Online Appendix for "Interest Rates Under Falling Stars"

Michael D. Bauer and Glenn D. Rudebusch
Federal Reserve Bank of San Francisco

December 27, 2019

1 Trend estimates

1.1 Overview of trend estimates

Table 1 provides an overview of all of the trend estimates used in our analysis. The first five estimates of r_t^* are obtained from models in published studies, and these are described below in Section 1.2 and shown in Figure 1. Our own four estimates of r_t^* are described in Section 1.3 and shown in Figure 2. In Section 3 of the paper, we use a "filtered" estimate and a "real-time" estimate of r_t^*, which are averages of the three filtered and six real-time estimates, respectively. The real-time estimate of i_t^* is the sum of the PTR estimate of π_t^* and the real-time estimate of r_t^*. The *ESE* model estimate of i_t^* is described at a high level in Section 4.2 of the paper, with details in Section 3.8 of this Online Appendix.

Our proxy estimate of the inflation trend π_t^* is the Fed's PTR measure, the perceived target rate (PTR) of inflation, see Federal Reserve Board of Governors (2019b). PTR, which is part of the Fed's FRB/US macroeconomic model, measures expectations for inflation in the price index of personal consumption expenditures (PCE). Consistent with this definition, our estimates of r_t^* are based on real interest rates relative to PCE inflation. Since 1979, PTR corresponds to long-run inflation expectations from the Survey of Professional Forecasters (SPF). Before 1979, it is based on estimates from a learning model for expected inflation. For details see Brayton and Tinsley (1996). Data are available at https://www.federalreserve.gov/econresdata/frbus/us-models-package.htm.

The estimates of r_t^* are described in the following. All of these estimates are based on the assumption that the real rate contains a stochastic trend, i.e., a stochastic endpoint.

1.2 External estimates of r_t^*

Del Negro et al. (2017) (DGGT) propose a number of Bayesian common-trend VARs to estimate r_t^* and its possible drivers. We focus on their baseline model in which three stochastic trends, including r_t^* and π_t^*, are estimated from five data series: (1) observed PCE inflation, (2) long-run inflation expectations from the PTR series, (3) the 3-month T-bill rate, (4) the 20-year Treasury yield and (5) long-run expectations of the 3-month yield. For details about the data and model specification see their section II.A. We estimate their model using data

Table 1: Overview of trend estimates

Trend	Source	Real-time	Filtered	Smoothed
r^*	Del Negro et al. (2017) (DGGT)	✓		✓
	Johannsen and Mertens (2016) (JM)	✓		✓
	Laubach and Williams (2016) (LW)		✓	✓
	Holston et al. (2017) (HLW)		✓	
	Kiley (2015)		✓	✓
r^*	UC	✓		✓
	Proxies	✓		✓
	SSM	✓		✓
	MA	✓		
π_t^*	PTR	✓		
i^*	Real-time	✓		
	ESE			✓

Overview of trend estimates used in the paper: five external sources of r^*, our own four estimates of r^*, the PTR estimate of π^*, and two i^* estimates. "Real-time" indicates (pseudo) real-time estimation of the trend proxy, "filtered" indicates estimates from a (one-sided) Kalman filter using full-sample parameter estimates, and "smoothed" indicates (two-sided) Kalman smoother or full-sample Bayesian estimate of the trend. All trend estimates are quarterly from 1971:Q4 to 2018:Q1.

up to 2018:Q1 and replicate their published results. The trends are smoothed (two-sided) estimates, as they are the posterior medians of the (MCMC) sampled trend series conditional on the full data set. In addition, we recursively estimate their model starting in 1971:Q4, expanding the data by adding one quarter at a time. The smoothed estimate of r_t^* is shown in the left panel of Figure 1, and the recursive, real-time estimate is shown in the right panel.

Johannsen and Mertens (2016, 2018) (JM) propose a time series model for interest rates with explicit treatment of the zero lower bound and stochastic volatility. For the version of their model reported in Johannsen and Mertens (2016), they generously provided us with updated estimates including both full-sample (smoothed/two-sided) and real-time estimated series of r_t^*. These are shown in Figure 1.

While the two estimation approaches above use the long-run definition of r_t^* that is the relevant one for the trend in nominal interest rates (see Section 2), the following three external estimates use a different definition and estimation approach. The prominent model of Laubach and Williams (2003, 2016)(LM), the slightly modified version of this model by Holston, Laubach, and Williams (2017)(HLW), and the version by Kiley (2015) all use a simple linearized New Keynesian macro model—essentially the Rudebusch and Svensson (1999) model—in which r_t^* is estimated as the *neutral* real interest rate at which monetary policy is neither expansionary nor contractionary.[1] But despite this difference with long-run r_t^*, these

[1]Compared to the LW model, the HLW model excludes relative price shocks from the Phillips curve, uses a simpler proxy for inflation expectations, and assumes a one-for-one effect of trend output growth on r_t^*. In

estimates are still worth considering in our context for several reasons. First, in practice, their "definition takes a 'longer-run' perspective, in that it refers to the level of real interest rates expected to prevail, say, five to ten years in the future, after the economy has emerged from any cyclical fluctuations and is expanding at its trend rate" (Laubach and Williams, 2016, p. 57), so the definitions are effectively quite close. Second, in all three of these models the neutral rate is a martingale, so at least as implemented within the models, the neutral real rate is also the long-run trend in the real rate. Third, these are among the most widely used estimates of r_t^* both in the academic literature and in practice. Accordingly, these estimates are useful additions to our analysis. A shortcoming of these estimates is that they are not available in real time. Of course, it would in principle be possible to recursively estimate these models—and for a limited time period towards the end of our sample Laubach and Williams (2016) have done this. But these macro-based estimates are by their nature very sensitive to the macroeconomic data, in particular real GDP, so that real-time data issues become a serious concern. We therefore only have filtered/one-sided estimates based on the full-sample parameter estimates, and smoothed/two-sided estimates, which are shown in the middle and right panels of Figure 1.

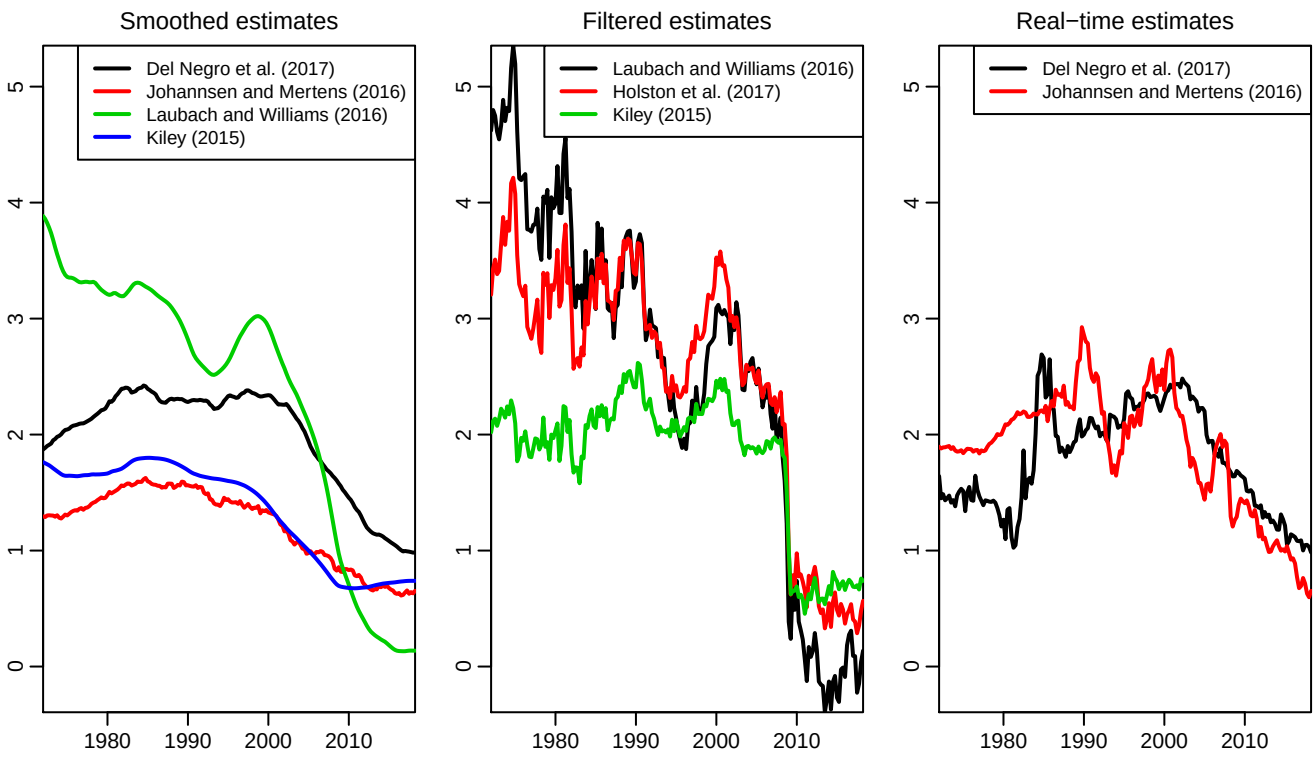
We exclude two other prominent estimates of r_t^*. Lubik and Matthes (2015) estimate r_t^* as the five-year forecast of the real rate from a time-varying parameter VAR model. This horizon is too short to be comparable with our long-run trend, and their implementation leads to a highly volatile estimate of r_t^*. Christensen and Rudebusch (2019) estimate r_t^* from yields on Treasury inflation-protected Securities (TIPS) using a dynamic term structure model that accounts for liquidity premia, but their sample only starts in 1998 with the introduction of TIPS.

