

Secular Economic Changes and Bond Yields

by Bruno Feunou and Jean-Sébastien Fontaine

Financial Markets Department
Bank of Canada, Ottawa, Ontario, Canada K1A 0G9
bfeunou@bankofcanada.ca, jfontaine@bankofcanada.ca

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.



Acknowledgements

We thank Paul Beaudry, Anna Cieslak, Òscar Jordà, Sharon Kozicki and Guillaume Roussellet for comments and suggestions. We also thank Maren Hansen for editorial assistance.

Abstract

We build a model for bond yields based on a small-scale representation of an economy with secular declines in inflation, the real rate and output growth. Long-run restrictions identify nominal shocks that influence long-run inflation but do not influence the long-run real rate or output growth. These nominal shocks have loadings that can change over time. The results show that, before the anchoring of inflation around the mid-1990s, nominal shocks lifted the output gap and inflation, leading to higher yields and a steeper yield curve via higher short-rate expectations and term premiums. The short rate peaked after several quarters but only after the responses of growth and inflation started to decline. With inflation anchored, however, nominal shocks have a short-lived impact on inflation, an insignificant impact on output and only a small impact on bond yields via the term premium.

Topics: Asset Pricing; Interest rates; Monetary policy and uncertainty; Potential output; Econometric and statistical methods

JEL codes: E43, G12

1 Introduction

What are the economic forces behind the secular decline in bond yields? In the early 1980s, bond yields exhibited cyclical changes around declines in the long-run level of inflation, π_t^* . This level stabilized at around 2 percent during the 1990s. More recently, bond yields exhibit cyclical changes around declines in the long-run level of the real rate, r_t^* . [Cieslak and Povala \(2015\)](#) and [Bauer and Rudebusch \(2020\)](#) show that it is essential to disentangle these cyclical and secular variations to understand what drives bond yields. Intuitively, ignoring the secular variations in π_t^* and r_t^* means that the continued decline in yields since the early 1980s is largely attributed to a decline in term premiums.

We contribute to a better understanding of the secular decline in bond yields. We implement the long-run restrictions pioneered by [Blanchard and Quah \(1989\)](#) to identify the nominal and real shocks in a model of bond yields that has, in the spirit of [Bauer and Rudebusch \(2020\)](#), both cyclical and secular components. Our approach is distinct because (i) we connect bond yields to macroeconomic variations, (ii) we let the cyclical and secular components share the same nominal and real shocks that we identify from the data and (iii) the shocks have loadings that vary over time. Hence, both the real and nominal shocks drive both the cyclical variation and the secular decline in bond yields via their impacts on π_t^* and r_t^* . In the results, when we let the loadings of these shocks vary over time, we find that nominal shocks play a large role in driving the economy and the bond yields early in our sample but that these impacts essentially disappear later in our sample, once inflation is anchored.

To understand these results, which we explain below, it is useful to first see the four building blocks of our approach. First, we use a small-scale representation of an economy that is based on the short rate, inflation and output, where each variable is the sum of one cyclical and one secular component, which are both unobserved.¹ The cyclical components

¹The presence of secular variations in output is also well-known and has been extensively studied since [Nelson and Plosser \(1982\)](#). See also the discussion and the references in [Cochrane \(1988\)](#).

are both correlated and related to each other. The secular components are also correlated. Similar to [Jordà and Taylor 2019](#), using this representation of an economy connects bond yields to underlying cyclical and secular macroeconomic changes that are based on a common set of assumptions.

Second, and in contrast with [Jordà and Taylor 2019](#), we let the cyclical and secular components share the same macroeconomic shocks. This commonality arises in equilibrium models that exhibit delays in adjustments to permanent shocks.² As in structural vector autoregression (VAR) analysis, the nature of the shocks follows from the economic content of identification restrictions. Our identification strategy is close in spirit to the long-run restrictions pioneered by [Blanchard and Quah \(1989\)](#). We implement the definition that nominal shocks are neutral: these shocks can influence cyclical fluctuations in output, inflation and interest rates as well as secular fluctuations in inflation but they leave unchanged the long-run levels of output growth and the real rate.

Third, we let the loadings of the nominal and real shocks on the secular and cyclical components change over time in a way that is similar to [Primiceri \(2005\)](#) but in a different context. Time-varying loadings means that the impact of shocks on an economy can change over time. In addition, having time-varying loadings implies that the volatilities of the short rate, inflation and output change over time, even if the shocks have constant variances. Indeed, [Stock and Watson \(2007\)](#) and [Wright \(2011\)](#) emphasize time-varying loadings in univariate reduced-form models of inflation.

Fourth, we derive bond prices using a reduced-form stochastic discount factor in the class introduced by [Monfort and Pegoraro \(2012\)](#). We assume that the nominal and real

²Section 3.4.1 in [Gali \(2008\)](#) discusses the standard New Keynesian model. For instance, permanent productivity shocks typically influence the potential output and the output gap, as well as the natural real rate of interest and the real-rate gap. Conversely, structural shocks that are relevant to the business cycle can also affect economic activity in the long run, a view recently summarized by [Blanchard \(2018\)](#). In the case of real GDP models, [Morley, Nelson, and Zivot \(2003\)](#) provide evidence against the zero-correlation assumption for permanent and transitory innovations.

shocks are sources of risk and that their risk prices are functions of the macro variables. Bond yields satisfy the no-arbitrage restriction and are functions of secular and cyclical components. However, we impose the restriction that the expected bond returns are stationary: they are not functions of secular components. Hence, consistent with the evidence of [Cieslak and Povala \(2015\)](#) and [Bauer and Rudebusch \(2020\)](#), our model implies that π_t^* and r_t^* enter into bond yields but not term premiums. This restriction makes the model parsimonious, improves the accuracy of the parameter estimates and also prevents the variance of the expected bond returns from diverging with the investment horizon.³ Finally, we follow earlier work and introduce in this representation an exogenous factor to capture the variations in bond yields that are uncorrelated with the macroeconomic shocks. We label this factor a “term structure factor” because, by design, it only influences the term premium and not the expectations component of bond yields. This prevents this exogenous shock from entering the dynamics of the short rate, output and inflation. In the results, this term structure factor explains a large share of the variations in bond yields at the quarterly horizon, but this share declines at longer horizons, which is consistent with the seminal macro-finance results of [Ang and Piazzesi \(2003\)](#).

Empirically, we find that the impacts of nominal shocks on bond yields are very different after inflation is anchored in the 1990s. We quantify three reasons for this change. First, the sensitivity of the secular inflation rate, π_t^* , to nominal shocks eventually disappears in our sample; that is, inflation becomes “anchored.” Second, the large and persistent impacts of nominal shocks on cyclical variations in output, inflation and the short rate also become smaller and transitory later in our sample. Third, the response of the term premium to nominal shocks also becomes much less persistent over time. We find that these changes essentially operate simultaneously. In the model, these changes are captured by the time-varying loadings for the real and nominal shocks. However, the

³We also restrict the time-varying loadings such that the volatility of the interest rates converges with the forecast horizon, while the volatility diverges in existing models of bond prices with secular changes.

results are not predetermined: how the loadings change is estimated from the data.

The first channel between nominal shocks and bond yields acts on long-term inflation. As expected, π_t^* estimates decline over time, starting in our sample at around 6 percent and reaching a bottom that is close to 2 percent in the mid-1990s. Nominal shocks explain around two thirds of this decline. The reason behind π_t^* 's bottoming is that the sensitivity of long-run inflation to nominal shocks gradually declines and essentially disappears by the 2000s. We pick two dates to illustrate this change: 1984Q1 and 2007Q1. In 1984, nominal shocks have a permanent impact on inflation. A one-standard-deviation nominal shock lifts π_t^* by around 0.1 percentage points. To put this number in context, the impact on actual inflation peaks above 0.3 percentage points. Move forward to 2007 and the response of π_t^* to nominal shocks becomes close to zero and insignificant.

The second channel between nominal shocks and bond yields acts on the cyclical responses of the economy. Early in our sample, there is a strong cyclical response. A one-standard-deviation nominal shock lifts inflation but it also lifts the output gap by 0.5 percentage points on impact; the effect peaks at around 0.7 and lasts a few years. In response, the short rate increases by around 0.45 percentage points, peaks at around 0.7 and the effect lasts a few years. By contrast, the same nominal shock only causes small, transitory and statistically insignificant responses in 2007. When we decompose the innovations to the cyclical component of the short rate, date by date, we find that nominal shocks play a dominant role until the mid-1990s but then become much less important later on. Overall, these changes are consistent with the theoretical argument in [Clarida, Gali, and Gertler \(2000\)](#) that the post-1979 policy responses to inflation stabilize output and inflation. These changes also reflect the empirical results in [Cogley and Sargent \(2005\)](#) that show that policy activism by the Federal Reserve in that period contributed to a fall in the level and persistence of inflation.

Combined, these channels imply that the response of bond yields to nominal shocks changes dramatically. Early in our sample, this response leads to a higher level of the

yield curve and a steeper slope between the short rate and longer-term yields. However, for the 2- and 10-year bonds, the initial impacts on the expectations components are around 0.6 and 0.4 percentage points, respectively. The difference between these two responses is due to the responses of the expectations about future short rates: they increase the curvatures and cause inversions of the yield curves between 2- and 10-year maturities. These inversions imply that the short rate will continue to rise but the growth and inflation rates will decrease. All of these effects of nominal shocks become much smaller and statistically insignificant later in the sample. Note that the observation that the expectations component drives the predictive content of a yield curve inversion for future growth is consistent with the results of [Ang, Piazzesi, and Wei \(2006\)](#). Also, the lack of inversion in the yield curve after a nominal shock we show later in our sample is consistent with the suggestion by [Haubrich \(2006\)](#) that the predictive content of a yield curve inversion has become weaker due to the Federal Reserve's rising credibility in taming inflation.

The third channel acts on the term premium. For the 2- and 10-year bonds, the impacts early in the sample are around 0.1 and 0.2 percentage points initially and peak at close to 0.25 and 0.4, respectively. This further steepens the yield curve but mutes its inversion. Overall, accounting for the expectations and term premium components, the responses of bond yields to nominal shocks considerably weaken after the mid-1990s. The initial all-in impact is almost halved on the 2-year yield, cut by one third on the 10-year yield and is much less persistent in both cases.

Given that secular inflation, π_t^* , has stabilized, the continued decline in bond yields during the 2000s is attributed to the decline in r_t^* , a point made clear by [Bauer and Rudebusch \(2020\)](#). The r_t^* estimate rises from 1993 to 2000, in line with existing estimates ([Laubach and Williams, 2003](#); [Holston, Laubach, and Williams, 2017](#)), but then initiates a gradual descent, reaching 2 percent around 2008 and 0.5 percent at the end of our sample. Nominal shocks are irrelevant by design for r_t^* . We find that changes to r_t^* in the first half

or so of our sample are driven by real shocks that are uncorrelated with changes to the long-run growth rate. However, the changes to r_t^* in the second half of the sample appear largely driven by shocks that are correlated with changes to potential output, suggesting that technology, capital accumulation and demographics play a larger role during that period.⁴

1.1 Literature

The earlier works of [Kozicki and Tinsley \(2001\)](#) and [Kozicki and Tinsley \(2006\)](#) introduce a shifting endpoint to capture secular changes in the short rate and the inflation dynamics, respectively. We use their definition of a shifting endpoint, which is based on the Beveridge-Nelson decomposition, to define the secular components in our small-scale economy. [Laubach and Williams \(2003\)](#) estimate a shifting endpoint for the real rate, and the evidence is expanded to other advanced economies by [Holston, Laubach, and Williams \(2017\)](#) as well as [Negro, Giannone, Giannoni, and Tambalotti \(2019\)](#), who document a common decline across advanced economies. This follows early evidence in [Garcia and Perron \(1996\)](#) of distinct regimes in the behavior of the US real interest rate. Our π_t^* and r_t^* estimates are similar to existing results. [Jordà and Taylor \(2019\)](#) investigate a small-scale representation of an economy with distinct endpoints for output, inflation and the real rate but they do not attempt to identify structural shocks. A joint representation of secular and cyclical components avoids the critique by [Coibion, Gorodnichenko, and Ulate \(2018\)](#) that many estimates of secular components exhibit transitory responses to measures of demand-side or monetary policy shocks, which could contaminate the identification of the shocks.

⁴[Laubach and Williams \(2003\)](#) also assume that the secular real rate component r^* is driven by the long-run output growth rate and an additional factor that leaves output unaffected. [Rachel and Smith \(2017\)](#) provide a detailed review of the potential drivers that could be underlying these changes in the long-run growth rate.

Van Dijk, Koopman, van der Wel, and Wright (2014) allow for a shifting endpoint in a model of bond prices that is used to improve interest rate forecasts. Cieslak and Povala (2015) provide compelling evidence that controlling for secular changes in expected inflation rates, based on either a discounted moving average or long-horizon survey forecasts, provides a more accurate measure of the bond risk premium. Christensen and Rudebusch (2019) recover r_t^* estimates using a dynamic term structure model (DTSM) for prices of individual inflation-indexed bonds. Joyce, Kaminska, and Lildholdt (2012) provide an early application to UK real bonds. Bauer and Rudebusch (2020) provide evidence that variations in the secular real interest rate are responsible for the very high persistence of interest rates, and these researchers offer a parsimonious DTSM of the yield curve, using three standard yield factors and one common stochastic trend. Adding to this work, we explore the macroeconomic determinants of secular changes in bond yields.

Primiceri (2005) and Cogley and Sargent (2005) use a vector auto-regression with time-varying parameters to study interactions between the US economy and monetary policy. Primiceri (2005) finds that the systematic responses of the short rate to inflation and unemployment trend toward more aggressive behavior over time. Cogley and Sargent (2005) document changes over time in the natural rate of unemployment, a core rate of inflation, the persistence in inflation and the degree of monetary policy activism. This work focuses on the periods before and after the Federal Reserve chairmanship of Paul Volcker. We focus on the responses of bond yields in the period starting with the chairmanship of Alan Greenspan, when the Federal Reserve maintained its degree of activism.

The rest of this paper is organized as follows. Section 2 introduces our small-scale model of a US economy with secular and cyclical components and details our specification of the bond prices. Section 3 provides details about the data and the estimation method, which is easy to implement and bypasses the Kalman filter. Section 4 presents the secular and cyclical components we recover, discusses the role of the different shocks and analyzes

the response of bond yields to these shocks. Section 5 concludes.

2 Model

We develop a small-scale representation of an economy where the short rate, the inflation rate and output have joint dynamics. Each variable has a stationary and a non-stationary component, which we label “cyclical” and “secular” in the following, respectively. One reason for modeling these components together is to recover internally consistent estimates of the secular changes in the nominal rate, the inflation rate and the real rate. Another reason is that the results exhibit internally consistent impulse response functions, avoiding the criticism in [Coibion, Gorodnichenko, and Ulate \(2018\)](#) that existing estimates of secular economic changes exhibit cyclical patterns.⁵

In the following, we first introduce a restricted version of the model that has constant loadings. Then, in a second step we introduce the time-varying loadings. Considering the restricted case first simplifies some of the exposition. It is useful to estimate both cases to understand the role of the time-varying loadings.

2.1 Notations

We label the short rate i_t , the inflation rate π_t and output y_t . Each of the macroeconomic variables has a Beveridge-Nelson decomposition, given by

$$\begin{aligned}i_t &= i_t^* + \tilde{i}_t \\ \pi_t &= \pi_t^* + \tilde{\pi}_t \\ y_t &= y_t^* + \tilde{y}_t.\end{aligned}\tag{1}$$

Equation 1 is a statistical representation of the observed macro variables in terms of their unobserved components. These components have no economic content yet. The following

⁵They find that several existing estimates of potential output exhibit predictable and transitory responses to demand-side or monetary policy shocks.

sections construct a system of dynamic equations where the variables with the (*) symbol representing the secular components and the variables with the (̂) symbol representing the cyclical components. This joint system provides the economic content and allows us to identify the unobserved components from the data. One feature of this system is that the components share the same set of shocks and can be correlated with each other. This contrasts with the assumption in most existing unobserved components models that innovations to the secular and cyclical components are uncorrelated.

2.2 Conditional Means

Stack the observed variables i_t , π_t and y_t in the vector M_t . This section specifies the joint conditional mean dynamics:

$$M_t \equiv \begin{bmatrix} i_t \\ \pi_t \\ y_t \end{bmatrix} = E_{t-1}[M_t] + \Sigma \varepsilon_t, \quad (2)$$

with shocks $\varepsilon_t \sim N(0, I)$. In the first step, we specify the joint conditional mean of the secular components. In the second step, we specify the joint conditional mean of the cyclical components.

Secular Components Define the expected growth rate $g_t^* \equiv E_t[y_{t+1}^* - y_t^*]$ and stack the variables i_t^* , π_t^* and g_t^* in the vector \bar{M}_t , for which we assume the following dynamics:

$$\bar{M}_t \equiv \begin{bmatrix} i_t^* \\ \pi_t^* \\ g_t^* \end{bmatrix} = \begin{bmatrix} i_{t-1}^* \\ \pi_{t-1}^* \\ g_{t-1}^* \end{bmatrix} + \begin{bmatrix} \sigma'_{i^*} \\ \sigma'_{\pi^*} \\ \sigma'_{g^*} \end{bmatrix} \varepsilon_t = \bar{M}_{t-1} + \bar{\Sigma} \varepsilon_t. \quad (3)$$

Following [Laubach and Williams \(2003\)](#) among others, we take that the growth of potential output Δy^* is integrated of order 1:

$$\Delta y_t^* = g_{t-1}^* + \sigma_{y^*}' \varepsilon_t. \quad (4)$$

This nests the case where the level of potential output y_t^* is integrated of order 1 if g_t^* is constant. Equation 3 means that each element of \bar{M}_t matches the definition of a shifting endpoint in the sense of [Kozicki and Tinsley \(2001\)](#). That is, for any element x_t^* of \bar{M}_t , we have that

$$x_t^* = \lim_{h \rightarrow \infty} E_t[x_{t+h}], \quad (5)$$

and that g_t^* is the shifting endpoint for output growth. Also, because we model these variables together, we can recover internally consistent estimates of i_t^* , π_t^* and $r_t^* = i_t^* - \pi_t^*$ (which is the secular component of the real rate).⁶

Cyclical Components We stack the cyclical components \tilde{i}_t , $\tilde{\pi}_t$ and \tilde{y}_t in the vector \tilde{M}_t , which we assume has standard stationary VAR dynamics, which are given by

$$\tilde{M}_t \equiv \begin{bmatrix} \tilde{i}_t \\ \tilde{\pi}_t \\ \tilde{y}_t \end{bmatrix} = \Phi(L) \begin{bmatrix} \tilde{i}_t \\ \tilde{\pi}_t \\ \tilde{y}_t \end{bmatrix} + \begin{bmatrix} \sigma_{\tilde{i}}' \\ \sigma_{\tilde{\pi}}' \\ \sigma_{\tilde{y}}' \end{bmatrix} \varepsilon_t = \Phi(L)\tilde{M}_t + \tilde{\Sigma}\varepsilon_t, \quad (6)$$

where $\Phi(\cdot)$ is a polynomial function, L is the lag operator and the unconditional mean is zero: $E[\tilde{M}_t] = 0$. The shocks $\varepsilon_t \sim N(0, I)$ are the same as in Equation 3.

