Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Forecasting and Trading Monetary Policy Effects on the Riskless Yield Curve with Regime Switching Nelson-Siegel Models

Massimo Guidolin and Manuela Pedio

Working Paper n. 639

This Version: January, 2019

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Forecasting and Trading Monetary Policy Effects on the Riskless Yield Curve with Regime Switching Nelson-Siegel Models[♣]

Massimo Guidolin
Bocconi University, Baffi-CAREFIN Centre, and IGIER

Manuela Pedio
Baffi-CAREFIN Centre and Bicocca University, Milan

January 2019

Abstract

We use monthly data on the US riskless yield curve for a 1982-2015 sample to show that mixing simple regime switching dynamics with Nelson-Siegel factor forecasts from time series models extended to encompass variables that summarize the state of monetary policy, leads to superior predictive accuracy. Such spread in forecasting power turns out to be statistically significant even controlling for parameter uncertainty and sample variation. Exploiting regimes, we obtain evidence that the increase in predictive accuracy is stronger during the Great Financial Crisis in 2007-2009, when monetary policy underwent a significant, sudden shift. Although more caution applies when transaction costs are accounted for, we also report that the increase in predictive power owed to the combination of regimes and of monetary variables that capture the stance of unconventional monetary policies is tradeable. We devise and test butterfly strategies that trade on the basis of the forecasts from the models and obtain evidence of risk-adjusted profits both per se and in comparisons to simpler models.

Key words: Term structure of interest rates, Dynamic Nelson-Siegel factors, regime switching, butterfly strategies, unconventional monetary policy.

1. INTRODUCTION

Understanding and forecasting the dynamics of the shape of the yield curve is crucial for many tasks, from pricing and valuation, to risk management and portfolio allocation. In the past thirty

[♣] We would like to thank Luca Benzoni, Yunjong Eo, and session participants at the 12th International Conference on Computational and Financial Econometrics (CFE 2018, in Pisa) for constructive comments. Linda Di Pietro and Adriana Troiano have provided excellent research assistance. Correspondence to Massimo Guidolin, massimo.guidolin@unibocconi.it.

years, an ever expanding literature has investigated the riskless term structure using a wide variety of models and methods. The most basic problem faced by these models is to synthesize the yield behaviour across different maturities. To this purpose, Nelson-Siegel (1987, henceforth NS) provide a parsimonious and flexible factor approach to fitting the cross section of yields that has become a dominant benchmark in applied work by academics, market practitioners, and central bankers (see Bank for International Settlements, 2005; European Central Bank, 2008). Interestingly however, NS approach is self-referential: past yield curve information enters the definition of the factors (level, slope, and curvature) which are then exploited to forecast the dynamics of yield curve, typically by using the simple but robust methods first illustrated by Diebold and Li (2006), that capture any structure in the time variation in the factors. Even though since the seminal work by Diebold, Rudebusch, and Aruoba (2006), we understand how to extend NS approach to observable macroeconomic variables, the evidence on whether these may improve the forecasting power of the typical dynamic NS (henceforth, DNS) models remains somewhat mixed (see Duffee, 2011). Oddly enough, such a lack of evidence that macroeconomic information ought to allow us to predict interest rates extends to monetary indicators as well, such policy rates, the base money aggregates, and the very size and maturity composition of the central bank's balance sheet, that has been recently the key lever of the so-called balance sheet policies used worldwide to tackle the effects of the Great Financial Crisis (see, e.g., Gambacorta, Hofmann, and Peersman, 2014).

In this paper, we use monthly data on the US riskless yield curve for a 1982-2015 sample to show that mixing a simple regime switching framework with DNS forecasts from time series models extended to encompass variables that summarize the stance of monetary policy leads to superior predictive accuracy. we also investigate the economic value that the forecasts may generate in trading strategies popular with prop trading desks, focussing on short, one-month prediction horizons that have proven to be more elusive in the literature that ignores regime shifts in DNS factors (see, e.g., Hordal, Tristani, and Vestin, 2006; Moench, 2008) Such spread in forecasting power often turns out to be statistically significant even controlling for parameter uncertainty and for sample variation. Exploiting regimes, we obtain evidence that the increase in predictive accuracy is stronger during the Great Financial Crisis (GFC, 2007-2009) when the methods and scope of monetary policy underwent a significant, sudden shift, as required by unsettled financial markets and spiralling deflationary dynamics. Although more caution needs to be used and transaction costs may severely affect these results, we also report that—especially with reference to the GFC sample and when trading consists of placing bets on the predicted changes in the slope of the yield curve—the increase in forecasting power owed to the combination of regimes and of monetary variables that capture the state of quantitative easing and of other unconventional monetary policies is also tradeable, to some extent. We devise and test commonly used butterfly strategies (combinations of well-known barbell and bullet trades) that trade on the basis of the forecasts derived by our models and obtain evidence of risk-adjusted profits that are sometimes statistically significant, per se and in comparisons to simpler models.