1.3 Details on our estimates of r_t^*

Our first estimate is based on a univariate unobserved components (UC) model for the real short-term interest rate, similar to Watson (1986). Recently, Fiorentini et al. (2018) have argued that such univariate models can infer r_t^* with greater precision than the LW model. We use an ex-ante real interest rate calculated as the difference between the three-month Treasury bill rate (Federal Reserve Board of Governors, 2019a) and core PCE inflation, that is, the four-quarter percent change in the price index for personal consumption expenditures excluding food and energy items (U.S. Bureau of Economic Analysis, 2019). (Inflation over the past year serves as a proxy estimate for expectations of inflation over the next quarter.) The real rate is decomposed into a random walk trend (r_t^*) and a stationary component (the real rate gap, specified as an AR(1) process with zero mean), which are the two state variables of the state-space model. The prior distributions for the parameters are uninformative, with the exception of the variance for the innovations to the random walk component, i.e., for changes in r_t^*. We use a tight prior around a low value for this variance, similar to DGGT. Specifically, the prior distribution for this variance is inverse-gamma, $IG(\alpha/2, \delta/2)$, with $\alpha = 100$ and $\delta = 0.01(\alpha + 2)$. This implies that the mode is 0.01, and the variance of the change in r_t^* over 100 years is 4, i.e., the standard deviation is 2 (percent). This is a slightly higher mode than used by DGGT (their mode implies a standard deviation over 100 years of one percent). In

Kiley's model, the IS curve is augmented with credit spreads.

Figure 1: External estimates of r_t^*



Estimates of r_t^* from existing models. Left panel: smoothed/two-sided estimates. Middle panel: filtered/one-sided estimates. Right panel: (pseudo) real-time estimates. The sample period is from 1971:Q4 to 2018:Q1.

our MCMC sampler, we draw the unobserved state variables using the simulation smoother of Durbin and Koopman (2002) and the parameters using standard Gibbs steps. For this model and for the following two models below, we use random starting values to initialize our MCMC chain, and we carefully monitor convergence of the MCMC sampler. For the full-sample estimation we use 100,000 MCMC draws. For our recursive estimation, we start in 1971:Q4 with 100,000 draws, and every time we add another observation, we obtain 20,000 more draws.

The second estimate (labeled "proxies") is from a multivariate model, which augments the UC model by two additional measurement equations relating r_t^* to two proxies that recent work has shown to be important correlates of the real rate trend. The first proxy is a ten-year moving average of quarterly real GDP growth, and the second proxy is a ten-year moving average of the quarterly growth rate in the total number of hours worked in the business sector, i.e., labor force hours. The choice of these proxies is motivated by the results in Lunsford and West (2019), who document strong low-frequency correlations between these series and the real rate over a post-war sample. While these macroeconomic data are subject to data revisions, in particular real GDP, the long moving averages and the use of these series in extracting the long-run trend in the real rate lessens concerns that data revisions would

materially affect our real-time trend estimate. In the additional measurement equations, the proxies are scaled by a parameter to be estimated and the measurement error is allowed to be serially correlated. The model has four state variables: r_t^*, the real-rate gap, and the two measurement errors (AR(1) processes with non-zero means). The prior distributions are uninformative, except again for the trend innovation variance, where the prior is the same as in the UC model. We design a hybrid MCMC sampler: The scaling parameters are drawn using random-walk Metropolis-Hastings steps with the state variables integrated out (i.e., using the Kalman filter to obtain the likelihood for calculation of the acceptance probability). Then, the state variables are sampled with the simulation smoother and the remaining parameters are drawn using Gibbs steps.

Our third estimate is from a state-space model (SSM) that is similar to the specification by DGGT in that it includes both inflation and the nominal short rate, and it estimates both r_t^* and π_t^*. The main differences are that we include neither survey expectations nor long-run yields, and that our measurement equations are somewhat more standard. Our three observation series are quarterly PCE inflation, the 3-month T-bill rate, and the PTR series for long-run inflation expectations. In addition to the trends, the two other state variables are the real-rate gap, r_t^g and the inflation gap, π_t^g, which follow a bivariate VAR with four lags. The measurement equations are

$$\begin{aligned} \pi_t &= \pi_t^* + \pi_t^g + e_t^\pi, \\ PTR_t &= \pi_t^* + e_t^{PTR}, \\ y_t^{(3m)} &= \pi_t^* + E_t\pi_{t+1}^g + r_t^* + r_t^g + e_t^y, \end{aligned}$$

where e_t^π, e_t^{PTR}, and e_t^y are *iid* measurement errors, and $E_t\pi_{t+1}^g$ is implied by the VAR. The priors are generally uninformative, and for the VAR we use the same Minnessota prior as DGGT. For the variance of the innovations to both r_t^* and π_t^* we use smoothness prior similar to the previous two models and the same prior modes for the variances as in DGGT (equivalent to one percent and two percent standard deviation, respectively, for changes in trends over 100 years). The MCMC sampler simply combines the simulation smoother for the state variables and Gibbs steps for the other parameters.

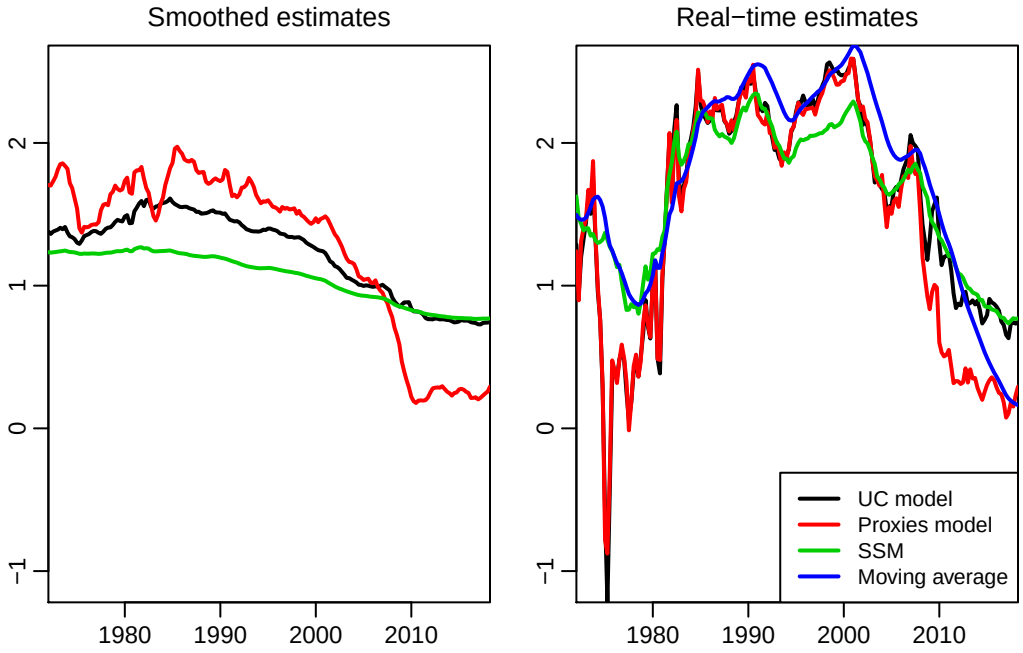
Finally, we also calculate a simple moving-average (MA) estimate of r_t^*. The observed real-rate series is the same as in the first two models above. Denoting this series by r_t, we calculate an exponentially-weighted moving average using the recursion $r_t^* = \alpha r_{t-1}^* + (1-\alpha)r_t$, which we start ten years before the beginning of our sample, in 1961:Q4, with $r_t^* = r_t$. We use $\alpha = 0.98$, a value in line with earlier work estimating macro trends.

1.4 Measurement error in macro trends

In establishing some stylized facts about a low-frequency macro-finance connection in Section 3, we treat the estimated macro trends as data. However, there is substantial model and estimation uncertainty attached to the various point estimates of r_t^*. Similarly, our survey-based measure of the long-run inflation trend, π_t^*, is also imprecise. How concerned should we be about measurement error?

Our results show that measures of macro trends are closely connected to the yield curve and

Figure 2: Our estimates of r_t^*



Our own four estimates of r_t^*. Left panel: smoothed/two-sided estimates. Right panel: (pseudo) real-time estimates. The sample period is from 1971:Q4 to 2018:Q1.

that they contain important information for predicting future yields and returns. The effect of measurement error in the trend estimates on our results depends on the structure of that error. One possibility is that the trends are subject to classical measurement error, an error that is orthogonal to the unobserved trend. In this case, it is well known that the coefficients in our regressions would be both less precise and biased toward zero. An alternative possibility is that the measurement error is orthogonal to the estimate of the trend, which may be more plausible since our trend proxies are estimates of the underlying trend based on the available information set (Mankiw and Shapiro, 1986). This type of measurement error would make our estimates noisy though not necessarily biased (Hyslop and Imbens, 2001). For either type of measurement error, the use of a noisy trend would likely lead to weaker estimated relationships than the use of the true underlying trend. Hence, accounting for measurement error suggests that our results should be viewed as a lower bound for the tightness of the connection between the yield curve and the *true* macro trends.

Finally, we note that our trend measures were not created on purpose to match the evolution of Treasury yields. Therefore, the connections we find could well be stronger if the trend estimates were optimized to exhibit a tight connection with or predictive power for long-term yields.

2 Additional results

2.1 Persistence and common trends

Here we report additional results for the presence of a common trend in nominal yields (Section 3.1 of the paper).

Table 2 documents the extremely high persistence of various measures of nominal Treasury yields. It reports summary statistics for the two-year yield, $y_t^{(8)}$, the ten-year yield, $y_t^{(40)}$, and the level of the yield curve, L_t, measured as the first principal component of yields (scaled so that the loadings add up to one). Summary statistics are also reported for our macro trends, including the inflation trend, our three estimates of the real rate trend—the filtered estimate, $r_t^{*,F}$, the real-time estimate, $r_t^{*,RT}$, and the moving-average estimate, $r_t^{*,MA}$—and our real-time proxy for the equilibrium nominal short rate, $i_t^* = \pi_t^* + r_t^{*,RT}$. To assess whether a simple difference between yields and trends can remove the extreme persistence and make yields stationary we also report summary statistics for yields detrended in this way. Along with standard deviations, the table reports two measures of persistence: the estimated first-order autocorrelation coefficient, $\hat{\rho}$, and the half-life, which indicates the number of quarters until half of a given shock has died out and is calculated as $\ln(0.5)/\ln(\hat{\rho})$. It also includes two tests for a unit root, the Augmented Dickey-Fuller (ADF) t-statistic and the non-parametric Phillips-Perron (PP) Z_α statistic. For the ADF test, we include a constant and k lagged difference in the test regression, where k is determined using the general-to-specific procedure suggested by Ng and Perron (1995), starting from four lags. For the PP test, we use a Newey-West estimator of the long-run variance with four lags. In the last column, we report the p-value for the null hypothesis of stationarity using the low-frequency stationarity test (LFST) of Müller and Watson (2013).