For comparability, we embed the cross-equation restrictions used by [Laubach and](#)

⁶Following [Laubach and Williams \(2003\)](#) again, we impose that the norms of the vectors σ_{y^*} and σ_{g^*} be proportional $\lambda_g \equiv \|\sigma_{g^*}\|/\|\sigma_{y^*}\|$. We use the value $\lambda_g = 0.053$ from [Holston, Laubach, and Williams \(2017\)](#). [Laubach and Williams \(2003\)](#) calibrate λ_g to the median unbiased estimator ([Stock and Watson, 1998](#)) to avoid the ‘‘pile-up’’ problem discussed in [Stock \(1994\)](#) and help pin down the potential output y_t^* and potential growth g_t^* .

Williams (2003). First, the reduced-form specification for inflation is given by

$$E_{t-1} [\tilde{\pi}_t] = b_1 \tilde{\pi}_{t-1} + b_2 \tilde{\pi}_{t-2,4} + b_y \tilde{y}_{t-1}, \quad (7)$$

where $\tilde{\pi}_{t-2,4}$ is the inflation rate computed between $t-2$ and $t-4$. Equation 7 ties inflation to the lags in both the inflation rate and the output gap. Second, the reduced-form specification of the output gap is given by

$$E_{t-1} [\tilde{y}_t] = a_1 \tilde{y}_{t-1} + a_2 \tilde{y}_{t-2} + \frac{a_r}{2} \sum_{j=1}^2 \tilde{r}_{t-j}, \quad (8)$$

where \tilde{r}_t is the cyclical component of the real rate $\tilde{r}_t = r_t - r_t^*$, which can be computed by using $\tilde{r}_t = \tilde{i}_t - E_t [\tilde{\pi}_{t+1}]$. Hence, Equation 8 ties the output gap to lags in the output gap and also to the monetary stance \tilde{r}_t . Third, we add to the framework of Laubach and Williams (2003) a reduced-form forward-looking specification of the short-rate gap, in the spirit of the Taylor rule (Clarida, Galí, and Gertler, 1998), given by

$$E_{t-1} [\tilde{i}_t] = \delta_\pi E_{t-1} [\tilde{\pi}_t] + \delta_y E_{t-1} [\tilde{y}_t] + \phi_z \left(\tilde{i}_{t-1} - \delta_\pi \tilde{\pi}_{t-1} - \delta_y \tilde{y}_{t-1} \right), \quad (9)$$

where the last term captures the inertia in the monetary policy (Rudebusch, 2006). Equations 7-9 imply that \tilde{M}_t follows the VAR(5) dynamics with the coefficients given in the appendix. This special case has the same number of parameters as an unrestricted VAR(1) specification, which is a natural comparison that we also estimate below. In the results, our baseline model is based on the cross-equation restrictions of Laubach and Williams (2003). Section 4.5 compares the results obtained using a VAR(1) for \tilde{M}_t and shows that the baseline cross-equation restrictions play an important role in obtaining realistic estimates of the potential output and the output gap.

2.3 Identifying Nominal Shocks

The conditional mean specification in Section 2.2 implies the following restrictions on the matrix Σ in Equation 2:

$$\Sigma \equiv \begin{bmatrix} \sigma'_{\tilde{z}} \\ \sigma'_{\tilde{\pi}} \\ \sigma'_{\tilde{y}} \end{bmatrix} + \begin{bmatrix} \sigma'_{i^*} \\ \sigma'_{\pi^*} \\ \sigma'_{y^*} \end{bmatrix} = \tilde{\Sigma} + \Sigma^*, \quad (10)$$

and, as in a standard structural VAR analysis, additional identification assumptions about Σ are needed to recover the shocks, ε_t , from the data. The traditional approach starts from the covariance matrix, Ω , of the reduced-form innovations $u_t = \Sigma\varepsilon_t$, which can be recovered directly from the data. Since it must be that $\Omega = \Sigma\Sigma'$, we need an additional $N(N - 1)/2$ restrictions as well as sign and ordering assumptions to identify the shocks (N is the size of matrix Σ). For instance, in the context of standard structural VARs, the “recursive” identification strategy relies on pinning down the signs and orders of the shocks and imposing $N(N - 1)/2$ exclusion restrictions on the Σ matrix.⁷

Nominal and real shocks The distinct feature of the approach we leverage to identify the nominal shocks is that the same shocks, ε_t , drive the secular and cyclical components of the macro variables. We use the standard definition of a nominal shock. This is a shock that may influence cyclical fluctuations in the output, inflation and interest rates but leaves unchanged the output growth and the real rate in the long run. This definition corresponds to the idea in structural models that a nominal shock does not influence the steady-state output and real rate that would prevail under flexible prices (i.e., the natural output and the natural real rate).⁸ Nominal shocks may capture monetary policy shocks as well as other shocks that affect the supply of inside money via banks or the financial system (Friedman and Schwartz, 1963; Brunnermeier and Sannikov, 2016), and we do not take a stand on what sources are more important.

⁷Kilian and Lütkepohl (2017) discuss identification in structural VAR models.

⁸See, e.g., Gali (2008) for a textbook discussion.

Based on this definition of the nominal shock, we introduce the following restrictions on the variance parameters:

$$e_1' \sigma_{r^*} = 0 \quad e_1' \sigma_{g^*} = 0 \quad e_2' \sigma_{g^*} = 0, \quad (11)$$

where e_j is a vector of zeros but with element j equal to 1 and $\sigma_{r^*} = \sigma_{i^*} - \sigma_{\pi^*}$. While we defer a formal discussion until Section 3.2, for now, we state that these restrictions are enough to identify Σ . The first two restrictions mean that (i) the nominal shock is ordered first and (ii) it is uncorrelated with the unexpected changes in r_t^* and g_t^* . These two restrictions implement the definition of a nominal shock.

The last restriction deals with the second and third shocks, both which can drive changes to the secular and cyclical components of output and the real rate. Hence, we label them as real shocks. However, we do not attempt to distinguish between these two shocks. Instead, the last restriction simply says that the second shock is uncorrelated with the unexpected changes in g_t^* . This is one of the possible orthogonalizations and has no economic content. For instance, it is not meaningful in an economic sense to distinguish between the impulse response functions with respect to these two shocks.

This identification of the nominal shock is close in spirit to the long-run restrictions of the demand and supply shocks introduced by [Blanchard and Quah \(1989\)](#). Their identification strategy pins down a “demand” shock that has no long-run impact on unemployment and output as well as a “supply” shock that can have a long-run impact on output (see [Dungey et al. \(2015\)](#) for a recent analysis of long-run restrictions).

Caveats [Blanchard and Quah \(1989\)](#) analyze plausible circumstances in which the interpretation in terms of demand and supply shocks is valid; however, they also discuss important caveats, which are worth repeating here. But note that, despite these caveats, we believe that our approach provides a useful perspective on the changing role of nominal and real shocks in the bond market.

One caveat is that the demand shock (or the nominal shock in our case) in the model may commingle a variety of actual demand disturbances with different dynamic effects on interest rates and output. This aggregation of different effects can influence the interpretation of impulse response functions. This caveat is an important reason why we use the label “nominal” instead of “demand” (as well as “real” instead of “supply”), since this aggregation may capture supply shocks that have little or no long-run impact.

Another caveat is that even demand disturbances may have long-run impacts on output. Demand disturbances that influence the savings rate may subsequently influence the long-run capital stock and output. Increasing returns or learning-by-doing would also mean that we may not be able to recover distinct demand and supply shocks by using long-run restrictions. In this case, the identification we use to find the nominal shocks would push these demand disturbances as part of the real shocks that we identify. This offers one interpretation of one of the two types of real shocks—the one that is uncorrelated with g_t^* —although there are other possibilities.

Even then, if demand has no long-term impact, its effects on capital accumulation may be indistinguishable from a truly permanent impact in finite-sample data. [Blanchard \(2018\)](#) voices this issue, given the slow recovery that followed the 2008-2009 financial crisis, when he asks whether potential output is really independent of nominal disturbances (or monetary policy) and whether there really is no long-run trade-off between inflation and output. [Jordà, Singh, and Taylor \(2020\)](#) provide international evidence that the capital stock and total factor productivity exhibit hystereses with respect to monetary policy shocks.

2.4 Time-varying Loadings

The class of models discussed in Section 2.2 has constant loadings parameters $\bar{\Sigma}$, $\tilde{\Sigma}$ and σ_{y^*} , but we are mostly interested in the specifications with time-varying loadings. There are several reasons for this. First, the consensus is that π_t^* is more volatile during the

1970s and 1980s relative to the period following the 1990s (see, e.g., Figure 2b in [Stock and Watson 2016](#)) but that r_t^* is relatively more volatile during the period after the mid-1990s. The existing estimates of π_t^* show a flat pattern starting sometime after the mid-1990s. By contrast, the existing estimates of r_t^* show larger changes and a declining pattern after the mid-1990s. There is also ample evidence that the cyclical component of the short rate, \tilde{i}_t , is much more sensitive early in recessions or around periods of financial distress ([Cieslak and Povala, 2016](#)).

These patterns suggest that sensitivities to macroeconomic shocks have changed over time. We interpret these as changes in the loadings of the shocks on the cyclical and secular components of the macro variables. The results in Section 4 show that the variations over time in these loadings improve the identification of the shocks in models that combine secular and cyclical components. These time-varying loadings also mean that the macro variables have time-varying volatilities.

Cyclical Components We consider an approach that is simple and parsimonious and preserves the connections with the established identification strategies in the literature. The loadings of the cyclical components on the shocks are given by the following:

$$\tilde{\Sigma}_t = \begin{bmatrix} v_{\tilde{i},t} & 0 & 0 \\ 0 & v_{\tilde{\pi},t} & 0 \\ 0 & 0 & v_{\tilde{y},t} \end{bmatrix} \begin{bmatrix} \sigma'_i \\ \sigma'_{\tilde{\pi}} \\ \sigma'_{\tilde{y}} \end{bmatrix} = \tilde{V}_t \tilde{\Sigma}, \quad (12)$$

where the matrix $\tilde{\Sigma}$ is unrestricted up to some identification assumptions we discuss below. Each vector $\sigma \cdot$ combines the shocks into fixed proportions while each scalar, $v_{\cdot,t}$, controls the scale of the innovations to each of the cyclical components over time. The dynamics for the time-varying factors, $v_{\cdot,t}$, are specified in Section 2.5 and are estimated jointly with other parameters of the model. Then, the cyclical components have dynamics given by

$$\tilde{M}_t = \Phi(L)\tilde{M}_t + \tilde{\Sigma}_{t-1}\varepsilon_t.$$

Secular Components We follow a similar route for the loadings of the shocks on the secular components. For \bar{M}_t , these are given by the following:

$$\bar{\Sigma}_t = \begin{bmatrix} v_{i^*,t} & 0 & 0 \\ 0 & v_{\pi^*,t} & 0 \\ 0 & 0 & v_{g^*,t} \end{bmatrix} \begin{bmatrix} \sigma'_{i^*} \\ \sigma'_{\pi^*} \\ \sigma'_{g^*} \end{bmatrix} = \bar{V}_t \bar{\Sigma}, \quad (13)$$

where the matrix $\bar{\Sigma}$ is unrestricted. For the potential output, we have

$$\sigma_{y^*,t} = v_{y^*,t} \sigma'_{y^*}, \quad (14)$$

but we restrict the scaling factor $v_{y^*,t} = v_{g^*,t}$. Again, the dynamics for each scalar, $v_{\cdot,t}$, are presented in Section 2.5 along with parameters that are estimated jointly with the rest of the model. These scalars, $v_{\cdot,t}$, vary the scale of the innovations to each of the secular components. However, the relative contribution of each shock is parameterized by the matrix $\bar{\Sigma}$. Then, the dynamics of the secular components are given by

$$\begin{aligned} \bar{M}_t &= \bar{M}_{t-1} + \bar{\Sigma}_{t-1} \varepsilon_t \\ \Delta y_t^* &= g_{t-1}^* + \sigma'_{y^*,t-1} \varepsilon_t. \end{aligned}$$

2.5 Loading Dynamics

The scaling factors must remain positive, and we use *EGARCH*(1, 1) dynamics for this purpose (see [Nelson, 1991](#)). To illustrate, in the case of the cyclical short rate \tilde{i}_t , we have that

$$\ln \left(v_{\tilde{i},t}^2 \right) = \omega_{\tilde{i}} + \beta_{\tilde{i}} \ln \left(v_{\tilde{i},t-1}^2 \right) + g \left(z_{\tilde{i},t} \right), \quad (15)$$

where $g \left(z_{\tilde{i},t} \right)$ is the innovation to the variance process given by a function of the normalized variable $z_{\tilde{i},t} \equiv \frac{\alpha'_i \varepsilon_t}{\sqrt{\alpha'_i \alpha_i}}$. The scaling factors $v_{\tilde{\pi},t}$ and $v_{\tilde{y},t}$ for the inflation and output gaps,

respectively, have similar dynamics. Therefore, the innovation is driven by the shocks ε_t , the loadings of each shock are given by the vectors of parameters $\alpha_{\tilde{i}}$ and the denominator $\sqrt{\alpha'_{\tilde{i}}\alpha_{\tilde{i}}}$ is a convenient normalization. All these parameters are estimated jointly. The function $g(\cdot)$ allows for asymmetry in the effects of $z_{\tilde{i},t}$ on $v_{\tilde{i},t}^2$, which is consistent with the evidence of the substantial asymmetry in the pattern observed in the output gap measures across expansion and recession phases recorded in the National Bureau of Economic Research (Morley and Piger, 2012).⁹

We introduce an important restriction to the process of the time-varying loadings for the secular components. The reason for this is that, in standard unobserved component models, the variance of the secular changes can diverge as the forecast horizon increases. This severely over-estimates the uncertainty around the future, which affects standard analytical tools like the variance's decomposition. To see what the issue is, consider the variance of $i_{t+\tau}^*$ some τ periods ahead:

$$Var_t [i_{t+\tau}^*] = (\sigma_{i^*}'\sigma_{i^*}) \left(\sum_{j=1}^{\tau} E_t [v_{i^*,t+j-1}^2] \right),$$

where the coefficient for each term in the sum is equal to 1 because of the unit root in the process for i^* . Proposition 1 provides sufficient conditions on the properties of $v_{i^*,t}^2$ for the convergence of $\lim_{\tau} Var_t [i_{t+\tau}^*]$.¹⁰

Proposition 1 Suppose the scalar $v_{i^*,t}^2$ has EGARCH(1,1) dynamics given by Equation (15), where $z_{i^*,t} \equiv \frac{\alpha_{i^*}'\varepsilon_t}{\sqrt{\alpha_{i^*}'\alpha_{i^*}}}$ and

$$g(z_{i^*,t}) = \sqrt{\alpha_{i^*}'\alpha_{i^*}} z_{i^*,t} + \kappa_{i^*} (|z_{i^*,t}| - E[|z_{i^*,t}|]).$$

⁹This function is given by $g(z_{\tilde{i},t}) = \sqrt{\alpha_{\tilde{i}}'\alpha_{\tilde{i}}} z_{\tilde{i},t} + \kappa_{\tilde{i}} (|z_{\tilde{i},t}| - E[|z_{\tilde{i},t}|])$.

¹⁰This approach can be generalized to other processes. A similar result is available from the authors in the context of the asymmetric GARCH(1,1) dynamics of Hentschel (1995).

If $\beta_{i^*} = 1$ and $\omega_{i^*} < \bar{\omega}_{i^*}$, where

$$\bar{\omega}_{i^*} \equiv \kappa_{i^*} \sqrt{\frac{2}{\pi}} - \ln \left(\frac{\exp \left(\frac{(\sqrt{\alpha'_{i^*} \alpha_{i^*} + \kappa_{i^*}})^2}{2} \right) \Phi_N \left(\kappa_{i^*} + \sqrt{\alpha'_{i^*} \alpha_{i^*}} \right)}{\exp \left(\frac{(\sqrt{\alpha'_{i^*} \alpha_{i^*} - \kappa_{i^*}})^2}{2} \right) \Phi_N \left(\kappa_{i^*} - \sqrt{\alpha'_{i^*} \alpha_{i^*}} \right)} \right),$$

then $Var_t [i_{t+\tau}^*]$ converges,

$$Var_t [i_{t+\tau}^*] \xrightarrow{\tau} \frac{(\sigma'_{i^*} \sigma_{i^*})}{\theta_{i^*}} v_{i^*,t}^2,$$

where $\theta_{i^*} \equiv 1 - e^{\omega_{i^*} - \bar{\omega}_{i^*}}$ and $\Phi_N(\cdot)$ is the standard normal distribution's cumulative distribution function.

The restrictions on the $EGARCH(1, 1)$ guarantee that $Var_t [i_{t+\tau}^*]$ converges to a positive scalar for arbitrarily large horizons, τ . Intuitively, the shocks that are very far away in the future eventually have no impact on the secular component (in expectations). [Stock and Watson \(2007\)](#) and [Wright \(2011\)](#) analyze unobserved component models with stochastic volatility (for the case of inflation), and the volatility dynamics that they use are obtained by imposing $\omega = 0, \beta = 1$ and $\kappa = 0$, which implies that $Var_t [\pi_{t+\tau}^*]$ diverges with the horizon in the model, since $\omega = \bar{\omega} = 0$. Absent the restriction that we introduce, the volatility of any of the variables that are influenced by the secular components would diverge as the forecast horizon increases, which would severely over-estimate the uncertainty and affect such standard analytical tools as the decomposition of the variance.

2.6 Bond Prices

Pricing Kernel This section develops a model for nominal yields where macro shocks are priced and have both secular and cyclical impacts on yields. We build on the assumption of no arbitrage, which guarantees the existence of a positive pricing kernel process, ξ_{t+1} , such that the price of any asset, V_t , that does not pay any dividends at time $t + 1$ satisfies

$V_t = E_t[\xi_{t+1}V_{t+1}]$. For tractability, we consider the following pricing kernel:

$$\ln(\xi_{t+1}) = -i_t + \lambda'_t \varepsilon_{t+1} + \frac{1}{2} \varepsilon'_{t+1} \Gamma_t \varepsilon_{t+1} - \chi_t + \lambda_{f,t} \varepsilon_{f,t+1}, \quad (16)$$

where χ_t is the convexity term given by

$$\chi_t = \ln E_t \left[\exp \left(\lambda'_t \varepsilon_{t+1} + (1/2) \varepsilon'_{t+1} \Gamma_t \varepsilon_{t+1} \right) \right] + \ln E_t \left[\exp \left(\lambda_{f,t} \varepsilon_{f,t+1} \right) \right].$$

The macroeconomic shocks in this economy are priced in the bond market, with the prices of risk λ_t to be defined. The pricing kernel also includes a quadratic term because the time-varying loadings generate the variance risk. This pricing kernel specification follows [Monfort and Pegoraro \(2012\)](#) to pin down the price of the variance risk Γ_t , also to be defined below. Finally, we follow [Ang and Piazzesi \(2003\)](#) and the subsequent literature in introducing one additional latent source of risk, $\varepsilon_{f,t+1}$, which is uncorrelated with the macro shocks. This new risk drives a latent factor, f_{t+1} , that is meant to capture the residual variations in nominal yields that are unexplained by the macro variables. We construct f_t such that it does not influence the dynamics of the short rate but can influence the term premium. Therefore, we label f_t a term structure factor and, for this purpose, we assume that f_t has simple AR(1) dynamics given by

$$f_{t+1} = \mu_f + \phi_f f_t + u_{f,t+1}, \quad (17)$$

where the innovation is given by $u_{f,t+1} = \sigma'_{fM} \tilde{\Sigma} \varepsilon_{t+1} + \sigma_f \varepsilon_{f,t+1}$, and where $\varepsilon_{f,t+1} \sim N(0, 1)$ is a term premium shock that is uncorrelated with the macro shocks.