The economic intuition of our key result stems from the plain observation that during the, GFC the US Treasury yield curve has been subject to unprecedented shifts; such dramatic twists and turns have come at a time in which the Fed has adopted unprecedented measures aimed at stabilizing financial markets and at supporting and re-inflating the economy to avoid a liquidity trap (what Borio and Disyatat, 2010, call *balance sheet policies*). It is hard to rationalize the thought that such measures—also because these were often driven by the very disorderly state of equity and bond markets in the US—may have not come to represent a major driver of the shape of the yield curve and its dynamics. Yet, if unconventional monetary policies (such as the maturity extension program, MEP, and several waves of quantitative easing, QE) came to become drivers of the yield curve in a causal sense, then it is legitimate to expect that one or more observable quantities capturing the stance of monetary policy during the GFC may have come to predict its movements. This is exactly what we test in this paper, finding that—especially during the GFC but to some extend also between 2009 and 2014—several measures related to MEP and QE do predict the level, slope, and/or curvature of US interest rates. In particular, our analysis tests a range of alternative variables, all representing imperfect proxies of unconventional monetary policies: two variables meant to proxy for the *size of the Fed's balance sheet*, to capture the effect of QE (the Fed's total assets and a Divisia monetary aggregate index); two variables that would like to measure the *composition of the Fed's balance sheet*, to capture the impact of the MEP (the percentage of Treasuries on the Fed's total assets and the average maturity of the Fed's portfolio of Treasuries); we also use a typical measure of the *Fed's interest rate policy* to proxy for the conventional monetary policy (the Effective Fed funds rate, FFR) and make sure that it is just unconventional policies that should have mattered during the GFC (also because the FFR hardly represented a policy tool available to policy-makers).

We contribute to at least three strands of literature. First, there is a literature that has spurred from Diebold and Li's (2006, henceforth DL) seminal paper, to show that DNS factors contain a rich time series structure that can be modelled and predicted using a simple, low-order AR process, which can be interpreted as restricted vector autoregressions (VAR). Levant and Ma (2017) introduce regimes in factor loadings and in volatility in the DNS framework. They also experiment with one specification in which the regimes are characterized by a latent Markov switching component which turns into the fourth latent factor. They find that the model with switching loading parameters gives the most accurate timing of regime duration in the term structure over their sample. Yet, both models provide a statistically superior in-sample fit vs. a single-state DNS VAR model. Chib and Kang (2013), Levant and Ma (2016) and Xiang and Zhu (2013) have recently extended DL's framework to account for Markov regimes. The former paper proposes linear affine term structure models (which nest DNS framework, as shown in Christensen, Diebold, and Rudebusch, 2011) with observable macroeconomic factors subject to non-recurrent structural breaks and employs Bayesian MCMC methods to show that the predictive performance of the model improves when regimes are taken into account. Levant and Ma incorporate regimes in the factor loadings of DL's model while we follow the strategy of the latter paper in which the DL's VAR model is subject to regime shifts. With reference to US data similar to the ones used in our paper, Xiang and Zhu report a significant improvement in forecasting performance deriving from the inclusion of regimes. We show that DL's forecasting results can be improved not only by just modelling regimes, but especially by combining Markov regime dynamics with a VAR framework that gives the state of monetary policy an explicit role. Moreover, something that had been harder without using data from the GFC (post-2008), we show that the predictive benefits from modelling regimes is exalted by the application of recursive pseudo out-of-sample (henceforth, OOS) designs that include the global crisis. Byrne, Cao, and Korobilis (2017) and Van Dijk, Koopman, Van der Wel, and Wright (2014) have incorporated up to 15 financial and macro variables in a large scale, time-varying VAR with stochastic volatility that represents a more general case than ours. Using Bayesian, dynamic model averaging techniques, they report that both parameter uncertainty and model uncertainty are important and in total account for one third of predictive variance and that macro-finance information is important to improve forecasting performance during recessions. Even though it is similar in spirit, we adopt a regime-switching view and look for evidence of predictability through economically-motivated loss function.