The yields are highly persistent, with a first-order autocorrelation coefficient of 0.97 and a half-life of around 22-27 quarters. Neither the ADF or PP tests reject a unit root, while the LFST p-values reject stationarity for each series. This evidence suggests that nominal yields can be effectively modeled as $I(1)$ processes.

The macro trends are even more persistent. For example, our real-time estimate of r_t^* has an autocorrelation coefficient of 0.98 and a half-life of about 36 quarters. The inflation trend and i_t^* have autocorrelation coefficients of 0.99 and half-lives of 85 and 60 quarters, respectively. Unsurprisingly, these macro trend proxies exhibit pronounced trend behavior, that is, they also behave like I(1) processes.

The naive detrending method of subtracting trend proxies from yields leads to time series with much lower persistence. If only the inflation trend is subtracted, the resulting series is less persistent but still sufficiently trending that the evidence from all three tests favors a unit root. When a real rate trend—$r_t^{*,F}$, $r_t^{*,RT}$, or $r_t^{*,MA}$—is also subtracted off together with π_t^*, the persistence drops further, and the evidence sometimes favors stationarity, depending on the specific interest rate, real rate trend proxy, and unit root test. This extremely simple detrending method, which makes the strong assumption that coefficients in the cointegration vector of yields and macro trends are all one (in absolute value), has some success, but only if both macro trends are accounted for. The evidence in Section 3.1 of the paper finds an even stronger macro-finance link when the cointegration coefficients are slightly larger than one.

Table 2: Persistence of interest rates, macroeconomic trends, and differences

Series	SD	$\hat{\rho}$	Half-life	ADF	PP	LFST
$y_t^{(8)}$	3.54	0.97	21.9	-1.25	-4.59	0.00
$y_t^{(8)} - \pi_t^*$	2.26	0.93	9.1	-2.07	-11.09	0.00
$y_t^{(8)} - \pi_t^* - r_t^{*,F}$	1.61	0.87	4.9	-3.21**	-22.58***	0.10
$y_t^{(8)} - \pi_t^* - r_t^{*,RT}$	2.00	0.92	7.8	-2.24	-13.23*	0.00
$y_t^{(8)} - \pi_t^* - r_t^{*,MA}$	2.05	0.91	7.8	-2.35	-13.83*	0.01
$y_t^{(40)}$	2.94	0.97	26.4	-1.13	-3.11	0.00
$y_t^{(40)} - \pi_t^*$	1.67	0.93	9.4	-2.32	-9.72	0.01
$y_t^{(40)} - \pi_t^* - r_t^{*,F}$	1.17	0.87	5.0	-3.61***	-22.87***	0.23
$y_t^{(40)} - \pi_t^* - r_t^{*,RT}$	1.33	0.90	6.7	-2.51	-15.16**	0.00
$y_t^{(40)} - \pi_t^* - r_t^{*,MA}$	1.36	0.90	6.8	-3.04**	-16.01**	0.01
L_t	3.13	0.97	26.6	-1.20	-3.42	0.00
$L_t - \pi_t^*$	1.84	0.93	9.8	-1.91	-9.69	0.00
$L_t - \pi_t^* - r_t^{*,F}$	1.24	0.87	4.8	-2.99**	-23.03***	0.20
$L_t - \pi_t^* - r_t^{*,RT}$	1.54	0.91	7.7	-2.19	-13.05*	0.00
$L_t - \pi_t^* - r_t^{*,MA}$	1.58	0.91	7.6	-2.36	-13.90*	0.01
π_t^*	1.60	0.99	85.4	-0.62	-1.16	0.00
$r_t^{*,F}$	1.03	0.97	27.2	-0.67	-1.16	0.00
$r_t^{*,RT}$	0.59	0.98	36.2	-0.77	-1.04	0.03
$r_t^{*,MA}$	0.70	0.98	43.4	-0.75	1.31	0.03
$i_t^* = \pi_t^* + r_t^{*,RT}$	1.72	0.99	59.7	-0.30	-0.28	0.00

Standard deviation (SD); first-order autocorrelation coefficient ($\hat{\rho}$); half-life ($\ln(0.5)/\ln(\hat{\rho})$); Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) unit root test statistics (with *,** and *** indicating significance at 10%, 5%, and 1% level) and p-values for Mueller-Watson low-frequency stationary test (LFST), for the two-year yield, $y_t^{(8)}$, the ten-year yield, $y_t^{(40)}$, the level of the yield curve (the first principal component of yields), L_t, the detrended yields, and macro trends. The r^*-estimates are the filtered ("F"), real-time ("RT") and moving-average ("MA") estimates described in Section 2 of the paper. The data are quarterly from 1971:Q4 to 2018:Q1.

Table 3: Cointegration regressions and tests for the level of the yield curve

	Level	(1)	(2)	(3)	(4)	(5)
constant	6.48	-0.23	-1.17	-2.60	-0.87	-3.06
	(0.55)	(0.62)	(0.35)	(0.44)	(0.42)	(0.37)
π_t^*		1.76	1.31	1.56	1.45	
		(0.13)	(0.11)	(0.08)	(0.09)	
r_t^*			1.13	1.90	1.12	
			(0.15)	(0.16)	(0.12)	
i_t^*						1.75
						(0.07)
R^2		0.83	0.93	0.96	0.97	0.94
Memo: r^*			filtered	real-time	mov. avg.	real-time
SD	2.94	1.39	1.10	0.85	1.09	0.81
$\hat{\rho}$	0.97	0.87	0.82	0.72	0.81	0.67
Half-life	26.4	5.1	3.5	2.1	3.3	1.8
ADF	-1.13	-2.68	-3.44	-3.92**	-3.45	-4.26**
PP	-3.11	-20.07*	-31.16**	-51.75***	-32.81**	-61.31***
LFST	0.00	0.04	0.28	0.37	0.10	0.84
Johansen $r = 0$		10.10	29.31	43.26***	38.11**	21.19**
Johansen $r = 1$		0.90	5.96	11.14	10.25	0.71
ECM $\hat{\alpha}$		-0.10	-0.16	-0.29	-0.33	-0.32
		(0.03)	(0.05)	(0.07)	(0.09)	(0.07)

Dynamic OLS regressions for the level of the yield curve, L_t, (the first principal component of yields) on macroeconomic trends, including four leads and lags of ΔL_t and differenced trend variables. Newey-West standard errors using six lags are in parentheses. The r^* estimates are described in Section 2 of the paper, and the long-run nominal short rate i_t^* is the sum of π_t^* and the real-time estimate of r_t^*. For the cointegration residuals (and, in the first column, for the yield level itself), the second panel reports standard deviations (SD), first-order autocorrelation coefficients ($\hat{\rho}$), half-lives ($\ln(0.5)/\ln(\hat{\rho})$), Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) unit root test statistics, and p-values for Mueller-Watson low-frequency stationary test (LFST). The table also reports the Johansen trace statistic which tests whether the cointegration rank (r) among the L_t and the macro trends is zero/one against the alternative that it exceeds zero/one, using four lags in the VAR. For ADF, PP and Johansen tests *,** and *** indicate significance at 10%, 5%, and 1% level. Estimates of the coefficient α (with White standard errors) on the cointegration residual in the error-correction model (ECM) for ΔL_t that also includes an intercept, four lags of ΔL_t, and four lags of differenced macro trends. The data are quarterly from 1971:Q4 to 2018:Q1.

Table 4: Excess return regressions for individual maturities

	Yields only				Yields and i^*				
	PC1	PC2	PC3	R^2	PC1	PC2	PC3	i^*	R^2
2y	0.06	0.05	-0.65	0.05	0.70	0.12	-0.26	-1.19	0.15
	(0.06)	(0.06)	(0.50)		(0.22)	(0.05)	(0.52)	(0.37)	[0.02]
5y	0.08	0.22	-1.72	0.08	1.53	0.38	-0.80	-2.74	0.18
	(0.11)	(0.12)	(0.87)		(0.41)	(0.10)	(0.89)	(0.70)	[0.00]
7y	0.08	0.34	-2.19	0.08	2.06	0.56	-0.94	-3.74	0.19
	(0.14)	(0.15)	(1.11)		(0.53)	(0.13)	(1.14)	(0.90)	[0.00]
10y	0.09	0.51	-2.74	0.09	2.90	0.83	-0.96	-5.31	0.20
	(0.20)	(0.20)	(1.56)		(0.71)	(0.18)	(1.57)	(1.23)	[0.00]
15y	0.08	0.83	-3.70	0.09	4.13	1.28	-1.15	-7.64	0.21
	(0.28)	(0.29)	(2.20)		(0.99)	(0.26)	(2.17)	(1.72)	[0.00]

Predictive regressions for quarterly excess returns on bonds with maturities 2, 5, 7, 10 and 15 years. The predictors are the first three principal components of yields (PC1, PC2, PC3) and, in the second specification, the long-run nominal short rate i_t^*, the sum of π_t^* and the real-time estimate of r_t^* (see Section 2 of the paper). Numbers in parentheses are White standard errors and in squared brackets are small-sample p-values for the spanning hypothesis (that the coefficient on i_t^* is zero) obtained with the bootstrap method of Bauer and Hamilton (2018). The data are quarterly from 1971:Q4 to 2018:Q1.

