + \delta_5 X_t + \varepsilon_{t\to t+3}^Z, \tag{4.3}$$

for $Z = 10$ and $Z = SLOPE$ for the pre-2000 and post-2000 samples. That is, we run horse race regressions to assess whether mortgage refinancing waves help explain why past changes in short-term interest rates forecast bond excess returns in the post-2000 data. We are interested in the coefficients on

[22] To obtain this duration measure at each point in time, Barclays uses its proprietary mortgage prepayment model to compute the expected cashflows on each mortgage-backed security in its index. Yield-to-worst is the internal rate of return that equates MBS price and the present value of expected cash flows. Barclays then computes the Macaulay duration of MBS treating expected cashflows as given and the duration of the index is just a valued-weighted average of individual security durations. Beginning in Jan-1989, Barclays reports an option-adjusted, effective duration for the MBS index which is the duration measure emphasized in Hanson (2014). In the post-2000 sample, this slightly more sophisticated measure has a 0.84 correlation in levels with the Macaulay duration measure we use here.

mortgage duration (X_t) and the change in the level of rates ($L_t - L_{t-6}$) and how these coefficients change when these two variables are included jointly in the regression as opposed to one at a time.

Panel A of Table 5 shows the results for estimating equation (4.2) using the refinancing disincentive ($X_t = y_t^M - c_t^M$) and shows that the refinancing disincentive is responsive to past changes in level and slope. Comparing columns (3) and (6), we see that it has become more responsive to past changes in level since 2000 (p-value = 0.06). Panel B reports the results for estimating this same equation using the duration of the Barclays MBS index ($X_t = DUR_t^M$) and delivers a similar message. Panels C and E report the results from estimating equation (4.3) using the refinancing disincentive and suggest that the refinancing channel helps explain why past changes in the level of rates predict high excess returns on long-term bonds and yield-curve flattening in the post-2000 data. In Panel C column (5), we see that the refinancing disincentive attracts a positive and significant coefficient when forecasting the 3-month excess returns on 10-year bonds, $rx_{t\to t+3}^{(10)}$, after 2000. By contrast, the corresponding coefficient in column (2) for the pre-2000 sample is near zero and insignificant. And, the difference between the coefficients in columns (2) and (5) is highly significant (p-value < 0.01). However, when we use both ($y_t^M - c_t^M$) and ($L_t - L_{t-6}$) to forecast $rx_{t\to t+3}^{(10)}$ in column (6), the coefficients on both variables decline noticeably relative to those in columns (4) and (5) where they are considered in isolation. This is precisely what we should expect if the refinancing channel plays an important role in explaining the short-lived excess sensitivity of long-term rates that we see in the post-2000 data.[23] Panels D and F show that DUR_t^M strongly forecasts $rx_{t\to t+3}^{(10)}$ and $rx_{t\to t+3}^{SLOPE}$ in the post-2000 sample with the expected signs.

Thus, the results in Table 5 suggest that the mortgage refinancing channel helps explain why long-term U.S. yields increasingly appear to temporarily overreact to movements in short rates.

4.2 Investor overextrapolation

Several recent papers, including Cieslak (2018), Giglio and Kelly (2018), Brooks et al. (2017) and D'Arienzo (2020) have argued that some bond investors make biased forecasts of future interest rates, generating rate-amplifying shifts in the demand for long-term bonds. Positive shocks to short rates lead overly-extrapolative investors to overestimate the future path of short-term interest rates and, therefore, to demand fewer long-term bonds. This means that the quantity of long-term bonds that must be held by unbiased investors rises when short rates rise, leading to an increase in term premium and generating excess sensitivity of long-term yields. If these expectational errors are transitory or if the arbitrage response to these rate-amplifying demand shocks is slow-moving, then investor overextrapolation will create a short-lived market imbalance, leading long-term yields to temporarily overreact to changes in short rates. Furthermore, the impact of investor overextrapolation on the U.S. bond market may have grown in recent decades since fixed-income mutual fund investors, who are thought to be quite prone to overextrapolation, have become far more important players in the bond market.

Modeling the mechanism. To model investor overextrapolation, we assume that there are some investors who have "diagnostic expectations" about short-term interest rates in the sense that they

[23] If mortgage refinancing was the only source of rate amplification in the U.S. bond market and there was no slow-moving arbitrage capital, then ($y_t^M - c_t^M$) should be a sufficient statistic for bond risk premia and should completely drive out ($L_t - L_{t-6}$) in a horse race specification. However, if mortgage refinancing was one of several amplification mechanisms, or if arbitrage capital was slow moving, then one expected both ($y_t^M - c_t^M$) and ($L_t - L_{t-6}$) to attract meaningful coefficients.

"overweight future outcomes that have become more likely in light of incoming data" (Bordalo et al., 2017). In contrast to most recent work on diagnostic expectations—which takes a representative agent approach—we adopt a heterogenous agent approach, enabling us to study the dynamic arbitrage response of unbiased bond investors to these rate-amplifying demand shocks.

Formally, we assume that fraction f of investors have diagnostic expectations about short rates. Diagnostic investors' demand for long-term bonds is $h_t = \tau(E_t^D [rx_{t+1}] / Var_t^D [rx_{t+1}])$ where $E_t^D [rx_{t+1}]$ denotes diagnostic investors' biased expectation of future bond excess returns. Fraction $(1-f)$ of a bond investors have fully rational expectations about short-term interest rates. Of these rational investors, fraction q are fast-moving with demands given by equation (3.3) and fraction $(1-q)$ are slow-moving and only rebalance the portfolios every $k \geq 2$ periods with demands given by (3.4). We assume the gross supply of long-term bonds is constant over time at $s_t = \bar{s}$.

Following Maxted (2020), we assume that diagnostic investors' expectation of the transient component of short-term rates ($i_{T,t}$) is

$$E_t^D [i_{T,t+1}] = \rho_T i_{T,t} + \theta \cdot [i_{T,t} - (\rho_T - \kappa_T) \sum_{j=0}^{\infty} \kappa_T^j i_{T,t-j-1}], \quad (4.4)$$

where $\theta \geq 0$ and $\kappa_T \in [0, \rho_T]$. The parameter θ governs the extent to which diagnostic expectations depart from full rationality ($\theta = 0$) and κ_T governs the persistence of investors' mistaken beliefs about short rates.[24] When $\theta > 0$ and $\kappa_T < \rho_T$, equation (4.4) says that diagnostic investors overestimate $i_{T,t+1}$ when $i_{T,t}$ has recently risen. Thus, investor overextrapolation leads to a model that is very similar to the reduced-form specification for net bond supply in equation (3.9). We adopt an analogous specification for diagnostic investors' expectations of the persistent component of short rates ($i_{P,t}$), but assume for simplicity that diagnostic investors form rational forecasts of all other relevant state variables.

The strength of the investor overextrapolation channel is given by $f\theta$—i.e., by the mass of diagnostic investors (f) times the extent to which their expectations depart from perfect rationality (θ). Why might $f\theta$ have risen in recent decades? While many bond investors may have a tendency to overextrapolate past changes in interest rates, it is natural to think that this tendency may be more pronounced amongst investors in fixed-income mutual funds. Indeed, there is a long literature arguing that mutual fund investors—who are predominantly households and smaller institutions—tend to be less financially sophisticated and more prone to common psychological biases than larger institutional investors (Barberis et al., 1998; Frazzini and Lamont, 2008; Dichev, 2007). Furthermore, mutual funds have become more important players in the U.S. bond market in recent decades. Based on data from Federal Reserve's Financial Accounts, mutual funds' share of Treasury and MBS holdings has gradually risen from roughly 5% in the early 1990s to nearly 10% today. More recently, mutual funds have rapidly gained share in the corporate bond market, rising from a 7% share in early 2009 to over 20% today. Thus, even if individual mutual fund investors have not become more extrapolative since 2000 (i.e., if the parameter θ in our model has not changed), this group of extrapolation-prone investors has become more important in the bond market (corresponding to a rise in f).

In this setting, we can demonstrate the following result:

[24] In the limit where $\kappa_T = 0$, $E_t^D [i_{T,t+1}] = \rho_T i_{T,t} + \theta \varepsilon_{T,t}$, so investors' mistakes ($\theta \varepsilon_{T,t}$) are serially uncorrelated over time. In the opposite limit where $\kappa_T = \rho_T$, $E_t^D [i_{T,t+1}] = \rho_T i_{T,t} + \theta i_{T,t}$, so investors' mistakes ($\theta i_{T,t}$) are just as persistent as $i_{T,t}$.

Proposition 3. ***Investor over-extrapolation model.*** *Consider the over-extrapolation model and for simplicity suppose $\rho_T = \rho_P$ and $\kappa_T = \kappa_P$. When $f\theta > 0$, long-term yields are excessively sensitive to short-term rates relative to the expectations hypothesis. When $f\theta > 0$ and $\kappa_T = \rho_T$, this excess sensitivity is only horizon-dependent—i.e., the model-implied regression coefficient β_h in equation (3.11) only declines with horizon h—when unbiased arbitrage capital is slow moving ($q < 1$). By contrast, when $f\theta > 0$ and $\kappa_P < \rho_T$, β_h declines with horizon h even if all arbitrage capital is fast-moving ($q = 1$).*

When $f\theta > 0$ and $\kappa_P < \rho_P$, overextrapolation generates transitory rate-amplifying demand shocks, giving rise to frequency-dependent excess sensitivity even without slow-moving capital. However, this frequency-dependent excess sensitivity becomes more pronounced when these demand shocks are met by a slow-moving arbitrage response from unbiased investors. Thus, we are best able to quantitatively match the post-2000 behavior of the β_h coefficients using calibrations of our overextrapolation model in which (i) $f\theta$ has risen from its pre-2000 level and (ii) arbitrage by unbiased investors is slow-moving.

In addition to explaining the behavior of the β_h coefficients, our model of investor overextrapolation predicts that: (1) the bond holdings of extrapolative investors, h_t, are low when interest rates have recently risen, (2) the level of extrapolative investors' bond holdings negatively predicts future excess returns on long bonds (i.e., $E_t[rx_{t+1}]$ is low when h_t is high), and (3) the level of extrapolative investors' bond holdings predicts subsequent yield-curve steepening (i.e., $E_t[\Delta S_{t+1}]$ is high when h_t is high).

Evidence from bond mutual fund flows. To assess whether investor overextrapolation has contributed to high-frequency excess sensitivity, we obtain monthly data from 1984 to 2019 on the total net assets of and the net dollar flows into taxable bond mutual funds from the Investment Company Institute. We then compute the 3-month cumulative percentage flow into bond funds, $\%FLOW_{t-3\to t}$.[25] Using bond fund flows as a proxy for the rate-amplifying demand of extrapolative investors in our model (h_t), we first estimate equation (4.2) with $X_t = \%FLOW_{t-3\to t}$ for the pre-2000 and post-2000 samples. The results are presented in Panel A of Table 6 and show that bond mutual funds tend to suffer investor outflows when short-term interest rates decline. This result is consistent with the vast literature on return-chasing behavior by mutual fund investors (Warther, 1995; Sirri and Tufano, 1998). Interesting, this relationship is actually stronger in the pre-2000 sample than in the post-2000 sample, consistent with other evidence that mutual fund flows have become less performance sensitive in recent years. However, the importance of mutual funds within the bond market has increased meaningfully since 2000.

In Panels B and C, we estimate equation (4.3) with $X_t = \%FLOW_{t-3\to t}$ for the pre-2000 and post-2000 samples. As shown in column (5), past mutual fund inflows predict low future excess returns on 10-year bonds (Panel B) and future yield-curve steepening (Panel C) in the post-2000 data. By contrast, as shown in column (2), there are no such relationships in the pre-2000 sample. However, when we use both $(L_t - L_{t-6})$ and $\%FLOW_{t-3\to t}$ to forecast returns in the in column (6), the coefficients on both variables decline meaningfully relative to the specifications shown in columns (4) and (5) where they are considered in isolation. As above, this is what one would expect if investor extrapolation plays a role in explaining why long-term yields temporarily overreact to short rates in the post-2000 data.[26]

[25] Formally, letting $\%FLOW_t = FLOW_t/TNA_{t-1}$ denote the percentage flow in month t, the 3-month cumulative percentage flow is $\%FLOW_{t-3\to t} = (1 + \%FLOW_t)(1 + \%FLOW_{t-1})(1 + \%FLOW_{t-2}) - 1$.

[26] The CRSP Mutual Funds Database does not have monthly TNA for most bond mutual funds until 1991. Therefore, we use ICI data to construct our main proxy for flows into bond mutual funds. However, we obtain similar results for the post-2000 period if we construct bond mutual fund flows using CRSP. In the post-2000 data, ICI divides taxable bond funds

In summary, the results in Table 6 suggest that investor overextrapolation helps explain why long-term U.S. yields have temporarily overreacted to movements in short rates in recent decades.

4.3 Investors who reach for yield

A final source of rate-amplifying demand that may help explain the horizon-dependent excess sensitivity we observe in the post-2000 data appeals to the idea that there is a growing set of investors who "reach for yield" when short rates decline. Specifically, according to the reaching-for-yield channel (Hanson and Stein, 2015), negative shocks to short rates boost the demand for long-term bonds from "yield-seeking investors" who care about the current yield on their portfolios over and above their expected portfolio returns. Thus, holding fixed the gross supply, the net supply of long-term bonds that must be held by fast- and slow-moving arbitrageurs declines when short rates fall, leading term premia to decline when short rates fall. More generally, low short rates may increase investors' risk appetites through a variety of nonstandard channels, either frictional (Rajan, 2005) or behavioral (Lian et al., 2017), thereby depressing term premia.

Modeling the mechanism. To model reaching for yield, we assume that fraction f of bond investors are "yield-seeking" and have non-standard preferences as in Hanson and Stein (2015). The idea is that, for either behavioral or institutional reasons, these investors care about the current yield on their portfolios over and above their expected portfolio returns. Specifically, yield-seeking investors' demand for long-term bonds is:

$$h_t = \tau \frac{y_t - i_i}{Var_t\left[rx_{t+1}\right]}. \tag{4.5}$$

Since $E_t\left[rx_{t+1}\right] = (y_t - i_i) - (\phi/(1-\phi)) \cdot E_t\left[y_{t+1} - y_t\right]$, equation (4.5) says that yield-seeking investors are only concerned with the current income or "carry" from holding long-term bonds and neglect any expected capital gains and losses from holding long-term bonds. And, because expectations-hypothesis logic implies that long-term yields are expected to rise when short rates are low, equation (4.5) implies that yield-seeking investors have an elevated demand for long-term bonds when short rates are low.[27] As above, a mass $(1-f)$ of a bond investors are expected-return-oriented and have standard mean-variance preferences. Of these standard expected-return-oriented investors, fraction q are fast-moving with demands given by equation (3.3) and fraction $(1-q)$ are slow-moving with demands given by (3.4). Finally, we assume the gross supply of long-term bonds is constant over time. Thus, the strength of the reaching-for-yield channel is given by f—i.e., the fraction of investors who are yield-seeking.

Why might investors' tendency to reach for yield—corresponding to a rise in f—have grown stronger since 2000? Lian et al. (2017) provide experimental evidence that there is a non-linear relationship between reaching-for-yield behavior and the prior level of rates—i.e., the tendency to take on greater risk when short-term riskless rates decline becomes more pronounced when the prior level of short rates is already low. Building on Prospect Theory (Kahneman and Tversky, 1979), they argue that reaching

into (i) investment grade, (ii) government, (iii) multisector, (iv) high yield, and (v) global bond funds. We obtain similar results post-2000 if we only use flows into investment grade funds which represent between 40% and 60% of total taxable bond funds or combined flows into categories (i), (ii), and (iii).

[27] Formally, $E_t\left[y_{t+1} - y_t\right] > 0$ when i_t is low, so $(y_t - i_i) > E_t\left[rx_{t+1}\right]$ when i_t is low. As a result, $h_t = \tau\left((y_t - i_i)/Var_t\left[rx_{t+1}\right]\right) > \tau\left(E_t\left[rx_{t+1}\right]/Var_t\left[rx_{t+1}\right]\right) = b_t$ when i_t is low. Conversely, $h_t < b_t$ when i_t is high.

for yield becomes more pronounced as rates fall further below some reference level that investors are accustomed to based on past experience.[28] In a similar vein, Campbell and Sigalov (2020) build a non-standard model of institutional portfolio choice which incorporates the kinds of "sustainable spending" constraints routinely employed by endowments, sovereign wealth funds, and trusts. They show that these institutional constraints generate reaching-for-yield behavior and that reaching-for-yield is strongest when the level of short-term rates is initially low. In summary, prior research suggests the reaching-for-yield channel may have grown stronger in recent years as interest rates have reached historically low levels.