Second, there is a literature that has included macroeconomic variables into both structural and empirical models of the term structure.[1] One of the seminal papers is Ang and Piazzesi (2003),

[1] The question as to the actual usefulness of macro variables in forecasting (as opposed to modelling and understanding) the US riskless yield curve does not seem to have been settled. While Coroneo, Modugno, and Giannone (2016) and Kim and Park (2013) (for unspanned macroeconomic risk) show that a term structure model augmented with a broad macro-finance information set can provide superior forecasts, Dewachter and Iania (2012) report that financial factors are more prominent than macro variables per se, Hordal, Tristani, and Vestin (2006) and Moench (2008) (for spanned macroeconomic risk factors) find that expanding the DNS factors to macro variables improves forecasting performance only for

who have analyzed the joint dynamics of yields and macroeconomic variables under no arbitrage restrictions, to show that models with macro factors are able to forecast better than models with only unobservable factors. Ang, Dong, and Piazzesi (2007) and Evans and Marshall (2007) allow for bidirectional macro-finance links and find that the amount of in-sample yield variation that can be attributed to macro factors depends on whether the system allows for bidirectional linkages indeed: only when the interactions are bidirectional the system attributes over half of the variance of long yields to macro factors. Diebold, Rudebusch, and Aruoba (2006) have estimated single-state DNS models in which the yield curve is described by a latent factor model integrated by a simple, nonstructural VAR representation of the observable macroeconomy, (manufacturing capacity utilization, the federal funds rate, and inflation). They reject both the hypothesis of ''no macro to yields'' and the hypothesis of ''no yields to macro'' links, which represents clear statistical evidence in favor of a bidirectional link between the macroeconomy and the yield curve. Yet, the stop short from comparing the predictive performance of "yield-only" vs. macro-extended models as they focus on the in-sample fit of the model.[2] At least with reference to the GFC and for the DNS-class of models, our paper sweeps away all existing doubts and distinctions as to whether any macroeconomic factors may matter to forecast the yield curve: monetary variables, especially those able to capture the direct effects of MEP and QE on the structure and size of the Fed's balance sheet, do matter.[3] Doshi, Jacobs, and Liu (2018) study the impact of long-run and short-run components of output growth and inflation on the term structure of yields in a no-arbitrage, linear affine framework. While in-sample, they use variance decompositions to show that the relative importance of macroeconomic predictors critically depends on the maturity of the bonds and the horizon, their OOS analysis focuses on the predictability of bond excess *returns* (which are approximately linear transformations of changes in yields) to demonstrate that incorporating long-run components in standard predictive regressions improves the performance relative the information in the current yield curve only. We develop instead a genuine OOS design for yield level forecasts in a DNS framework.

intermediate and long prediction horizons, and Ullah, Tsukuda, and Matsuda (2013) find mixed evidence on Japanese data. See the reviews in Duffee (2013) and Gürkaynak and Wright (2012).

[2] Levant and Ma (2016) have extended these results to UK data also by explicitly accounting for a structural break during the early 90s at the time the UK exited the European Monetary System. Rudebusch and Wu (2007) have extended this modelling strategy to include a small-scale structure model that features a monetary policy reaction function, an output Euler equation, and an inflation equation within an affine, no-arbitrage dynamic specification.

[3] However, as we shall document in Sections 5 and 6, which specific Markov switching VAR model actually delivers the most accurate forecasts and profitable trading signals may itself be subject to some time variation.

More generally, there is a third literature, which finds its roots in the classical work by Eugene Fama on the efficient markets hypothesis, on the actual and purported links between macroeconomic variables and asset prices. The modelling of the riskless yield curve has long been a typical example of the disconnect between the macro and finance literatures, see e.g., Estrella and Hardouvelis (1991) and more recently, Hordal, Tristani, and Vestin (2006): in this body of research, it would be sensible to look for predictive signals of future macroeconomic conditions in the yield curve (e.g., the classical term spread indicator), but not the opposite—current macro variables would not have any forecasting bite for interest rates, see for instance the discussion in Balfoussia and Wickens (2007). Duffee (2011) uncovers a "hidden" factor that has opposite effects on expected future interest rates and bond risk premia, large impact on the dynamics of yields but on which any snapshot of the time-t yield curve conveys no information. Such a factor is shown to explain half of the dynamics of interest rates but measures of macroeconomic activity explain only a small fraction of its variation; similarly, the ability of the hidden factor to forecast excess returns is not captured by any macroeconomic variables. Although debating these classical questions does not represent our main goal, a Reader may use our results to argue that—at least with reference to the US term structure and during the unconventional monetary policy experiment—regimes exist in which at least monetary factors predict the yield curve which, as a result, would fail to include all relevant information on the future path of the economy.