Prices of Risk For ease of exposition, we specify the prices of risk in the case where \tilde{M}_t has VAR(1) dynamics. Following [Duffee \(2002\)](#), we consider prices of risk that are a linear

functions of the state variables:

$$\begin{bmatrix} \lambda_{\tilde{i},t} \\ \lambda_{\tilde{\pi},t} \\ \lambda_{\tilde{y},t} \\ \lambda_{f,t} \end{bmatrix} = \begin{bmatrix} (I - \Gamma_t)\tilde{\Sigma}_t^{-1} & 0_{N \times 1} \\ 0_{1 \times N} & 1 \end{bmatrix} \times \left(\begin{bmatrix} \lambda_{0,\tilde{i}} \\ \lambda_{0,\tilde{\pi}} \\ \lambda_{0,\tilde{y}} \\ \lambda_{0,f} \end{bmatrix} + \begin{bmatrix} \lambda'_{\tilde{i}} \\ \lambda'_{\tilde{\pi}} \\ \lambda'_{\tilde{y}} \\ \lambda'_f \end{bmatrix} \begin{bmatrix} \tilde{i}_t \\ \tilde{\pi}_t \\ \tilde{y}_t \\ f_t \end{bmatrix} \right), \quad (18)$$

with one important difference due to the first term on the right-hand side of this equation.¹¹ This term is different in our case because the sources of risks, ε_t , have time-varying loadings on macroeconomic variables, while in [Duffee \(2002\)](#) all state variables are Gaussian with constant variance. In the specification that we detail below, the price of variance risk Γ_t collapses to zero when all state variables have constant variances, hence matching the specification in [Duffee \(2002\)](#).

In addition to the assumption of linearity, we embed two additional restrictions in the specification of the prices of risk. First, we restrict the price of risk of the term structure factor in Equation 18 as follows:

$$\lambda'_f = [0'_{3 \times 1} \quad \lambda_{ff}]. \quad (19)$$

This assumption ensures that the financial market factor does not affect the dynamics of i_t , does not enter the expectations component of bond yields, but only affects bond yields through term premiums.

The second restriction in Equation 18 is implicit: we exclude the secular components from the prices of risk. This is to guarantee that the expected returns from holding bonds are stationary. This restriction also makes the model parsimonious and improves the accuracy of the parameter estimates. Most existing DTSMs use stationary pricing

¹¹In the case where \tilde{M}_t has general VAR(p) dynamics, the second term on the right-hand side of Equation 18 also includes $p - 1$ lags of \tilde{M}_{t-1} , as in [Joslin et al. \(2013a\)](#).

factors. There are many lines of evidence supporting this assumption. [Hall, Anderson, and Granger \(1992\)](#) provide early evidence that bond yields are cointegrated and that the bond risk premium is stationary (see also [Engsted and Tanggaard, 1994](#)). Consistent with this observation, [Cieslak and Povala \(2015\)](#) find that, for bonds, the predictability of excess returns that are based on current yields essentially doubles once the inflation trend is removed from these yields. In fact, they report that the predictable component of bond returns has a half-life of 10 months (yields have half-lives of five years or more) and that this is orthogonal to the inflation trend. [Bauer and Rudebusch \(2020\)](#) examine an important dynamic term structure model where bond yields share a common trend given by i_t^* , which corrects for inflation and the real rate trends. They emphasize that describing bond yields as cointegrated processes is a useful description in small samples.

Unspanned Volatility To be consistent with the observation that the yields do not span the conditional volatilities of the yields ([Collin-Dufresne and Goldstein, 2002](#); [Collin-Dufresne, Goldstein, and Jones, 2009](#)), we assume the following price of the variance risk:

$$\Gamma_t = I - \tilde{\Sigma}'_t (\tilde{\Sigma} \tilde{\Sigma}')^{-1} \tilde{\Sigma}_t. \quad (20)$$

This term disappears if the scaling factors $v_{\cdot,t}$ are constant and equal to 1, in which case the conditional variance $\tilde{\Sigma}_t$ is constant and the coefficient Γ_t is zero. Hence, our approach uses multiple unspanned volatility factors and is similar to earlier works by [Creal and Wu \(2017\)](#) and, more recently, [Hansen \(2019\)](#) that use unspanned volatility factors to analyze macroeconomic risks in bond yields.

Bond Prices Given these assumptions about the pricing kernel, using the standard recursive argument, it follows that the price of a nominal bond with n periods until maturity, $P_t^{(n)} = E_t[\xi_{t+1} P_{t+1}^{(n-1)}]$, is given by

$$P_t^{(n)} = \exp(A_n - n i_t^* + B'_n \tilde{M}_t + B_{f,n} f_t). \quad (21)$$

The bond yields are given by

$$Y_t^{(n)} = i_t^* + a_n + b'_n \tilde{M}_t + b_{f,n} f_t, \quad (22)$$

where the coefficients are given in the appendix and with the initial conditions $a_1 = 0$, $b_1 = e_1$ and $b_{f,1} = 0$ so that $Y_t^{(1)} = i_t = i_t^* + e'_1 \tilde{M}_t$. Equation 22 means that bond yields that have a stationary risk premium share one common cointegration relationship, where any interest rate spread is stationary. The constant a_n captures the average term premium (e.g., the constant is 1.66 percent for the 10-year yield in our results).

Risk-Neutral Dynamics The existence of the positive pricing kernel ξ_{t+1} is equivalent to the existence of an equivalent martingale measure (or risk-neutral measure) \mathbb{Q} , such that $E_t^{\mathbb{Q}}[V_{t+1}] = E_t[e^{i_t} \xi_{t+1} V_{t+1}]$. It is useful to look at the dynamics under this risk-neutral measure:

$$\begin{aligned} f_t &= \phi_f^{\mathbb{Q}} f_t + u_{f,t}^{\mathbb{Q}} \\ \tilde{M}_t &= K + \Phi^{\mathbb{Q}} \tilde{M}_{t-1} + \Phi_{Mf}^{\mathbb{Q}} f_{t-1} + \tilde{\Sigma} \varepsilon_t^{\mathbb{Q}} \\ \bar{M}_t &= \bar{K} + \bar{M}_{t-1} + \bar{\Sigma} \varepsilon_t^{\mathbb{Q}}, \end{aligned} \quad (23)$$

where $\Phi_{Mf}^{\mathbb{Q}}$ is a 3×1 vector. One key feature is that the autoregressive matrix for \bar{M}_t under the risk-neutral measure \mathbb{Q} is the identity matrix, exactly as under the physical or historical measure. This is what guarantees the stationarity of the bond risk premium. To see this, consider a more general case where this autoregressive matrix is given by $\bar{\Phi}^{\mathbb{Q}}$. Applying the conditional expectations operators to the definition of the log of the excess

bond returns, we find that the (log) bond risk premium is given by the following:

$$\begin{aligned}
brp_{t \rightarrow t+n}^{(\tau)} = & A_{\tau-n} + A_n - A_\tau + \bar{B}'_{\tau-n} \left(I - \left(\bar{\Phi}^Q \right)^n \right) \bar{M}_t \\
& + B'_{\tau-n} \left\{ E_t \left[\tilde{M}_{t+n} \right] - \left(\Phi^Q \right)^n \tilde{M}_t \right\} \\
& + B_{f,\tau-n} E_t [f_{t+n}] - (B_{f,\tau} - B_{f,n}) f_t,
\end{aligned} \tag{24}$$

which immediately shows that the risk premium is stationary if and only if $\bar{\Phi}^Q = I$.¹²

3 Estimation

3.1 Data

Macro Data We estimate the model using US data. The sample period starts in 1983 and ends in 2019. The sampling frequency is quarterly. Figure 2a shows the macroeconomic data. For the short rate, i_t , we use the secondary market rate for 3-month Treasury bills, in percentage terms. For the inflation rate π_t , we use the compounded rate of change in the personal-consumption-expenditures index, excluding food and energy, annualized, in percentage terms and seasonally adjusted. For output y_t , we use real gross domestic product, in log of billions and seasonally adjusted. We use the macro data available from the website of the Federal Reserve Bank of St. Louis.

Yield and Survey Data Figure 2b shows the yields for bonds with 2-, 5- and 10-year maturities. For our estimation, we select zero-coupon yields with annual maturities between one and ten years, in annualized percentage terms, from the GSW database ([Gurkaynak, Sack, and Wright, 2006](#)). We also use data on the rates of inflation swaps

¹²This expression for the bond risk premium follows from the definition of excess returns, and from the price of a zero-coupon nominal bond with n periods to maturity in the general case, given by:

$$P_t^{(n)} = \exp \left(A_n + \bar{B}'_n \bar{M}_t + B'_n \tilde{M}_t + B_{f,n} f_t \right),$$

with coefficients given by standard recursions, starting at $A_1 = 0$, $\bar{B}_1 = B_1 = -e_1$ and $B_{f,1} = 0$.

with 2-, 5- and 10- year maturities.

In the spirit of [Kim and Orphanides \(2012\)](#), we use survey data to help with the precision of the estimation. Long-horizon survey forecasts can improve the identification of secular and cyclical components. [Orphanides and Wei \(2012\)](#) also use surveys as a way to capture structural changes in a stationary VAR that they estimate recursively. [Cieslak and Povala \(2015\)](#) find that using survey-based expectations of inflation from a variety of sources to control for the expected inflation in predictive regressions of bond excess returns consistently strengthens the return predictability.

Using survey data plays an important role in disciplining the model. The Beveridge-Nelson decomposition relies on the difference in the persistence between components. In practice, it is difficult to pin down this difference in a short sample when using only the likelihood. Intuitively, increasing the variance of the secular changes may improve the fit of the one-period forecasts, which is the basis of the likelihood, but this can also cause over-fitting and produce secular component estimates that are implausible. However, at very long horizons, the model forecasts depend on the secular components but not on the cyclical components. Hence, long-term surveys help to identify variations in the secular components as well as obtain stable and sensible estimates of the secular components.

We use the estimates of the long-horizon forecasts for the short rate, inflation and GDP growth available from [Crump, Eusepi, and Moench \(2016\)](#). The idea behind their estimates is that different surveys offer forecasts at different frequencies and horizons, which can be pooled together to obtain smoothed estimates at regular frequencies and fixed horizons. Clearly, the measurement errors surrounding the smoothed estimates are not negligible, and this uncertainty will be embedded in the estimation procedure, which we discuss below. Therefore, different model specifications may deliver a range of plausible but different estimates, even when survey data are used.

Figure 2c shows long-horizon survey forecasts for the short rate, inflation and output growth in our sample period. The secular changes are visually apparent. There is a slow

decline for inflation, which exhibits a plateau starting at around 1998. There is also a slow decline as well as a plateau starting at around 1998 for the short rate, but this decline resumes after 2009. This renewed decline is absent from the survey forecasts for inflation and have been attributed to declines in the real neutral rate. Finally, the long-horizon forecasts show output growth rising to 3 percent for much of the decade after 1998 but declining to below 2 percent by the end of our sample period.

3.2 Identification

We proceed in steps for clarity. In the first step, we take as given the identification of Σ_t from $\Sigma_t \Sigma_t' = \Omega_t = (\tilde{\Sigma}_t + \Sigma_t^*)(\tilde{\Sigma}_t + \Sigma_t^*)'$, date by date, to discuss the parameter identification in a statistical sense. For this discussion, we group the parameters into two blocks. The first block contains the parameters that drive the time-varying loadings given in Section 2.4. The second block contains the parameters that drive the conditional means given in Section 2.2. In the second step, we discuss the identification of Σ_t .

If Σ_t is identified date by date and if the block of conditional mean parameters is identified, then the block of conditional variance parameters is also identified. This block contains the parameters ω_{\cdot} , β_{\cdot} , α_{\cdot} and κ_{\cdot} given in Equation 15 as well as the parameters ω_{\cdot^*} , α_{\cdot^*} and κ_{\cdot^*} given in Proposition 1. The parameters can be estimated using the [Nelson \(1991\)](#) EGARCH recursions, which can be started at the initial values $v_{\cdot,0} = 1$, without loss of generality, and using the conditional mean parameter to obtain updates of the shocks in the EGARCH recursions.

The second block contains the parameters σ_{g^*} , $\tilde{\Sigma}$ and Σ^* in Equation 10, as well as Φ_j for $1 \leq j \leq p$ in Equation 6.¹³ If Σ_t is identified, then we can analyze the identification of these parameters based on the moving-average representation of $\Delta M_t \equiv M_t - M_{t-1}$:

$$\Delta M_t = g_0^* e_3 + \Sigma_{t-1} \varepsilon_t + \sum_{j=1}^{t-1} \Theta_{t-1,j} \varepsilon_{t-j}, \quad (25)$$

¹³For this discussion, we consider the general VAR(p) dynamics, which are that $\Phi(L) = \sum_{j=1}^p \Phi_j L^j$.

where the coefficients $\Theta_{t-1,j}$ are predetermined with respect to ΔM_t . This means that, when updating, it is feasible to alternate between the EGARCH recursions and the conditional mean recursions. The parameter in this moving-average representation can be used to identify the structural parameters. These coefficients are given by

$$\Theta_{t-1,j} = (\Psi_j - \Psi_{j-1}) \tilde{\Sigma}_{t-1} + e_3 \sigma'_{g^*,t-1}, \quad (26)$$

where the coefficients can be obtained recursively from the structural Φ_k using the following:

$$\Psi_j = \sum_{k=1}^p \Phi_k \Psi_{j-k}, \quad (27)$$

for $j = 1, 2, \dots$ with $\Psi_0 = I_3$ and $\Psi_j = 0$ for any $j < 0$.¹⁴ Based on Equations 25-27, the online appendix establishes the following lemma.

Lemma 1 *If Σ_t is identified, then $\tilde{\Sigma}$, Σ^* , σ_{g^*} and Φ_j for $1 \leq j \leq p$ are also identified.*

This leaves the question of how to identify the parameter Σ_t of the reduced-form innovations. This is the central identification problem in a standard structural VAR with constant loadings. However, introducing time-varying loadings means that the identification of the shocks becomes somewhat more involved. [Primiceri \(2005\)](#) analyzes the issue in a multivariate autoregressive system where all of the parameters can change over time. The intuition from this more general case applies here: the identification of the shocks requires $N(N - 1)/2$ restrictions for each date t . In our case, this requires three restrictions on the matrix Σ_t , date by date.

To ease the computational burden of estimating the parameters, we first identify Σ_t with the Cholesky decomposition of Ω_t , which is unique. Once we obtain the parameter estimates, we can rotate the estimated shocks, ε_t , into new shocks, $\hat{\varepsilon}_t$, that satisfy the

¹⁴The coefficients Ψ_j are those in the equivalent infinite-horizon moving-average representation for \tilde{M}_t , $\tilde{M}_t = \sum_{j=0}^{\infty} \Psi_j \tilde{\Sigma}_{t-j-1} \varepsilon_{t-j}$, where we set \tilde{M}_0 to zero, which is its population mean.

desired identification restrictions by using one set of equations for every date. Lemma 2 states this result formally.

Lemma 2 *There exists a unique matrix R_t such that the macroeconomic shocks given by $\hat{\varepsilon}_{t+1} = R_t^{-1} \varepsilon_{t+1}$ verify the identification restrictions in Equation 11, date by date. These shocks have a loading matrix given by*

$$\begin{aligned}\hat{\Sigma}_t &= \tilde{V}_t \tilde{\Sigma} R_t \\ \hat{\Sigma}_t^* &= \bar{V}_t \Sigma^* R_t,\end{aligned}$$

and these new shocks are observationally equivalent to $\hat{\Sigma}_t \hat{\varepsilon}_{t+1} = \Sigma_t \varepsilon_{t+1}$, where $\hat{\Sigma}_t = \hat{\Sigma}_t + \hat{\Sigma}_t^*$.

The online appendix provides the proof and the algebra to construct this rotation matrix. Different identification restrictions correspond to different rotation matrices. [Primerici \(2005\)](#) follows a similar two-step identification strategy. This two-step approach differs from a more traditional one that estimates the reduced-form innovations, u_t , and the covariance matrix, Ω , in a first step, say via ordinary least squares, and then identifies Σ from Ω in a second step to recover the shocks. The reason for this difference is practical. Estimation of the model with time-varying loadings requires the shocks to update the loadings when using the recursions such as in Equation 15. Hence, given the estimates, we can use matrix R_t to compute the new shocks, a new loading matrix that determines the macro dynamics as well as important objects such as impulse response functions and variance decompositions.

3.3 Likelihood

Estimation via maximum likelihood is straightforward and much of the details are relegated to the online appendix. We do not attempt to implement the model using real-time data, which subjects us to the critique in [Clark and Kozicki \(2005\)](#) that notes that estimates of secular components in real-time data are influenced by data revisions. The likelihood

is available up to a recursion and does not need a Kalman filter because the innovations are available from the observed macro data, $u_{t+1} = M_{t+1} - E_t[M_{t+1}]$. The likelihood is given by

$$\mathcal{L}_{M,t+1} = -\frac{3}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\Omega_t)) - \frac{1}{2} u'_{t+1} \Omega_t^{-1} u_{t+1}, \quad (28)$$

where the innovations, u_t , and the conditional variances are available recursively in the appendix.

[Cieslak and Povala \(2015\)](#) find that long-term survey forecasts also do a good job of stripping yields from the trend in the expected inflation and pinning down the cyclical bond risk premium (see their Table 10). In addition, they report that survey-based measures of expected inflation bring in additional information in this context, relative to bond yields. [Kim and Orphanides \(2012\)](#) propose using survey forecasts of short-term interest rates as an additional input to the estimation of the persistent component of yields, albeit their analysis focuses on stationary models. In the same spirit, we add measurement equations based on the survey forecasts, \bar{S}_t , of macro variables that are available from [Crump, Eusepi, and Moench \(2016\)](#). These survey forecasts correspond exactly to the very-long-horizon forecasts \bar{M}_t in the model:

$$\bar{S}_t = \bar{M}_t + e_{S,t}, \quad (29)$$

where the measurement errors, $e_{S,t}$, have independent normal distributions. The likelihood of the survey data is then given by the following:

$$\mathcal{L}_{S,t+1} = -\frac{3}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\Omega_S)) - \frac{1}{2} e'_{S,t+1} \Omega_S^{-1} e_{S,t+1}. \quad (30)$$

We calibrate the matrix Ω_S to the standard deviations of the measurement errors obtained from [Crump, Eusepi, and Moench \(2016\)](#), which capture the sampling uncertainty that surrounds their estimates of \bar{S}_t that are based on the available data. In this way, the long-

term surveys provide important information about the unobserved components, but the standard errors that we calibrate are sufficiently wide that, at times, our baseline estimates disagree with the survey forecasts.