Using this model, we can then show:

Proposition 4. ***Investor reaching for yield model.*** *Consider the reaching-for-yield model and for simplicity suppose $\rho_T = \rho_P$. When $f > 0$, long-term yields are excessively sensitive to short-term rates when judged relative to the expectations hypotheisis. However, this excess sensitivity is only horizon-dependent—i.e., the model-implied regression coefficient β_h in equation (3.11) only declines with horizon h—when arbitrage capital is slow moving ($q < 1$).*

As above, our model of investor reaching-for-yield also predicts that: (1) the bond holdings of yield-seeking investors, h_t, are low when interest rates are high, (2) the level of yield-seeking investors' bond holdings negatively predicts future excess returns on long-term bonds, and (3) the level of yield-seeking investors' bond holdings predicts subsequent yield-curve steepening.

Since these nonstandard shifts in demand are tied to the level of short interest rates as opposed to the recent change in short rates, reaching for yield generates persistent shifts in rate-amplifying investor demand. Thus, while reaching-for-yield can generate excess sensitivity relative to the expectations hypothesis (Hanson and Stein, 2015), in the absence of slow-moving capital, it does not generate horizon-dependent excess sensitivity. And, while the combination of reaching-for-yield and slow-moving capital generates horizon-dependent sensitivity, our calibrations of our reaching-for-yield model struggle to quantitatively match the profile of β_h witnesses in the post-2000 data. Thus, it is not obvious that investor reaching-for-yield can explain why excess sensitivity has become so pronounced at high frequencies in recent decades. Going even further, it seems natural to posit that reaching-for-yield itself may be a slow-moving phenomenon—i.e., investors may only gradually take on greater portfolio risk following a decline in short-term interest rates. If true, this would further weaken the ability of reaching-for-yield to explain why the excess sensitivity of long-term yields has become most pronounced at very high frequencies.

Evidence from sectoral bond market flows. To empirically assess this reaching-for-yield explanation for our findings, we use quarterly data from the Federal Reserve's Financial Accounts on the aggregate net bond acquisitions by insurers (life plus property-casualty), pension funds (private plus state and local), and banks to construct empirical proxies for the bond demand of yield-seeking investors, h_t. We focus on these highly-regulated financial intermediaries since prior research has argued that they are most likely to be concerned about the current yield on their portfolios and, therefore, to reach for yield when interest rates decline.[29] For intermediaries in sector i, we compute the percentage bond flows

[28] Relatedly, Lian et al. (2017) argue that, because people tend to think in proportions as opposed to in differences small return differentials loom larger in investors' minds when short rates are lower (see Bordalo et al. (2013)).

[29] Insurers and banks are generally not required to include changes in the mark-to-market value of their portfolios in their reported earnings, which may give way to yield-seeking behavior. For prior work on reaching-for-yield by insurers, see Becker and Ivashina (2015). For pension funds, see Lu et al. (2019). For banks, see Maddaloni and Peydró (2011), Hanson and Stein (2015), and Drechsler et al. (2018).

in quarter t as $\%FLOW_{i,t} = FLOW_{i,t}/HOLD_{i,t-1}$, where $FLOW_{i,t}$ denotes net bond acquisitions by intermediaries in sector i during quarter t and $HOLD_{i,t-1}$ is bond holdings at the end of quarter $t-1$. Bonds here include the sum of U.S. Treasury securities, agency debt and GSE-guaranteed mortgage-backed securities, and corporate bonds. Thus, our construction of these sector-level flows roughly mimics the construction of bond mutual fund flows above.

In Table 7, we then estimate quarterly regressions that are analogous to equations (4.2) and (4.3) using these sector-level bond flows $\%FLOW_{i,t}$ as X_t. We report the results separately for the pre-2000 and post-2000 samples. As shown in Panel A of Table 7, in the post-2000 data, we find little evidence that recent increase in short-term rates lead to a reduction in bond purchases by insurers, pensions, and banks. Furthermore, in Panels B and C, we find little evidence that bond purchases by these intermediaries predict low excess returns on long-term bonds in the following quarter or subsequent yield-curve steepening as would be suggested by a reaching-for-yield explanation for our findings.

In summary, we find little evidence that reaching-for-yield plays a major role in driving the kind of short-lived overreaction of long-term yields to changes in short rates that we see since 2000. To be sure, this lack of evidence does not imply that reaching-for-yield plays an unimportant role in determining financial market risk premia more generally, especially at lower frequencies. This negative conclusion only applies to the ability of reaching-for-yield to explain the sorts of transitory fluctuations in bond risk premia that underpin the horizon-dependent excess sensitivity observed in recent decades.

5 Implications

5.1 High-frequency identification

Our findings have clear implications for identification approaches based on the high-frequency responses of long-term yields to macroeconomic news and monetary policy announcements. Specifically, papers in the vast event-study literature often implicitly assume that one can directly infer the expected long-run effects of news shocks by looking at the high-frequency reactions of long-term asset prices—see e.g., Ederington and Lee (1993) and MacKinlay (1997). And, several recent papers—see e.g., Nakamura and Steinsson (2018) and Hördahl et al. (2015)—have made this assumption more explicitly. Intuitively, if changes in long-term yields during a tight window around a macro news announcement are governed by the expectations hypothesis, then the high-frequency reaction of long-term yields directly reveals the expected long-run effect of the associated news shock.[30] For instance, if the 10-year forward Treasury rate fell by 20 basis points in a short window around a Federal Reserve monetary policy announcement, then one would infer that this announcement led expected future short rates in 10 years time to drop by 20 basis points.

However, our evidence casts serious doubt on this assumption. If, as we argue, a large portion of the impact of news shock on long-term yields reflects rapidly-reverting shifts in term premia, then the short-run impact of news shocks on long-term yields will differ meaningfully from their expected long-run impact on future short rates. As a result, the high-frequency responses of long-term yields are likely to provide

[30] Thus, the event-study approach provides approximately unbiased estimates if bond risk premia only fluctuate at business-cycle frequencies since, in that case, the term premium component of long-term yields would remain largely unchanged in short windows around macro news announcements.

a highly biased estimate of the longer-run impact of news announcements. Fortunately, it is relatively straightforward to eliminate this bias: one needs to use an methodology that does not assume that we can directly infer the expected long-run effects of news shocks simply by looking the high-frequency reactions of long-term asset prices.[31] Of course, these unbiased approaches lead to far less precise estimates, so economists face a steep bias-variance trade-off. Specifically, the short-run market impact of news on long-term yields can be estimated very precisely, but these are likely to be biased estimates of the longer-run impact that is typically of greatest interest to macroeconomists and policymakers.

Still, it is conceivable that changes in 1-year yields that are associated with news announcements are different, and do not trigger transient movements in term premia, as argued by Nakamura and Steinsson (2018) and Hördahl et al. (2015). To get some direct evidence on this question, we form an "economic news index" for month t, $News_t$, by cumulating daily changes in 1-year yields within month t on days with important macroeconomic news announcements. Our data on the timing of macroeconomic news announcements comes from Money Market Services/Action Economics and begins in 1980. The announcements we consider are: FOMC announcements, the employment situation report, retail sales, durable goods orders, new and existing home sales, housing starts, CPI, and PPI. We then estimate the following predictive regression for the subsequent change in 10-year forward rates:

$$f_{t+h}^{(10)} - f_t^{(10)} = \delta_0 + \delta_1 L_t + \delta_2 S_t + \delta_3 (L_t - L_{t-1}) + \delta_4 (S_t - S_{t-1}) + \lambda \cdot News_t + \varepsilon_{t+h}, \quad (5.1)$$

where L_t and S_t denote the level and slope of the yield curve at the end of month t. In other words, equation (5.1) simply adds $News_t$ to the Jorda (2005) local projections that we previously estimated in equation (2.3). Table 8 shows the results for both pre- and post-2000 samples and for $h =$ 3-, 6-, 9-, and 12- month future changes in forward rates.

In Panel A, we omit $News_t$, so the estimates are (essentially) the same as those in Table 3.[32] As previously shown, past increases in short-term rates are associated with predictable future declines in long-term forward rates in the post-2000 data, but there is no such tendency in the pre-2000 data. In Panel B, we add $News_t$, but omit the prior changes in level and slope. We see that positive values of the news index predict subsequent declines in long-term forwards in the post-2000 data. Indeed, the coefficients on $News_t$ in Panel B are similar to those on $L_t - L_{t-1}$ in Panel A.

In Panel C, we include $News_t$ as well as the prior changes in level and slope as an independent variables. The goal is to see if shifts in short-term rates on announcement and non-announcement days have different implications for the expected future change in long-term forward rates. Once we control for the total change in short rates in month t, $L_t - L_{t-1}$, we find that the coefficient on $News_t$ is small and insignificant, indicating that the response of long-term forwards rates on announcement days is just as likely to reverse as the response on non-announcement days.

In Panel D, we break $News_t$ into two pieces—one reflecting changes in short-term rates on FOMC

[31]For instance, one could estimate the long-run effects of a shock using a Structural VAR in which high-frequency asset prices movements are used as external instruments for monetary policy or other shocks. And, then IV-estimates of the SVAR would be used to trade out the long-run dynamic effects of the shock—see, e.g., Gertler and Karadi (2015) and Eberly et al. (2020). Similarly, one could estimate the long-run effects of news shocks using Jorda (2005) style "local projections" in which one regresses outcomes at future horizons on high-frequency market reactions to news.

[32]Since the pre-2000 sample in Table 8 runs from 1980-1999, the pre-2000 results in Table 8 (Panel A) differ slightly from those in Table 3 (Panel B) where the data begins in 1971.

announcement days ($News_{t,FOMC}$) and one for all other news announcements ($News_{t,Other}$)—to see if FOMC announcements differ from other macro news announcements. We exclude the 1980-81 monetary targeting regime and thus FOMC announcement dates begin in 1982. As in Panel C, we include $L_t - L_{t-1}$ as an independent variable. If anything, the results in Panel D suggest that, since 2000, changes in short rates on FOMC announcement days are *more* likely to be followed by subsequent reversals in long-term forward rates than changes on non-announcement days.

In summary, we conclude that, since 2000, the high-frequency response of long-term rates to economic news appears to dissipate meaningfully at lower frequencies, posing challenges to interpreting high-frequency yield curve responses in more recent data.

5.2 Monetary policy transmission

Our results also have important implications for the transmission of monetary policy. Central banks conduct conventional monetary policy by adjusting short-term nominal rates. According to the standard New Keynesian view (Gali, 2008), changes in short-term nominal rates affect short-term real rates because of nominal rigidities. And, the resulting shifts in short-term real rates affects long-term real rates via the expectations hypothesis, which in turn influence household spending and firm investment. Stein (2013) points out that the excess sensitivity of long-term yields—whereby shocks to short rates move term premia on long-term bonds in the same direction—should strengthen the effects of monetary policy relative to the canonical view. Stein (2013) refers to this as the "recruitment" channel of monetary transmission.

In the framework we developed in Section 3, the strength of this recruitment channel at business-cycle frequencies (e.g., over a 1 to 3-year horizon) depends on (i) the relative strength of the relevant supply-and-demand-based amplification mechanisms (i.e., the size of C relative to investor risk tolerance τ) and (ii) the persistence of the associated supply-and-demand shocks. Specifically, when ρ_s is well below ρ_T as under the mortgage refinancing interpretation of C, the associated shifts in term premia would be quite transient and would likely have only modest effects on investment and spending at medium-run frequencies. (A caveat here is that reductions in short rates that trigger mortgage refinancing waves may only temporarily lower term premia, but the effect of refinancing waves on distribution of household savings, and hence consumption, may persist long after term premia have reverted in heterogeneous agent settings.[33]) By contrast, when $\rho_s \approx \rho_T$ as under the reaching-for-yield interpretation of C, the shifts in term premia would be more persistent and likely to have larger effects on aggregate demand.

Our empirical results indicate that a significant part of the influence of short-term rates on term premia is quite transitory. Thus, our findings suggest that recruitment channel may be smaller than one would conclude based on a simplistic extrapolation of the high-frequency response of term premia to policy shocks documented by Hanson and Stein (2015), Gertler and Karadi (2015), and Gilchrist et al. (2015). More generally, our findings suggest that central banks should heed the way that monetary policy impacts financial conditions at business-cycle frequencies, but should focus less on the immediate market response to their announcements since, in the presence of slow-moving capital, much of the latter may be

[33]Indeed, to the extent that mortgage refinancing plays an important role in U.S. monetary policy transmission as in recent heterogeneous agent models (Beraja et al., 2018; Berger et al., 2018; Eichenbaum et al., 2018; Wong, 2019), then even short-lived excess sensitivity may make monetary policy more potent than in a world where long rates are not excessively sensitive. Specifically, the refinancing channel suggests that reductions in short rates may trigger *larger* refinancing waves and, hence provide greater monetary stimulus, than one would expect in a world where term premia are constant.

quite transitory. In this way, our findings lend support to the argument in Stein and Sunderam (2018) that the Federal Reserve has become too focused on high-frequency asset price movements.

5.3 Bond market "conundrums"

Third, our findings can help explain the rising prevalence of bond market episodes like the one that former Federal Reserve Chairman Greenspan famously called the "conundrum"—the period after June 2004 when the Fed raised short-term rates, but longer-term yields declined. This "conundrum" was first noted in Greenspan (2005) and has been explored in many papers, including Backus and Wright (2007).

Consistent with the weaker low-frequency sensitivity of long-term rates in recent years, "conundrum" episodes—defined as 6-month periods where short- and long-term rates move in *opposite* directions—have grown increasingly common. Specifically, since 2000, 1- and 10-year nominal Treasury yields have moved in the *opposite* direction in 37% of all 6-month periods. By contrast, from 1971 to 1999, the corresponding figure was 18%, and the difference is statistically significant (p-val < 0.001).

Here we show that the non-Markovian dynamics documented in Section 2—the fact that past changes in the level of rates increasingly predict a future flattening of the yield curve—help explain several noteworthy "conundrums." Figure 7 plots 1-year and 10-year Treasury rates around three widely discussed "conundrums": Greenspan's original 2004 "conundrum," 2008 which was a "conundrum in reverse," and the 2017 "conundrum." In all three cases, 1-year and 10-year yields moved in opposite directions.

Consider Greenspan's original 2004 "conundrum." To draw the link between non-Markovian yield-curve dynamics and this "conundrum," we use the system of predictive equations for level and slope from Table 2. Starting in May 2004, we simulate the counterfactual path of 10-year yields that would have prevailed if, in the post-2000 sample, the slope of the yield curve had not responded to past changes in the level. To do so, we take the unrestricted estimates of the predictive equation (2.1b) for slope from column (6) in Table 2 and the restricted estimates from column (4) which constrain past changes to have no effect ($\delta_{3S} = \delta_{4S} = 0$). Starting in May 2004, we generate the counterfactual path of 10-year yields that would have obtained if $\delta_{3S} = \delta_{4S} = 0$. We hold the level factor at its actual value and use the residuals from the unrestricted regression in column (6), but set the parameters for the slope equation to their estimated values from the *restricted* regression in column (4).

The top panel of Figure 7 plots the actual 1- and 10-year yields over this 2004 conundrum period along with the 10-year yield under this counterfactual scenario. Had the slope not responded to lagged changes in the level of the yield curve, Figure 7 shows that, instead of falling, 10-year yields would have risen in 2004. The next two panels repeat this exercise for the 2008 "conundrum in reverse" (starting in December 2007) and the 2017 "conundrum" (starting in November 2016). If the slope had not responded to past changes in level, 10- and 1-year yields would have moved in the same direction in both cases.

5.4 Affine term structure models

Finally, we explore the implications of our results for affine term structure models which are a widely-used, reduced-form tools for understanding the term structure of bond yields (Duffee, 2002; Duffie and Kan, 1996). In these models, the n-year zero coupon yield is $y_t^{(n)} = \alpha_{0(n)} + \boldsymbol{\alpha}'_{1(n)}\mathbf{x}_t$, where $\mathbf{x}_t$ is a vector of state variables and the $\alpha_{0(n)}$ and $\boldsymbol{\alpha}_{1(n)}$ coefficients satisfy a set of recursive equations. In the Internet

Appendix, we apply the estimation methodology of Adrian et al. (2013) and fit affine term structure models using the first K principal components of 1- to 10-year yields as the state variables $\mathbf{x}_t$. We show that standard affine models—models that are Markovian with respect to these current yield-curve factors—cannot fit our key finding that the sensitivity of long rates to short rates β_h declines so strongly with horizon h in the post-2000 data. This remains so even if we estimate models that include many (e.g., $K = 5$) current yield-curve factors as state variables. However, we show that our key finding is consistent with non-Markovian term structure models in which past lags of the yield-curve factors are treated as "unspanned state variables."[34]

6 Conclusion

The strong sensitivity of long-term interest rates to changes in short rates is a long-standing puzzle. In this paper, we have shown that since 2000 this sensitivity has become even stronger at high frequencies. By contrast, this sensitivity has fallen significantly when looking at low-frequency changes. As a result, in the post-2000 data, the sensitivity of long-term rates to changes in short-term rates declines steeply with the horizon over which these changes are computed.