Table 3 shows the results for a cointegration analysis similar to the one carried out in Section 3.1 of the paper, here applied to the level of the yield curve. The results are qualitatively similar to those we reported in the paper (Table 1) for the ten-year yield.

2.2 Predicting excess returns

Here we report additional results for excess return predictions using macro trends (Section 3.2 of the paper).

For assessing the predictive power of macro trends for excess bond returns, the paper focuses on returns that are averaged across bond maturities from two to 15 years. Table 4 shows (full-sample) estimates for individual maturites. It compares the usual yields-only specification to the specification that also includes our estimate of the long-run nominal short rate i_t^*. For all maturities, the addition of i_t^* (i) substantially raises R^2, (ii) makes both the level (PC1) and slope (PC2) strongly significant predictors, and (iii) leads to a large and strongly significant coefficient on i_t^* itself. This trend coefficient has the opposite sign and a somewhat larger magnitude than the coefficient on the level. These findings all closely parallel those in reported in the paper (Table 2). In addition, the results here show that the coefficients on both yield predictors and i_t^* rise with maturity, due to the fact that return volatility scales with bond maturity. Finally, it is noteworthy that for the two-year maturity, excess returns appear unpredictable with yield information alone, but become strongly predictable once our trend proxy is added.

In the presence of persistent predictors, it is generally difficult to interpret the magnitude of

R^2 as a measure of predictive accuracy, because persistent predictors that are truly irrelevant in population can substantially increase R^2 in small samples (Bauer and Hamilton, 2018). To be able to better interpret the R^2 we use our bootstrap to generate small-sample distributions of R^2 under the spanning hypothesis (that is, under the null that the three PCs of yields contain all the relevant information for predicting returns). We then compare the statistics obtained in the actual data to the quantiles of these bootstrap distributions. If the level or change of a regression R^2 is outside of the 95%-bootstrap interval then the regression results are inconsistent with the spanning hypothesis and suggest that additional predictors beyond yields alone are statistically significant.

The top panel of Table 5 reports this comparison for the specifications with (i) only yields, (ii) yields and π_t^*, (iii) yields, π_t^* and the real-time estimate of r_t^*, and (iv) yields and i_t^*. Adding π_t^* to the regression increases R^2 by 7 percentage points, while the 95%-bootstrap interval indicates that under the null hypothesis it would be uncommon to observe an increase in R^2 of more than 5 percentage points. Adding the real-time r_t^* estimate increases R^2 to 21%, and the increase relative to the yields-only specification is 12 percentage points, while the bootstrap suggests an increase of at most 7 percentage points would be plausible under the null. Adding just i_t^* also increases R^2 by 12 percentage points—much more than is plausible under the null. In the post-1985 subsample, the increase in R^2 from only adding π_t^* is not statistically significant, whereas the increases from adding either both macro trends or only i_t^* are significant.

3 Details on dynamic term structure model

3.1 Dynamic system

The evolution of the state variables under the real-world (or physical) probability measure, denoted as the $\mathbb{P}$-measure, is

$$P_t = \bar{P} + \gamma\tau_t + \tilde{P}_t, \quad \tau_t = \tau_{t-1} + \eta_t, \quad \tilde{P}_t = \Phi\tilde{P}_{t-1} + \tilde{u}_t, \tag{1}$$

and this is the common trends representation of the dynamic system (Stock and Watson, 1988). The state variables $Z_t = (\tau, P_t')'$ are cointegrated: $\beta' Z_t \sim I(0)$, with $\beta = (-\gamma, I_N)'$ an obvious choice for the N cointegration vectors. Because yields are just linear combinations of the risk factors, they are also cointegrated with a single common trend.

To derive the VAR representation of the dynamic system, substitute for $\tilde{P}_t$ and $\tilde{P}_{t-1}$ in the VAR equation, using $\tilde{P}_t = P_t - \bar{P} - \gamma\tau_t$. This yields

$$\begin{aligned} P_t &= (I_N - \Phi)\bar{P} + \Phi P_{t-1} + \gamma\tau_t - \Phi\gamma\tau_{t-1} + \tilde{u}_t \\ &= (I_N - \Phi)\bar{P} + (I_N - \Phi)\gamma\tau_{t-1} + \Phi P_{t-1} + u_t, \end{aligned}$$

where the last equation defines the innovations to P_t, $u_t = \gamma\eta_t + \tilde{u}_t$, which have covariance matrix $\Omega = E(u_t u_t') = \gamma\gamma'\sigma_\eta^2 + \widetilde{\Omega}$. Thus, the VAR representation for Z_t is

$$Z_t = \mu_Z + \Phi_Z Z_{t-1} + v_t, \quad v_t = (\eta_t, u_t')', \tag{2}$$

Table 5: Predictive power: R^2 with small-sample bootstrap intervals

	Full sample: 1971:Q4–2018:Q1		Subsample: 1985:Q1–2018:Q1	
Predictors	R^2	ΔR^2	R^2	ΔR^2
Yields only	0.09		0.08	
	[0.03, 0.19]		[0.03, 0.18]	
Yields and π_t^*	0.16	0.07	0.10	0.02
	[0.04, 0.20]	[0.00, 0.05]	[0.04, 0.19]	[0.00, 0.05]
Yields, π_t^* and r_t^*	0.21	0.12	0.19	0.10
	[0.05, 0.21]	[0.00, 0.07]	[0.05, 0.20]	[0.00, 0.07]
Yields and i_t^*	0.21	0.12	0.16	0.07
	[0.04, 0.20]	[0.00, 0.04]	[0.04, 0.19]	[0.00, 0.04]
Yields detrended by π_t^*	0.15	0.07	0.08	-0.00
	[0.03, 0.19]	[-0.03, 0.04]	[0.03, 0.18]	[-0.03, 0.04]
Yields detrended by π_t^* and r_t^*	0.18	0.09	0.17	0.09
	[0.04, 0.19]	[-0.04, 0.05]	[0.03, 0.18]	[-0.04, 0.05]
Yields detrended by i_t^*	0.20	0.12	0.14	0.06
	[0.03, 0.19]	[-0.03, 0.03]	[0.03, 0.18]	[-0.03, 0.03]

Predictive power of regressions for quarterly excess bond returns, averaged across two- to 15-year maturities. The predictors are three principal components (PCs) of yields, the PTR estimate of the inflation trend π_t^*, our real-time estimate of the equilibrium real rate r_t^*, and the equilibrium nominal short rate i_t^* taken as the sum of these inflation and real-rate trend estimates. The last three specifications use detrended yields, that is, three PCs of the yield residuals in regressions on (i) π_t^*, (ii) π_t^* and r_t^*, or (iii) i_t^*. Increase in R^2 (ΔR^2) is reported relative to the first specification with only PCs of yields. Numbers in square brackets are 95%-bootstrap intervals obtained by calculating the same regressions statistics in 5,000 bootstrap data sets generated under the (spanning) null hypothesis that only yields have predictive power for bond returns, using the bootstrap method of Bauer and Hamilton (2018).

with

$$\mu_Z = \begin{pmatrix} 0 \\ (I_n - \Phi)\bar{P} \end{pmatrix},$$

$$\Phi_Z = \begin{pmatrix} 1 & 0_{1\times N} \\ (I_N - \Phi)\gamma & \Phi \end{pmatrix},$$

and innovation covariance matrix

$$\Omega_v = E(v_t v_t') = \begin{pmatrix} \sigma_\eta^2 & \gamma'\sigma_\eta^2 \\ \gamma\sigma_\eta^2 & \Omega \end{pmatrix}.$$

This is obviously a cointegrated VAR with one common trend and N cointegration vectors since Φ_Z has exactly one eigenvalue equal to unity.

The vector-error-correction (VEC) representation is

$$\Delta Z_t = \mu_Z + \alpha\beta' Z_{t-1} + v_t, \quad \alpha = \begin{pmatrix} 0 \\ \Phi - I_N \end{pmatrix}, \ \beta = \begin{pmatrix} -\gamma' \\ I_N \end{pmatrix},$$

where $\Phi_Z - I_{N+1} = \alpha\beta'$. Since the intercept can be written as $\mu_Z = -\alpha\beta'(0, \bar{P}')'$, the cointegration residual is $\beta'\left(Z_{t-1} - (0, \bar{P}')'\right) = \tilde{P}_{t-1}$. It Granger-causes the yield factors according to the last N equations of the VEC representation, which can be written as

$$\Delta P_t = (\Phi - I_N)\tilde{P}_{t-1} + u_t.$$

That is, deviations of the time-t yield factors from their equilibrium, $P_t = \bar{P} + \gamma\tau_t$, are reduced over time by future changes in yields.

3.2 Risk-neutral dynamics and stochastic discount factor

The dynamic system under the risk-neutral measure $\mathbb{Q}$ is

$$P_t = \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} P_{t-1} + u_t^{\mathbb{Q}}, \tag{3}$$

where the innovations $u_t^{\mathbb{Q}}$ are *iid* normal with covariance matrix Ω. We assume that under $\mathbb{Q}$ the yield factors follow a *stationary* VAR. This specification avoids the counterfactual implication of a unit root under $\mathbb{Q}$ that yields and forward rates diverge to minus infinity with maturity—in that case the conditional variance of future short rates, and hence the convexity in yields, would be unbounded, as discussed below in 3.3.