To compute the likelihood of bond yields, and following [Chen and Scott \(1993\)](#) and [Joslin, Le, and Singleton \(2013b\)](#), we assume that one linear combination of yields is measured perfectly. This allows us to recover the latent factor, f_t , directly from the cross-section of bond yields, given the parameters of the models, and also to bypass a Kalman filter step. This likelihood is given in the online appendix. Finally, the inflation swap rate, $\bar{\pi}^{(n)}$, is derived as the risk-neutral expectations of the average inflation rate over some horizon, n . The likelihood of the swap data is then given by the following:

$$\mathcal{L}_{sw,t} = -\frac{3}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\Omega_{sw})) - \frac{1}{2} u'_{sw,t} (\Omega_{sw})^{-1} u_{sw,t}, \quad (31)$$

where $u_{sw,t}$ is the vector of the measurement errors for the inflation swap rates.

4 Results

We estimate several versions of the model, with a focus on understanding nominal shocks. Our baseline model incorporates the cross-equation restrictions given in Section 2.2 as well as time-varying loadings for shocks. We also estimate the case with constant loadings, which is nested in the baseline model. We find that it is important to let the shock loadings vary over time as this allows us to capture how the impact of nominal shocks changes over time. We also estimate the case with unrestricted VAR(1) dynamics. We find that the cross-equation restrictions from [Laubach and Williams \(2003\)](#) help pin down the output gap and the potential output and that the VAR(1) estimates are implausible, for instance when compared with the Congressional Budget Office (CBO) estimates.

4.1 Real and Nominal Shocks

The shocks that we recover, especially the nominal shocks, are key outputs of the model. Panel (a) of Figure 2 shows the shocks in our baseline model and Panel (b) reports the shocks in a model with constant loadings. The results from both models suggest that nominal shocks are prevalent in every recession in our sample but that they are less prevalent during the protracted recovery since 2009.

However, there are some differences among the shocks in these two models. To gauge the differences, Panels (c)-(e) of Figure 2 show scatter plots for each type of shock in these two models (the units in the scatter plots are standard deviations). The correlations range from 0.61 and 0.72 for the two real shocks and up to 0.79 for the nominal shocks (this translates into univariate regression R^2 s of 37, 52 and 62 percent, respectively). Some of the differences are large. One notable example is the large negative shocks that occur late in 2008. The negative nominal shock is much larger in the model with time-varying loadings, while the real shocks play a bigger role in the model with constant loadings. This difference corresponds to the outlier that we can see to the bottom left of Panel (e).

Overall, this shows that the assumption of constant loadings influences the identifications of the nominal and real shocks. In the following section, we show that this leads to different interpretations of some of the cyclical and secular changes in interest rates. In this particular case, using constant loadings leads to the conclusion that the great financial crisis has a large secular impact on inflation and the real rate and, therefore, on bond yields.

4.2 Endpoint Estimates

The endpoints π_t^* and r_t^* play an important role in “de-trending” bond yields within the model, as in [Cieslak and Povala \(2015\)](#) and [Bauer and Rudebusch \(2020\)](#). This section shows that assuming time-varying or constant shock loadings leads to different endpoint estimates. More importantly, using time-varying loadings reveals different nominal

shocks.

Inflation Endpoint Figure 3 presents the inflation endpoint, π_t^* , from the baseline model and the model with constant loadings. The overall pattern is unsurprising. Early in our sample, the inflation endpoint is around 6 percent, but it quickly falls and remains on a plateau of around 4 percent for most of the 1980s. Starting in 1992, the endpoint of inflation gradually declines again and reaches a lower plateau of slightly above 2 percent around 1998; it stays there for the remainder of the sample. The models closely agree with each other but the baseline model produces a much smoother path because of the time-varying loading $v_{\pi^*,t}$. To see the effect of this loading, Panel (b) of Figure 3 shows the time-series of the one-year-ahead volatility for the shifting endpoint $\text{var}_t(\pi_{t+4}^*)$. This shows that, in our sample, the inflation endpoint becomes anchored. Early on, the annual volatility is around 30 basis points, but it dramatically declines after 1990, reaching a very low level of between 1 and 3 basis points after 2000 (for comparison, in our sample, the standard deviation of the observed inflation rate is 52 basis points). In the model with constant volatility, the conditional variance $\text{var}_t(\pi_{t+4}^*)$ is constant at 23 basis points, which is why the endpoint estimate continues to fluctuate in this model throughout our sample period.

One key benefit of our framework is that we can document the contribution of nominal shocks to the inflation endpoint π_t^* , shown in Panel (c) of Figure 3. We also report the total contribution of both real shocks. The baseline results show that both nominal and real shocks influence secular inflation. Overall, across the sample, around two thirds of the decline is attributed to the cumulative impact of the nominal shocks. The baseline results also show that the impact of both shocks declines toward zero during the sample, which is due to changes in the $v_{\pi^*,t}$ loading. Early in the sample, the loading of the nominal shocks is around 0.1, meaning that π_t^* tends to absorb one tenth of the nominal shocks during that period, but this loading declines toward zero over the course of our sample. The loadings of the real shocks exhibit similar changes because of changes to $v_{\pi^*,t}$.

How the shock loadings change over time plays an essential role in capturing the anchoring of π_t^* . To see this, in Panel (d) we report the contribution of each shock in the model with constant loadings. In this restricted case, the nominal shocks play a much smaller role throughout the sample. In fact, in this model, most of the disinflationary period until the mid-1990s is attributed to the real shocks. Overall in the sample, around 60 percent of the decline in π_t^* is attributed to the real shocks. This model's interpretation clashes with the general view that the decline in secular inflation during this period is largely due to the actions of central banks. In addition, the real shocks in this model continue to cause substantial updates to π_t^* throughout the sample. This is inconsistent with the general view that variations in secular inflation are very small toward the end of the sample. For instance, the episode of the great financial crisis shows a large negative impact of real shocks on π_t^* but that this reverts after only a few years.

Real-rate Endpoint Panel (a) of Figure 4 presents estimates of the real-rate endpoint r_t^* . The overall pattern shows three phases that are broadly in line with the results reported by [Laubach and Williams \(2003\)](#) and [Holston, Laubach, and Williams \(2017\)](#) for the US. First, r_t^* exhibits a gradual decline over the ten years between 1983 and 1993, ending soon after the 1990-1991 recession. This is a period when the long-term forecasts show declines for the short rate that are more pronounced than the declines for inflation. Then, the r_t^* estimate rises from 1993 to 2000, which corresponds to Fed chairman Greenspan's prediction for an acceleration in productivity growth. During this period, output growth accelerates but inflation remains subdued, and these are attributed to the long-awaited impact of information technology on productivity. [Meyer \(2003\)](#) offers a detailed review of the policy discussions of that period. However, part of the decline in r_t^* may be due to real benefits from the lower level and lower volatility of inflation (see, e.g., [Barro \(2013\)](#) and [Al-Marhubi \(1998\)](#) , respectively). After this episode, the neutral rate resumes its decline, reaching 2 percent around 2008 and 0.5 percent in our baseline model at the end of our sample.

Our results shed some light on the volatility of the neutral real rate. Panel (b) shows that the neutral rate's one-year-ahead volatility is 14 basis points in the model with a constant volatility. By contrast, after 2005, the estimated volatility of the real-rate endpoint in the baseline model is less than 5 basis points. Therefore, the decline in the baseline real-rate endpoint estimates after 2008 is due to a succession of correlated shocks that push r_t^* in the same direction and not to the incidence of a few large shocks.

Panel (c) reports the contribution of both real shocks separately (nominal shocks play no role here). We find that the first type of real shock plays a greater role early in the sample but that the second type of real shock becomes more important during the 2000s. The first type is orthogonal to the innovations in g^* , while the second can influence both r^* and g^* . Hence, changes to the secular real rate tend to be uncorrelated to changes in the potential growth, g_t^* , in the first half of the sample up until roughly 1998.¹⁵ However, the recent decline in r_t^* after 2008-2009 is largely driven by shocks that are correlated with changes to potential shocks, suggesting that technology, capital accumulation and demographics play larger roles during that period. Finally, Panel (d) reports the contribution of real shocks in the model with constant loadings. The difference is stark. The model with constant loadings shows no changes in the nature of the real shocks that are driving r_t^* and no compression in the dispersion of these shocks.

4.3 Cyclical Short Rate

Given the secular components π_t^* and r_t^* , the variations in bond yields due to the expectations component are determined by the dynamics of the cyclical component of the short rate \tilde{i}_t . Panel (a) of Figure 5 shows estimates of \tilde{i}_t in our baseline model as well as in the

¹⁵This raises the question of whether the variations in r_t^* early in our sample can be interpreted as variations in the neutral real rate, because standard models would associate changes to the neutral real rate to changes in potential output growth. We note that our estimates for r_t^* are close to the estimates in [Laubach and Williams \(2003\)](#). We leave for future research analyses whether further economic restrictions have impacts on the identification of shocks to the neutral rate.

case with constant loadings. Overall, the estimates align very closely across models. For instance, the estimates are close to zero toward the end of our sample, at around minus 0.5 percentage points, which indicates that the policy tightening was almost finished by the end of 2018.

Panel (b) shows the one-year-ahead volatility of the cyclical rate $var_t(\tilde{i}_{t+4})$, the volatility of its endpoint $var_t(i_{t+4}^*)$ and the covariance term, which together sum to the conditional volatility of the observed short rate. When this volatility is higher, this corresponds to periods when the expectations component is likely to play a large role. The results show three large spikes in the volatility of the cyclical rate around 1983-1984, 1998 and 2007 as well as more modest increases around 1990 and 1994. The spikes are consistent with the findings of [Cieslak and Povala \(2016\)](#) that the short-rate expectations become more volatile before recessions and around asset-market volatility bouts. Outside of these spikes the cyclical volatility is much lower and close to the secular volatility.

Panels (c)-(d) of Figure 5 compare the contribution of each macroeconomic shock to \tilde{i} in models with constant and time-varying coefficients, respectively. The baseline model draws a consistent picture of the impact nominal shocks have on inflation and the cyclical component of the short rate. We find that the contribution of nominal shocks to the short rate are large early in the sample period, which corresponds to the period when nominal shocks also contribute to secular changes to inflation (Figure 3) and when the impulse response functions show a strong policy response to nominal shocks (see Figure 6 below). By contrast, the nominal shocks' contributions to \tilde{i}_t almost disappear in the second half of the sample, when inflation becomes anchored and shocks to the cyclical short rate are smaller. This appears to be consistent with the conclusion of [Ramey \(2016\)](#), in a review of the empirical evidence, that monetary policy has become more systematic over time.¹⁶

Impulse Response Function The disappearing impact of nominal shocks on π_t^* and \tilde{i}_t coincides with changes in the impulse response functions of every observed macro vari-

¹⁶Since it is plausible that monetary policy shocks are captured by nominal shocks.

able. We find that the responses of the short rate, inflation and output to nominal shocks decrease between the beginning and end of the sample period. To see the changes, Figure 6 compares the impulse response functions of the short rate, inflation and output at two dates in time. One date is the first quarter of 1984, which is the period when Figure 3 shows an important disinflationary role for nominal shocks. The other date is the first quarter of 2007, which is the period when Figure 3 shows that the impact on secular inflation is tiny. We choose 2007Q1 because this precedes the great financial crisis and avoids the extended period that follows it, when the 3-month rate hovers just above zero. This way, the changes that we report cannot be associated with the potential changes that are due to the lower bound being below nominal yields. Focusing on two dates makes it easy to see the contrasts. Choosing different dates in these two periods does not change the message.

The results in Figure 6 show that, early in the sample, a nominal shock leads the short rate to increase by 0.45 percentage points on impact and peak at 0.6 points after two years. This is associated with an increase in the inflation rate of around 0.2 percentage points, which is essentially permanent, and an increase in output growth of around 0.5, which is short-lived. One interpretation of these results is that, when inflation is unanchored, nominal shocks lead to the economy over-heating and a persistent increase in inflation, along with a large and persistent policy rate response. However, the responses are very different later in our sample. The impact on the short rate peaks at only 0.2 percentage points and is less persistent, the impact on inflation is large but quickly disappears, and the impact on output growth is small and also quickly disappears. One interpretation is that, with inflation anchored, nominal shocks may cause substantial inflation, but this increase is transitory, with little economic overheating and only a small short rate response. This pattern is consistent with the theoretical argument in [Clarida, Gali, and Gertler \(2000\)](#) that the post-1979 policy responses to inflation stabilize output and inflation. It is also consistent with the empirical results in [Cogley and Sargent \(2005\)](#) that policy activism

by the Federal Reserve after the chairmanship of Paul Volcker contribute to the fall in inflation as well as to changes in its persistence.

The decompositions in terms of the secular and cyclical components offered in Panel (b) differ from the results of [Jordà and Taylor \(2019\)](#), who suggest that the cyclical component of the real rate explains, at most, 40 percent of the variability and, therefore, that the secular component explains the majority of the variations in the real rate. This difference is not due to the distinction between the real and nominal short rate. Our results indicate that the shocks to \tilde{i}_t largely translate to \tilde{r}_t on impact. Instead, one reason for the differences between our results and theirs is that we use a common set of shocks that drive the unobserved cyclical and secular components. This choice puts a cap on the degrees of freedom. Another reason is that we also use survey data to help pin down secular variations. Finally, we restrict the role played by secular components in long-term yields because we require that bond risk premiums be stationary. These differences limit the variability that the model can attribute to the unobserved secular components.

4.4 Bond Yields

Given the substantial changes in how secular and cyclical components of the short rate, i_t , respond to nominal shocks during our sample period, there should be no surprise that we find that the responses of bonds to nominal shocks also change.

Bond Yield Impulse Response Function Figure 7 shows the impulse response functions of bond yields with 3-month, 2-year and 10-year maturities with respect to nominal shocks. We show the responses at these three maturities on two dates, as in the previous section. The first row exhibits the responses of the bond yields, the expectations components and the term premium for the first quarter of 1984. The second row exhibits the same responses but for the first quarter of 2007.

Early in the sample, the initial impact on the short rate is around 0.45 percentage points and peaks close to 0.6 (which are the same as in Panel (a) of Figure 6). The initial

impact on the 2-year yield is around 0.7 percentage points and peaks close to 0.8, while the initial impact on the 10-year yield is around 0.6 percentage points and peaks close to 0.7. Therefore, bond yields strongly respond to nominal shocks in the first half of the sample. Nominal shocks lead to (i) a higher level of bond yields, (ii) a steeper slope between the short rate and longer-term yields and (iii) an inversion of the yield curve between 2- and 10-year maturities.

Panel (b) shows that the expectations component dominates the response of bond yields to nominal shocks. This is consistent with the interpretation that, when inflation is unanchored, nominal shocks lead to strong policy responses. The initial impact on the expectations component is around 0.6 and 0.4 percentage points for the 2- and 10-year bonds, respectively, which explains the yield curve inversion. This impact gradually declines afterward but some of it is permanent because nominal shocks lift π_t^* (compared with Figure 7b). The impact on the term premium component is smaller, around 0.1 and 0.2 percentage points for the 2- and 10-year bonds, respectively, and peaks close to 0.25 and 0.4.

This inversion between the 2- and 10-year bond yields arises because the response of the short rate increases with the horizon until after the responses of the growth and inflation rates decrease. Note that the inversion predicts slower growth in the baseline model. The observation that the expectations component drives the predictive content of a yield curve inversion for future growth is consistent with the results of [Ang, Piazzesi, and Wei \(2006\)](#).

The second row in Figure 7 shows the response in 2007Q1, which is typical for the second half of our sample. Stated simply, here, the responses of bond yields to nominal shocks are muted. The impact on the 2-year yield is close to 0.4 percentage points, which is half the peak impact in 1984Q1, and it is more quickly reversed. We find that the responses of the expectations components drive the more muted responses of the bond yields. The impacts on the term premiums are still substantial but dissipate faster than

during the earlier period.

Term Premium Panel (a) of Figure 8 shows the term premium for the 10-year zero-coupon yield estimated in our model compared with the estimates of [Adrian, Crump, and Moench \(2013\)](#) (ACM hereafter), available from the Federal Reserve Bank of New York, as well as the estimates of [Crump, Eusepi, and Moench \(2016\)](#) (CEM hereafter). The ACM estimate is widely cited by analysts and practitioners. While ACM do not explicitly account for secular variations in i_t^* , the model specifications and estimation procedures deliver term premium estimates with smaller trends than the estimates from a typical maximum likelihood estimation. The CEM estimate relies on a rich panel of survey data to correct for the influence of secular variations in i_t^* and does not rely on the no-arbitrage restriction.

Figure 8 shows that the three estimates share broad similarities, which is of course reassuring. There are two differences worth underlining. First, our baseline term premium estimate appears stationary. By contrast, both the ACM and CEM estimates show a downward trend from 4 percent at the beginning of our sample period to less than 0 percent at the end of our sample period. Hence, these estimates attribute a greater share of the high interest rates early in our sample to a higher term premium. Second, our baseline estimate exhibits stronger cyclical patterns throughout the sample. For instance, the baseline results exhibit a sharp drop in the term premium starting in the third quarter of 2008 at the peak of the crisis, while the drop is more modest based on other models. While there are several differences in the purposes and assumptions underlying these three estimates, the key difference between these two observations is the combination of (i) allowing for secular changes in bond yields and (ii) the restriction embedded in the baseline estimates that the term premium be stationary.

Perhaps unsurprisingly, the term premium shock is the most important driver of the term premium. Panel (b) reports the impulse response function of the 2- and 10-year term premia relative to the shock, $\varepsilon_{f,t+1}$. This is substantially larger than the impact reported

in Figure 7 for nominal shocks. However, the effect dies out relatively quickly. For the 10-year bond, a one-standard-deviation term premium shock raises the yield by close to 0.6 percentage points over one quarter, which dissipates to around 0.4 and 0.2 after one and two years, respectively.

To illustrate the magnitude of this impact relative to all other macroeconomic shocks, Panel (c) of Figure 8 shows the share of the one-year-ahead variance of the term premium that is explained by each shock. This share changes over time because of the time-varying loadings. Overall, the term premium shock almost always explains two thirds or more of the term premium variations at this horizon. This is consistent with the common observation in the literature that macroeconomic variables explain a small share of the term premium variability. Nonetheless, and perhaps unsurprisingly, the macro shocks are more important than the term premium shock when monetary policy changes its target rate more actively. The share attributed to macroeconomic shocks reaches beyond 50 percent in a handful of cases, for instance around the 1990 recession, during the recovery after the 2001 recession and in 2008.