Before 2000, long-term interest rates were quite sensitive to short-term interest rates because inflation expectations were relatively unanchored, making short rates highly persistent. Since 2000, the sensitivity of long-term rates has become horizon-dependent and arises because past increases in short-term rates temporarily raise the term premium component of long-term yields, leading long-term yields to temporarily overreact to changes in short rates. Consistent with this view, we show that, controlling for current yields, past changes in short rates predict (i) future yield-curve flattening, (ii) future declines in long-term yields and forwards, and (iii) high future excess returns on long-term bonds in the post-2000 data.

We proposed a simple model that can explain these puzzling facts. In our model, the post-2000 tendency of long-term yields to temporarily overreact to changes in short-term rates is explained by the combination of (i) rate-amplifying shifts in the demand for long-term bonds and (ii) a gradual arbitrage response to these demand shifts. We have presented evidence that two specific rate-amplifying demand mechanisms—mortgage refinancing waves and investor overextrapolation of past changes in short rates—each help explain the horizon-dependent excess sensitivity of long-term yields that we see since 2000.

Our findings have important implications for the recruitment channel of monetary policy transmission (Stein, 2013). In recent years this channel appears far more short-lived than one might conclude based on a simplistic reading of high-frequency evidence. More broadly, part of the high-frequency response of long rates to shocks to short rates represents term premium movements that tend to wear off quickly. Consequently, it is important to remember that event-study approaches only measure high-frequency responses of long-term rates to macroeconomic news and that the impact may often be more muted at the lower frequencies that are typically of greatest interest to macroeconomists and policymakers.

[34] An unspanned state variable is a variable that is useful for forecasting future bond yields and returns but that has no impact on the current yield curve (Duffee, 2002). To be clear, we do not argue that the past increase in the level of rates is *literally* unspanned. Instead, as discussed in the Internet Appendix, we think this variable is *close* to being unspanned.

References

ABRAHAMS, M., T. ADRIAN, R. K. CRUMP, E. MOENCH, AND R. YU (2016): "Decomposing real and nominal yield curves," *Journal of Monetary Economics*, 84, 182–200.

ADRIAN, T., R. K. CRUMP, AND E. MOENCH (2013): "Pricing the term structure with linear regressions," *Journal of Financial Economics*, 110, 110–138.

ANDREWS, D. W. (1993): "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, 821–856.

BACKUS, D. AND J. H. WRIGHT (2007): "Cracking the conundrum," *Brookings Papers on Economic Activity*, 1, 293–316.

BARBERIS, N., A. SHLEIFER, AND R. VISHNY (1998): "A model of investor sentiment," *Journal of Financial Economics*, 49, 307–343.

BECKER, B. AND V. IVASHINA (2015): "Reaching for yield in the bond market," *Journal of Finance*, 70, 1863–1902.

BEECHEY, M. J. AND J. H. WRIGHT (2009): "The high-frequency impact of news on long-term yields and forward rates: Is it real?" *Journal of Monetary Economics*, 56, 535–544.

BENNETT, P. R. P. AND S. PERISTIANI (2001): "Structural change in the mortgage market and the propensity to refinance," *Journal of Money, Credit, and Banking*, 33, 955–975.

BERAJA, M., A. FUSTER, E. HURST, AND J. VAVRA (2018): "Regional heterogeneity and the refinancing channel of monetary policy," *Quarterly Journal of Economics*, 134, 109–183.

BERGER, D. W., K. MILBRADT, F. TOURRE, AND J. VAVRA (2018): "Mortgage prepayment and path dependent effects of monetary policy," NBER Working Paper 25157.

BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2013): "Salience and consumer choice," *Journal of Political Economy*, 121, 803–843.

——— (2017): "Diagnostic expectations and credit cycles," *Journal of Finance*, 73, 199–227.

BROOKS, J., M. KATZ, AND H. LUSTIG (2017): "Post-FOMC announcement drift in U.S. bond markets," Working Paper.

CAMPBELL, J. R., C. L. EVANS, J. D. FISHER, AND A. JUSTINIANO (2012): "Macroeconomic effects of Federal Reserve forward guidance," *Brookings Papers on Economic Activity*, 2012, 1–80.

CAMPBELL, J. Y. AND R. J. SHILLER (1988): "Stock prices, earnings, and expected dividends," *Journal of Finance*, 43, 661–676.

——— (1991): "Yield spreads and interest rate movements: A bird's eye view," *The Review of Economic Studies*, 58, 495–514.

CAMPBELL, J. Y. AND R. SIGALOV (2020): "Portfolio Choice with Sustainable Spending: A Model of Reaching for Yield," NBER Working Paper 27025.

CHO, C.-K. AND T. J. VOGELSANG (2017): "Fixed-b inference for testing structural change in a time series regression," *Econometrics*, 5, 2.

CHOW, G. C. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591–605.

Cieslak, A. (2018): "Short-rate expectations and unexpected returns in Treasury bonds," *Review of Financial Studies*, 31, 3265–3306.

Cochrane, J. H. and M. Piazzesi (2002): "The Fed and interest rates—a high-frequency identification," *American Economic Review: Papers and Proceedings*, 92, 90–95.

——— (2005): "Bond risk premia," *American Economic Review*, 95, 138–160.

D'Arienzo, D. (2020): "Maturity increasing over-reaction and bond market puzzles," Working paper, Bocconi University.

Dichev, I. D. (2007): "What are stock investors' actual historical returns? Evidence from dollar-weighted returns," *American Economic Review*, 97, 386–401.

Domanski, D., H. S. Shin, and V. Shushko (2017): "The hunt for duration: Not waving but drowning?" *IMF Economic Review*, 65, 113–153.

Drechsler, I., A. Savov, and P. Schnabl (2018): "A model of monetary policy and risk premia," *Journal of Finance*, 73, 317–373.

Duffee, G. (2002): "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57, 405–443.

Duffee, G. R. (2013): "Forecasting interest rates," in *Handbook of Economic Forecasting, Volume 2*, ed. by G. Elliott and A. Timmermann, Elsevier.

Duffie, D. (2010): "Asset price dynamics with slow-moving capital," *Journal of Finance*, 65, 1238–1268.

Duffie, D. and R. Kan (1996): "Yield factor models of interest rates," *Mathematical Finance*, 64, 379–406.

Eberly, J. C., J. H. Stock, and J. H. Wright (2020): "The Federal Reserve's Current Framework for Monetary Policy: A Review and Assessment," *International Journal of Central Banking*, 16, 5–71.

Ederington, L. H. and J. H. Lee (1993): "How markets process information: News releases and volatility," *Journal of Finance*, 48, 1161–1191.

Eichenbaum, M., S. Rebelo, and A. Wong (2018): "State dependent effects of monetary policy: The Refinancing Channel," NBER Working Paper 25152.

Evans, C. L. and D. A. Marshall (1998): "Monetary policy and the term structure of nominal interest rates: Evidence and theory," *Carnegie-Rochester Conference Series on Public Policy*, 49, 53–111.

Fama, E. and R. R. Bliss (1987): "The information on long-maturity forward rates," *American Economic Review*, 77, 680–692.

Feunou, B. and J.-S. Fontaine (2014): "Non-Markov Gaussian term structure models: The case of inflation," *Review of Finance*, 18, 1953–2001.

Frazzini, A. and O. A. Lamont (2008): "Dumb money: Mutual fund flows and the cross-section of stock returns," *Journal of Financial Economics*, 88, 299–322.

Fuster, A., M. Plosser, P. Schnabl, and J. Vickery (2019): "The role of technology in mortgage lending," *Review of Financial Studies*, 32, 1854–1899.

Gali, J. (2008): *Monetary Policy, Inflation, and the Business Cycle*, Princeton, New Jersey: Princeton University Press.

Gertler, M. and P. Karadi (2015): "Monetary policy surprises, credit costs, and economic activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

Giglio, S. and B. Kelly (2018): "Excess volatility: Beyond discount rates," *Quarterly Journal of Economics*, 133, 71–127.

Gilchrist, S., D. López-Salido, and E. Zakrajšek (2015): "Monetary policy and real borrowing costs at the zero lower bound," *American Economic Journal: Macroeconomics*, 7, 77–109.

Greenspan, A. (2005): "Federal Reserve Board's semiannual Monetary Policy Report to the Congress," Testimony on February 16, 2005 before the Senate Committee on Banking, Housing, and Urban Affairs.

Greenwood, R., S. G. Hanson, and G. Y. Liao (2018): "Asset price dynamics in partially segmented markets," *Review of Financial Studies*, 31, 3307–3343.

Greenwood, R. and D. Vayanos (2014): "Bond supply and excess bond returns," *Review of Financial Studies*, 27, 663–713.

Grossman, S. J. and M. H. Miller (1988): "Liquidity and market structure," *Journal of Finance*, 43, 617–633.

Gürkaynak, R. S., B. Sack, and E. T. Swanson (2005): "The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models," *American Economic Review*, 95, 425–436.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007): "The U.S. Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54, 2291–2304.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2010): "The TIPS yield curve and inflation compensation," *American Economic Journal: Macroeconomics*, 2, 70–92.

Gurkaynak, R. S., B. P. Sack, and E. T. Swanson (2005): "Do actions speak louder than words? The response of asset prices to monetary policy actions and statements," *International Journal of Central Banking*, 1, 55–93.

Hansen, L. P. and R. J. Hodrick (1980): "Forward exchange rates as optimal predictors of future spot rates: An econometric analysis," *Journal of Political Economy*, 88, 829–853.

Hanson, S. G. (2014): "Mortgage convexity," *Journal of Financial Economics*, 113, 270–299.

Hanson, S. G. and J. C. Stein (2015): "Monetary policy and long-term real rates," *Journal of Financial Economics*, 115, 429–448.

Hördahl, P., E. M. Remolona, and G. Valente (2015): "Expectations and risk premia at 8:30AM: Macroeconomic announcements and the yield curve," BIS working paper 527.

Jorda, O. (2005): "Estimation and inference of impulse responses by local projections," *American Economic Review*, 95, 161–182.

Joslin, S., M. Preibsch, and K. J. Singleton (2014): "Risk premiums in dynamic term structure models with unspanned macro risks," *Journal of Finance*, 69, 1197–1233.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An analysis of decision under risk," *Econometrica*, 47, 263–292.

Kiefer, N. M. and T. J. Vogelsang (2005): "A new asymptotic theory for heteroskedasticity-autocorrelation robust tests," *Econometric Theory*, 21, 1130–1164.

Kuttner, K. N. (2001): "Monetary policy surprises and interest rates: evidence from the Fed Funds futures market," *Journal of Monetary Economics*, 47, 523–544.

Lian, C., Y. Ma, and C. Wang (2017): "Low interest rates and risk taking: Evidence from individual investment decisions," .

Litterman, R. and J. Scheinkman (1991): "Common factors affecting bond returns," *Journal of Fixed Income*, 1, 54–61.

Lu, L., M. Pritsker, A. Zlate, K. Anadu, and J. Bohn (2019): "Reach for yield by U.S. public pension funds," .

Lucca, D. O. and F. Trebbi (2009): "Measuring central bank communication: an automated approach with application to FOMC statements," .

MacKinlay, A. C. (1997): "Event studies in economics and finance," *Journal of Economic Literature*, 35, 13–39.

Maddaloni, A. and J.-L. Peydró (2011): "Bank risk-taking, securitization, supervision, and low interest rates: Evidence from the euro-area and the US lending standards," *Review of Financial Studies*, 24, 2121–2165.

Malkhozov, A., P. Mueller, A. Vedolin, and G. Venter (2016): "Mortgage risk and the yield curve," *Review of Financial Studies*, 29, 1220–1253.

Mankiw, N. G. and L. H. Summers (1984): "Do long term interest rates overreact to short-term interest rates?" *Brookings Papers on Economic Activity*, 1, 223–242.

Maxted, P. (2020): "A macro-finance model with sentiment," .

Nakamura, E. and J. Steinsson (2018): "High-frequency identification of monetary non-neutrality: The information effect," *Quarterly Journal of Economics*, 133, 1283–1330.

Newey, W. K. and K. D. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Piazzesi, M., J. Salomao, and M. Schneider (2015): "Trend and cycle in bond premia," Working Paper, Stanford University.

Rajan, R. (2005): "Has financial development made the world riskier?" *Jackson Hole Economic Symposium Proceedings*, 313–369.

Samuelson, P. A. (1947): *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.

Shiller, R. J. (1979): "The volatility of long-term interest rates and expectations models of the term structure," *Journal of Political Economy*, 87, 1190–1219.

Shiller, R. J., J. Y. Campbell, and K. L. Schoenholtz (1983): "Forward rates and future policy: Interpreting the term structure of interest rates," *Brookings Papers on Economic Activity*, 1, 173–217.

Shin, H. S. (2017): "How much should we read into shifts into long-dated yields?" U.S. Monetary Policy Forum, speech.

Sirri, E. R. and P. Tufano (1998): "Costly search and mutual fund flows," *Journal of Finance*, 53, 1589–1622.

Stein, J. C. (2013): "Yield-oriented investors and the monetary transmission mechanism," .

Stein, J. C. and A. Sunderam (2018): "The Fed, the bond market, and gradualism in monetary policy," *Journal of Finance*, 1015–1060.

Stock, J. H. and M. W. Watson (2007): "Why has U.S. inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39, 3–33.

Swanson, E. T. and J. C. Williams (2014): "Measuring the effect of the zero lower bound on medium- and longer-term interest rates," *American Economic Review*, 104, 3154–3185.

Vayanos, D. and J.-L. Vila (2020): "A preferred-habitat model of the term structure of interest rates," *Econometrica*, Forthcoming.

Warther, V. A. (1995): "Aggregate mutual fund flows and security returns," *Journal of Financial Economics*, 39, 209–235.

Wong, A. (2019): "Refinancing and the transmission of monetary policy to consumption," Working paper, Princeton University.

Table 1: Regressions of changes in long-term rates on short-term rates. This table reports the estimated regression coefficients from equations (1.1) and (1.2) for each reported sample. The dependent variable is the change in the 10-year U.S. Treasury zero-coupon yield or forward rate, either nominal, real or their difference (IC, or inflation compensation). The independent variable is the change in the 1-year nominal U.S. Treasury zero-coupon yield in all cases. Changes are considered with daily data, and with monthly data using monthly ($h = 1$), quarterly ($h = 3$), semi-annual ($h = 6$) and annual ($h = 12$) horizons. In the 1971-1999 monthly sample, time t runs from Aug-1971 to Dec-1999 and the number of monthly observations is 341 irrespective of h. In the 2000-2019 monthly sample, t runs from Jan-2000 to Dec-2019, so the number of monthly observations runs 239 from for $h = 1$ to 228 for $h = 12$. For $h > 1$, we report Newey-West (1987) standard errors are in brackets, using a lag truncation parameter of $\lceil 1.5 \times h \rceil$; for $h = 1$, we report heteroskedasticity robust standard errors. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Significance is computed using the asymptotic theory of Kiefer and Vogelsang (2005) which has better finite sample properties than traditional asymptotic theory.

Panel A. 10-year zero coupon yields and IC

	(1) Nominal	(2) Nominal	(3) Real	(4) IC
Daily	0.56***	0.87***	0.54***	0.33***
	[0.02]	[0.03]	[0.03]	[0.02]
Monthly	0.46***	0.66***	0.38***	0.26***
	[0.04]	[0.11]	[0.09]	[0.09]
Quarterly	0.48***	0.44***	0.22**	0.22*
	[0.04]	[0.07]	[0.10]	[0.12]
Semi-annual	0.50***	0.34***	0.21**	0.13
	[0.04]	[0.07]	[0.08]	[0.09]
Yearly	0.56***	0.23***	0.15**	0.08
	[0.05]	[0.05]	[0.06]	[0.05]
Sample	1971-1999	2000-2019	2000-2019	2000-2019

Panel B. 10-year instantaneous forward yields and IC

	(1) Nominal	(2) Nominal	(3) Real	(4) IC
Daily	0.39***	0.49***	0.31***	0.17***
	[0.03]	[0.04]	[0.03]	[0.03]
Monthly	0.29***	0.26*	0.18**	0.06
	[0.04]	[0.14]	[0.08]	[0.09]
Quarterly	0.31***	0.06	0.09*	-0.03
	[0.05]	[0.09]	[0.05]	[0.05]
Semi-annual	0.33***	-0.02	0.04	-0.06
	[0.06]	[0.08]	[0.04]	[0.05]
Yearly	0.39***	-0.13**	-0.02	-0.11**
	[0.07]	[0.06]	[0.04]	[0.04]
Sample	1971-1999	2000-2019	2000-2018	2000-2019

Table 2: Estimates of predictive equations for level and slope. This table reports the estimated regression coefficients from monthly predictive equations (2.1a) and (2.1b) for the Aug-1971 to Dec-1999 and Jan-2000 to Dec-2019 samples. Dependent variables are the level ($L_t \equiv y_t^{(1)}$) and slope ($S_t \equiv y_t^{(10)} - y_t^{(1)}$) of the U.S. Treasury zero-coupon yield curve. Heteroskedasticity robust standard errors are in brackets. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The table also shows AIC and BIC values (to be minimized) for each possible specification of the system of two equations. Lastly, the implied β_1 and β_{12} coefficients from equation (2.2) for each possible specification of the system are reported.