From the short rate equation

$$i_t = \delta_0 + \delta_1' P_t. \tag{4}$$

and the $\mathbb{Q}$-dynamics in (3) it follows that the long-run mean of the short rate under $\mathbb{Q}$ is constant and given by

$$E^{\mathbb{Q}}(i_t) = \delta_0 + \delta_1' E^{\mathbb{Q}}(P_t) = \delta_0 + \delta_1'(I_N - \Phi^{\mathbb{Q}})^{-1}\mu^{\mathbb{Q}}.$$

It further follows from our specification that the trend τ_t is unspanned by yields, meaning

that there is no deterministic mapping from P_t (or model-implied yields) to τ_t. Formally, the requirement for τ_t to be an unspanned risk factor is

$$E^{\mathbb{Q}}(i_{t+h}|P_t, \tau_t) = E^{\mathbb{Q}}(i_{t+h}|P_t), \quad \text{for all } h,$$

see Joslin et al. (2014). Since neither the short rate i_t (see equation 4) nor expectations of future yield factors P_{t+h} (see equation 3) depend on τ_t once we condition on P_t, this requirement is trivially satisfied. Note that the $\mathbb{Q}$-dynamics of τ_t are not identified because yields and bond prices are not sensitive to movements in τ_t. If we had allowed for a unit eigenvalue of $\Phi^{\mathbb{Q}}$, so that the long-run risk-neutral mean of the short rate would be time-varying, this $\mathbb{Q}$-endpoint would be spanned by P_t and not deterministically related to τ_t.

Two key features of our risk-neutral distribution for P_t are its stationarity and its independence of τ_t. We now explain how these risk-neutral dynamics are reconciled with the non-stationary unit root dynamics under the real-world probability measure that are central to our paper. As in Joslin et al. (2014), the stochastic discount factor (SDF) for the bond market is the projection of the economy-wide SDF on the risk factors driving bond prices, P_t. Our specification implies that the SDF is exponentially affine:

$$M_{t+1} = \exp(-i_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}) \tag{5}$$

with $\varepsilon_t = \Omega^{-1/2}u_t$, where $\Omega^{1/2}$ is the Cholesky decomposition of Ω.[2] Shocks to τ_t do not affect the SDF because these risks are unspanned. The risk prices λ_t depend on all risk factors—including τ_t—according to the affine function

$$\lambda_t = \Omega^{-1/2}(\lambda_0 + \lambda_1 Z_t), \tag{6}$$

where λ_0 is a N-vector and λ_1 is a $N \times (N+1)$ matrix of risk sensitivities. To see how this SDF reconciles our real-world and risk-neutral dynamic specifications, and to find λ_0 and λ_1, we derive the risk-neutral distribution of P_{t+1} conditional on P_t and τ_t from the real-world distribution and the SDF, using standard tools (Ang and Piazzesi, 2003; Le et al., 2010; Bauer, 2018). The conditional Laplace-transform (that is, the conditional moment-generating function) under $\mathbb{Q}$ is

$$\begin{aligned}
E^{\mathbb{Q}}(\exp(u'P_{t+1})|Z_t) &= E(\exp(u'P_{t+1} - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1})|Z_t) \\
&= E(\exp(u'(\mu_P + \Phi_{PZ}Z_t + \Omega^{1/2}\varepsilon_{t+1}) - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1})|Z_t) \\
&= \exp(u'(\mu_P + \Phi_{PZ}Z_t) + \frac{1}{2}u'\Omega u - u'\Omega^{1/2}\lambda_t) \\
&= \exp(u'(\mu_P - \lambda_0 + (\Phi_{PZ} - \lambda_1)Z_t) + \frac{1}{2}u'\Omega u).
\end{aligned}$$

The first equality changes the probability measure using our SDF and the fact that for any t+1-

[2] That is, $\Omega^{1/2}$ is lower triangular and satisfies $\Omega = \Omega^{1/2}(\Omega^{1/2})'$.

measurable random variable X_{t+1} we have $E_t^{\mathbb{Q}}(X_{t+1}) = E_t(M_{t+1}X_{t+1})/E_t(M_{t+1})$.[3] The second equality plugs in the dynamic process for P_{t+1} from equation (2), with μ_P denoting the last N elements of μ_Z and Φ_{PZ} denoting the last N rows of Φ_Z. The third equality uses the mean of a log-normal distribution. The fourth and last equality substitutes for $\Omega^{1/2}\lambda_t$ from equation (6). Since the moment-generating function for a multivariate normal distribution with mean μ and covariance matrix Σ is $\exp(u'\mu + 0.5u'\Sigma'u)$ we see that under $\mathbb{Q}$ the conditional distribution of P_{t+1} is Gaussian with mean $\mu_p - \lambda_0 + (\Phi_{PZ} - \lambda_1)Z_t$ and covariance matrix Ω. That is, $E_t^{\mathbb{Q}}P_{t+1} = E_tP_{t+1} - \lambda_0 - \lambda_1 Z_t = E_tP_{t+1} - \Omega^{1/2}\lambda_t$. In light of equations (3) and (2) we conclude that

$$\lambda_0 = \mu_P - \mu^{\mathbb{Q}} = (I_n - \Phi)\bar{P} - \mu^{\mathbb{Q}} \tag{7}$$

and

$$\lambda_1 = \Phi_{PZ} - \left[0_{N\times 1}, \Phi^{\mathbb{Q}}\right] = \left[(I_N - \Phi)\gamma, \Phi - \Phi^{\mathbb{Q}}\right]. \tag{8}$$

By construction, the trend τ_t does not affect the risk-neutral conditional expectations of future P_{t+1}. Thus, the risk sensitivities for τ_t (in the first column of λ_1) have to exactly equal the sensitivities of real-world conditional expectations of P_{t+1} to τ_t (in the first column of Φ_{PZ}). Some authors (e.g. Duffee, 2011; Bauer and Rudebusch, 2017) speak of "knife-edge" restrictions when constraints lead to such cancellations and render certain factors unspanned. These restrictions imply that movements in a factor that is unspanned (τ_t in our case) lead to changes in term premia that exactly offset the changes in expectations, leaving yields unchanged. The risk sensitivities for P_t (in the last three columns of λ_1) are unconstrained, except that the implied $\Phi^{\mathbb{Q}}$ must have eigenvalues less than unity in absolute value.

3.3 Affine loadings

Prices of zero-coupon bonds are exponentially affine. Specifically, if $p_t^{(n)}$ denotes the log-price of an n-period bond, $p_t^{(n)} = \mathcal{A}_n + \mathcal{B}_n'P_t$, and the coefficients can be found using the pricing equation $\exp\left(p_t^{(n+1)}\right) = E_t^{\mathbb{Q}}\exp\left(-i_t p_{t+1}^{(n)}\right)$. They follow the usual recursions (e.g., Ang and Piazzesi, 2003):

$$\mathcal{A}_{n+1} = \mathcal{A}_n + \mathcal{B}_n'\mu^{\mathbb{Q}} + \frac{1}{2}\mathcal{B}_n'\Omega\mathcal{B}_n - \delta_0 \tag{9}$$

$$\mathcal{B}_{n+1} = (\Phi^{\mathbb{Q}})'\mathcal{B}_n - \delta_1 \tag{10}$$

[3] The density of the risk-neutral measure with respect to the real-world measure, also known as the Radon-Nikodym derivative, is

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{M_{t+1}}{E_t(M_{t+1})} = \exp\left(-\frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}\right).$$

with initial conditions $\mathcal{A}_0 = 0$ and $\mathcal{B}_0 = 0$.[4] Yields are affine functions of the factors:

$$y_t^{(n)} = A_n + B_n' P_t \quad A_n = -\frac{1}{n}\mathcal{A}_n, \quad B_n = -\frac{1}{n}\mathcal{B}_n.$$

If we denote the vector of J yields that are included in our model as Y_t, we can write $Y_t = A + BP_t$ for J-vector $A = (A_{n_1}, \ldots, A_{n_J})'$, and $J \times N$ matrix $B = (B_{n_1}, \ldots, B_{n_J})'$.

It is often useful to work with forward rates instead of yields, and we use them here to develop intuition about the role of the trend τ_t. We consider one-period forward rates $f_t^{(n)} = p_t^{(n)} - p_t^{(n+1)}$. Yields are averages of forward rates, $y_t^{(n)} = n^{-1}\sum_{j=0}^{n-1} f_t^{(n)}$. Forward rates depend on the risk factors as follows:

$$\begin{aligned} f_t^{(n)} &= \mathcal{A}_n - \mathcal{A}_{n+1} + (\mathcal{B}_n - \mathcal{B}_{n+1})' P_t \\ &= \underbrace{-\frac{1}{2}\mathcal{B}_n'\Omega\mathcal{B}_n}_{\text{convexity}} + \underbrace{\delta_0 - \mathcal{B}_n'\mu^{\mathbb{Q}} + \delta_1'(\Phi^{\mathbb{Q}})^n P_t}_{\text{expected short rate under } \mathbb{Q}} \\ &= \text{constant} + \delta_1'(\Phi^{\mathbb{Q}})^n \gamma\tau_t + \delta_1'(\Phi^{\mathbb{Q}})^n \tilde{P}_t \end{aligned} \tag{11}$$

The first equality simply uses the definition of the forward rate and the affine form for log bond price. The second equality follows from the recursions (9)-(10). It shows that forward rates are equal to a (negative) convexity term plus risk-neutral expected future short rates, $E_t^{\mathbb{Q}} i_{t+n} = \delta_0 + \delta_1' E_t^{\mathbb{Q}} P_{t+n}$. The last equality substitutes for P_t from (1) to show the loadings of forward rates on the trend and cycle factors.

The convexity term in the expression for forward rates is $-\frac{1}{2}\mathcal{B}_n'\Omega\mathcal{B}_n$. Yield convexity is the average of forward rate convexity. Since the $\mathbb{Q}$-dynamics are stationary, $\mathcal{B}_n$ converges to a finite limit as n tends to infinity, and so does the convexity in yields and forward rates. If, by contrast, $\Phi^{\mathbb{Q}}$ had an eigenvalue on the unit circle, convexity and therefore yields and forward rates would diverge to minus infinity. This fact, discussed in Campbell et al. (1997, p. 433), is an important reason why term structure models are usually specified with stationary risk-neutral dynamics, and we follow this tradition here.