The rest of the paper is structured as follows. Section 2 describes the empirical methodologies adopted in the paper and its research design. Section 3 introduces the data, provides summary statistics and deals in some details with pros and cons of our proxy for unconventional monetary policy. Section 4 reports on model selection and sketches the most notable features of the (very heavy volume of) estimation results. Section 5 is the key step of the paper and reports on comparative forecasting performance and on tests of equal predictive accuracy. Section 6 tests whether the predictive performances described in Section 5 can be used to support profitable trading strategies, also when realistic transaction costs are applied, and may be read as an extension of Section 5 to a more realistic, directly relevant (at least to traders) loss function, in spirit of Leitch and Tanner (1991). Section 6 concludes and outlines a few directions for extensions and additional research.

2. ECONOMETRIC METHODOLOGY

2.1. *Baseline Dynamic Nelson-Siegel mode*

Our starting point to model and forecast the yield curve is represented by the now classical DL's

(2006) two-step approach.[4] By applying restrictions on the factor loadings DL re-write the classical Nelson–Siegel Laguerre's approximation (for the forward rates) as

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right), \quad (1)$$

where $y_t(\tau)$ is the yield to the maturity τ at time t and β_{1t}, β_{2t}, β_{3t} are latent factors that entirely determine the dynamics of $y_t(\tau)$ for all τ; 1, $\frac{1-e^{-\lambda\tau}}{\lambda\tau}$, and $\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}$ are the factor loadings that instead define the cross section of the yield curve for any t; λ is a fixed decay parameter that affects the shape of the curve. Because the factor loading associated to β_{1t} is constant at 1, as the maturity τ tends to infinity, all yields tend to β_{1t}. Hence β_{1t} is often thought of as a long-term factor. The loading on β_{2t} starts instead at 1 (as $\tau \to 0$ from the right) and decays monotonically to zero: therefore β_{2t} has a maximum impact at short maturities and a minimal one on long maturities, so it can be interpreted as a short-term factor. Finally, it can be checked that the loading on β_{3t} starts at zero, then increases and finally decays to zero, with the global maximum depending on the coefficient λ; because for the typical values of λ estimated in the literature, its loading reaches such a maximum at intermediate maturities, we can refer to β_{3t} as a medium-term factor.

To emphasize the interpretation of the DNS factors, as in DL (2006), we rename them according to their effect on the shape of the yield curve: β_{1t} is referred to as the Level factor (L_t), β_{2t} as the (the negative of the) Slope factor (S_t), and β_{3t} as the Curvature factor (C_t). Finally, the parameter λ represents the rate of change of the factor loadings with maturity. Because of its critical role as well as its tricky econometrics (see, e.g., Diebold, Rudebusch, and Aruoba, 2006, who anyway estimate a similar value of 0.077 close to DL's original calibration; Moench, 2012, reports a Bayesian MCMC posterior mode estimate of 0.067; Xiang and Zhu, 2013, report a slightly lower mean estimate, between 0.04 and 0.05), we follow Byrne, Cao, and Korobilis (2017), Coroneo, Modugno, and Giannone (2016), Diebold and Li (2006), and Van Dijk, Koopman, Van der Wel, and

[4] Of course, it would be sensible to extend our research to arbitrage-free DNS models, à la Christensen, Diebold, and Rudebusch (2009). Our choice not to follow an arbitrage-free approach is supported by two main arguments. Firstly, Diebold, Rudebusch, and Aruoba (2006) have argued, if the market were to generate arbitrage-free data, then the DNS curve should capture this feature reasonably well even without imposing theoretical restrictions on the model. Moreover, there may be a loss of efficiency in not imposing the restriction of no arbitrage if it is valid, but this must be weighed against the possibility of misspecification if transitory arbitrage opportunities are not immediately traded away. In addition, the literature shows that unrestricted DNS models often have better forecasting power than arbitrage-free ones especially *over short prediction horizons*, see Duffee (2002, 2008), Diebold and Rudebusch (2013), and when compared to random walk forecasts, see Carriero and Giacomini (2011) and Guidolin and Thornton (2018). Moreover, Coroneo, Nyholm and Vidova-Koleva (2011) and Tu and Chen (2018) have emphasized that when estimated on US data, a standard DNS model may be close to being arbitrage-free even when it does not explicitly impose these restrictions.