	Pre-2000			Post-2000		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Level						
L_t	0.98***	0.97***	0.96***	0.97***	0.98***	0.98***
	[0.02]	[0.02]	[0.02]	[0.01]	[0.01]	[0.01]
S_t	0.00	-0.01	-0.02	-0.02*	-0.01	-0.00
	[0.04]	[0.04]	[0.04]	[0.01]	[0.01]	[0.01]
$L_t - L_{t-6}$		-0.01	0.05		0.08***	0.06**
		[0.04]	[0.05]		[0.03]	[0.03]
$S_t - S_{t-6}$			0.13*			-0.03*
			[0.07]			[0.02]
Dependent Variable: Slope						
L_t	0.01	0.01	0.01	0.00	-0.01	-0.01
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
S_t	0.96***	0.96***	0.97***	0.98***	0.96***	0.96***
	[0.03]	[0.02]	[0.03]	[0.02]	[0.02]	[0.02]
$L_t - L_{t-6}$		0.00	-0.03		-0.11***	-0.12***
		[0.02]	[0.03]		[0.02]	[0.03]
$S_t - S_{t-6}$			-0.08			-0.02
			[0.05]			[0.03]
N	341	335	335	239	239	239
Implied β_1	0.46	0.46	0.46	0.66	0.71	0.71
Implied β_{12}	0.52	0.51	0.58	0.59	0.38	0.30
AIC	-5720.3	-5607.9	-5608.7	-4529.0	-4567.0	-4565.6
BIC	-5697.3	-5577.4	-5570.5	-4508.1	-4539.2	-4530.8
Sample	1971-1999	1972-1999	1972-1999	2000-2019	2000-2019	2000-2019

Table 3: Predictable yield-curve dynamics following an impulse to short-term interest rates. This table reports the estimated regression coefficients in equation (2.3) for the Aug-1971 to Dec-1999 and Jan-2000 to Dec-2019 samples. For $h = 3$, 6, 9, and, 12-months changes, we show results for 10-year yields ($z_t = y_t^{(10)}$), 10-year forward rates ($z_t = f_t^{(10)}$), level ($z_t = L_t$), and slope ($z_t = S_t$). We report Newey-West standard errors in brackets using a lag truncation parameter of $\lceil 1.5 \times h \rceil$. Significance: $^{*}p < 0.1$, ** $p < 0.05$, $^{***}p < 0.01$. Significance is computed using the asymptotic theory of Kiefer and Vogelsang (2005) which has better finite sample properties than traditional asymptotic theory.

	Pre-2000				Post-2000			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. var with h	3	6	9	12	3	6	9	12
Dependent variable: $y_{t+h}^{(10)} - y_t^{(10)}$								
$L_t - L_{t-1}$	-0.03	0.10	0.04	0.23*	-0.11	-0.36*	-0.41*	-0.20
	[0.12]	[0.13]	[0.11]	[0.12]	[0.13]	[0.19]	[0.20]	[0.20]
$S_t - S_{t-1}$	0.03	0.30	0.34	0.55**	-0.09	-0.16	-0.11	-0.16
	[0.18]	[0.19]	[0.20]	[0.26]	[0.15]	[0.17]	[0.15]	[0.20]
L_t	-0.06*	-0.11*	-0.17*	-0.23**	-0.07**	-0.14***	-0.19***	-0.23***
	[0.03]	[0.06]	[0.08]	[0.10]	[0.03]	[0.05]	[0.06]	[0.07]
S_t	-0.13*	-0.27**	-0.45***	-0.62***	-0.10**	-0.20**	-0.28**	-0.35**
	[0.07]	[0.11]	[0.16]	[0.19]	[0.05]	[0.09]	[0.11]	[0.12]
Adj.R^2	0.03	0.08	0.14	0.20	0.05	0.11	0.16	0.20
Dependent variable: $f_{t+h}^{(10)} - f_t^{(10)}$								
$L_t - L_{t-1}$	-0.08	0.09	-0.05	0.12	-0.30***	-0.52***	-0.71***	-0.74***
	[0.12]	[0.11]	[0.10]	[0.12]	[0.11]	[0.17]	[0.14]	[0.15]
$S_t - S_{t-1}$	0.01	0.18	0.16	0.34	-0.14	-0.11	-0.02	-0.17
	[0.18]	[0.19]	[0.20]	[0.22]	[0.16]	[0.20]	[0.19]	[0.21]
L_t	-0.04	-0.09	-0.14*	-0.19**	-0.03	-0.05	-0.06	-0.06
	[0.03]	[0.05]	[0.08]	[0.09]	[0.03]	[0.05]	[0.06]	[0.07]
S_t	-0.16**	-0.31***	-0.48***	-0.66***	-0.09*	-0.17*	-0.27**	-0.35**
	[0.06]	[0.11]	[0.15]	[0.19]	[0.05]	[0.09]	[0.11]	[0.14]
Adj.R^2	0.04	0.10	0.17	0.24	0.04	0.09	0.17	0.23
Dependent variable: $L_{t+h} - L_t$								
$L_t - L_{t-1}$	0.07	0.26	0.18	0.60***	0.66***	0.86**	1.33**	1.70***
	[0.24]	[0.21]	[0.20]	[0.19]	[0.17]	[0.38]	[0.53]	[0.59]
$S_t - S_{t-1}$	0.13	0.64**	0.57*	1.07**	-0.23**	-0.38*	-0.59*	-0.58
	[0.30]	[0.30]	[0.31]	[0.49]	[0.11]	[0.21]	[0.29]	[0.39]
L_t	-0.09*	-0.18**	-0.26**	-0.36***	-0.08**	-0.19**	-0.28***	-0.38**
	[0.05]	[0.08]	[0.11]	[0.12]	[0.03]	[0.07]	[0.11]	[0.15]
S_t	-0.00	-0.09	-0.24	-0.33	-0.03	-0.07	-0.04	-0.01
	[0.11]	[0.15]	[0.20]	[0.25]	[0.04]	[0.09]	[0.15]	[0.21]
Adj.R^2	0.03	0.08	0.12	0.17	0.20	0.26	0.36	0.42
Dependent variable: $S_{t+h} - S_t$								
$L_t - L_{t-1}$	-0.10	-0.16	-0.14	-0.37***	-0.77***	-1.23***	-1.74***	-1.90***
	[0.16]	[0.12]	[0.12]	[0.11]	[0.16]	[0.33]	[0.39]	[0.51]
$S_t - S_{t-1}$	-0.10	-0.34*	-0.23	-0.51	0.14	0.22	0.48*	0.43
	[0.17]	[0.18]	[0.16]	[0.30]	[0.12]	[0.22]	[0.27]	[0.32]
L_t	0.03	0.07*	0.10**	0.13***	0.01	0.05	0.09	0.15
	[0.03]	[0.03]	[0.04]	[0.04]	[0.03]	[0.07]	[0.10]	[0.14]
S_t	-0.12*	-0.18*	-0.21*	-0.29**	-0.07*	-0.13	-0.24	-0.34
	[0.07]	[0.10]	[0.11]	[0.13]	[0.04]	[0.09]	[0.13]	[0.20]
Adj.R^2	0.08	0.16	0.20	0.28	0.16	0.24	0.37	0.43
N	340	340	340	340	237	234	231	228
Sample	1971-1999	1971-1999	1971-1999	1971-1999	2000-2019	2000-2019	2000-2019	2000-2019

Table 4: Estimates of predictive equations for bond excess returns. This table reports the estimated regression coefficients in equation (2.5) using monthly data from the Aug-1971 to Dec-1999 and Jan-2000 to Dec-2019 samples. We report results various return forecast horizon (k). Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. For $k = 1$-month returns, we report heteroskedasticity robust standard errors are in brackets. For $k = 3$ and 6-month returns, we report Newey and West (1987) standard errors in brackets, using a lag truncation parameter of 5 and 9 months, respectively. In this case, p-values are computed using the asymptotic theory of Kiefer and Vogelsang (2005) which has better finite sample properties than traditional asymptotic theory.

Panel A: Pre-2000 sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var for k	1	1	1	3	3	3	6	6	6
Dependent Variable: $rx^{(10)}_{t\to t+k}$									
L_t	0.17	0.18	0.22*	0.53*	0.54	0.65*	0.99*	1.01	1.25**
	[0.13]	[0.13]	[0.12]	[0.31]	[0.33]	[0.34]	[0.57]	[0.62]	[0.60]
S_t	0.55**	0.61**	0.66***	1.57**	1.91***	2.07***	3.17***	3.62***	3.98***
	[0.24]	[0.24]	[0.24]	[0.64]	[0.65]	[0.64]	[1.04]	[1.13]	[1.09]
$L_t - L_{t-6}$		0.10	-0.18		0.59	-0.22		0.79	-1.00
		[0.19]	[0.29]		[0.49]	[0.54]		[0.69]	[0.90]
$S_t - S_{t-6}$			-0.57			-1.67**			-3.67***
			[0.44]			[0.81]			[1.30]
Adj.R^2	0.02	0.01	0.02	0.05	0.06	0.07	0.11	0.12	0.17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable: $rx^{LEVEL}_{t\to t+k}$									
L_t	-0.04*	-0.04*	-0.04**	-0.08*	-0.09*	-0.10**	-0.11**	-0.11**	-0.13**
	[0.02]	[0.02]	[0.02]	[0.04]	[0.05]	[0.05]	[0.05]	[0.05]	[0.05]
S_t	-0.06	-0.07*	-0.08**	-0.15*	-0.21**	-0.24***	-0.23***	-0.26***	-0.29***
	[0.04]	[0.03]	[0.04]	[0.09]	[0.08]	[0.08]	[0.09]	[0.09]	[0.09]
$L_t - L_{t-6}$		-0.01	0.04		-0.10	0.03		-0.04	0.09
		[0.03]	[0.04]		[0.08]	[0.08]		[0.09]	[0.09]
$S_t - S_{t-6}$			0.11*			0.26**			0.27*
			[0.06]			[0.11]			[0.14]
Adj.R^2	0.01	0.01	0.02	0.04	0.06	0.08	0.09	0.10	0.12
Dependent Variable: $rx^{SLOPE}_{t\to t+k}$									
L_t	0.02	0.02	0.03**	0.03	0.04	0.04	0.01	0.01	0.01
	[0.01]	[0.01]	[0.01]	[0.02]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
S_t	0.00	0.01	0.01	-0.01	0.02	0.03	-0.09	-0.11	-0.12
	[0.03]	[0.02]	[0.03]	[0.06]	[0.05]	[0.06]	[0.08]	[0.08]	[0.08]
$L_t - L_{t-6}$		0.00	-0.03		0.04	-0.00		-0.04	0.01
		[0.02]	[0.03]		[0.05]	[0.06]		[0.04]	[0.05]
$S_t - S_{t-6}$			-0.06			-0.10			0.11
			[0.05]			[0.09]			[0.10]
Adj.R^2	0.01	0.01	0.01	0.02	0.03	0.03	0.04	0.05	0.06
N	341	335	335	341	335	335	341	335	335
Sample	1971-1999	1972-1999	1972-1999	1971-1999	1972-1999	1972-1999	1971-1999	1972-1999	1972-1999

Panel B: Post-2000 sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var for k	1	1	1	3	3	3	6	6	6
Dependent Variable: $rx_{t \to t+k}^{(10)}$									
L_t	0.30***	0.32***	0.28**	0.73***	0.81***	0.71**	1.37***	1.50***	1.34**
	[0.11]	[0.11]	[0.11]	[0.27]	[0.28]	[0.30]	[0.44]	[0.47]	[0.51]
S_t	0.56***	0.63***	0.53**	1.41***	1.65***	1.42**	2.71***	3.07***	2.73***
	[0.21]	[0.20]	[0.21]	[0.48]	[0.52]	[0.58]	[0.83]	[0.93]	[0.95]
$L_t - L_{t-6}$		0.33	0.62*		0.98**	1.66***		1.33*	2.39**
		[0.24]	[0.33]		[0.46]	[0.61]		[0.73]	[1.07]
$S_t - S_{t-6}$			0.48			1.12			1.76
			[0.35]			[0.72]			[1.22]
Adj.R^2	0.03	0.03	0.03	0.08	0.10	0.12	0.16	0.18	0.20
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable: $rx_{t \to t+k}^{LEVEL}$									
L_t	-0.03***	-0.03***	-0.03***	-0.09***	-0.08***	-0.07***	-0.12***	-0.11***	-0.10***
	[0.01]	[0.01]	[0.01]	[0.03]	[0.02]	[0.02]	[0.04]	[0.03]	[0.03]
S_t	-0.04***	-0.03**	-0.03**	-0.10***	-0.07**	-0.05*	-0.13***	-0.08*	-0.05
	[0.01]	[0.01]	[0.01]	[0.03]	[0.03]	[0.03]	[0.04]	[0.04]	[0.04]
$L_t - L_{t-6}$		0.05**	0.03		0.12**	0.06		0.19***	0.12
		[0.02]	[0.02]		[0.06]	[0.06]		[0.07]	[0.07]
$S_t - S_{t-6}$			-0.04**			-0.09**			-0.11**
			[0.02]			[0.04]			[0.05]
Adj.R^2	0.06	0.10	0.11	0.13	0.21	0.23	0.21	0.37	0.39
Dependent Variable: $rx_{t \to t+k}^{SLOPE}$									
L_t	0.01	-0.00	0.00	0.02	-0.00	0.00	-0.01	-0.05	-0.04
	[0.01]	[0.01]	[0.01]	[0.03]	[0.03]	[0.03]	[0.06]	[0.04]	[0.05]
S_t	-0.01	-0.03*	-0.03	-0.05	-0.10**	-0.10*	-0.16*	-0.26***	-0.24***
	[0.02]	[0.02]	[0.02]	[0.05]	[0.05]	[0.05]	[0.08]	[0.08]	[0.08]
$L_t - L_{t-6}$		-0.09***	-0.10***		-0.24***	-0.25***		-0.36***	-0.40***
		[0.02]	[0.03]		[0.05]	[0.06]		[0.07]	[0.11]
$S_t - S_{t-6}$			-0.01			-0.02			-0.07
			[0.03]			[0.07]			[0.13]
Adj.R^2	0.00	0.06	0.06	0.03	0.19	0.18	0.08	0.28	0.28
N	239	239	239	237	237	237	234	234	234
Sample	2000-2019	2000-2019	2000-2019	2000-2019	2000-2019	2000-2019	2000-2019	2000-2019	2000-2019

Table 5: The role of mortgage refinancing: Evidence from mortgage-related quantities. Panels A and B report the estimated coefficients for equation (4.2) where the dependent variable is the mortgage refinancing disincentive: $X_t = y_t^M - c_t^M$ or the duration of the Barclay's MBS index: $X_t = DUR_t^M$. Panels C-F report the estimated coefficients using equation (4.3) to forecast 3-month returns using either the mortgage refinancing disincentive or the duration. All regressions are estimated using monthly data for the Jan-1976 to Dec-1999 and Jan-2000 to Dec-2019 samples. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. We report Newey and West (1987) standard errors in brackets, using a lag truncation parameter of 9 months in Panels A and B and 5 months in Panels C-F. p-values are computed using the asymptotic theory of Kiefer and Vogelsang (2005) which has better finite sample properties than traditional asymptotic theory.