Of particular interest are the loadings of forward rates on the trend, which are $\delta_1'(\Phi^{\mathbb{Q}})^n\gamma$. Figure 3 plots these loadings for a range of maturities from zero to 400 using the estimates of the *OSE* model. For the short rate, $i_t = f_t^{(0)}$, the loading is one, since we have normalized $\delta_1'\gamma = 1$ in order to identify τ_t as the equilibrium short rate i_t^*. The loadings increase quickly to a peak near 1.8 around the five-year horizon, and then slowly decline. Because the $\mathbb{Q}$-dynamics are stationary, the loadings converge to zero with increasing maturity, as do all yield and forward rate loadings—the limiting-maturity forward rate/bond yield is a constant. Since the largest eigenvalue of $\Phi^{\mathbb{Q}}$ is below but very close to one, the forward rate loadings

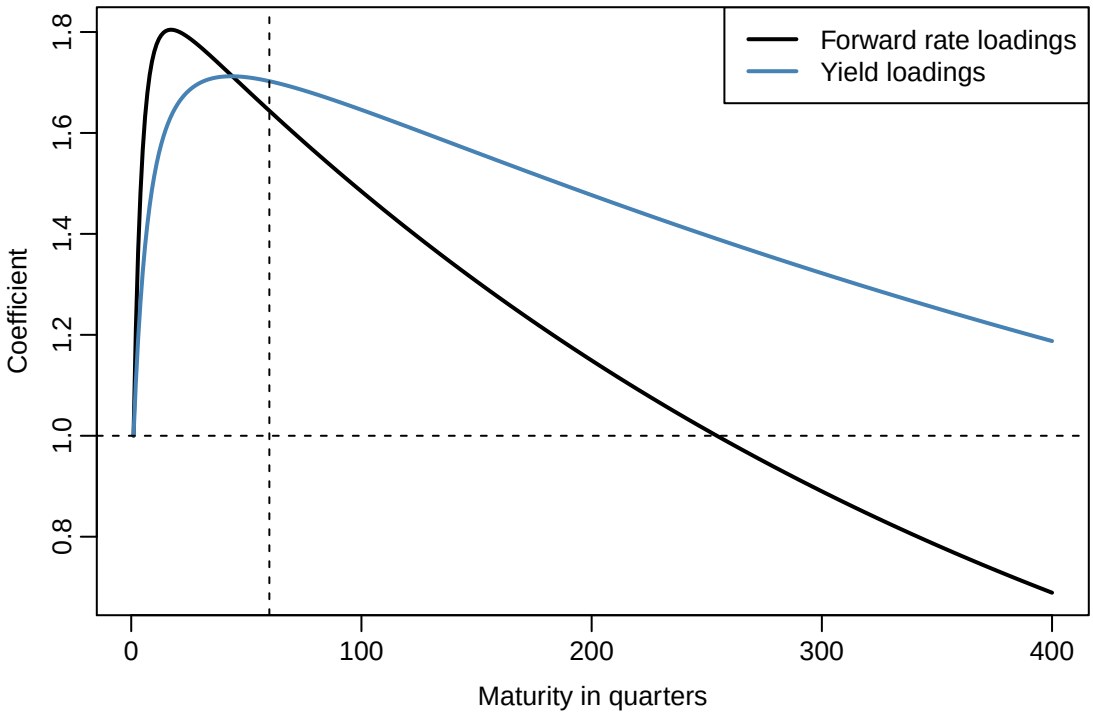
[4] The closed-form solution for $\mathcal{B}_n$ is

$$\mathcal{B}_n = -\left[\sum_{j=0}^{n-1}\left(\Phi^{\mathbb{Q}}\right)^j\right]'\delta_1 = -\left[\left(I_N - \Phi^{\mathbb{Q}}\right)^{-1}\left(I_N - (\Phi^{\mathbb{Q}})^n\right)\right]'\delta_1, \quad \lim_{j\to\infty}\mathcal{B}_n = -\left[\left(I_N - \Phi^{\mathbb{Q}}\right)^{-1}\right]'\delta_1,$$

where the second equality and the existence of the limit rely on our assumption that $\Phi^{\mathbb{Q}}$ has no eigenvalues on the unit circle.

decline only very gradually, and don't fall below one until $n = 255$.

Figure 3: Loadings of forward rates and yields on i_t^*



Loadings of forward rates and yields on the long-run trend $\tau_t = i_t^*$, based on *OSE* model estimates. Dashed vertical line indicates the longest observed maturity in our sample (15 years).

Loadings of yields on the trend are also shown in Figure 3. Since these are averages of forward rate loadings, they naturally peak later, around the ten-year maturity, and decline even more slowly thereafter. For the yields included in our data, the loadings on τ_t are substantially *above one*, consistent with a positive effect of the trend on the term premium to be discussed below. Only for extremely long maturities do the loadings approach one and ultimately decline towards zero, the infinite-maturity limit. In other words, stationary risk-neutral dynamics imply that infinite-maturity yields are constant, but our estimates imply that at the maturities of practical interest yields are sufficiently volatile and vary more than one-for-one in response to movements in the long-run trend.

If instead the model had a unit root under $\mathbb{Q}$, it would be possible to have yield loadings on i_t^* that have a positive limit, so that loadings of the term premium could have a limit other than minus one. However, such a specification would have other undesirable features, namely divergent convexity. In the end, either specification can be consistent with observed bond yields, though neither seems wholly satisfactory in the infinite-maturity limit.

3.4 JSZ normalization

We follow Joslin et al. (2014), who also estimate a model with unspanned risk factors, in imposing the normalization of Joslin et al. (2011)(JSZ) on the yield factors P_t. This normalization is convenient because it imposes all identifying restrictions on the $\mathbb{Q}$-dynamics, so that the real-world dynamics can be estimated without restrictions.

The JSZ normalization works as follows: First, start with N latent factors X_t and impose normalizing restrictions on their $\mathbb{Q}$-dynamics. Second, use the affine mapping

$$P_t = WY_t = WA_X + WB_X X_t, \tag{12}$$

where A_X and B_X are the affine yield loadings on X_t, to obtain the $\mathbb{Q}$-parameters and yield loadings for P_t in terms of the parameters/loadings of X_t.

The normalizations are (i) $i_t = \iota' X_t$, (ii) $\mu_X^{\mathbb{Q}} = (k^{\mathbb{Q}}, 0, 0)'$ and (iii) $\Phi_X^{\mathbb{Q}} = diag(\lambda^{\mathbb{Q}})$, with real, distinct, descending-ordered diagonal elements $\lambda^{\mathbb{Q}}$ that are assumed to be less than one in absolute value.[5] The loadings of yields on P_t, A and B, are easily calculated from A_X and B_X using the mapping in equation (12). Alternatively, one can first calculate the "rotated" parameters $\mu^{\mathbb{Q}}$, $\Phi^{\mathbb{Q}}$, δ_0 and δ_1, and then use the recursions (9)-(10) to calculate the loadings A and B. Importantly, the yield loadings only depend on the parameters $k^{\mathbb{Q}}$, $\lambda^{\mathbb{Q}}$ and Ω. The normalization ensures the consistency conditions $WA = 0$ and $WB = I_N$ so that we indeed have $WY_t = P_t$.

3.5 Excess bond returns

Model-implied excess bond returns are

$$\begin{aligned} rx_{t+1}^{(n)} &= p_{t+1}^{(n-1)} - p_t^{(n)} - i_t = \mathcal{A}_{n-1} + \mathcal{B}_{n-1}' P_{t+1} - \mathcal{A}_n - \mathcal{B}_n' P_t - \delta_0 - \delta_1' P_t \\ &= -\frac{1}{2}\mathcal{B}_{n-1}'\Omega\mathcal{B}_{n-1} + \mathcal{B}_{n-1}'\left(P_{t+1} - \mu^{\mathbb{Q}} - \Phi^{\mathbb{Q}} P_t\right) \\ &= -\frac{1}{2}\mathcal{B}_{n-1}'\Omega\mathcal{B}_{n-1} + \mathcal{B}_{n-1}'\left(E_t P_{t+1} - E_t^{\mathbb{Q}} P_{t+1}\right) + \mathcal{B}_{n-1}' u_{t+1}. \end{aligned}$$

where the second line uses the coefficient recursions (9)-(10) and the third line uses conditional expecations of P_{t+1} under the real-world and risk-neutral mesures (from equations 2 and 3). Expected excess returns, adjusted for the convexity term, are thus

$$\begin{aligned} E_t rx_{t+1}^{(n)} + \frac{1}{2} Var_t rx_{t+1}^{(n)} &= \mathcal{B}_{n-1}'\Omega^{1/2}\lambda_t = \mathcal{B}_{n-1}'(\lambda_0 + \lambda_1 Z_t) \\ &= \mathcal{B}_{n-1}'\left[\lambda_0 + (I_N - \Phi)\gamma\tau_t + (\Phi - \Phi^{\mathbb{Q}})P_t\right] \qquad (13) \\ &= \mathcal{B}_{n-1}'\left[\lambda_0 + (\Phi - \Phi^{\mathbb{Q}})\bar{P} + (I_N - \Phi^{\mathbb{Q}})\gamma\tau_t + (\Phi - \Phi^{\mathbb{Q}})\tilde{P}_t\right], \qquad (14) \end{aligned}$$

where the first equality uses the fact that the prices of risk capture the difference between real-world and risk-neutral conditional expectations, the second line uses the risk sensitivities

[5] JSZ show that it is possible to allow for repeated and complex eigenvalues.

given in (8), and the third line substitutes for P_t to show the loadings of expected excess returns on the ternd and cycle factors.

Again, the loadings on the trend τ_t are most interesting. Equation (13) shows that when holding the yield factors P_t constant, the loadings on τ_t are $\mathcal{B}'_{n-1}(I_N - \Phi)\gamma$. Our model estimates imply substantially negative values for these loadings, across all n, in line with the regression estimates in Section 3.2 of the paper. An increase in the trend *ceteris paribus* means that yields will adjust upward causing negative excess returns for bond holders. This shows the crucial role of the unspanned trend for bond risk premia.