Panel (d) shows the variance decomposition for longer horizons, up to 40 quarters and fixes the date to 2007Q1. The share of the term premium variance attributed to macroeconomic shocks increases to 50 percent at horizons between 2 and 3 years and rises beyond that. At these longer horizons, the real shocks together explain the majority of the term premium variations while, as expected, the nominal shocks explain only a small share. This pattern is common across the sample. The real shocks typically dominate term premium variations at lower frequencies. Given that the term premium only depends on cyclical variables \tilde{i}_t , $\tilde{\pi}_t$ and \tilde{y}_t , this immediately implies that, based on our identification strategy (see the discussion in Section 2.3), real shocks have more-persistent impacts on business cycle variations than nominal shocks do for most of our sample period.

4.5 VAR(1) Dynamics

One important question is whether the cross-equation restrictions in Section 2.2 play an important role in the results. Figure 9 reports the key results that highlight the differences. Panel (a) shows estimates of the potential output from our baseline model, from the model with constant loadings and from the model with constant loadings and VAR(1) dynamics. The VAR(1) produces a potential output that is much smoother than in the other models. This is not innocuous since it affects the output gap estimates and, indirectly, the shocks recovered from the model.

To see this, Panel (b) shows the output gap from the same models as well as the estimates from the CBO.¹⁷ The estimates from the baseline model and the model with the constant loadings closely agree with each other. By the end of the sample period, the estimated output gap is around -1 percent and the CBO estimate is close to zero. We also find that the CBO output gap is not as deep following the 2001 and 2009 recessions and tends to recover more quickly. We should expect some differences between these two estimates since they use different methods: the CBO estimate relies on the “production-function” method instead while the approach here relies on both cross-equation restrictions and the statistical properties of the secular versus the cyclical components (following [Blanchard and Quah 1989](#)). [Coibion, Gorodnichenko, and Ulate \(2018\)](#) provide a detailed discussion of different approaches to estimating potential output and their relative advantages. Like us, they find that the CBO estimates imply a quicker decline in the output gap than the results obtained when following [Blanchard and Quah \(1989\)](#). However, our results and those of the CBO agree on the timing of the peaks and troughs. All of these estimates indicate a largely U-shaped recovery toward potential output. They also indicate that the slowdown in potential growth started before 2008. These results are consistent with the

¹⁷We use the log difference between real GDP from the Bureau of Economic Analysis and the potential output from the CBO, both available from the the Federal Reserve Bank of St. Louis FRED database. This suggests the troughs of the output gap were around -2, -4, -3 and - 6 percent around 1987, 1993, 2002 and 2009, respectively.

evidence on the Great Recession presented by [Eo and Morley \(2020\)](#), which are based on a regime-switching model for US output.

The output gap estimates that are based on the VAR(1) are very different, which suggests that the cross-equation restrictions play an important role in identifying the output gap. In fact, the estimates are implausible. The estimated output gap is between 1 and 5 percent during the period between 1997 and 2007, which includes the 2001 recession. Moreover, the estimated output gap that is based on the VAR(1) reaches -10 percent after 2009 and remains close to this level toward the end of our sample, in 2018. Hence, the results show that the restrictions from [Laubach and Williams \(2003\)](#) that we implement in our baseline model are useful for pinning down the output gap and the potential output, to a large extent, because our Equation 8 connects the output gap to the monetary stance, \tilde{r}_t . Note that these are relevant for applications in which the system includes unobserved cyclical and secular components, as in our case. The results do not bear on the widespread use of the reduced-form VAR(1) dynamics to capture the relationship between the observed macro variables when there are no unobserved components.

5 Conclusion

We provide a framework where secular and cyclical changes to the short rate, inflation and output are jointly determined by a small set of shocks with loadings that change over time. This framework incorporates pricing equations for bond yields in terms of macro variables in addition to a latent term structure factor. We use restrictions on the correlations between innovations to the secular and cyclical changes to identify and analyze the impact of nominal shocks. We find that the role and impact of nominal shocks on macro variables and bond yields dramatically change during our sample period. Our approach paves the road to identifying the distinct impact of other types of structural shocks on the cyclical and secular changes in bond yields. In particular, our baseline model ignores the impact of regulatory or fiscal changes and does not distinguish between the roles of technology,

demography and capital accumulation behind the evolution of potential output.

Figure 1: US Macro, Bond and Survey Data

Panel (a) shows the quarterly inflation rate, π_t , as the compounded rate of change of the personal-consumption-expenditures index excluding food and energy, in percentage terms, annualized and seasonally adjusted, and the output growth rate, g_t , as the log-change of the seasonally adjusted real gross domestic product. Panel (b) shows the nominal three-month short rate, i_t , and the 2-, 5- and 10-year nominal zero-coupon rates (for clarity, other maturities are not reported). Panel (c) shows long-horizon survey forecasts of i_t , π_t and g_t obtained from [Crump, Eusepi, and Moench \(2016\)](#).

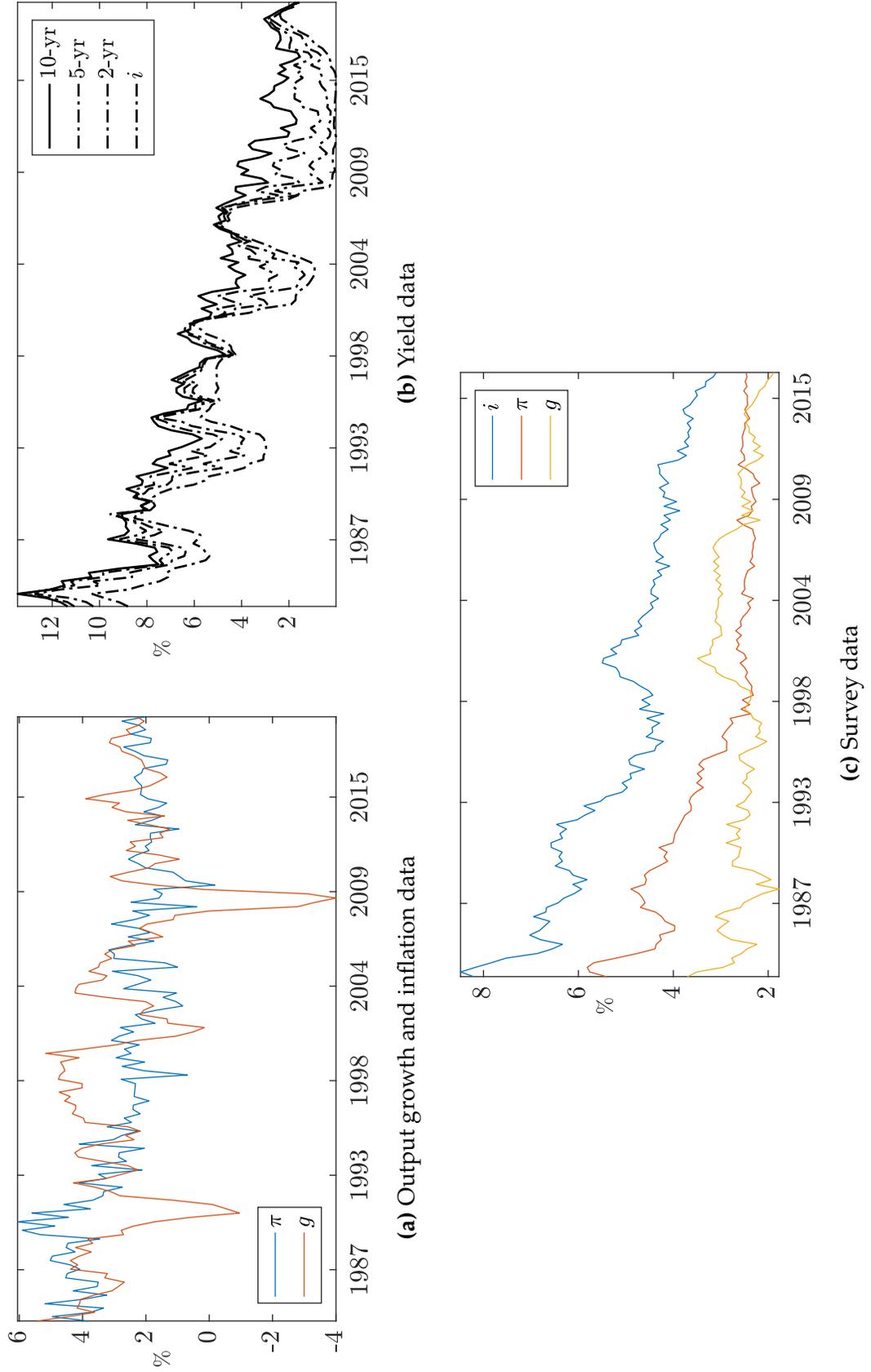


Figure 2: Estimated Real and Nominal Shocks

Nominal and real shocks identified in a model with time-varying loadings (Panel a) and constant loadings (Panel b). Panels (c)-(e) show scatter plots of the nominal shock and both real shocks with time-varying and constant loadings (x-axis and y-axis, respectively). Quarterly data, 1983–2019. Units are multiples of standard deviations in every panel.

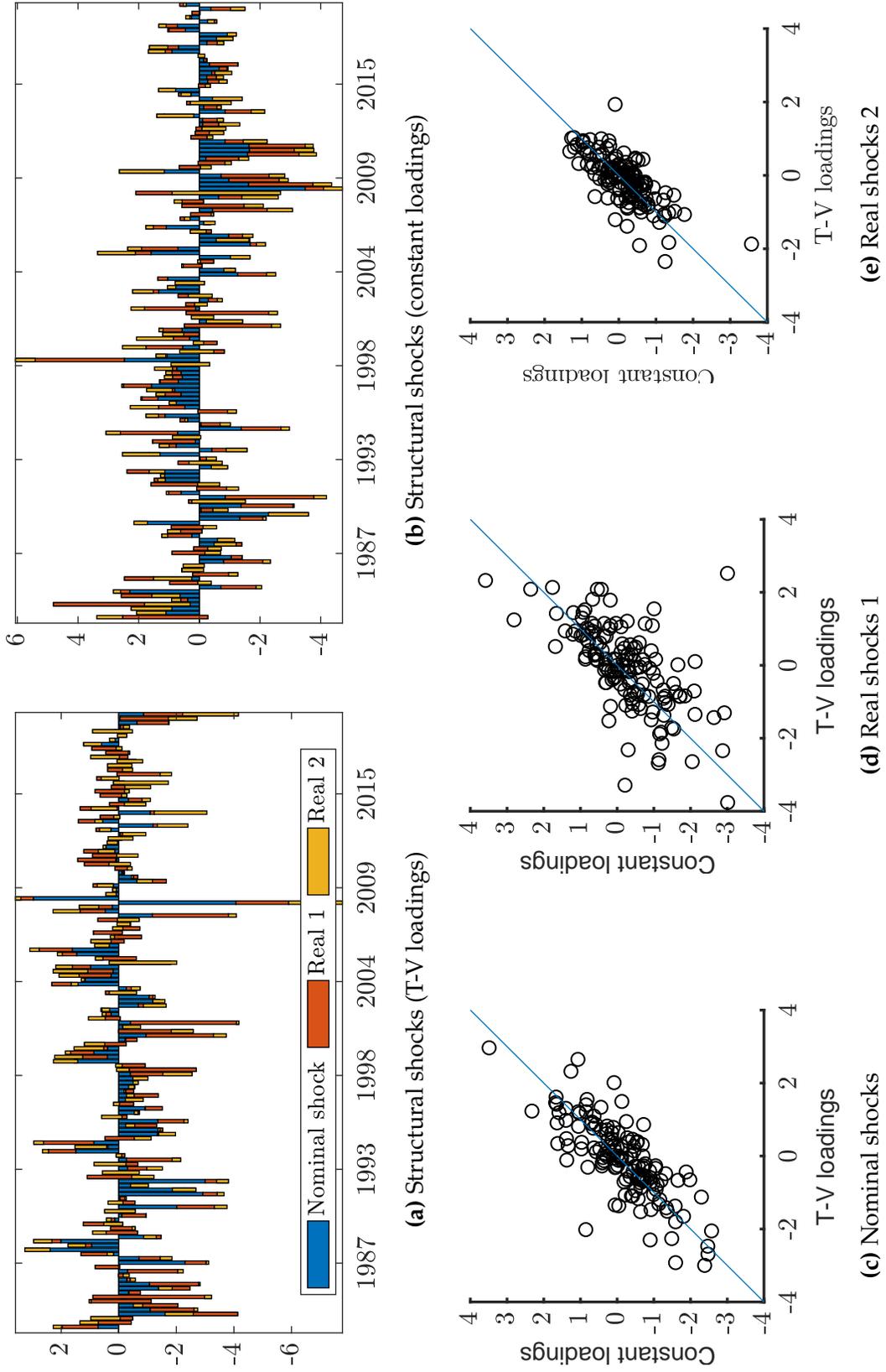


Figure 3: Inflation Endpoint π_t^*

Panel (a) shows two series of inflation endpoint estimates, π_t^* , from models with constant or time-varying loadings (the baseline model), respectively, in percent. Panel (b) shows two series of estimated one-year-ahead conditional volatility of π_t^* from the same models, in percent. Panels (c)-(d) show the contribution of each shock to the period-by-period update of π_t^* in models with time-varying or constant loadings, respectively, where the units are standard deviations.

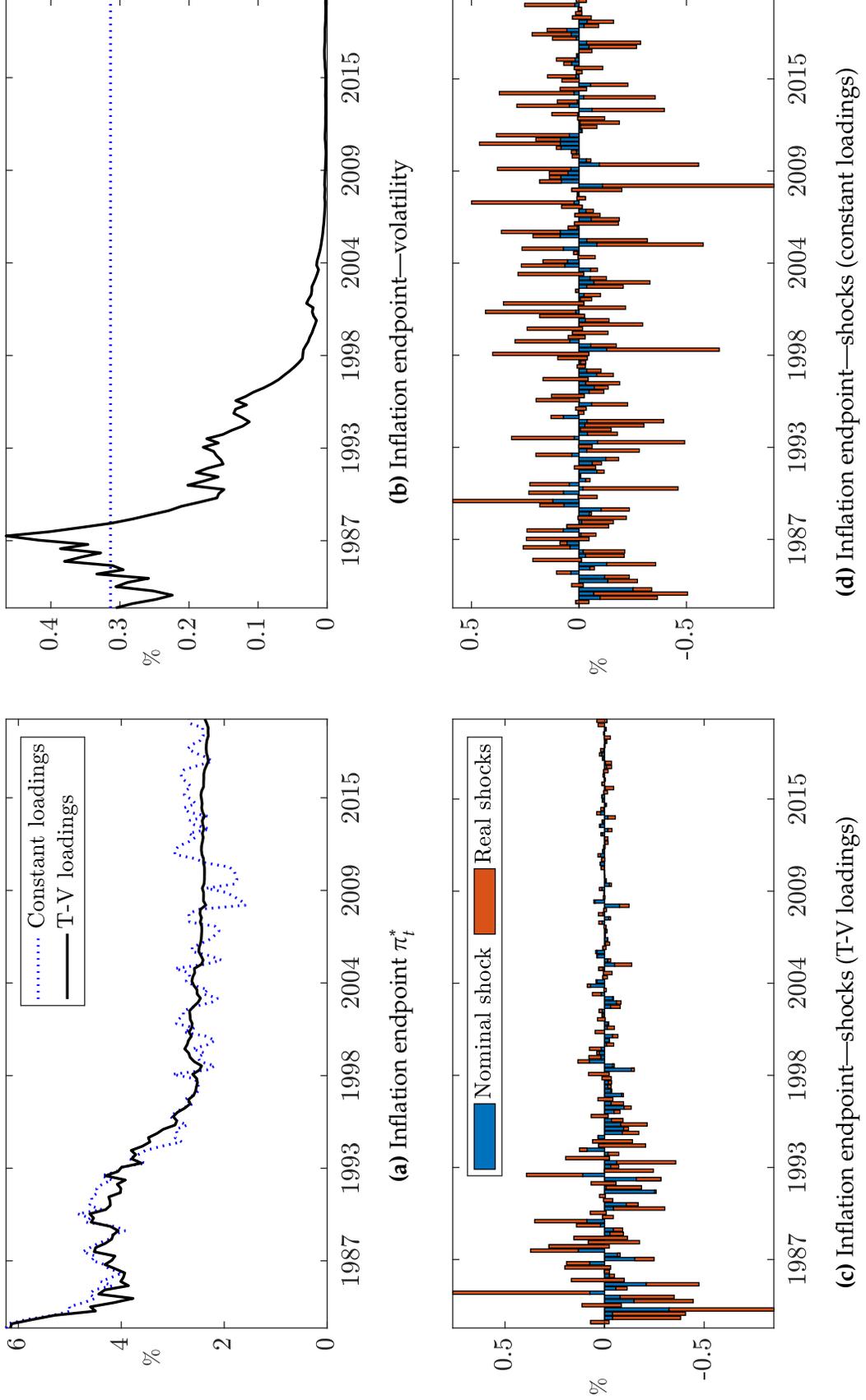
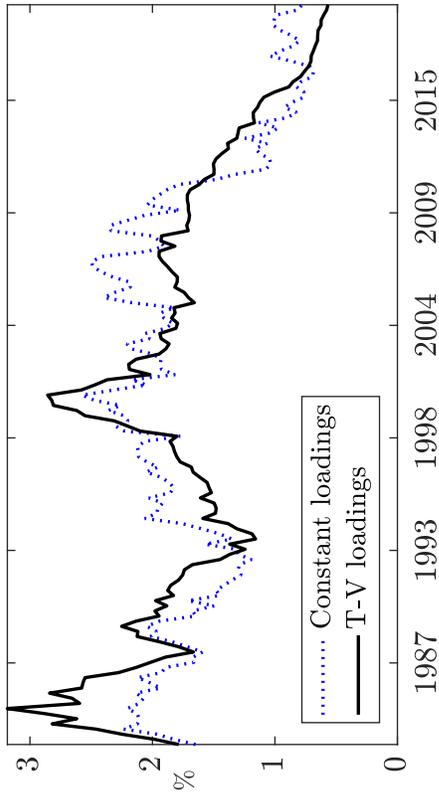
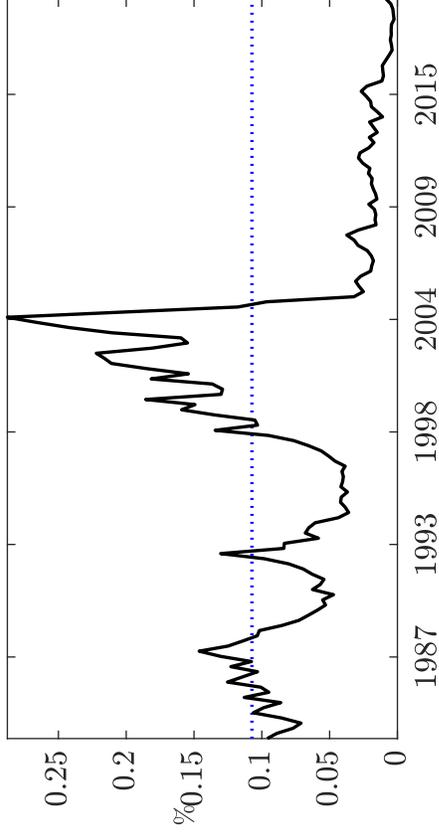


Figure 4: Real-rate Endpoint r_t^*

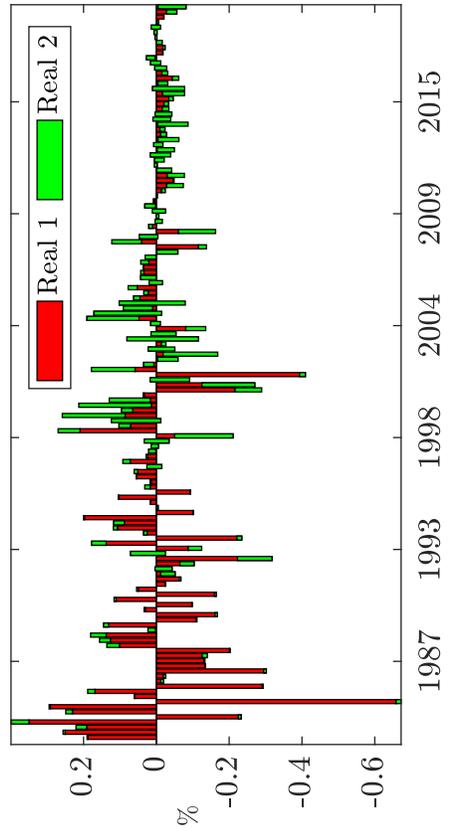
Panel (a) shows two series of real-rate endpoint estimates, r_t^* , from models with constant or time-varying loadings (the baseline model), respectively. Panel (b) shows the estimated one-year-ahead conditional volatility of r_t^* from the same two models. Panels (c)-(d) show the contribution of real shocks to the period-by-period update in these two models (by construction, the nominal shocks have no impact). All units are annualized in percentages.