	(1)	(2)	(3)	(4)	(5)	(6)		(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Dependent Variable: $y_t^M - c_t^M$							Panel B. Dependent Variable: DUR_t^M						
L_t	0.76***	0.76***	0.72***	0.23***	0.24***	0.20***		0.16***	0.16***	0.15***	0.05	0.08	0.01
	[0.07]	[0.06]	[0.06]	[0.05]	[0.05]	[0.05]		[0.04]	[0.04]	[0.04]	[0.09]	[0.09]	[0.08]
S_t	0.17	0.13	0.05	-0.03	0.00	-0.08		0.15**	0.08	0.06	-0.06	0.03	-0.11
	[0.10]	[0.11]	[0.10]	[0.07]	[0.06]	[0.06]		[0.08]	[0.08]	[0.08]	[0.14]	[0.12]	[0.12]
$L_t - L_{t-6}$		-0.07	0.18*		0.17**	0.42***			-0.13**	-0.06		0.44***	0.90***
		[0.08]	[0.10]		[0.08]	[0.07]			[0.05]	[0.06]		[0.13]	[0.12]
$S_t - S_{t-6}$			0.56***			0.42***				0.17**			0.76***
			[0.16]			[0.06]				[0.08]			[0.12]
Adj.R^2	0.86	0.86	0.88	0.52	0.55	0.67		0.23	0.29	0.30	0.02	0.16	0.36
Panel C. Dependent Variable: $rx_{t\to t+3}^{(10)}$							Panel D. Dependent Variable: $rx_{t\to t+3}^{(10)}$						
L_t	0.67*	1.10*	0.86	0.71**	0.21	0.34		0.67*	0.82**	0.88**	0.71**	0.68**	0.70**
	[0.36]	[0.59]	[0.59]	[0.30]	[0.30]	[0.34]		[0.36]	[0.36]	[0.36]	[0.30]	[0.28]	[0.30]
S_t	2.20**	1.83**	2.21***	1.42**	1.50***	1.57***		2.20***	1.98***	2.28***	1.42**	1.51***	1.55***
	[0.75]	[0.77]	[0.76]	[0.58]	[0.49]	[0.57]		[0.75]	[0.71]	[0.73]	[0.58]	[0.50]	[0.58]
$L_t - L_{t-6}$	-0.30		-0.25	1.66***		0.88		-0.30		-0.38	1.66***		0.71
	[0.56]		[0.60]	[0.61]		[0.79]		[0.56]		[0.55]	[0.61]		[0.83]
$S_t - S_{t-6}$	-1.70*		-1.55	1.12		0.35		-1.70*		-1.45	1.12		0.31
	[0.92]		[1.08]	[0.72]		[0.80]		[0.92]		[0.90]	[0.72]		[0.85]
$y_t^M - c_t^M$		-0.74	-0.27		2.34***	1.82*	DUR_t^M		-1.77*	-1.45*		1.31***	1.05**
		[0.71]	[0.77]		[0.80]	[0.98]			[0.94]	[0.81]		[0.35]	[0.51]
Adj.R^2	0.07	0.05	0.06	0.12	0.13	0.14		0.07	0.07	0.08	0.12	0.14	0.14
Panel E. Dependent Variable: $rx_{t\to t+3}^{SLOPE}$							Panel F. Dependent Variable: $rx_{t\to t+3}^{SLOPE}$						
L_t	0.04	0.06	0.04	0.00	0.08**	0.05*		0.04	0.05	0.05*	0.00	0.02	0.00
	[0.03]	[0.05]	[0.05]	[0.03]	[0.04]	[0.03]		[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
S_t	0.03	-0.00	0.03	-0.10**	-0.06	-0.12**		0.03	0.01	0.04	-0.10*	-0.06	-0.11**
	[0.06]	[0.06]	[0.06]	[0.05]	[0.04]	[0.05]		[0.06]	[0.06]	[0.06]	[0.05]	[0.05]	[0.05]
$L_t - L_{t-6}$	-0.00		-0.00	-0.25***		-0.15**		-0.00		-0.01	-0.25***		-0.16*
	[0.06]		[0.06]	[0.06]		[0.07]		[0.06]		[0.06]	[0.06]		[0.08]
$S_t - S_{t-6}$	-0.11		-0.11	-0.02		0.09		-0.11		-0.10	-0.02		0.06
	[0.10]		[0.11]	[0.07]		[0.08]		[0.10]		[0.10]	[0.07]		[0.08]
$y_t^M - c_t^M$		-0.04	0.00		-0.30***	-0.25**	DUR_t^M		-0.13	-0.10		-0.15***	-0.11**
		[0.06]	[0.07]		[0.10]	[0.10]			[0.09]	[0.08]		[0.04]	[0.05]
Adj.R^2	0.04	0.02	0.03	0.18	0.11	0.22		0.04	0.04	0.04	0.18	0.11	0.21
Sample	1976-1999	1976-1999	1976-1999	2000-2019	2000-2019	2000-2019		1976-1999	1976-1999	1976-1999	2000-2019	2000-2019	2000-2019

Table 6: The role of investor over-extrapolation: Evidence from bond mutual fund flows. Data on flows into taxable bond mutual funds is from the Investment Company Institute. Panels A reports the estimated regression coefficients for equation (4.2) using $X_t = \%\Delta FLOW_{t-3\to t}$. Panels B and C report the estimated regression coefficients when we use $X_t = \%\Delta FLOW_{t-3\to t}$ in equation (4.3) to forecast 3-month returns. We estimate these regressions using monthly data for the Apr1984 to Dec-1999 and Jan-2000 to Dec-2019 subsamples. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. We report Newey and West (1987) standard errors in brackets, using a lag truncation parameter of 9 months in Panel A and 5 months in Panels B and C. p-values are computed using the asymptotic theory of Kiefer and Vogelsang (2005) which has better finite sample properties than traditional asymptotic theory.

	Pre-2000			Post-2000		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: $FLOW_{t-3\to t}$						
L_t	1.96**	2.15***	2.24***	0.06	0.01	0.04
	[0.89]	[0.67]	[0.69]	[0.23]	[0.20]	[0.23]
S_t	4.08**	3.24**	3.92**	0.60	0.44	0.51
	[1.64]	[1.27]	[1.45]	[0.37]	[0.36]	[0.39]
$L_t - L_{t-6}$		-4.00***	-5.29***		-0.77**	-1.00**
		[1.04]	[1.29]		[0.32]	[0.46]
$S_t - S_{t-6}$			-4.31**			-0.38
			[2.15]			[0.45]
Adj.R^2	0.24	0.43	0.50	0.09	0.16	0.16
N	189	189	189	240	240	240
Dependent Variable: $rx^{(10)}_{t\to t+3}$						
L_t	1.27***	1.13***	1.04***	0.71**	0.76***	0.72**
	[0.30]	[0.29]	[0.35]	[0.30]	[0.27]	[0.31]
S_t	1.88***	1.37**	1.47*	1.42**	1.63***	1.55**
	[0.66]	[0.64]	[0.74]	[0.58]	[0.50]	[0.60]
$L_t - L_{t-6}$	-0.01		0.54	1.66***		1.42**
	[0.61]		[0.89]	[0.61]		[0.64]
$S_t - S_{t-6}$	-1.01		-0.56	1.12		1.02
	[0.88]		[0.98]	[0.72]		[0.73]
$FLOW_{t-3\to t}$		0.07	0.10		-0.34**	-0.24
		[0.08]	[0.11]		[0.14]	[0.15]
Adj.R^2	0.16	0.17	0.17	0.12	0.10	0.12
N	189	189	189	237	237	237
Dependent Variable: $rx^{SLOPE}_{t\to t+3}$						
L_t	0.01	0.01	0.02	0.00	0.01	-0.00
	[0.04]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
S_t	-0.04	-0.01	-0.03	-0.10**	-0.08*	-0.12**
	[0.05]	[0.05]	[0.05]	[0.05]	[0.04]	[0.05]
$L_t - L_{t-6}$	-0.04		-0.06	-0.25***		-0.22***
	[0.05]		[0.07]	[0.06]		[0.06]
$S_t - S_{t-6}$	0.10		0.08	-0.02		-0.01
	[0.10]		[0.10]	[0.07]		[0.07]
$FLOW_{t-3\to t}$		-0.00	-0.00		0.05***	0.03**
		[0.01]	[0.01]		[0.01]	[0.01]
Adj.R^2	0.03	-0.01	0.02	0.18	0.08	0.20
N	189	189	189	237	237	237
Sample	1984-1999	1984-1999	1984-1999	2000-2019	2000-2019	2000-2019

Table 7: The role of reaching-for-yield: Evidence from sectoral bond market flows. Data on sectoral-level bond market flows are from the Federal Reserve's Financial Accounts. Bond holdings include the sum of Treasury Securities (Table 210), Agency and GSE-Backed Securities (Instrument Table 211), and Corporate Bonds (Table 213). Our series for "Insurers" combines together data for Property-Casualty Insurance Companies (Table 115) and Life Insurance Companies (Table 116); "Pensions" combines together data for Private Pension Funds (Table 118) and State and Local Government Employee Retirement Funds (Table 120); and "Banks" uses data for U.S.-chartered depository institutions (Table 111). For intermediary sector i, we then compute the percentage bond flow in quarter t as $\%FLOW_{i,t} = FLOW_{i,t}/HOLD_{i,t-1}$, where $FLOW_{i,t}$ denotes net bond acquisitions by intermediaries in sector i during quarter t and $HOLD_{i,t-1}$ is bond holdings at the end of quarter $t-1$. Panel A reports the estimated regression coefficients for equation (4.2) using $X_t = \%FLOW_{i,t}$ for each sector i. We estimate these regressions using quarterly data for the 1971Q3-1999Q4 and 2000Q1-2019Q4 samples. Panels B and C report the estimated coefficients for equation (4.3). We report heteroskedasticity robust standard errors in brackets. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Pre-2000			Post-2000		
	(1)	(2)	(3)	(4)	(5)	(6)
Sector (i):	Insurance	Pensions	Banks	Insurance	Pensions	Banks
Dependent Variable: $FLOW_{i,t}$						
L_t	0.24***	0.27***	0.33***	0.23***	-0.17	0.06
	[0.05]	[0.07]	[0.12]	[0.07]	[0.19]	[0.16]
S_t	0.71***	-0.25*	0.67**	0.47***	-0.00	0.42
	[0.11]	[0.15]	[0.28]	[0.11]	[0.29]	[0.29]
$L_t - L_{t-2}$	-0.04	-0.28**	-0.53**	0.16	1.44***	0.05
	[0.09]	[0.11]	[0.26]	[0.17]	[0.42]	[0.41]
Adj.R^2	0.25	0.24	0.12	0.18	0.16	0.01
N	112	112	112	80	80	80
Dependent Variable: $rx^{(10)}_{t\to t+1}$						
L_t	0.39	0.68*	0.68*	0.95**	0.90**	0.85**
	[0.46]	[0.41]	[0.39]	[0.39]	[0.35]	[0.36]
S_t	0.99	1.34*	1.86**	1.83**	1.66***	1.68**
	[0.92]	[0.78]	[0.83]	[0.74]	[0.61]	[0.64]
$FLOW_{t-1\to t}$	0.54	-0.57	-0.50***	-0.56	0.24	-0.20
	[0.59]	[0.40]	[0.15]	[0.73]	[0.18]	[0.32]
Adj.R^2	0.02	0.03	0.06	0.06	0.07	0.06
N	114	114	114	79	79	79
Dependent Variable: $rx^{SLOPE}_{t\to t+1}$						
L_t	0.03	0.04	0.03	0.02	0.01	0.01
	[0.03]	[0.03]	[0.03]	[0.04]	[0.04]	[0.04]
S_t	-0.04	-0.03	-0.04	-0.05	-0.06	-0.06
	[0.09]	[0.08]	[0.08]	[0.07]	[0.06]	[0.06]
$FLOW_{t-1\to t}$	0.02	-0.03	0.02	-0.00	-0.03*	0.03
	[0.05]	[0.03]	[0.02]	[0.07]	[0.02]	[0.03]
Adj.R^2	0.01	0.01	0.02	0.00	0.04	0.02
N	114	114	114	79	79	79
Sample	1971-1999	1971-1999	1971-1999	2000-2019	2000-2019	2000-2019

Table 8: Economic news and subsequent changes in forward rates. This table reports the regression coefficents in equation (5.1) using monthly data from the Aug1971 to Dec-1999 and Jan-2000 to Dec2019 samples. Newey-West (1987) standard errors are in brackets, using a lag truncation parameter of $\lceil 1.5 \times h \rceil$. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Significance is computed using the asymptotic theory of Kiefer and Vogelsang (2005) which has better finite sample properties than traditional asymptotic theory.

	Pre-2000				Post-2000			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$f^{(10)}_{t+h} - f^{(10)}_{t}$ with h	3	6	9	12	3	6	9	12
Panel A								
$L_t - L_{t-1}$	-0.09	0.11	-0.04	0.12	-0.30***	-0.52***	-0.71***	-0.74***
	[0.12]	[0.12]	[0.12]	[0.13]	[0.11]	[0.17]	[0.16]	[0.15]
$S_t - S_{t-1}$	0.03	0.29	0.19	0.35	-0.14	-0.11	-0.02	-0.17
	[0.19]	[0.20]	[0.22]	[0.24]	[0.16]	[0.20]	[0.20]	[0.21]
L_t	-0.04	-0.08	-0.13	-0.18*	-0.03	-0.05	-0.06	-0.06
	[0.03]	[0.06]	[0.08]	[0.09]	[0.03]	[0.05]	[0.06]	[0.07]
S_t	-0.11	-0.24*	-0.37**	-0.52**	-0.09*	-0.17*	-0.27**	-0.35**
	[0.07]	[0.13]	[0.17]	[0.21]	[0.05]	[0.09]	[0.11]	[0.14]
Adj.R^2	0.02	0.05	0.09	0.14	0.04	0.09	0.17	0.23
Panel B								
$News_t$	-0.29	-0.12	-0.80***	-0.46	-0.45**	-0.63*	-0.82***	-0.67**
	[0.21]	[0.30]	[0.24]	[0.26]	[0.18]	[0.34]	[0.27]	[0.30]
L_t	-0.04	-0.08	-0.12	-0.17*	-0.04	-0.05	-0.06	-0.06
	[0.03]	[0.06]	[0.08]	[0.09]	[0.03]	[0.05]	[0.06]	[0.07]
S_t	-0.10	-0.23*	-0.37*	-0.52**	-0.10**	-0.18**	-0.26**	-0.35**
	[0.07]	[0.12]	[0.17]	[0.20]	[0.04]	[0.08]	[0.11]	[0.13]
Adj.R^2	0.02	0.05	0.11	0.14	0.04	0.08	0.15	0.21
Panel C								
$News_t$	-0.11	-0.11	-0.99**	-0.72**	-0.28	-0.05	0.15	0.38
	[0.34]	[0.35]	[0.42]	[0.29]	[0.29]	[0.47]	[0.36]	[0.29]
$L_t - L_{t-1}$	-0.06	0.14	0.19	0.29	-0.19	-0.50*	-0.77***	-0.89***
	[0.16]	[0.15]	[0.16]	[0.17]	[0.17]	[0.24]	[0.22]	[0.18]
$S_t - S_{t-1}$	0.03	0.29	0.23	0.38	-0.15	-0.11	-0.02	-0.15
	[0.19]	[0.20]	[0.20]	[0.23]	[0.16]	[0.20]	[0.18]	[0.21]
L_t	-0.04	-0.08	-0.12	-0.17*	-0.03	-0.05	-0.06	-0.06
	[0.03]	[0.06]	[0.08]	[0.09]	[0.03]	[0.05]	[0.06]	[0.07]
S_t	-0.11	-0.24*	-0.37*	-0.52**	-0.09*	-0.17*	-0.27**	-0.35**
	[0.07]	[0.13]	[0.17]	[0.21]	[0.05]	[0.09]	[0.11]	[0.14]
Adj.R^2	0.01	0.05	0.10	0.14	0.04	0.08	0.16	0.23
N	240	240	240	240	237	234	231	228
Sample	1980-1999	1980-1999	1980-1999	1980-1999	2000-2019	2000-2019	2000-2019	2000-2019
Panel D								
$News_{t,\text{FOMC}}$	0.11	-0.01	-1.89	-0.44	-0.99**	-0.74*	-1.18	0.09
	[0.77]	[1.04]	[1.56]	[1.56]	[0.46]	[0.39]	[0.75]	[0.56]
$News_{t,\text{Other}}$	-0.25	-0.17	-0.45	-0.32	-0.02	0.21	0.65	0.49
	[0.33]	[0.45]	[0.48]	[0.43]	[0.29]	[0.57]	[0.41]	[0.41]
$L_t - L_{t-1}$	0.09	0.03	0.03	-0.00	-0.20	-0.51**	-0.79***	-0.89***
	[0.18]	[0.21]	[0.24]	[0.27]	[0.16]	[0.24]	[0.20]	[0.18]
$S_t - S_{t-1}$	0.20	0.30	0.33	0.55**	-0.17	-0.13	-0.05	-0.16
	[0.24]	[0.27]	[0.23]	[0.25]	[0.16]	[0.20]	[0.18]	[0.21]
L_t	-0.07*	-0.13*	-0.18*	-0.20*	-0.03	-0.05	-0.06	-0.06
	[0.03]	[0.06]	[0.09]	[0.10]	[0.03]	[0.05]	[0.06]	[0.07]
S_t	-0.07	-0.17	-0.28	-0.42*	-0.10**	-0.18*	-0.28**	-0.36**
	[0.07]	[0.14]	[0.19]	[0.21]	[0.05]	[0.09]	[0.11]	[0.14]
Adj.R^2	0.03	0.08	0.14	0.16	0.05	0.09	0.18	0.23
N	216	216	216	216	237	234	231	228
Sample	1982-1999	1982-1999	1982-1999	1982-1999	2000-2019	2000-2019	2000-2019	2000-2019

Figure 1: Regressions of changes in long-term yields on short-term rates. This figure plots the estimated regression coefficients β_h from equation (1.1) versus horizon (h) for the pre-2000 and post-2000 sample: $y_{t+h}^{(10)} - y_t^{(10)} = \alpha_h + \beta_h(y_{t+h}^{(1)} - y_t^{(1)}) + \varepsilon_{t,t+h}$. The dependent variable is the h-month change in the 10-year nominal zero-coupon U.S. Treasury yield and the independent variable is the h-month change in the 1-year nominal zero-coupon U.S. Treasury yield. Changes are considered with daily data (plotted as $h = 0$ in the figure) and with monthly data using $h = 1$, ..., 12-month changes.