Equation (14) shows the separate roles of trends and cycles for bond risk premia. If we hold the yield cycles $\tilde{P}_t$ constant, the loadings on τ_t are $\mathcal{B}'_{n-1}(I_N - \Phi^{\mathbb{Q}})\gamma$. These loadings tend to be an order of magnitude smaller (in absolute value) than the loadings obtained with yields—consistent with the findings in Section 3.2 of the paper that predictive regressions including only detrended yields capture most of the predictability of excess bond returns. These loadings are positive because risk premia are positively related to movements in the trend, based on our estimates.[6] Additional unreported results show that predictive regressions with both detrended yields and trend proxies lead to coefficients on the trend that are insignificantly different from both zero and the model-based loadings, and thus consistent with the implications of the model.

3.6 Term premium

The term premium is the difference of a long-term interest rate and that rate's expectations component. Section 2 of the paper showed for simplicity the most common definition of the term premium and the expectations component:

$$y_t^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} E_t i_{t+j} + TP_t^{(n)}, \tag{15}$$

so that the term premium is the excess log-return of holding an n-period bond to maturity instead of rolling over short-term bonds. But in the context of no-arbitrage term structure models, it is advantageous to instead use "risk-neutral rates" which, like yields, account for the convexity of bond prices. Specifically, the risk-neutral yield on an n-period bond is

$$\tilde{y}_t^{(n)} = -\log E_t \exp\left(-\sum_{j=0}^{n-1} i_{t+j}\right)/n,$$

which differs from the average expected future short rate $\sum_{j=0}^{n-1} E_t i_{t+j}/n$ by a Jensen's inequality term. The risk-neutral yields are affine in Z_t, and the loadings are calculated using recursions similar to those in equations (9)- (10), with μ_Z and Φ_Z replacing $\mu^{\mathbb{Q}}$ and $\Phi^{\mathbb{Q}}$, Ω_v replacing Ω, and an $N+1$ vector $(0, \delta_1')'$ replacing δ_1. Crucially, risk-neutral yields depend on

[6] It is only possible for the trend to drop out completely from expected excess returns and term premia if the model also contained a unit root under $\mathbb{Q}$. In that case, certain parameter restrictions imply zero loadings for all risk premia on the trend, and equivalently unit loadings of all yields and forward rates on the trend.

both τ_t and P_t, since τ_t affects real-world expectations of future short rates. Using risk-neutral yields, the term premium on an n-period bond is then defined as

$$\widetilde{TP}_t^{(n)} = y_t^{(n)} - \tilde{y}_t^{(n)}.$$

Because the real-world dynamics feature a unit root, risk-neutral rates diverge to minus infinity (and the term premium to infinity) as maturity increases. This issue is neither of theoretical nor practical concern: First, no tradeable securities have payoffs tied to these quantities. Second, at the maturities we focus on, due to the high persistence of the risk-neutral dynamics, the convexity in actual and risk-neutral yields is quite similar, and estimates of the term premium are not noticeably affected by convexity.

We again turn to forward rates for intuition. Using the notation $\tilde{p}_t^{(n)} = \tilde{\mathcal{A}}_n + \tilde{\mathcal{B}}_n Z_t$ for log risk-neutral bond prices and their loadings, the risk-neutral forward rate is

$$\begin{aligned}
\tilde{f}_t^{(n)} &= \tilde{p}_t^{(n)} - \tilde{p}_t^{(n+1)} = \tilde{\mathcal{A}}_n - \tilde{\mathcal{A}}_{n+1} + (\tilde{\mathcal{B}}_n - \tilde{\mathcal{B}}_{n+1})' Z_t \\
&= -\frac{1}{2}\tilde{\mathcal{B}}_n' \Omega_v \tilde{\mathcal{B}}_n + \delta_0 - \tilde{\mathcal{B}}_n' \mu_Z + (0, \delta_1') \Phi_Z^n Z_t \\
&= -\frac{1}{2}\tilde{\mathcal{B}}_n' \Omega_v \tilde{\mathcal{B}}_n + \delta_0 + (0, \delta_1') E_t Z_{t+n} \\
&= -\frac{1}{2}\tilde{\mathcal{B}}_n' \Omega_v \tilde{\mathcal{B}}_n + \delta_0 + \delta_1' (\bar{P} + \gamma \tau_t + E_t \tilde{P}_{t+n}) \\
&= -\frac{1}{2}\tilde{\mathcal{B}}_n' \Omega_v \tilde{\mathcal{B}}_n + \tau_t + \delta_1' \Phi^n \tilde{P}_t, \qquad (16)
\end{aligned}$$

where the second line follows from the recursions for the risk-neutral loadings, the third line recognizes and replaces the conditional expectations for Z_{t+n}, the fourth line uses the fact that $(0, \delta_1') E_t Z_{t+n} = \delta_1' E_t P_{t+n}$ and substitutes for $E_t P_{t+n}$, and the last line uses $\delta_0 + \delta_1' \bar{P} = 0$, $\delta_1' \gamma = 1$ and $E_t \tilde{P}_{t+n} = \Phi^n \tilde{P}_t$. These derivations clearly show that $\tilde{f}_t^{(n)} = E_t i_{t+n} + \text{convexity}$, and that these expectations are of course affected one-for-one by the trend $\tau_t = i_t^*$. Equations (11) and (16) imply that the forward term premium is

$$ftp_t^{(n)} = f_t^{(n)} - \tilde{f}_t^{(n)} = \text{constant} + \delta_1' \left[(\Phi^{\mathbb{Q}})^n - I_N \right] \gamma \tau_t + \delta_1' \left[(\Phi^{\mathbb{Q}})^n - \Phi^n \right] \tilde{P}_t. \qquad (17)$$

The loadings on τ_t have a limit of -1 as n tends to infinity: the limiting-maturity forward rate is constant, as the limiting-maturity forward term premium exactly offsets movements in the limiting-maturity risk-neutral rate. However, for the maturities relevant in practice our model estimates imply that the loadings of $ftp_t^{(n)}$ on τ_t are positive. That is, for the interest rate maturities we focus on (up to 15 years), forward rates have loadings on τ_t that are above one, as discussed above in 3.3, because forward term premia have loadings on τ_t that are above zero. Since yields are just averages of forward rates, the same holds for yields and the yield term premium: yield loadings on the trend are larger than one, consistent with the cointegration regressions reported in Section 3.1 of the paper, and term premia have loadings on the trend that are positive.

3.7 Estimation with observed shifting endpoint

In this implementation of our shifting-endpoint DTSM, we use an empirical proxy for $\tau_t = i_t^*$ and treat this state variable as observable. Furthermore, P_t is assumed to be observable, in which case, $WY_t^o = WY_t$ or $We_t = 0$. Thus, there are effectively only $J - N$ independent measurement errors (see Joslin et al., 2011). We use the loadings for the first three principal components of observed yields to construct the three factors P_t, as is commonly done (e.g., Joslin et al., 2011). That is, the rows of W contain the first three eigenvectors of the sample covariance matrix of Y_t^o.

Because all of the state variables in Z_t are observable, estimation is quite simple: Not only can the model's likelihood function be evaluated without the Kalman filter, we can also concentrate out the parameters Φ and σ_e^2 from the likelihood. That is, for given values of $k^{\mathbb{Q}}$, $\lambda^{\mathbb{Q}}$, γ, $\bar{P}$, Ω, and σ_η^2, we can analytically obtain the values of Φ and σ_e^2 that maximize the likelihood function. Computationally then, we maximize the log-likelihood over 15 parameters ($k^{\mathbb{Q}}$, 3 in $\lambda^{\mathbb{Q}}$, 2 in γ, 2 in $\bar{P}$, 6 in Ω, and σ_η^2) instead of over 25 parameters.

Alternatively, it could be assumed that the state variables are unobservable, which would require estimation of the state-space form of the model using the Kalman filter. We used this alternative method to estimate our model and obtained results essentially identical to those reported for the version with observable state variables. Still, in some applications the Kalman filter may necessary. For example, it can accommodate missing observations or include multiple measurement equations to pin down i_t^* using several different proxies. Using the Kalman filter is straightforward but computationally intensive (although excellent starting values for the model parameters can be obtained from a first-step estimation with observable state variables.)

3.8 Bayesian estimation with estimated shifting endpoint

For our *ESE* model we use a Bayesian framework with a block-wise MCMC algorithm for estimation. We choose the first yield factor in P_t to be the short rate (i.e., the three-month yield, as the model is estimated on quarterly data). In this case, our normalization of the trend implies that $\gamma_1 = 1$ and $\bar{P}_1 = 0$, similar to the standard identification in multivariate unobserved components models (Harvey, 1990). This choice of normalization simplifies our MCMC estimation as it implies less prior dependence across parameter blocks than would be the case with principal components as yield factors. We choose the other two yield factors to be the two- and ten-year yields.

The prior distributions are as follows:

- For the trend innovation variance, σ_η^2, we assume an inverse-gamma distribution, specifically, $IG(\alpha_\eta/2, \delta_\eta/2)$ with $\alpha_\eta = 100$ and $\delta_\eta = 0.06^2/400(\alpha_\eta + 2)$. This parametrization implies a tight prior distribution around a mode of $0.06^2/400$, which corresponds to a standard deviation of six percent for changes in τ_t over 100 years. We view this prior as a conservative choice that is justified in large part by consideration of the slow-moving macroeconomic drivers underlying π_t^* and r_t^* and by the circumscribed evolution of the various available estimates of those macro trends. Our prior for σ_η^2 is very similar to

DGGT, who also use their model's prior to limit "the amount of variation that it attributes to the trends" (DGGT, p. 249).