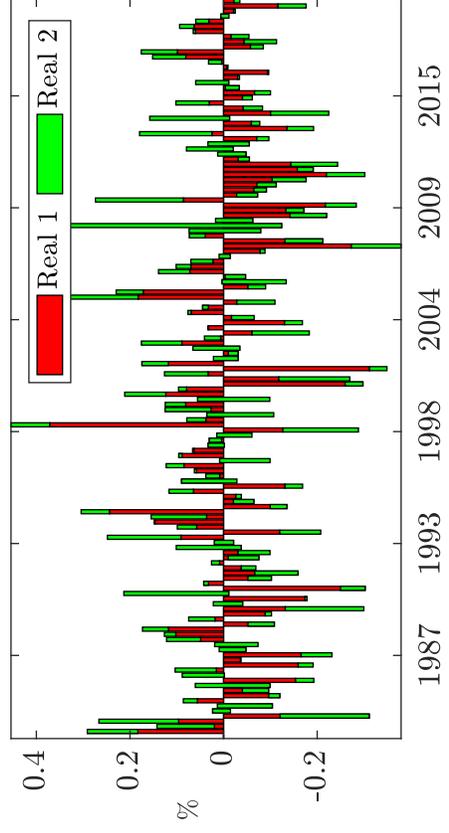
(a) Real-rate endpoint r_t^*



(b) Real-rate endpoint—volatility



(c) Real-rate endpoint—shocks (T-V loadings)



(d) Real-rate endpoint—shocks (constant loadings)

Figure 5: Cyclical Short Rate \tilde{i}_t

Panel (a) compares the cyclical short rate, \tilde{i}_t , from models with constant or time-varying loadings (the baseline model), respectively. Panel (b) compares the one-year-ahead volatility for both \tilde{i}_t and i_t^* as well as their covariance from these two models. The dotted lines report the results from the model with constant loadings, matching the volatility components by color. Panels (c)-(d) show the period-by-period contribution of the nominal and real shocks to the cyclical short rate from these two models, respectively. All units are annualized percentages.

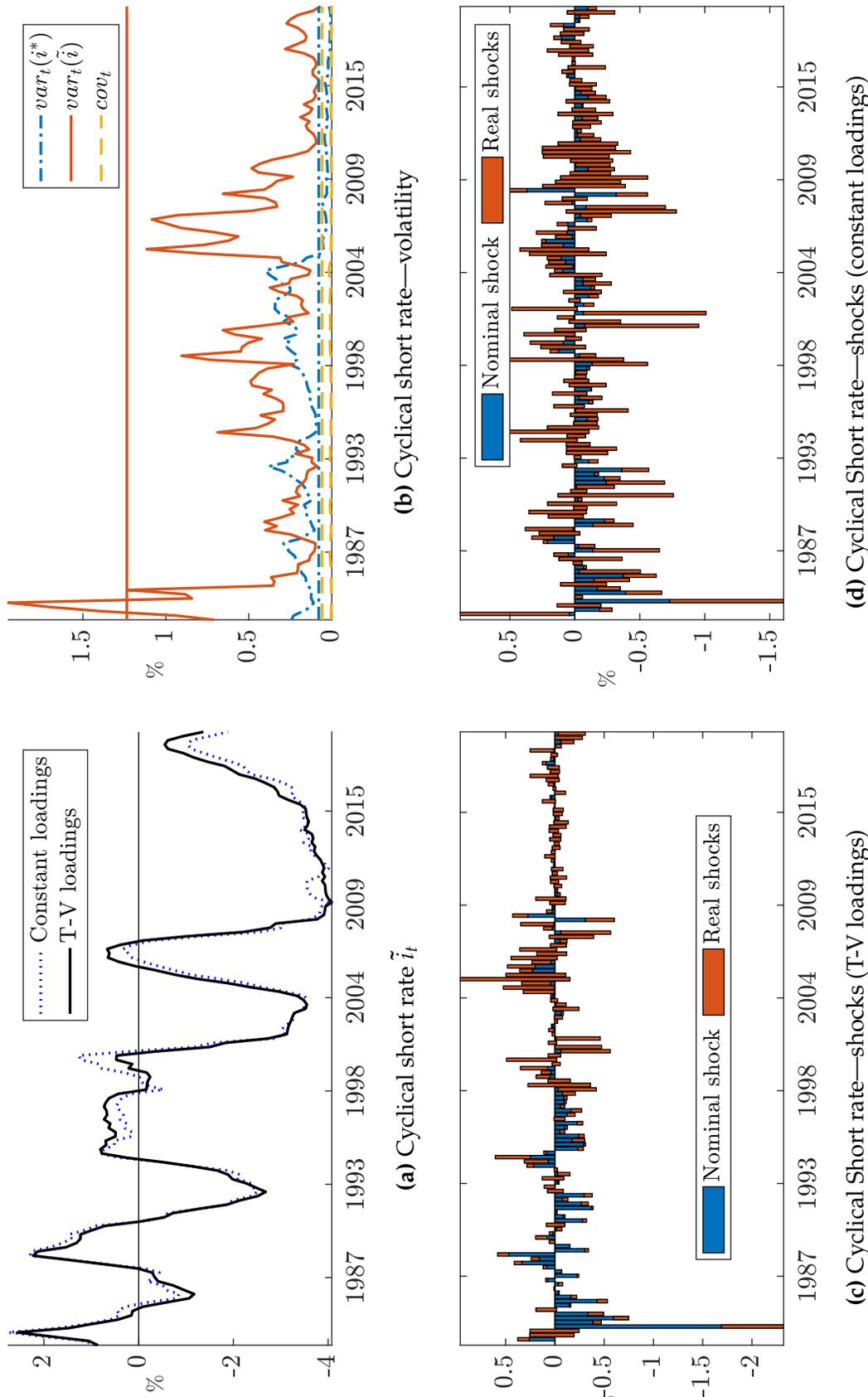


Figure 6: Nominal Shocks Impulse Response Functions – Macro Variables

Impulse response functions with respect to the nominal shocks at horizons from 1 to 40 quarters, from the model with time-varying loadings (the baseline model), on two different dates in our sample. Panel (a) shows the impulse responses function of the short rate i_t . Panel (b) shows the responses of the inflation rate π_t . Panel (c) shows the responses of the output growth rate g_t . Panel (d) shows the responses of the output gap.

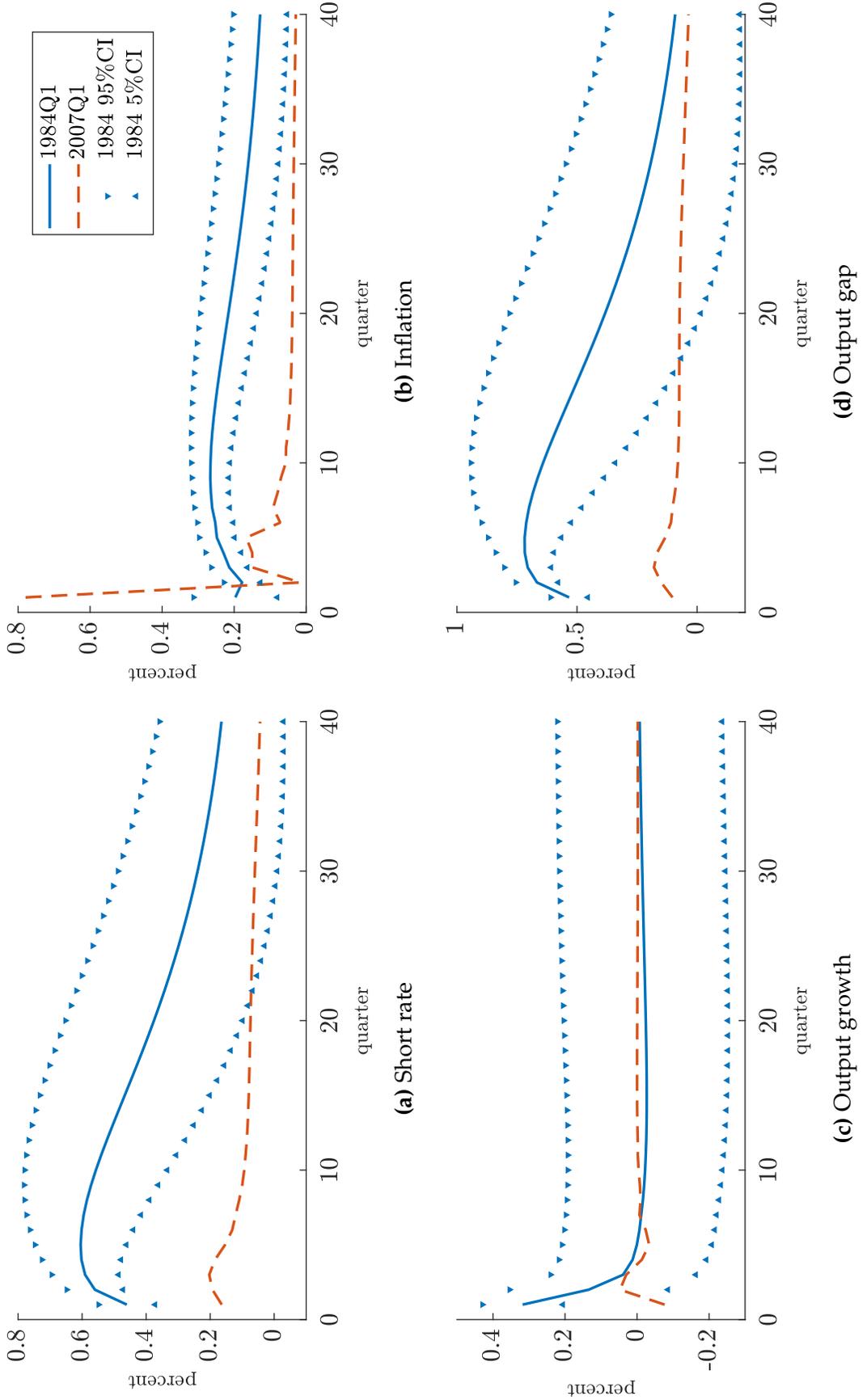


Figure 7: Nominal Shocks Impulse Response Functions – Bond Yields

Impulse response functions with respect to nominal shocks at horizons from 1 to 40 quarters at two different dates in our sample. Panels (a)-(c) show the responses of the yields, expectations and term premium components for bonds with 3 months (the short rate), 2 years and 10 years to maturity in 1984Q1. Panels (d)-(f) show the same responses but for 2007Q1. All units are in percentage points.

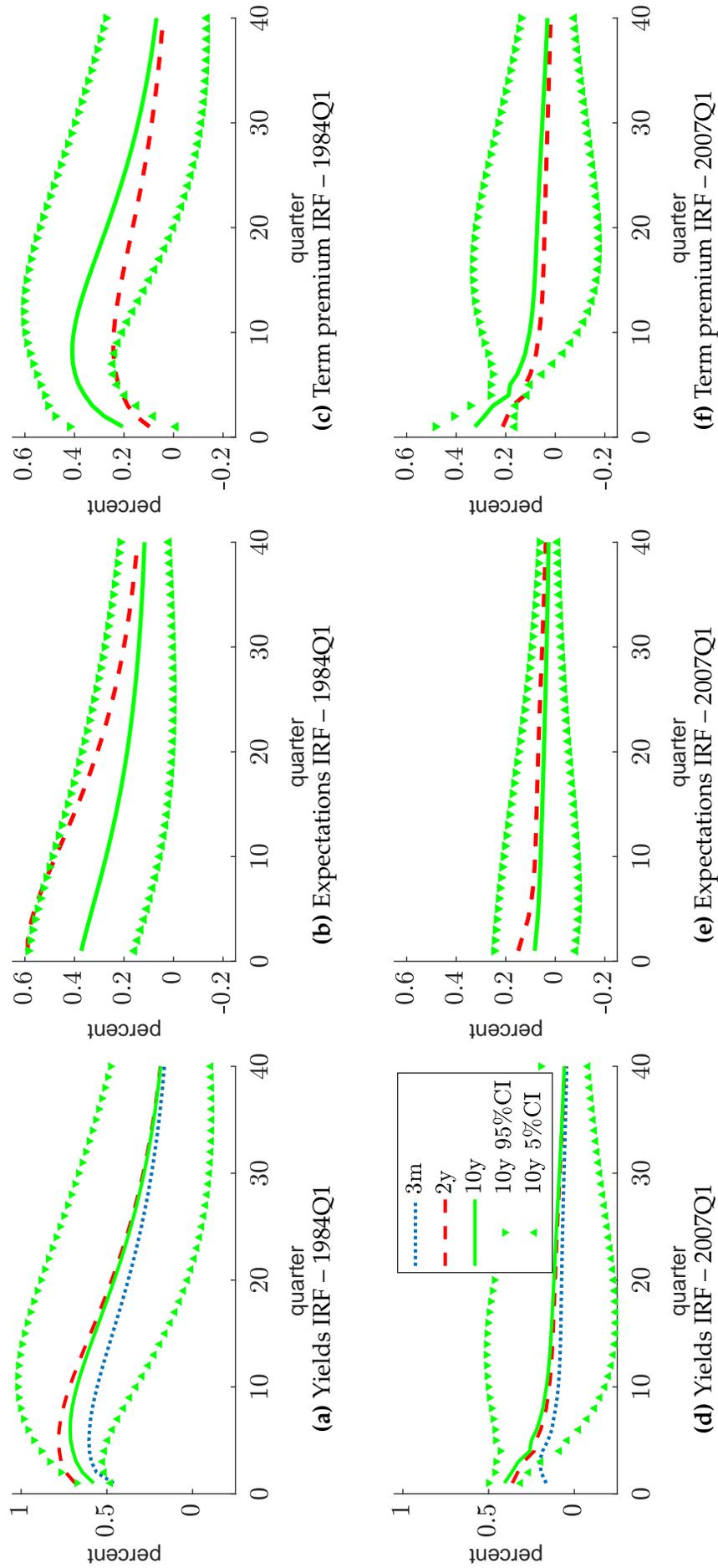
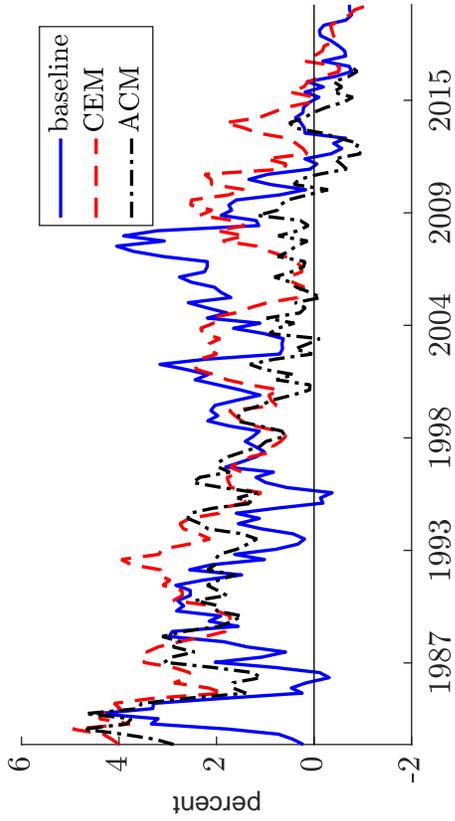
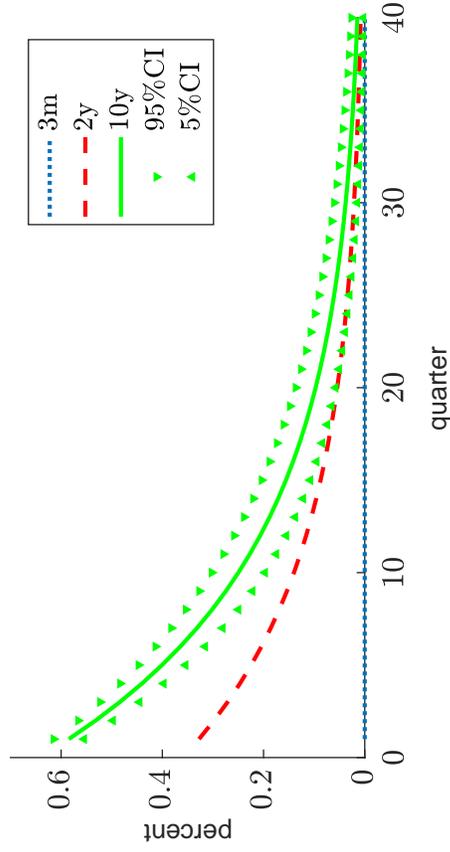


Figure 8: Term Premium

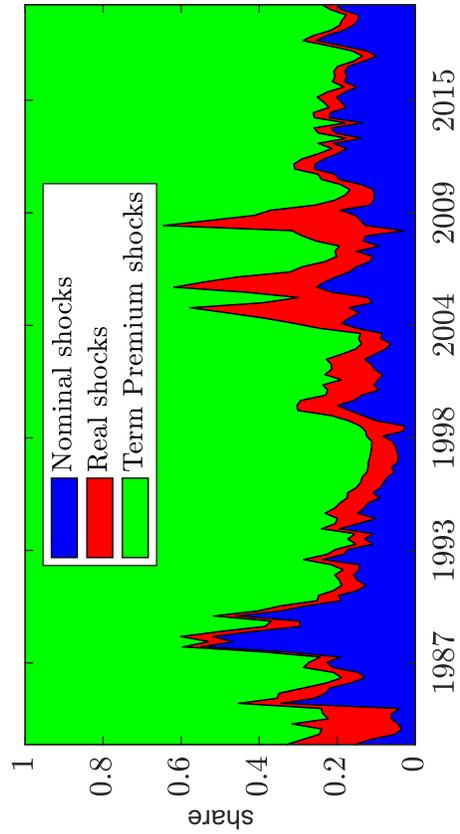
Panel (a) shows term premium estimates in percentages from the baseline model, from [Adrian, Crump, and Moench \(2013\)](#) and from [Crump, Eusepi, and Moench \(2016\)](#), labeled ACM and CEM, respectively. Panel (b) shows the term premium impulse response function with respect to term premium shocks $\varepsilon_{f,t+1}$ from our baseline specification. Panel (c) shows the term premium variance decomposition at the 4-quarter horizon, over time. Panel (d) shows the variance decomposition in the first quarter of 2007 at horizons between 1 and 40 quarters.