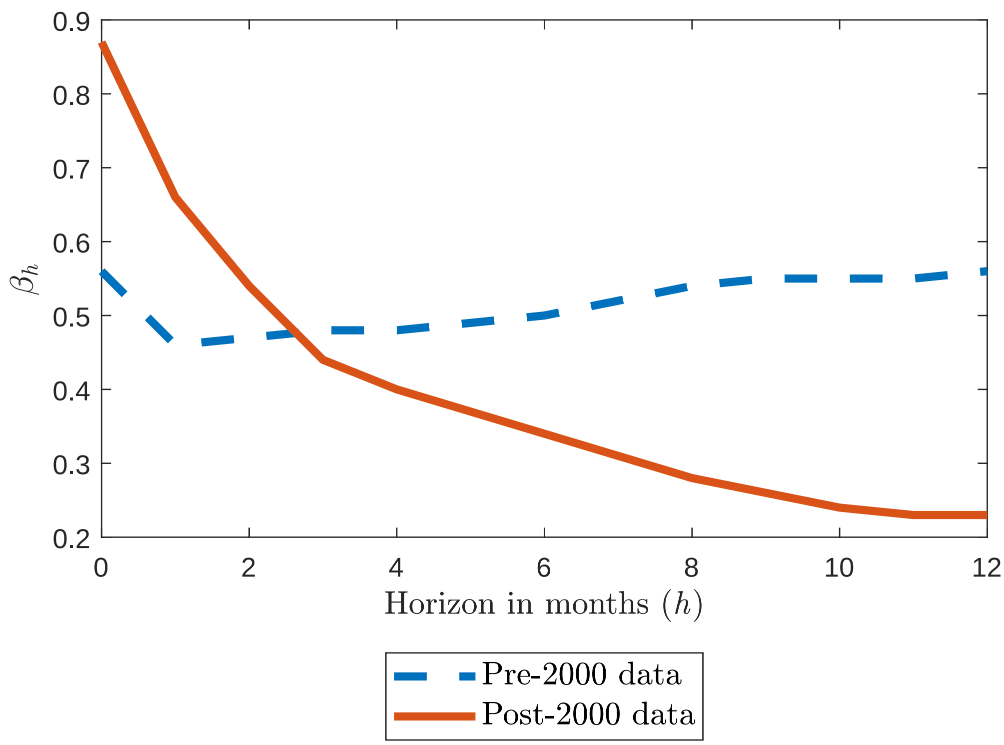


Figure 2: Rolling regression estimates of equations (1.1) and (1.2) This figure plots rolling estimates of the slope coefficients in equations (1.1) and (1.2) with $h = 12$-month changes using 10-year rolling windows for estimation. Results are plotted against the midpoint of the 10-year rolling window. 95% confidence intervals are included (shaded areas), formed using Newey-West standard errors with a lag truncation parameter of 18 and 95% critical values from the asymptotic theory of Kiefer and Vogelsang (2005). Specifically, the 95% confidence interval is ± 2.41 times the estimated standard errors as opposed to ± 1.96 under traditional asymptotic theory.

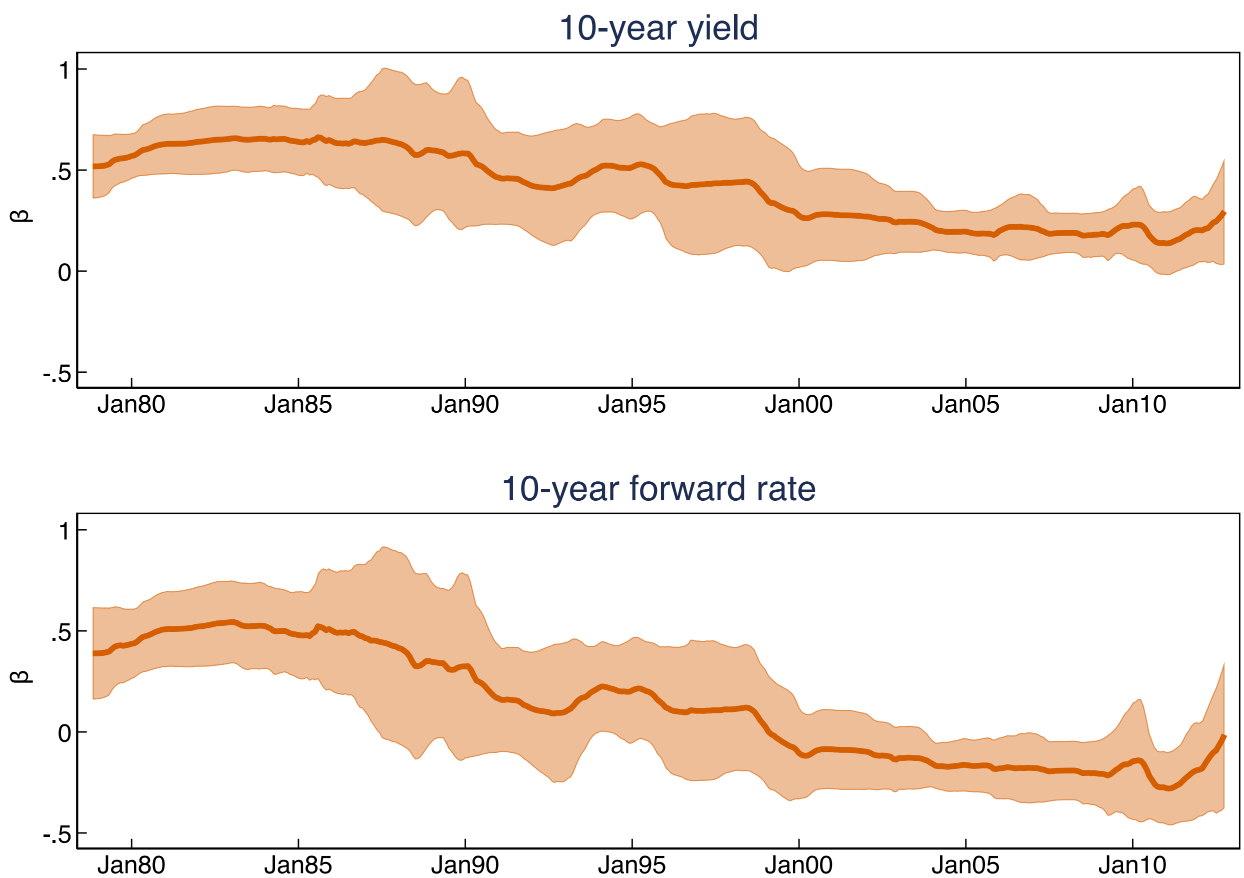


Figure 3: Break tests for equations (1.1) and (1.2) This figure plots the Wald test statistic for each possible break date in equations (1.1) and (1.2) with $h = 12$-month changes from a fraction 15% of the way through the sample to 85% of the way through the sample. The horizontal red dashed lines denote 10%, 5%, and 1% critical values for the maximum of these Wald statistics as in Andrews (1993). Our Wald tests use a Newey and West (1987) variance matrix with a lag truncation parameter of 18. To address the tendency for tests based on the Newey-West variance estimator to over-reject in finite samples, we use the Cho and Vogelsang (2017) critical values for a null of no structural break. The Cho and Vogelsang (2017) critical values are based on the asymptotic theory of Kiefer and Vogelsang (2005) and are slightly larger than the traditional critical values from Andrews (1993).

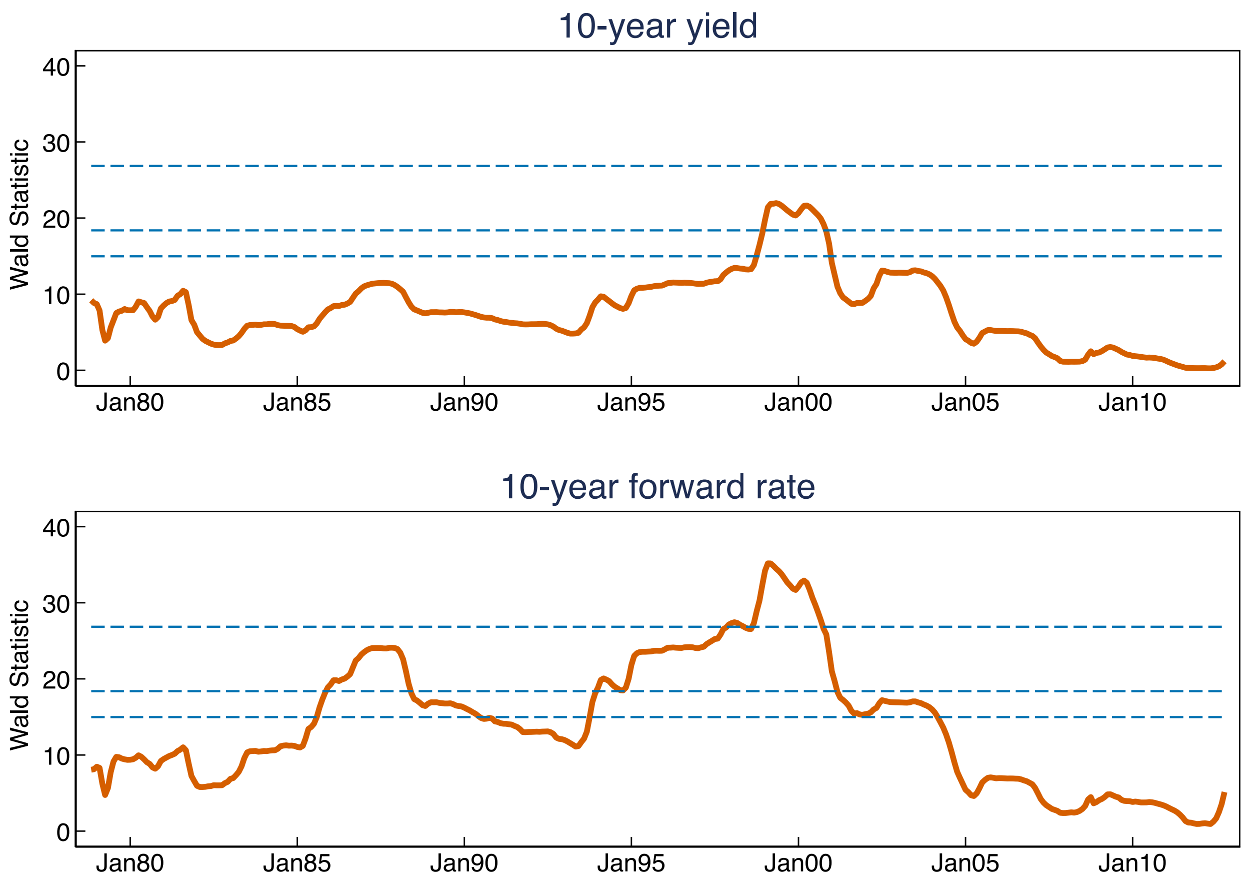


Figure 4: Predictable yield-curve dynamics following an impulse to short-term interest rates. The figures plot the coefficients $\delta_3^{(h)}$ versus horizon h from estimating equations (2.3) for various horizons $h = 1..., 12$-months in the pre-2000 and post-2000 samples. We show results for 10-year yields ($z_t = y_t^{(10)}$), 10-year forward rates ($z_t = f_t^{(10)}$), level ($z_t = L_t$) and slope ($z_t = S_t$). 95% confidence intervals are shown as dashed lines, formed using Newey-West standard errors and 95% critical values from the asymptotic theory of Kiefer and Vogelsang (2005). We use a Newey-West lag truncation parameter of 0 for $h = 1$ and $\lceil 1.5 \times h \rceil$ for $h > 1$.

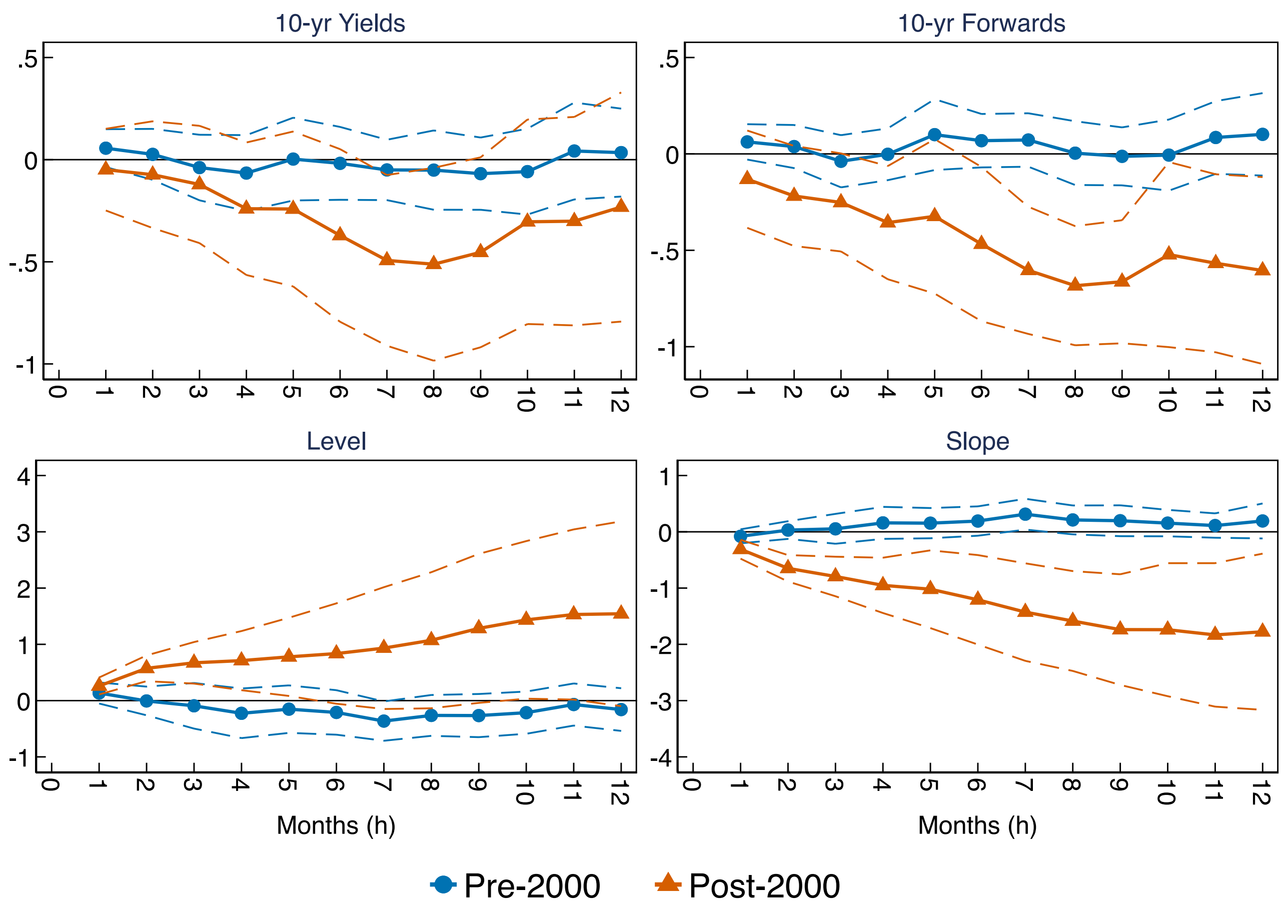


Figure 5: Model-implied coefficients β_h versus horizon (h) in months. The first figure shows the model-implied β_h coefficients from equation (3.11) for the pre-2000 and post-2000 calibrations discussed in the text. The second figure isolates the role of slow-moving capital in the post-2000 calibration, alternately setting $q = 100\%$ ("No slow-moving capital") and $q = 30\%$ ("With slow-moving capital").