- For Φ and $\widetilde{\Omega}$, we specify a Minnesota-type independent normal/inverse-Wishart. We use a simple data-based procedure to scale our prior distributions: We regress P_t on the observable i_t^* proxy, and collect the standard deviations of the residuals in a vector s. The prior for $\widetilde{\Omega}$ is inverse-Wishart, $IW(\kappa, \Psi)$ with $\kappa = N + 2$ and $\Psi = diag(s)$. The prior mean for $\widetilde{\Omega}$ is $\Psi/(\kappa - N - 1) = \Psi$, and the prior mode is $\Psi/(\kappa + N + 1)$. That is, we use the lowest value of κ for the mean to exist, so that the distribution is very disperse. The prior for Φ is a Minnessota prior, but centered around a zero matrix. The prior standard deviation for the (i, j)-element of Φ is $\lambda s_i / s_j$. For the hyperparameter controlling the tightness we use $\lambda = 0.2$ like DGGT and others. We also restrict the eigenvalues of Φ to be inside the unit circle.

- The priors for the unrestricted (second and third) elements of γ are independent Gaussian distributions with means of one and variances of 0.2. The unrestricted elements of $\bar{P}$ have independent Gaussian priors with means of zero and variances of 0.05^2. These mildly informative priors help improve the efficiency of the MCMC sampler without imposing too much prior information.

- For the measurement error variance σ_e^2, we use an inverse-gamma distribution with $\alpha_e = 4$ and $\delta_e = 0.001^2(\alpha_e + 2)$, implying a prior mode for the variance of 0.001^2 which corresponds to a standard deviation of 10 basis points. This prior mode is motivated by the fact that sufficiently flexible DTSMs can generally achieve a good fit to observed yields with only about 5-10 basis points root-mean-squared error.

- The priors for the remaining parameters are completely uninformative.

We use a state-space formulation in terms of the state variables $\tilde{Z}_t = (\tau_t, \tilde{P}_t')'$, which is of course equivalent to using τ_t and P_t. The measurement equation is

$$Y_t^o = A + B\bar{P} + B\gamma\tau_t + B\tilde{P}_t,$$

and the state equation is

$$\tilde{Z}_t = \begin{pmatrix} 1 & 0 \\ 0 & \Phi \end{pmatrix} \tilde{Z}_{t-1} + \begin{pmatrix} \eta_t \\ \tilde{u}_t \end{pmatrix}.$$

Our MCMC algorithm is a block-wise Metropolis-Hastings (M-H) sampler (Chib and Greenberg, 1995). First, note that the log-likelihood is the sum of the cross-sectional log-likelihood (from the measurement equation) and the "dynamic" log-likelihood (from the transition equation). The former, called the (log of the) $\mathbb{Q}$-likelihood by Joslin et al. (2011), for observation t is

$$\log f(Y_t^o | P_t, k^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Omega, \sigma_e^2) \propto -\frac{1}{2} |Y_t^o - A - BP_t| / \sigma_e^2,$$

where A and B depend on $(k^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Omega)$ and $|\cdot|$ denotes the $\mathcal{L}^2$ (Euclidean) norm. The latter,

the $\mathbb{P}$-likelihood, for observation t is

$$\log f(Z_t|Z_{t-1}, \bar{P}, \gamma, \Phi, \Omega, \sigma_\eta^2) = -\frac{1}{2}\left(\log|\widetilde{\Omega}| + \tilde{u}_t'\widetilde{\Omega}^{-1}\tilde{u}_t + \log(\sigma_\eta^2) + \eta_t^2/\sigma_\eta^2\right),$$

where $\widetilde{\Omega} = \Omega - \gamma\gamma'\sigma_\eta^2$, $\tilde{u}_t = \tilde{P}_t - \Phi\tilde{P}_{t-1}$, $\tilde{P}_t = P_t - \gamma\tau_t - \bar{P}$, and $\eta_t = \Delta\tau_t$.

The first block is the sampling of the state variables using the simulation smoother of Durbin and Koopman (2002), and we condition on Z_t when drawing the other blocks. The parameter blocks are (1) $(k^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$, (2) Ω, (3) g (which determines γ), (4) p (which determines $\bar{P}$), (5) σ_η^2, (6) Φ, and (7) σ_e^2. The sampling of Φ and σ_e^2 is straightforward: Gibbs steps can be employed since the full conditional posterior is available. For Φ, the draw is only accepted if the matrix has all eigenvalues inside the unit circle, to ensure stationarity of $\tilde{P}_t$. For blocks (1)-(5), the conditional posterior distributions cannot be sampled; therefore, we must use M-H steps. Only the parameters in blocks (1) and (2) affect the loadings A and B and hence the $\mathbb{Q}$-likelihood. The parameters in blocks (3)-(5) only affect the $\mathbb{P}$-likelihood, which saves on computational costs. For our M-H steps, we use tailored independence proposal distributions similar to Chib and Ergashev (2009). Specifically, conditional on all other parameters and the state variables, we use numerical optimization to find the mode and Hessian of the conditional posterior distribution for a given parameter block. Our proposal distribution is then a multivariate t-distribution (with five degrees of freedom) that is centered around this mode and has a covariance matrix equal to the inverse of this Hessian. This way of constructing M-H proposal distributions has the great benefit that no fine-tuning of the proposal distributions is necessary, and the efficiency of the sampler is generally much better than using random walk proposals (Chib and Ergashev, 2009).

We run our sampler to obtain five different chains from random starting values (around the posterior mode that we find numerically), in each case obtaining 100,000 observations and discarding the first half as burn-in samples. This gives us an MCMC sample of 250,000 iterations for our posterior analysis.

References

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, 745–787.

Bauer, M. D. (2018): "Restrictions on Risk Prices in Dynamic Term Structure Models," *Journal of Business & Economic Statistics*, 36, 196–211.

Bauer, M. D. and J. D. Hamilton (2018): "Robust Bond Risk Premia," *Review of Financial Studies*, 31, 399–448.

Bauer, M. D. and G. D. Rudebusch (2017): "Resolving the Spanning Puzzle in Macro-Finance Term Structure Models," *Review of Finance*, 21, 511–553.

Brayton, F. and P. A. Tinsley (1996): "A Guide to FRB/US: A Macroeconomic Model of

the United States," Finance and Economics Discussion Series 1996-42, Board of Governors of the Federal Reserve System.

Campbell, J. Y., A. W.-C. Lo, and A. C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.

Chib, S. and B. Ergashev (2009): "Analysis of Multifactor Affine Yield Curve Models," *Journal of the American Statistical Association*, 104, 1324–1337.

Chib, S. and E. Greenberg (1995): "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327–335.

Christensen, J. H. E. and G. D. Rudebusch (2019): "A New Normal for Interest Rates? Evidence from Inflation-Indexed Debt," *Review of Economics and Statistics*, 101.

Del Negro, M., D. Giannone, M. P. Giannoni, and A. Tambalotti (2017): "Safety, Liquidity, and the Natural Rate of Interest," *Brookings Papers on Economic Activity*, 2017, 235–316.

Duffee, G. R. (2011): "Information In (and Not In) the Term Structure," *Review of Financial Studies*, 24, 2895–2934.

Durbin, J. and S. J. Koopman (2002): "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89, 603–616.

Federal Reserve Board of Governors (2019a): "3-Month and 6-Month Treasury Bill rates [TB3MS and TB6MS]," Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/TB3MS and https://fred.stlouisfed.org/series/TB6MS (accessed December 27, 2019).

——— (2019b): "Data for FRB/US Model," https://www.federalreserve.gov/econres/us-models-package.htm (accessed December 27, 2019).

Fiorentini, G., A. Galesi, G. Perez-Quiros, and E. Sentana (2018): "The Rise and Fall of the Natural Interest Rate," Documentos de Trabajo 1822, Banco de Espana.

Harvey, A. C. (1990): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

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.

Hyslop, D. R. and G. W. Imbens (2001): "Bias from Classical and Other Forms of Measurement Error," *Journal of Business & Economic Statistics*, 19, 475–481.

Johannsen, B. K. and E. Mertens (2016): "The Expected Real Interest Rate in the Long Run: Time Series Evidence with the Effective Lower Bound," Feds notes, Board of Governors of the Federal Reserve System.

——— (2018): "A Time Series Model of Interest Rates with the Effective Lower Bound," BIS Working Papers 715, Bank for International Settlements.

Joslin, S., M. Priebsch, and K. J. Singleton (2014): "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance*, 69, 1197–1233.

Joslin, S., K. J. Singleton, and H. Zhu (2011): "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, 24, 926–970.

Kiley, M. T. (2015): "What Can the Data Tell Us about the Equilibrium Real Interest Rate," Finance and Economics Discussion Series 2015-077, Board of Governors of the Federal Reserve System.

Laubach, T. and J. C. Williams (2003): "Measuring the Natural Rate of Interest," *Review of Economics and Statistics*, 85, 1063–1070.

——— (2016): "Measuring the Natural Rate of Interest Redux," *Business Economics*, 51, 57–67.

Le, A., K. J. Singleton, and Q. Dai (2010): "Discrete-Time Affine Term Structure Models with Generalized Market Prices of Risk," *Review of Financial Studies*, 23, 2184–2227.

Lubik, T. A. and C. Matthes (2015): "Calculating the Natural Rate of Interest: A Comparison of Two Alternative Approaches," Economic Brief EB15-10, Federal Reserve Bank of Richmond.

Lunsford, K. G. and K. D. West (2019): "Some Evidence on Secular Drivers of US Safe Real Rates," *American Economic Journal: Macroeconomics*, 11, 113–139.

Mankiw, N. G. and M. D. Shapiro (1986): "News or Noise? An Analysis of GNP Revisions," Working Paper 1939, National Bureau of Economic Research.

Müller, U. K. and M. W. Watson (2013): "Low-Frequency Robust Cointegration Testing," *Journal of Econometrics*, 174, 66–81.

Ng, S. and P. Perron (1995): "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268–281.

Rudebusch, G. D. and L. E. O. Svensson (1999): "Policy Rules for Inflation Targeting," in *Monetary Policy Rules*, University of Chicago Press, 203–262.

Stock, J. H. and M. W. Watson (1988): "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

U.S. Bureau of Economic Analysis (2019): "Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index) [PCEPILFE]," Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/PCEPILFE (accessed November 30, 2019).

WATSON, M. W. (1986): "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, 18, 49–75.