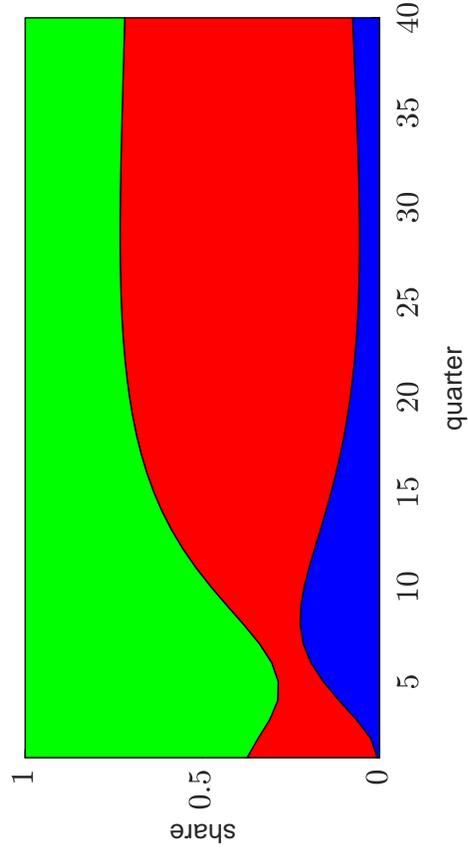
(a) 10-year term premium



(b) Term premium IRF



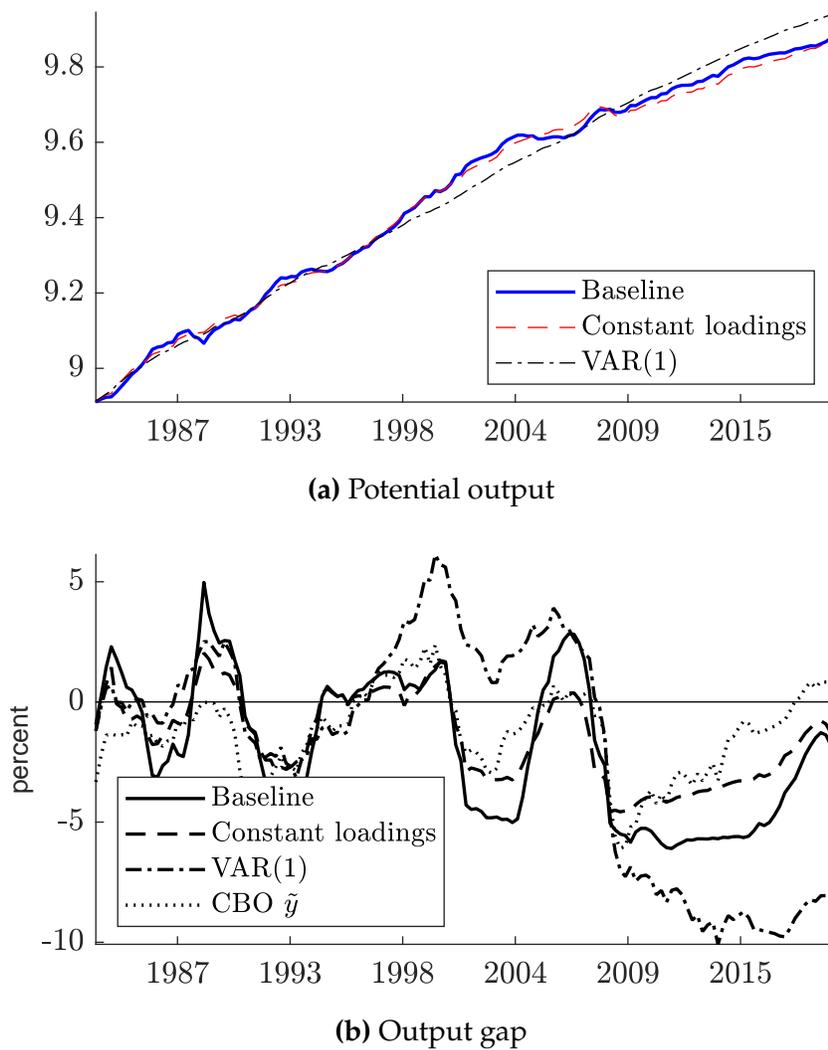
(c) Variance decomposition over time (one-year horizon)



(d) Variance decomposition across horizons (2007Q1)

Figure 9: Estimates from a Model with VAR(1) Dynamics

Results in a model with VAR(1) cyclical dynamics and constant loadings. Panel (a) compares the potential output y_t^* estimates (units are in log-dollars) from the baseline model, the model with constant loadings and the VAR(1) model with constant loadings. Panel (b) compares the output gap, \tilde{y} , estimates from the same model as well as the CBO estimates (units are in percentage points).



References

- Adrian, T., R. K. Crump, and E. Moench (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics* 110(1), 110–138.
- Al-Marhubi, F. (1998). Cross-country evidence on the link between inflation volatility and growth. *Applied Economics* 30(10), 1317–1326.
- Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50(4), 745–787.
- Ang, A., M. Piazzesi, and M. Wei (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics* 131, 359–403.
- Barro, R. J. (2013). Inflation and economic growth. *Annals of Economics & Finance* 14(1).
- Bauer, M. and G. Rudebusch (2020). Interest rates under falling stars. *American Economic Review* 110(5), 1316–1354.
- Blanchard, O. (2018). Should we reject the natural rate hypothesis? *Journal of Economic Perspectives* 32(1), 97–120.
- Blanchard, O. J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review* 79(4), 655–673.
- Brunnermeier, M. K. and Y. Sannikov (2016). The i theory of money. Technical Report 22533, National Bureau of Economic Research.
- Chen, R.-R. and L. Scott (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *The Journal of Fixed Income* 3, 14–31.
- Christensen, J. H. E. and G. D. Rudebusch (2019). A new normal for interest rates? Evidence from inflation-indexed debt. *The Review of Economics and Statistics* 101(5), 933–949.
- Cieslak, A. and P. Povala (2015). Expected returns in treasury bonds. *The Review of Financial Studies* 28(10), 2859–2901.
- Cieslak, A. and P. Povala (2016). Information in the term structure of yield curve volatility. *The Journal of Finance* 71(3), 1393–1436.
- Clarida, R., J. Galí, and M. Gertler (1998). Monetary policy rules in practice: Some international evidence. *European Economic Review* 42(6), 1033 – 1067.
- Clarida, R., J. Gali, and M. Gertler (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly Journal of Economics* 115(1), 147–180.
- Clark, T. E. and S. Kozicki (2005). Estimating equilibrium real interest rates in real time. *The North American Journal of Economics and Finance* 16(3), 395 – 413.

- Cochrane, J. H. (1988). How big is the random walk in GNP? *The Journal of Political Economy* 96(5), 893–920.
- Cogley, T. and T. J. Sargent (2005). Drifts and volatilities: monetary policies and outcomes in the post WW II US. *Review of Economic Dynamics* 8(2), 262–302.
- Coibion, O., Y. Gorodnichenko, and M. Ulate (2018). The cyclical sensitivity in estimates of potential output. *Brookings Papers on Economic Activity*, 343.
- Collin-Dufresne, P. and R. S. Goldstein (2002). Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. *The Journal of Finance* 57(4), 1685–1730.
- Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones (2009). Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics* 94(1), 47–66.
- Creal, D. D. and J. C. Wu (2017). Monetary policy uncertainty and economic fluctuations. *International Economic Review* 58(4), 1317–1354.
- Crump, R. K., S. Eusepi, and E. Moench (2016). The term structure of expectations and bond yields. Working Paper 775, Federal Reserve Bank of New York.
- Duffee, G. (2002). Term premia and interest rate forecasts in affine model. *The Journal of Finance* 57, 405–443.
- Dungey, M., J. P. Jacobs, J. Tian, and S. Van Norden (2015). Trend in cycle or cycle in trend? New structural identifications for unobserved-components models of US real GDP. *Macroeconomic Dynamics* 19(4), 776–790.
- Engsted, T. and C. Tanggaard (1994). Cointegration and the US term structure. *Journal of Banking and Finance* 18(1), 167–181.
- Eo, Y. and J. Morley (2020). Why has the US economy stagnated since the Great Recession? *The Review of Economics and Statistics*. forthcoming.
- Friedman, M. and A. J. Schwartz (1963). *A Monetary History of the US 1867-1960*. Princeton University Press.
- Gali, J. (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.
- Garcia, R. and P. Perron (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics* 78(1), 111–125.
- Gurkaynak, R., B. Sack, and J. Wright (2006). The U.S. Treasury curve: 1961 to present. Technical Report 2006-28, Federal Reserve Board.
- Hall, A. D., H. M. Anderson, and C. W. J. Granger (1992). A cointegration analysis of Treasury bill yields. *The Review of Economics and Statistics* 74(1), 116–126.

- Hansen, A. L. (2019). Yield curve volatility and macroeconomic risk. Working paper, University of Copenhagen.
- Haubrich, J. G. (2006). Does the yield curve signal recession? *Federal Reserve Bank of Cleveland* 15.
- Hentschel, L. (1995). All in the family nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* 39(1), 71 – 104.
- Holston, K., T. Laubach, and J. C. Williams (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics* 108, S59 – S75. 39th Annual NBER International Seminar on Macroeconomics.
- Jordà, Ò., S. R. Singh, and A. M. Taylor (2020). The long-run effects of monetary policy. Technical Report 26666, National Bureau of Economic Research.
- Jordà, Ò. and A. M. Taylor (2019). Riders on the storm. Technical Report 26262, National Bureau of Economic Research.
- Joslin, S., A. Le, and K. J. Singleton (2013a). Gaussian macro-finance term structure models with lags. *Journal of Financial Econometrics* 11(4), 581–609.
- Joslin, S., A. Le, and K. J. Singleton (2013b). Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs. *Journal of Financial Economics* 109(3), 604–622.
- Joyce, M. A., I. Kaminska, and P. Lildholdt (2012). Understanding the real rate conundrum: An application of no-arbitrage models to the UK real yield curve. *Review of Finance* 16(3), 837–866.
- Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.
- Kim, D. H. and A. Orphanides (2012). Term structure estimation with survey data on interest rate forecasts. *Journal of Financial and Quantitative Analysis* 47(1), 241–272.
- Kozicki, S. and P. Tinsley (2001). Shifting endpoints in the term structure of interest rates. *Journal of Monetary Economics* 47(3), 613–652.
- Kozicki, S. and P. Tinsley (2006). Survey-based estimates of the term structure of expected U.S. inflation. Working Paper 2006-46, Bank of Canada.
- Laubach, T. and J. C. Williams (2003). Measuring the natural rate of interest. *The Review of Economics and Statistics* 85(4), 1063–1070.
- Meyer, L. H. (2003). What I learned at the Fed. In *Macroeconomics, Monetary Policy, and Financial Stability. 2003 conference: A Festschrift in Honor of Charles Freedman*.
- Monfort, A. and F. Pegoraro (2012). Asset pricing with second-order Esscher transforms. *Journal of Banking & Finance* 36(6), 1678 – 1687.

- Morley, J. and J. Piger (2012). The asymmetric business cycle. *The Review of Economics and Statistics* 94(1), 208–221.
- Morley, J. C., C. R. Nelson, and E. Zivot (2003). Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different? *Review of Economics and Statistics* 85(2), 235–243.
- Negro, M. D., D. Giannone, M. P. Giannoni, and A. Tambalotti (2019). Global trends in interest rates. *Journal of International Economics* 118, 248–262.
- Nelson, C. R. and C. R. Plosser (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10(2), 139–162.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59(2), 347–370.
- Orphanides, A. and M. Wei (2012). Evolving macroeconomic perceptions and the term structure of interest rates. *Journal of Economic Dynamics and Control* 36(2), 239–254.
- Primiceri, G. E. (2005). Time-varying structural vector autoregressions and monetary policy. *The Review of Economic Studies* 72(3), 821–852.
- Rachel, L. and T. D. Smith (2017). Are low real interest rates here to stay? *International Journal of Central Banking* 13(3), 1–42.
- Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In *Handbook of Macroeconomics*, Volume 2, pp. 71–162. Elsevier.
- Rudebusch, G. (2006). Monetary policy inertia: Fact or fiction? *International Journal of Central Banking* 2, 85–135.
- Stock, J. (1994). Unit roots, structural breaks, and trends. In *Handbook of Macroeconomics*, pp. 2740–2843. Elsevier.
- Stock, J. H. and M. W. Watson (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *Journal of the American Statistical Association* 93(441), 349–358.
- Stock, J. H. and M. W. Watson (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking* 39, 3–33.
- Stock, J. H. and M. W. Watson (2016). Core inflation and trend inflation. *Review of Economics and Statistics* 98(4), 770–784.
- Van Dijk, D., S. J. Koopman, M. van der Wel, and J. H. Wright (2014). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics* 29(5), 693–712.
- Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *The American Economic Review* 101(4), 1514–1534.

A Online Appendix

Not intended for publication.

A.1 Cyclical Dynamics—VAR(5) Coefficients

Given the relationships in Equation 7, Equation 8 and Equation 9, the cyclical components follow VAR(5) dynamics with the following coefficients:

$$\Phi_1 = \begin{bmatrix} \left(\phi_z + \frac{a_r \delta_y}{2} \right) & \left(\delta_\pi b_1 - \frac{a_r b_1 \delta_y}{2} - \phi_z \delta_\pi \right) & \left(\delta_\pi b_y + \delta_y \left(a_1 - \frac{a_r b_y}{2} \right) - \phi_z \delta_y \right) \\ 0 & b_1 & b_y \\ \frac{a_r}{2} & -\frac{a_r b_1}{2} & \left(a_1 - \frac{a_r b_y}{2} \right) \end{bmatrix},$$

$$\Phi_2 = \begin{bmatrix} \frac{\delta_y a_r}{2} & \left(\frac{\delta_\pi b_2}{3} - \frac{\delta_y a_r}{2} \left(\frac{b_2}{3} + b_1 \right) \right) & \delta_y \left(a_2 - \frac{a_r b_y}{2} \right) \\ 0 & \frac{b_2}{3} & 0 \\ \frac{a_r}{2} & -\frac{a_r}{2} \left(\frac{b_2}{3} + b_1 \right) & \left(a_2 - \frac{a_r b_y}{2} \right) \end{bmatrix}, \quad \Phi_5 = \begin{bmatrix} 0 & -\frac{\delta_y a_r b_2}{2 \cdot 3} & 0 \\ 0 & 0 & 0 \\ 0 & -\frac{a_r b_2}{2 \cdot 3} & 0 \end{bmatrix},$$

$$\Phi_3 = \begin{bmatrix} 0 & \left(\frac{\delta_\pi b_2}{3} - \delta_y a_r \frac{b_2}{3} \right) & 0 \\ 0 & \frac{b_2}{3} & 0 \\ 0 & -a_r \frac{b_2}{3} & 0 \end{bmatrix}, \quad \Phi_4 = \begin{bmatrix} 0 & \left(\frac{\delta_\pi b_2}{3} - \delta_y a_r \frac{b_2}{3} \right) & 0 \\ 0 & \frac{b_2}{3} & 0 \\ 0 & -a_r \frac{b_2}{3} & 0 \end{bmatrix}.$$

A.2 Convergence of the Volatility of Secular Components

We derive conditions for the convergence of long-horizon volatility of secular components. Recall that the volatility of each secular component is given by

$$\ln(v_t^2) = \omega + \beta \ln(v_{t-1}^2) + g(z_t),$$

where we ignore the subscript since we obtain the same conditions for each secular component and where

$$z_t \equiv \frac{\alpha' u_t}{\sqrt{\alpha' \alpha}}, \quad g(z_t) = \sqrt{\alpha' \alpha} z_t + \kappa (|z_t| - E[|z_t|]).$$

We use the following result:

$$\ln(E[\exp(g(z)v)]) = -\kappa \sqrt{\frac{2}{\pi}} v + \ln \left(\frac{\exp\left(\frac{(\sqrt{\alpha' \alpha} + \kappa)^2 v^2}{2}\right) \Phi\left(\left(\kappa + \sqrt{\alpha' \alpha}\right) v\right) + \exp\left(\frac{(\sqrt{\alpha' \alpha} - \kappa)^2 v^2}{2}\right) \Phi\left(\left(\kappa - \sqrt{\alpha' \alpha}\right) v\right)}{2} \right).$$