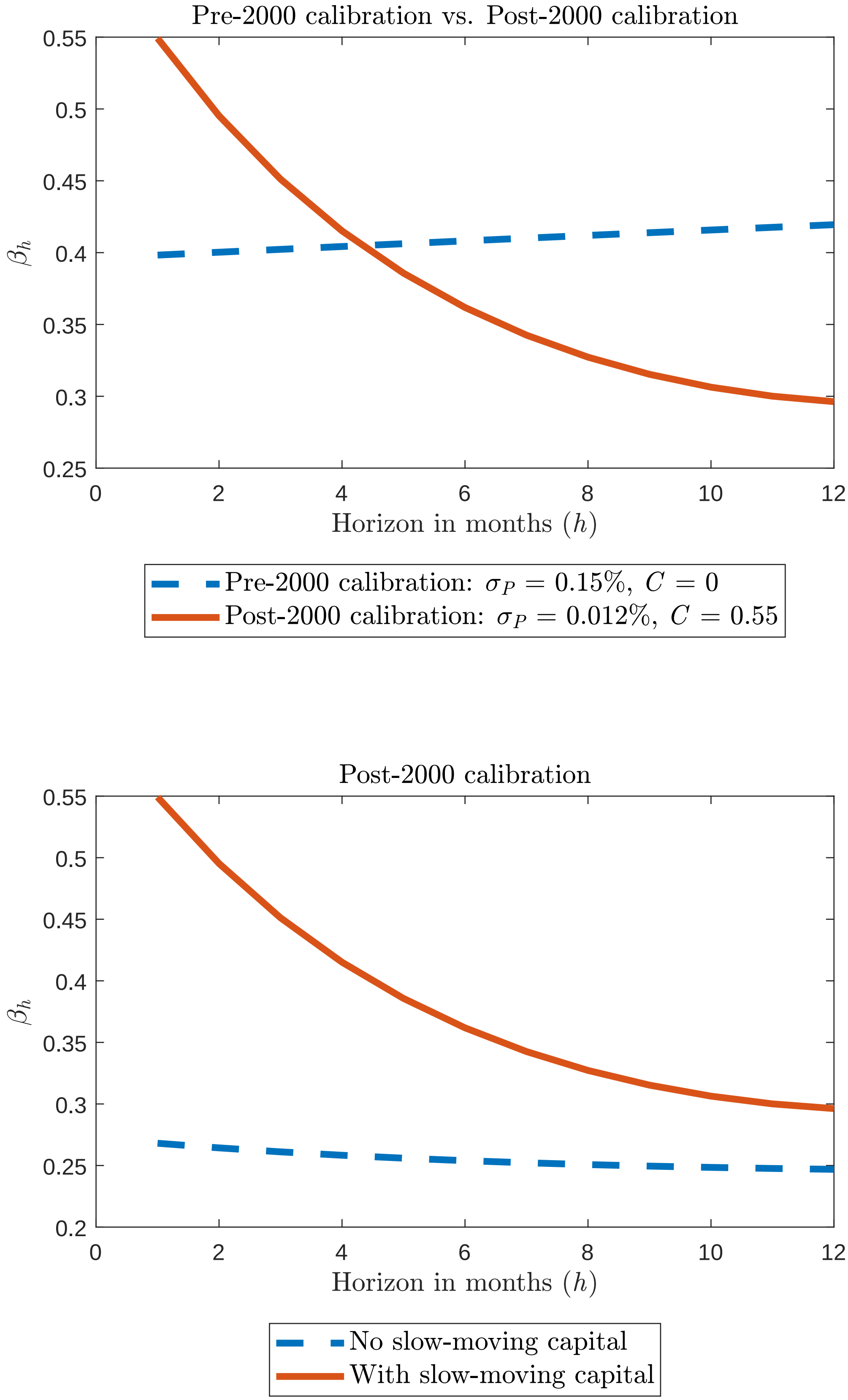


Figure 6: Model-implied impulse response functions for the post-2000 calibration. For the post-2000 calibration, we show the response of short-term and long-term interest rates following a one-time shock to short-term interest rates. We plot short-term nominal interest rates (i_t), long-term nominal yields (y_t), the term spread ($y_t - i_t$), and the term premium (tp_t). Initially, short-term nominal rates are at their steady-state level of $\bar{\imath} = 3\%$ and the term premium on long-term nominal bonds is at a steady-level of 2%. We then assume there is a 50 bp shock to both the persistent and transient components of the short rate that lands at $t = 12$, leading short-term nominal rates to jump from 3% to 4%.

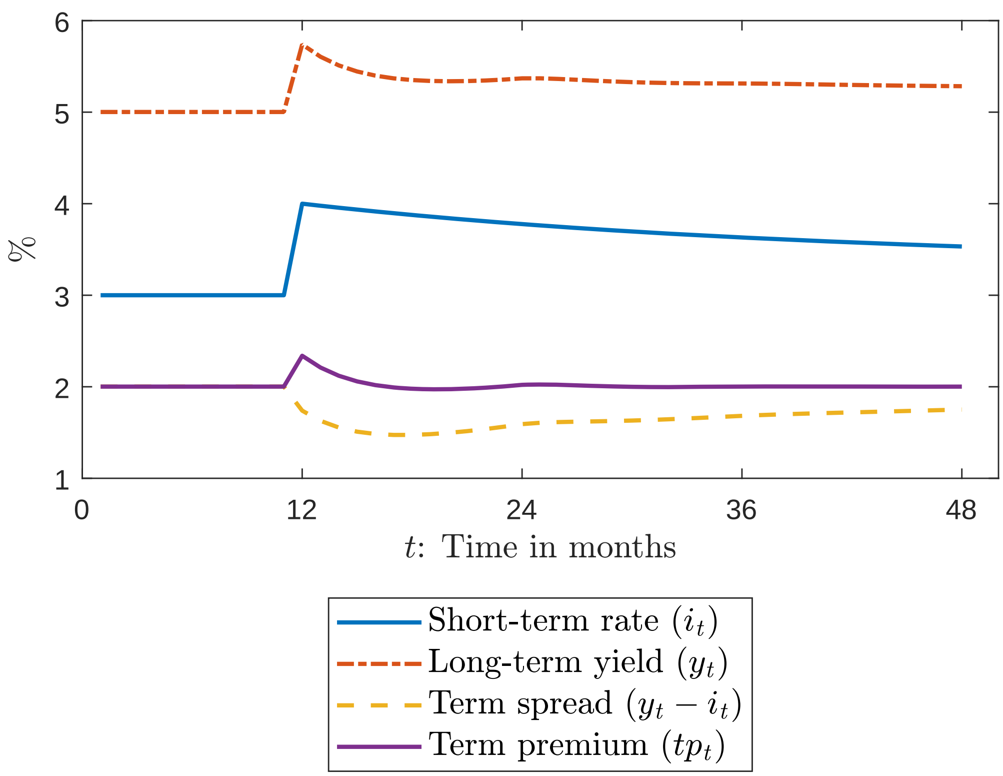
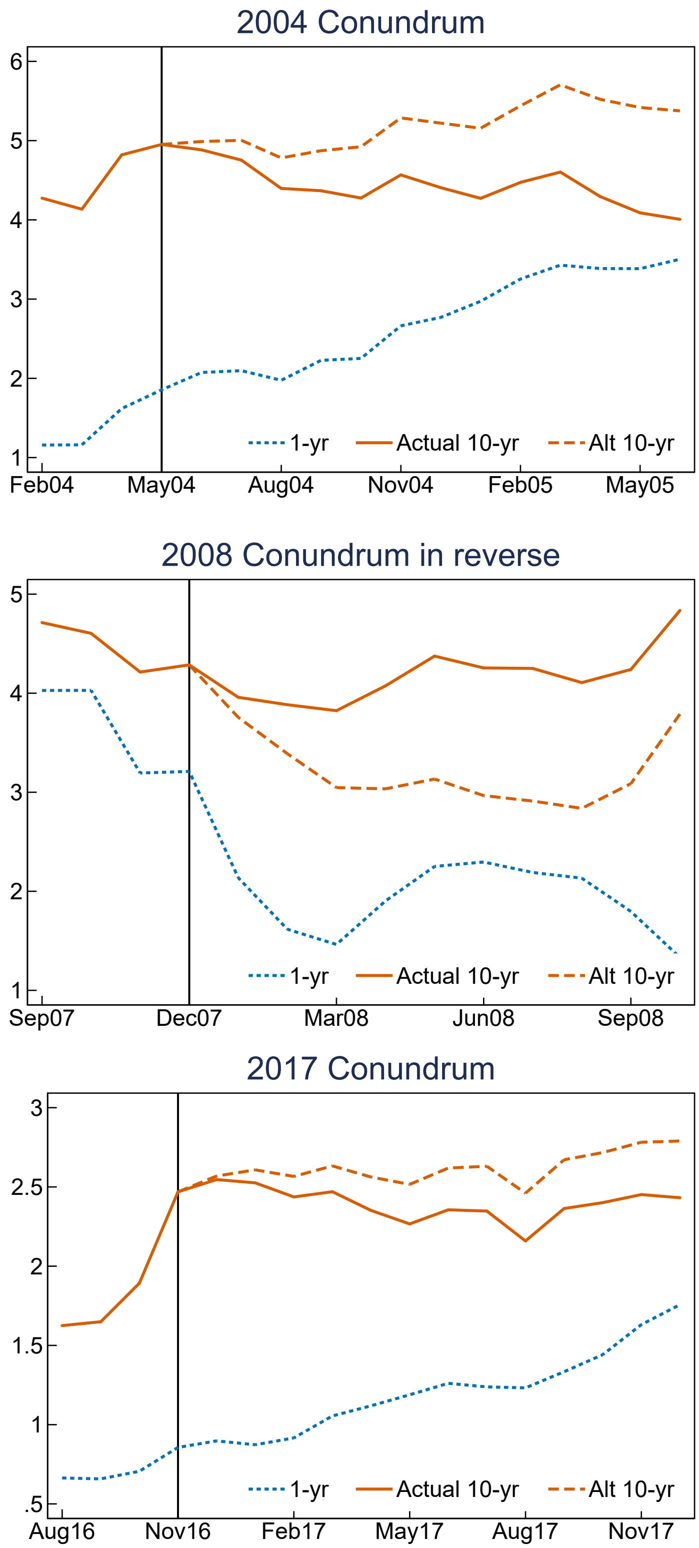


Figure 7: Counterfactual paths of ten-year yields in selected "conundrum" episodes. This figure plots 1- and 10-year yields in the original 2004 "conundrum" episode, the 2008 "conundrum in reverse" episode and the "2017 conundrum." As described in the text, we also plot counterfactual 10-year yields (Alt 10-yr) generated from restricting the slope to depend on lags of level and slope, but not also on lagged changes in level and slope.



A Appendix

In this Appendix, we formalize the three supply-and-demand driven channels of rate amplification discussed in the main text: the mortgage refinancing channel, the investor overextrapolation channel, and the reaching-for-yield channel.

A.1 The mortgage refinancing channel

The model setup follows Malkhozov et al. (2016).[35] There is a constant face value M of outstanding long-term, *fixed-rate* mortgages with an embedded prepayment option. The primary mortgage rate, denoted y_t^M, equals the long-term bond yield, y_t, plus a constant spread, λ: $y_t^M = y_t + \lambda$. (This constant spread play no role in the resulting analysis and can be set to zero without loss of generality.) Let c_t^M denote the average coupon on outstanding mortgages at the beginning of time t. We assume that c_t^M evolves according to the following law of motion:

$$c_{t+1}^M - c_t^M = -\eta \cdot (c_t^M - y_t^M), \tag{A.1}$$

where $\eta \in [0, 1]$. The difference between the beginning-of-period average mortgage coupon c_t^M and the current primary mortgage rate y_t^M is called the "refinancing incentive." Thus, according to equation (A.1), when the refinancing incentive is higher at time t, more households refinance their existing high-coupon mortgages at time t, leading the average mortgage coupon to fall from t to $t + 1$. Iterating on equation (A.1) and making use of the fact that $y_t^M = y_t + \lambda$, we obtain:

$$c_t^M = \textstyle\sum_{j=0}^{\infty} \eta\,(1-\eta)^j\, y_{t-1-j}^M = \sum_{j=0}^{\infty} \eta\,(1-\eta)^j\, y_{t-1-j} + \lambda. \tag{A.2}$$

Thus, the average mortgage coupon is just a backward-looking, geometric average of past long-term yields plus a constant. While clearly a simplification, this is a good empirical description of the average coupon on outstanding mortgages.[36]

We assume the *effective* gross supply of long-bonds that bond investors must hold at time t is

$$s_t = M \cdot DUR_t^M, \tag{A.3}$$

where M is face value of outstanding mortgages and DUR_t^M is the average "duration" or effective maturity of outstanding mortgages at time t.[37] When s_t is high, bond investors must collectively bear greater interest rate risk in equilibrium. We assume that average mortgage duration at time t is

$$DUR_t^M = \overline{DUR}^M - N \cdot \left(c_t^M - y_t^M\right), \tag{A.4}$$

where $N > 0$ is the so-called "negative convexity" of the average mortgage. Intuitively, when the refinancing incentive $(c_t^M - y_t^M)$ is high, many households are likely to refinance their mortgages in the near-term, implying that the average mortgage behaves more like a short-term bond—i.e., DUR_t^M is low and bond investors must bear less interest rate risk. By contrast, when the refinancing incentive is low, households are less likely to refinance and the typical mortgage behaves more like a long-term bond. Again, this is a good empirical description of DUR_t^M (Hanson, 2014; Malkhozov et al., 2016).[38]

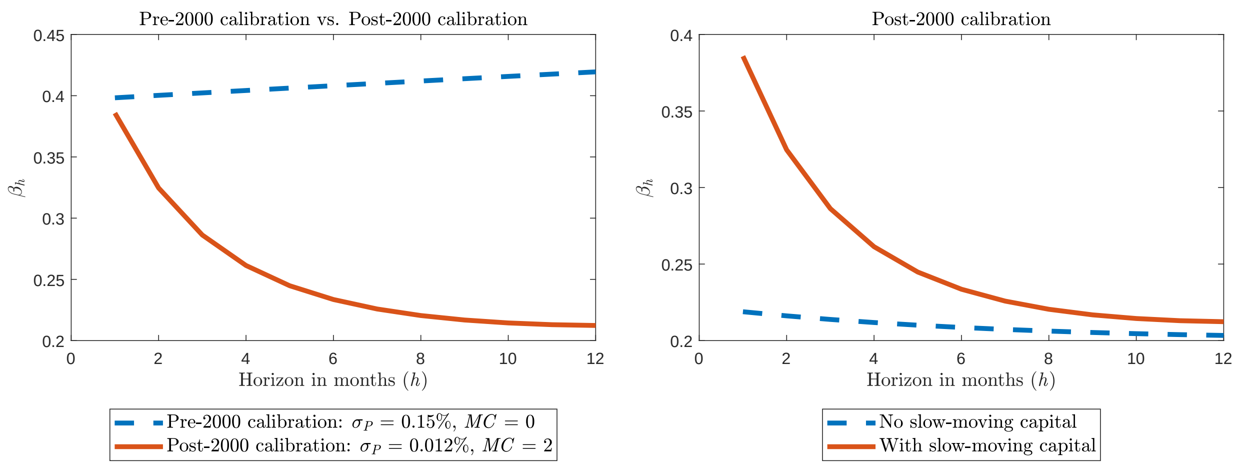
[35] Hanson (2014) explores the mortgage refinancing channel in a two period model. We follow the modelling approach in Malkhozov et al. (2016) since this allows us to speak to the dynamics which are our primary focus here.

[36] A realistic elaboration would incorporate state-dependence in the elasticity of refinancing with respect to the incentive. For instance, one might assume $c_{t+1}^M - c_t^M = -\eta\left[c_t^M - y_t^M\right] \cdot (c_t^M - y_t^M)$ where $\eta[\cdot] > 0$ and $\eta'[\cdot] > 0$, implying that the average coupon falls more when $c_t^M > y_t^M$ than when $c_t^M < y_t^M$—see e.g., Berger et al. (2018); Eichenbaum et al. (2018)

[37] Formally, duration is the semi-elasticity of a bond's price with respect to its yield. Thus, the longer a bond's duration, the greater is its exposure to movements in interest rates.

[38] As detailed in Hanson (2014), there are two key reasons why movements in expected mortgage refinancing temporarily alters the aggregate amount of interest rate risk that specialized bond investors must bear. First, households only *gradually*

Figure 8: Illustrative calibration of mortgage refinancing model.