Hence,

$$\begin{aligned} E_t[\exp(v \ln(v_{t+1}^2))] &= E_t[\exp(v(\omega + \beta \ln(v_t^2) + g(z_{t+1})))] \\ &= \exp(v\omega + v\beta \ln(v_t^2) + \ln(E[\exp(g(z)v)])) \\ &= \exp(v\omega + \ln(E[\exp(g(z)v)]) + v\beta \ln(v_t^2)), \end{aligned}$$

so we can write

$$E_t[\exp(v \ln(v_{t+\tau}^2))] = \exp(a_\tau(u) + b_\tau(u) \ln(v_t^2)),$$

and at longer horizons

$$\begin{aligned} E_t \left[\exp \left(\nu \ln \left(v_{t+\tau+1}^2 \right) \right) \right] &= E_t \left[\exp \left(a_\tau(u) + b_\tau(u) \ln \left(v_{t+1}^2 \right) \right) \right] \\ &= \exp \left(a_\tau(u) + b_\tau(u) \omega + \ln \left(E \left[\exp \left(g(z) b_\tau(u) \right) \right] \right) + b_\tau(u) \beta \ln \left(v_t^2 \right) \right), \end{aligned}$$

where

$$\begin{aligned} a_0(u) &= 0; \quad b_0(u) = u \\ a_{\tau+1}(u) &= a_\tau(u) + b_\tau(u) \omega + \ln \left(E \left[\exp \left(g(z) b_\tau(u) \right) \right] \right) \\ b_{\tau+1}(u) &= b_\tau(u) \beta, \end{aligned}$$

and hence

$$\begin{aligned} b_\tau(u) &= \beta^\tau u \\ a_\tau(u) &= \sum_{j=0}^{\tau-1} \left[b_j(u) \omega + \ln \left(E \left[\exp \left(g(z) b_j(u) \right) \right] \right) \right] \\ &= \sum_{j=0}^{\tau-1} \left[\beta^j u \omega + \ln \left(E \left[\exp \left(g(z) \beta^j u \right) \right] \right) \right] \\ &= u \omega \sum_{j=0}^{\tau-1} \beta^j + \sum_{j=0}^{\tau-1} \left\{ -\kappa \sqrt{\frac{2}{\pi}} \beta^j u + \ln \left(\frac{\exp \left(\frac{(\sqrt{\alpha'\alpha} + \kappa)^2}{2} (\beta^j u)^2 \right) \Phi \left((\kappa + \sqrt{\alpha'\alpha}) \beta^j u \right)}{+ \exp \left(\frac{(\sqrt{\alpha'\alpha} - \kappa)^2}{2} (\beta^j u)^2 \right) \Phi \left((\kappa - \sqrt{\alpha'\alpha}) \beta^j u \right)} \right) \right\} \\ &= \left\{ \left(\left(\omega - \kappa \sqrt{\frac{2}{\pi}} \right) u + \ln \left(\frac{\exp \left(\frac{(\sqrt{\alpha'\alpha} + \kappa)^2}{2} u^2 \right) \Phi \left((\kappa + \sqrt{\alpha'\alpha}) u \right)}{+ \exp \left(\frac{(\sqrt{\alpha'\alpha} - \kappa)^2}{2} u^2 \right) \Phi \left((\kappa - \sqrt{\alpha'\alpha}) u \right)} \right) \right\}^\tau \text{ if } \beta = 1, \end{aligned}$$

Focusing on the case $\beta = 1$, we have that

$$\begin{aligned} E_t \left[v_{t+j}^2 \right] &= E_t \left[\exp \left(\ln \left(v_{t+j}^2 \right) \right) \right] = \exp \left(a_j(1) + b_j(1) \ln \left(v_t^2 \right) \right) \\ &= v_t^2 \exp \left[\left(\omega - \kappa \sqrt{\frac{2}{\pi}} + \ln \left(\frac{\exp \left(\frac{(\sqrt{\alpha'\alpha} + \kappa)^2}{2} \right) \Phi \left(\kappa + \sqrt{\alpha'\alpha} \right)}{+ \exp \left(\frac{(\sqrt{\alpha'\alpha} - \kappa)^2}{2} \right) \Phi \left(\kappa - \sqrt{\alpha'\alpha} \right)} \right) \right] j, \end{aligned}$$

and therefore,

$$\frac{E_t \left[v_{t+j+1}^2 \right]}{E_t \left[v_{t+j}^2 \right]} = \exp \left[\omega - \kappa \sqrt{\frac{2}{\pi}} + \ln \left(\frac{\exp \left(\frac{(\sqrt{\alpha'\alpha} + \kappa)^2}{2} \right) \Phi \left(\kappa + \sqrt{\alpha'\alpha} \right)}{+ \exp \left(\frac{(\sqrt{\alpha'\alpha} - \kappa)^2}{2} \right) \Phi \left(\kappa - \sqrt{\alpha'\alpha} \right)} \right) \right],$$

and, taking the limit:

$$\lim_{j \rightarrow \infty} \frac{E_t \left[v_{t+j+1}^2 \right]}{E_t \left[v_{t+j}^2 \right]} = \left(\exp \left(\frac{(\sqrt{\alpha'\alpha} + \kappa)^2}{2} \right) \Phi \left(\kappa + \sqrt{\alpha'\alpha} \right) + \exp \left(\frac{(\sqrt{\alpha'\alpha} - \kappa)^2}{2} \right) \Phi \left(\kappa - \sqrt{\alpha'\alpha} \right) \right) \exp \left[\omega - \kappa \sqrt{\frac{2}{\pi}} \right].$$

Hence if $\beta = 1$ and the following condition holds:

$$\omega - \kappa\sqrt{\frac{2}{\pi}} + \ln\left(\exp\left(\frac{(\sqrt{\alpha'\alpha} + \kappa)^2}{2}\right)\Phi(\kappa + \sqrt{\alpha'\alpha}) + \exp\left(\frac{(\sqrt{\alpha'\alpha} - \kappa)^2}{2}\right)\Phi(\kappa - \sqrt{\alpha'\alpha})\right) < 0,$$

then $\sum_{j=1}^{\tau} E_t \left[v_{t+j-1}^2 \right]$ converges to

$$\frac{v_t^2}{1 - \left(\exp\left(\frac{(\sqrt{\alpha'\alpha} + \kappa)^2}{2}\right)\Phi(\kappa + \sqrt{\alpha'\alpha}) + \exp\left(\frac{(\sqrt{\alpha'\alpha} - \kappa)^2}{2}\right)\Phi(\kappa - \sqrt{\alpha'\alpha})\right) \exp\left[\omega - \kappa\sqrt{\frac{2}{\pi}}\right]}.$$

A.3 Bond Yields

We derive coefficients for the bond yield equation. We consider the general case where nominal bond prices depend on secular and cyclical components and we derive restrictions such that the term premium is stationary for bonds with arbitrary maturities. Define \bar{M}_t as follows:

$$\bar{M}_t' = [i^* \quad \pi^* \quad g^*]'$$

We assume the following risk-neutral dynamics for the macro components:

$$\begin{aligned} \bar{M}_{t+1} &= K_0^* + \Phi^* \bar{M}_t + \bar{\Sigma} \varepsilon_{t+1}^Q \\ \tilde{M}_{t+1} &= \tilde{K}_0 + \tilde{\Phi} \tilde{M}_t + \phi_{Mf} f_t + \tilde{\Sigma} \varepsilon_{t+1}^Q, \end{aligned} \quad (32)$$

where the scalar f_t follows stationary $AR(1)$ dynamics under risk-neutral and physical probability measures:

$$f_{t+1} = \mu_{Q,f} + \phi_{Q,f} f_t + u_{f,t+1}^Q \quad (33)$$

$$f_{t+1} = \phi_f f_t + u_{f,t+1}^P, \quad (34)$$

where $u_{f,t+1}^Q \sim^Q N(0, \sigma_{f,Q}^2)$ and $u_{f,t+1}^P = \sigma'_{fM} \tilde{\Sigma} \varepsilon_{t+1} + \sigma_f \varepsilon_{f,t+1}$ and where $\varepsilon_{f,t+1}$ is *i.i.d* $N(0, 1)$. Since the short rate can be written as

$$i_t = e_1' \left(\bar{M}_t + \tilde{M}_t \right), \quad (35)$$

it follows that the yield to maturity of a zero-coupon bond is given by

$$\begin{aligned} Y_t^{(n)} &= -\frac{\ln(P_t^{(n)})}{n} = -\frac{A_n}{n} - \frac{B_n^{*'}}{n} \bar{M}_t - \frac{\tilde{B}_n'}{n} \tilde{M}_t - \frac{B_{f,n}}{n} f_t \\ &= a_n + b_n^{*'} \bar{M}_t + \tilde{b}_n' \tilde{M}_t + b_{f,n} f_t, \end{aligned} \quad (36)$$

where for $n > 1$, the coefficients are given by the following recursions and the initial conditions are $A_1 = 0$, $B_1^* = \tilde{B}_1 = -e_1$ and $B_{f,1} = 0$:

$$\begin{aligned}
A_{n+1} &= A_n + B_n^{*'} K_0^* + \tilde{B}_n' \tilde{K}_0 + \frac{1}{2} \begin{pmatrix} B_n^* \\ \tilde{B}_n \end{pmatrix}' \Omega_Q \begin{pmatrix} B_n^* \\ \tilde{B}_n \end{pmatrix} + \frac{\sigma_{f,Q}^2 B_{f,n}^2}{2} \\
B_{n+1}^{*'} &= B_n^{*'} \Phi_Q^* - e_1' = -e_1' \sum_{j=0}^n (\Phi_Q^*)^j \\
\tilde{B}_{n+1}' &= \tilde{B}_n' \tilde{\Phi}_Q - e_1' = -e_1' \sum_{j=0}^n (\tilde{\Phi}_Q)^j \\
B_{f,n+1} &= B_{f,n} \phi_{Q,f} + \tilde{B}_n' \Phi_{Q,Mf} = -e_1' \left[\sum_{j=0}^n \left(\frac{1 - \phi_{Q,f}^{n-j}}{1 - \phi_{Q,f}} \right) (\tilde{\Phi}_Q)^j \right] \Phi_{Q,Mf}.
\end{aligned}$$

A.4 Identification—Lemma 1

Using Equation 26 we have

$$\Psi_j = I + \left(\sum_{k=1}^j \tilde{\Theta}_k \right) \tilde{\Sigma}^{-1}. \quad (37)$$

where we define $\tilde{\Theta}_j \equiv \Theta_j - e_3 \sigma_{g^*}'$. We can then use the recursion for Ψ_j in Equation 27 to recover the parameter Φ_j for $j = 1, \dots, p$:

$$\begin{aligned}
\Phi_1 &= \Psi_1 \\
\Phi_j &= \Psi_j - \sum_{k=1}^{j-1} \Phi_{j-k} \Psi_k, \quad j = 2, \dots, p,
\end{aligned} \quad (38)$$

which implies that we can write Φ_j for $j = 1, \dots, p$ as functions of the $\tilde{\Theta}_j$ and $\tilde{\Sigma}$:

$$\Phi_j = \Phi_j \left(\tilde{\Theta}_{j,1 \leq j \leq q}; \tilde{\Sigma} \right). \quad (39)$$

Using Equations 27 and 37 with $j = p + 1$, we can establish that

$$\left(\sum_{k=1}^p \Phi_k \Psi_{p+1-k} \right) \tilde{\Sigma} - \tilde{\Sigma} = \left(\sum_{k=1}^{p+1} \tilde{\Theta}_k \right), \quad (40)$$

which together (39) characterize $\tilde{\Sigma}$ as a function of the parameters $\tilde{\Theta}_j$:

$$\tilde{\Sigma} = \tilde{\Sigma} \left(\tilde{\Theta}_{j,1 \leq j \leq q+1} \right). \quad (41)$$

Finally, Equations 27 and 37 with $j = p + 2$ together yield an equation with only σ_{g^*} as the unknown parameter. Given that we have assumed the identification of Σ , we get Σ^* as the difference: $\Sigma^* = \Sigma - \tilde{\Sigma}$. Here, we have solved for the complete set of structural parameters as functions of the reduced-form parameters, hence the identification.

A.5 Identification—Lemma 2

We define the matrix R that satisfies the properties stated in Lemma 2. Define R as follows:

$$R = \Sigma_1^{-1} \Sigma_2,$$

so that $RR' = I$ by construction. The matrix Σ_1 and Σ_2 are given in the following. For this purpose, we rewrite the estimates of σ_{r^*} and σ_{g^*} using cartesian coordinates:

$$\sigma_{r^*} = \|\sigma_{r^*}\| \begin{pmatrix} \cos(\phi_r^*) \\ \sin(\phi_r^*) \cos(\theta_r^*) \\ \sin(\phi_r^*) \sin(\theta_r^*) \end{pmatrix} \quad \sigma_{g^*} = \|\sigma_{g^*}\| \begin{pmatrix} \cos(\phi_g^*) \\ \sin(\phi_g^*) \cos(\theta_g^*) \\ \sin(\phi_g^*) \sin(\theta_g^*) \end{pmatrix},$$

where $\phi_r^*, \phi_g^* \in [0, \pi]$ and $\theta_r^*, \theta_g^* \in [0, 2\pi]$. We first define Σ_2 :

$$\Sigma_2 = \begin{pmatrix} e_1^\top \\ \hat{\sigma}_{r^*}^\top \\ \hat{\sigma}_{g^*}^\top \end{pmatrix}$$

with

$$e_1^\top = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \hat{\sigma}_{r^*} = \|\sigma_{r^*}\| \begin{pmatrix} 0 \\ \cos(\hat{\theta}_r^*) \\ \sin(\hat{\theta}_r^*) \end{pmatrix} \quad \hat{\sigma}_{g^*} = \|\sigma_{g^*}\| \begin{pmatrix} 0 \\ \cos(\hat{\theta}_g^*) \\ \sin(\hat{\theta}_g^*) \end{pmatrix},$$

where $\hat{\theta}_g^* = \pi/2$ and $\hat{\theta}_r^* = \pi/2 - \arccos(\rho)$ with $\rho \equiv \frac{\sigma_{g^*}^\top \sigma_{r^*}}{\|\sigma_{r^*}\| \|\sigma_{g^*}\|}$. Next we define Σ_1 :

$$\Sigma_1 = \begin{pmatrix} \sigma^\top \\ \sigma_{r^*}^\top \\ \sigma_{g^*}^\top \end{pmatrix},$$

where

$$\sigma = \begin{pmatrix} \cos(\phi) \\ \sin(\phi) (\cos(\theta - \theta_r^*) \cos(\theta_r^*) - \sin(\theta - \theta_r^*) \sin(\theta_r^*)) \\ \sin(\phi) (\sin(\theta - \theta_r^*) \cos(\theta_r^*) + \cos(\theta - \theta_r^*) \sin(\theta_r^*)) \end{pmatrix},$$

so that we need to provide $\cos(\phi)$, $\sin(\phi)$, $\cos(\theta - \theta_r^*)$ and $\sin(\theta - \theta_r^*)$ to complete the construction of R . First, if $\cos(\phi_r^*) \geq 0$, then

$$\cos(\phi) = \frac{-\sin(\phi_r^*) \cos(\theta - \theta_r^*)}{\sqrt{\sin^2(\phi_r^*) \cos^2(\theta - \theta_r^*) + \cos^2(\phi_r^*)}}$$

$$\sin(\phi) = \frac{\cos(\phi_r^*)}{\sqrt{\sin^2(\phi_r^*) \cos^2(\theta - \theta_r^*) + \cos^2(\phi_r^*)}}.$$

Second, if $\cos(\phi_r^*) < 0$, then

$$\begin{aligned}\cos(\phi) &= \frac{\sin(\phi_r^*) \cos(\theta - \theta_r^*)}{\sqrt{\sin^2(\phi_r^*) \cos^2(\theta - \theta_r^*) + \cos^2(\phi_r^*)}} \\ \sin(\phi) &= \frac{-\cos(\phi_r^*)}{\sqrt{\sin^2(\phi_r^*) \cos^2(\theta - \theta_r^*) + \cos^2(\phi_r^*)}}.\end{aligned}$$

Third, if $\frac{\cos(\phi_g^*)}{\cos(\phi_r^*)} \geq 0$, then

$$\begin{aligned}&\cos(\theta - \theta_r^*) \\ &= \frac{\cos(\phi_r^*) \sin(\phi_g^*) \sin(\theta_r^* - \theta_g^*)}{\sqrt{\left(\cos(\phi_r^*) \sin(\phi_g^*) \cos(\theta_r^* - \theta_g^*) - \cos(\phi_g^*) \sin(\phi_r^*)\right)^2 + \cos^2(\phi_r^*) \sin^2(\phi_g^*) \sin^2(\theta_r^* - \theta_g^*)}} \\ &\sin(\theta - \theta_r^*) \\ &= \frac{\cos(\phi_r^*) \sin(\phi_g^*) \cos(\theta_r^* - \theta_g^*) - \cos(\phi_g^*) \sin(\phi_r^*)}{\cos(\phi_r^*) \sin(\phi_g^*) \sin(\theta_r^* - \theta_g^*)} \cos(\theta - \theta_r^*).\end{aligned}$$

Fourth, if $\frac{\cos(\phi_g^*)}{\cos(\phi_r^*)} < 0$ then

$$\begin{aligned}&\cos(\theta - \theta_r^*) \\ &= \frac{-\cos(\phi_r^*) \sin(\phi_g^*) \sin(\theta_r^* - \theta_g^*)}{\sqrt{\left(\cos(\phi_r^*) \sin(\phi_g^*) \cos(\theta_r^* - \theta_g^*) + \cos(\phi_g^*) \sin(\phi_r^*)\right)^2 + \cos^2(\phi_r^*) \sin^2(\phi_g^*) \sin^2(\theta_r^* - \theta_g^*)}} \\ &\sin(\theta - \theta_r^*) \\ &= \frac{\left(\cos(\phi_r^*) \sin(\phi_g^*) \cos(\theta_r^* - \theta_g^*) + \cos(\phi_g^*) \sin(\phi_r^*)\right)}{\cos(\phi_r^*) \sin(\phi_g^*) \sin(\theta_r^* - \theta_g^*)} \cos(\theta - \theta_r^*).\end{aligned}$$

A.6 Likelihood

Macro Data The likelihood of the macro data is available recursively. For the given time- t values of \tilde{M}_t , \bar{M}_t and y_t^* , we compute $E_t[M_{t+1}]$ using Equations 3, 4 and 6 in Section 2.2, and we compute the innovations $u_{t+1} = M_{t+1} - E_t[M_{t+1}]$. Then, for the given time- t values for $\tilde{\Sigma}_t$ and $\bar{\Sigma}_t$, the (log) likelihood of the macro variables is given by Equation 28. To increment the recursion, we update the conditional mean and conditional variance components using the laws of motion discussed in Sections 2.2-2.4. We estimate the initial values for \bar{M}_0 and y_0^* and we set $\bar{M}_0 = E[\bar{M}_t] = 0$ and the starting values for $v_{.,0} = 1$. For ease of comparison between the different models, we keep \bar{M}_0 and y_0^* fixed to the estimated values in the baseline model. The case with constant loadings is easily nested by fixing the values $v_{.,t} = 1$.

Bond Yields We assume that one linear combination of yields is measured to perfectly recover the latent factor f_t directly from the cross-section of bond yields, given the parameters of the models, and we compute

the likelihood of the bond yields. Consider the portfolio of yields WY_t that are measured without errors, where W is a row vector with size J . Then, it follows that

$$f_t = (1/B_{f,W}) \left[WY_t - \left(A_W + (W\iota_J) i_t^* + \tilde{B}'_W \tilde{M}_t \right) \right], \quad (42)$$

where ι_J is a $J \times 1$ vector of 1, $A_W = WA$, $B_W = WB$ and $B_{f,W} = WB_f$, and where A , B and B_f stack the yield coefficients. In practice, W is the equally weighted portfolio of the J yields. The remaining $J - 1$ portfolios of yields $Y_t^e \equiv W_e Y_t$ are measured with errors where the matrix stacking W and W_e has full rank and the likelihood is given by the following:

$$\mathcal{L}_{Y^e,t} = -\frac{J-1}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\Omega_{Y^e})) - \frac{1}{2} u'_{Y^e,t} (\Omega_{Y^e})^{-1} u_{Y^e,t}, \quad (43)$$

where $u_{Y^e,t}$ are the $J - 1$ portfolios of the yield pricing errors, Ω_{Y^e} is a diagonal matrix of the standard deviations of the measurement errors and the matrix stacking W and W_e has full rank.