Combining equations (A.2), (A.3), and (A.4), the *effective* supply of long-bonds at time t is

$$s_t = M \cdot \overline{DUR}^M + MN \cdot (y_t - \textstyle\sum_{j=0}^{\infty} \eta \, (1-\eta)^j \, y_{t-1-j}). \tag{A.5}$$

In other words, bond investors must bear greater interest rate risk when the long-term yield is currently high relative to its backward-looking, geometric average—i.e., when interest rates have recently risen.

In Internet Appendix C, we solve and calibrate this model of the mortgage refinancing channel. In this version of the model, there are two reasons why shocks to short-term interest rates give rise to transitory movements in the term premium component of long-term yields. First, when $\eta > 0$, mortgage refinancing waves trigger *temporary* shifts in the effective supply of long-term bonds—i.e., these effective supply shocks are less persistent than the underlying shocks to short-term interest rates. Second, these supply shocks are met by a slow-moving arbitrage response. This combination of transitory supply shocks and a slow-moving arbitrage response creates short-lived imbalances in the market for long-term bonds, leading long-term yields to temporarily overreact to short-term rates.[39]

Naturally, this version of the model can match the key stylized fact we have documented, namely that $\beta_h = \beta_h = Cov\left[y_{t+h} - y_t, i_{t+h} - i_t\right] / Var\left[i_{t+h} - i_t\right]$ is a sharply declining function of h in the post-2000 data but not in the pre-2000 data. As an illustrative calibration, we assume that $MN = 2$ in the post-2000 data and $MN = 0$ in the pre-2000 calibration. In other words, we assume that the mortgage refinancing channel is operative in the post-2000 period, but was not operative in the pre-2000 period. We assume $\eta = 0.15$ in both periods. The values of all other model parameters, including $q = 0.30$ and $k = 12$, are the same as those in the calibrations in Section 3.

refinance their mortgages following a decline in primary mortgage rates. Second, household borrowers do not alter their asset-side holdings of long-term bonds to hedge the time-varying interest rate risk they are bearing on the liability side. In combination, these features mean that households are effectively borrowing shorter term during refinancing waves when the refinancing incentive, $(c_t^M - y_t^M)$, is high. As a result, households bear greater interest rate risk during refinancing waves, while bond investors bear less risk. In summary, refinancing waves function like shocks to the effective supply of long-term bonds because risk sharing between households and specialized bond investors is imperfect and varies over time.

[39] One simplification of this model is that all bond investors hold mortgage-backed securities (MBS) and, thus, bear a time-varying amount interest rate risk. In practice, two different kinds of investors own MBS. One set of MBS investors—e.g., mortgage banks and the government sponsored enterprises—"delta-hedge" the embedded prepayment option and, thus, bear a (relatively) constant amount of interest rate risk over time. Other MBS investors do not delta-hedge and bear a time-varying amount of risk. As discussed in Hanson (2014), in the first instance, it does not matter whether some MBS holders delta-hedge the prepayment option since the relevant hedging flows correspond one-for-one with changes in the aggregate quantity of duration risk. However, a slow-moving arbitrage response to refinancing waves arguably becomes more relevant to the extent that some MBS investors delta-hedge their time-varying interest rate exposure.

A.2 Investor overextrapolation channel

Recalling that $i_t = i_{P,t} + i_{T,t}$, we assume that diagnostic investors make biased forecasts of the persistent and transitory components of short-term interest rates. Following Maxted (2020), we assume the expectations of diagnostic investors are given by:

$$E_t^D [i_{P,t+1}] = \bar{\imath} + \rho_P (i_{P,t} - \bar{\imath}) + \theta \cdot m_{P,t}, \tag{A.6a}$$
$$E_t^D [i_{T,t+1}] = \rho_T i_{T,t} + \theta \cdot m_{T,t}, \tag{A.6b}$$

where

$$m_{P,t} = \kappa_P m_{P,t-1} + \varepsilon_{P,t} = (i_{P,t} - \bar{\imath}) - (\rho_P - \kappa_P) \sum\nolimits_{j=0}^{\infty} \kappa_P^j (i_{P,t-j-1} - \bar{\imath}), \tag{A.7a}$$
$$m_{T,t} = \kappa_T m_{T,t-1} + \varepsilon_{T,t} = i_{T,t} - (\rho_T - \kappa_T) \sum\nolimits_{j=0}^{\infty} \kappa_T^j i_{T,t-j-1}, \tag{A.7b}$$

$\theta \geq 0$, $\kappa_P \in [0, \rho_P]$, and $\kappa_T \in [0, \rho_T]$. When $\theta = 0$, diagnostic expectations coincide with rational expectations, which we continue to denote using $E_t[\cdot]$. When $\theta > 0$, equations (A.6) and (A.7) imply that diagnostic investors tend to overestimate future short-term rates when short rates have recently risen. And, the κ_P and κ_T parameters govern the persistence of their mistaken beliefs about short rates.[40] While diagnostic investors make biased forecasts of short rates, we assume for simplicity that they form rational forecasts of all other relevant state variables.

A mass f of bond investors have diagnostic expectations and their demand for long-term bonds is:

$$h_t = \tau \frac{E_t^D [rx_{t+1}]}{Var_t^D [rx_{t+1}]} = \tau \frac{E_t^D [rx_{t+1}]}{Var_t [rx_{t+1}]}, \tag{A.8}$$

where $E_t^D [rx_{t+1}]$ denotes diagnostic investors' biased expectation of bond excess returns.[41] There is a mass $(1-f)$ of a bond investors with rational expectations. Of these rational investors, fraction q are fast-moving with demands $b_t = \tau (E_t [rx_{t+1}] / Var_t [rx_{t+1}])$ and fraction $(1-q)$ are slow-moving and only rebalance the portfolios every k periods. The demand for long-term bonds from the subset of slow-moving investors who are active at time t is $d_t = \tau(E_t[\sum_{j=1}^k rx_{t+j}]/Var_t[\sum_{j=1}^k rx_{t+j}])$. We assume the *gross supply* of long-term bonds is *constant* over time and equal to $\bar{s}$. Thus, the market clearing condition for long-term bonds at time t is:

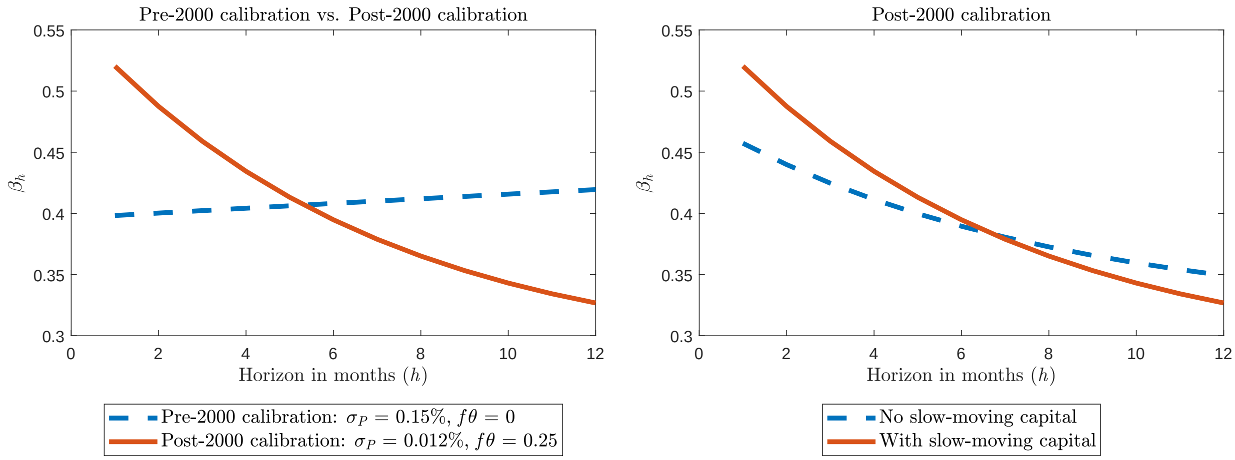
$$\overbrace{f h_t + (1-f) q b_t + (1-f)(1-q) k^{-1} d_t}^{\text{Active demand}} = \overbrace{\bar{s} - (1-f)(1-q) k^{-1} \sum\nolimits_{i=1}^{k-1} d_{t-i}}^{\text{Active supply}}. \tag{A.9}$$

In Internet Appendix C, we solve for equilibrium in this setting. When short rates have recently fallen, diagnostic investors underestimate future short-term interest rates and, as a result, want to hold more long-term bonds. To accommodate this induced demand shock, rational investors must reduce their holdings of long-term bonds, pushing down the term premium compensation required by rational investors. Since the biases of diagnostic investors are tied to recent changes in short rates, this model is nearly isomorphic to our reduced-form specification where the short-rate-driven shocks to bond supply are more transient than short rates. In particular, this investor overextrapolation channel leads long-

[40] As shown in Maxted (2020), these κ parameters are a simple way of parameterizing the "background context" that diagnostic investors use to assess the "representativeness" of incoming data for future states. Specifically, in the limit where κ_P and $\kappa_T \to 0$, the background context when making forecasts at time t is what diagnostic investors knew at time $t-1$ as in Bordalo et al. (2017) and the resulting expectational errors are very short-lived. In the opposite limit where $\kappa_P \to \rho_P$ and $\kappa_T \to \rho_T$, the background context at time t is the unconditional distribution of short rates as in D'Arienzo (2020) and the resulting expectational errors are far more persistent.

[41] We have $Var_t^D [rx_{t+1}] = Var_t [rx_{t+1}]$ since, as shown by Maxted (2020), diagnostic investors perceive the same conditional variance of future short-term interest rates as rational investors.

Figure 9: Illustrative calibration of the investor overextrapolation model.



term rates to *temporarily* overreact to movements in short rates due to the combination of (i) transitory shifts in non-fundamental demand for long-term bonds that are triggered by short rate shocks and (ii) a slow-moving arbitrage response to these non-fundamental demand shifts.

As shown in the illustrative calibration below, this model in which investors overextrapolate changes in short-term interest rates can match the key stylized facts we document. In the post-2000 period, we assume that $f = 50\%$ of investors have diagnostic expectations with parameter $\theta = 0.5$ (we set $\kappa_P = \kappa_T = 0.8$) and that $q = 0.15$ and $k = 18$, so there is a fairly slow-moving arbitrage response to the resulting non-fundamental shifts in demand for long-term bonds. In the pre-2000 period, we assume that $f = 0$. As discussed in the main text, the rise in f is meant to capture the growing importance of extrapolation-prone bond fund investors in the U.S. bond market in recent decades. The values of all other model parameters are the same as those in the calibrations in Section 3.

A.3 Reaching-for-yield channel

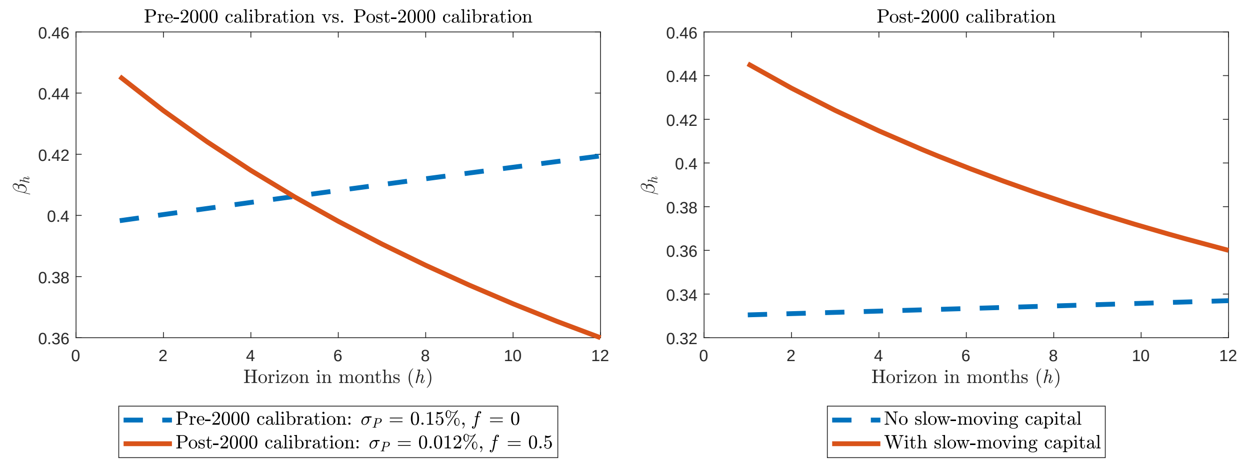
We assume that fraction f of bond investors are "yield-seeking" and have non-standard preferences as in Hanson and Stein (2015). The idea is that, for either frictional or behavioral reasons, these investors care about the *current yield* on their portfolios over and above expected portfolio returns. Specifically, yield-seeking investors' demand for long-term bonds is:

$$h_t = \tau \frac{y_t - i_i}{V^{(1)}}. \tag{A.10}$$

Since $E_t [rx_{t+1}] = (y_t - i_i) - (\phi/(1-\phi)) \cdot E_t [y_{t+1} - y_t]$, equation (A.10) implies that yield-seeking investors are only concerned with the current income or carry from holding long-term bonds and neglect any expected capital gains and losses from holding long-term bonds. A mass $(1-f)$ of a bond investors are expected-return-oriented and have standard mean-variance preferences. Of these expected-return-oriented investors, fraction q are fast-moving investors with demands $b_t = \tau (E_t [rx_{t+1}] / Var_t [rx_{t+1}])$ and fraction $(1-q)$ are slow-moving. The demand for long-term bonds from the subset of slow-moving investors who are active at time t is $d_t = \tau(E_t[\sum_{j=1}^{k} rx_{t+j}]/Var_t[\sum_{j=1}^{k} rx_{t+j}])$. We assume the *gross supply* of long-term bonds is *constant* over time and equal to $\overline{s}$. Thus, the market clearing condition for long-term bonds at time t is the same as in equation (A.9).

In Internet Appendix C, we solve for equilibrium in this setting. To build intuition, first consider the case where there is no slow-moving capital. In this case where $q = 1$, our model is simply an infinite-horizon version of the 2-period model in Hanson and Stein (2015). Because expected mean reversion in short rates implies that the yield curve is steep when short rates are low, yield-seeking investors' demand for long-term bonds is higher when short rates are lower. To accommodate this induced demand shock,

Figure 10: Illustrative calibration of the investor reaching-for-yield model.



expected-return-oriented investors must reduce their holdings of long-term bonds when short rates are low, pushing down the term premium compensation they require. Thus, in the absence of slow-moving capital, long-term rates are excessively sensitive to short rates because short rates and term premium move in the same direction. However, without slow-moving capital, this excess sensitivity is *not* greater at short horizons—i.e., changes in short rates do not create temporary market imbalances.

When $q < 1$, our model adds a slow-moving arbitrage response to the price-pressure created by yield-seeking investors. This means that the excess sensitivity of long-term rates to short-term rates will be greatest at short horizons. The intuition is simple. Suppose there is an decline in short rates which steepens the yield curve, thereby boosting yield-seeking investors' demand for long-term bonds. In the short-run, the only expected-return-oriented investors can absorb this induced demand shock for long-term bonds are the fast-moving ones and the slow-moving ones who initially happen to be active. However, the mass of slow-moving investors who can absorb this induced demand shock grows over time. As a result, the excess sensitivity of long-term rates to movements in short-term rates is greatest at high frequencies and diminishes at lower frequencies.

This model of the reaching-for-yield channel can qualitatively match the key stylized facts we have documented. Specifically, in the post-2000 period, we assume that $f = 0.5$, $q = 0.1$, and $k = 24$—i.e., we assume a good deal of reaching for yield and a fairly sluggish arbitrage response. By contrast, we assume that $f = 0$ in the pre-2000 period. The rise in f is consistent with the idea that yield-seeking investor behavior has become stronger in recent decades. The values of all other model parameters are the same as those in the calibrations in Section 3. As shown below, we see that β_h is decreasing in h in the post-2000 calibration, but is increasing in h—and far less variable—in the pre-2000 calibration.

While the combination of reaching-for-yield and slow-moving capital generates horizon-dependent sensitivity, our calibrations struggle to *quantitatively* match the profile of β_h seen in the post-2000 data. Specifically, comparing the calibration of the reaching-for-yield model in Figure 10 with those for the mortgage refinancing and investor overextrapolation models in Figures 8 and 9, respectively, we see that the former model struggles generate quantitatively match the steep post-2000 profile of β_h. This is because the reaching-for-yield channel generates highly persistent shifts rate-amplifying net supply, whereas the refinancing and overextrapolation channels generate transitory shifts in net supply. And, as we have emphasized throughout, strongly horizon-dependent excess sensitivity is most likely to arise when transitory rate-amplifying supply-and-demand shocks are met by a slow-moving arbitrage response.