Wright (2014) and fix λ at 0.0609. In correspondence to such a value, the loading on the Curvature factor is maximized at exactly 30 months, which seems reasonable and appears consistent with the arguments in Yu and Zivot (2011).[5]

As it has been recognized since Diebold, Rudebusch and Aruoba (2006), DL's model can be re-written in state-space form. Such a form emphasizes the distinction between the measurement equations (which relate the cross-section of yields to the three latent factors), a set of transition equations (which describes the time evolution of the factors), and give a role to potential measurement and specification errors (to account for time-varying liquidity, credit risk, and mispricing factors), $\varepsilon_{t,\tau}$. In particular, the vector of measurement equations is

$$\begin{bmatrix} y_{t,\tau_1} \\ y_{t,\tau_2} \\ \vdots \\ y_{t,\tau_N} \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \dfrac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \dfrac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \dfrac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \dfrac{1-e^{-\lambda\tau_n}}{\lambda\tau_N} & \dfrac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{bmatrix} \begin{bmatrix} L_t \\ S_t \\ C_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t,\tau_1} \\ \varepsilon_{t,\tau_2} \\ \vdots \\ \varepsilon_{t,\tau_N} \end{bmatrix}, \qquad (2)$$

where N is the number of available maturities $(\tau_1, \tau_2, \dots, \tau_N)$ and the $N\times1$ vector $\boldsymbol{\varepsilon}_t$ collects maturity-specific errors and is also interpretable as an idiosyncratic source of maturity-specific risks. We assume that $\boldsymbol{\varepsilon}_t$ is IID over time and simultaneously cross-sectionally uncorrelated (as in Diebold, Rudebusch, and Aruoba, 2006; Levant and Ma, 2016), even though this assumption may be weakened, if desirable. With three factors, the vector of transition equations specifies the dynamics of Level, Slope, and Curvature over time. In our baseline analysis, we generalize DL's set up to the case in which the factors follow a multivariate Gaussian VAR(p) process:

$$\begin{aligned} \boldsymbol{f}_{t+1} = \begin{bmatrix} L_{t+1} \\ S_{t+1} \\ C_{t+1} \end{bmatrix} &= \begin{bmatrix} \mu_L \\ \mu_S \\ \mu_C \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \phi_{11}^j & \phi_{12}^j & \phi_{13}^j \\ \phi_{21}^j & \phi_{22}^j & \phi_{23}^j \\ \phi_{31}^j & \phi_{32}^j & \phi_{33}^j \end{bmatrix} \begin{bmatrix} L_{t+1-j} \\ S_{t+1-j} \\ C_{t+1-j} \end{bmatrix} + \begin{bmatrix} \eta_{L,t+1} \\ \eta_{S,t+1} \\ \eta_{C,t+1} \end{bmatrix} \\ &= \boldsymbol{\mu} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j \boldsymbol{f}_{t+1-j} + \boldsymbol{\eta}_{t+1} \end{aligned} \qquad (3)$$

where $\boldsymbol{\mu}$ is the vector of intercept terms, $\boldsymbol{\Phi}_j$ is a 3×3 matrix that governs the factor dynamics, and $\boldsymbol{\eta}$ is an innovation vector process, such that $\boldsymbol{\eta}_{t+1} \sim N(0, \boldsymbol{\Sigma})$. Because $\boldsymbol{\Sigma}$ is potentially a full matrix,

[5] Yu and Zivot (2011) have examined the OOS performance of one-step state-space vs. two-step dynamic DL models to find that, surprisingly, a formal state-space approach that encompasses the estimation of λ fails to improve the OOS predictions for Treasury yields. This result is confirmed by Coroneo, Modugno, and Giannone (2016) when latent macroeconomic are introduced in the model. Within a wide range of values, Nelson and Siegel (1987) find that the goodness-of-fit of the yield curve is not very sensitive to the specific value of λ.

shocks to different factors may carry a non-zero contemporaneous correlation—in this we differ from Xiang and Zhu (2013).

The estimation of the model in (2)-(3) is performed through the two-step procedure introduced by DL (2006), who in fact specialize on a restricted version of (3) in which p = 1 and $\mathbf{\Phi}_1$ is diagonal. However, the economic value results in Caldeira, Moura, and Santos (2016) emphasize that an unrestricted $\mathbf{\Phi}_1$ may offer benefits. Although other, single-step approaches (for instance, based on the Kalman filter) have been developed, because of its simplicity, convenience and numerical reliability the two-step estimation maintains a strong appeal.[6] In the first step, we estimate the latent factors via OLS to fit at each point in time the (cross-sectional) riskless yield curve. We thus obtain a three-dimensional time series of yield factors, $\{\hat{L}_t, \hat{S}_t, \hat{C}_t\}_{t=1}^{T}$. In the second step, we model the factors' time series as driven by a VAR(p) process.[7]

2.2. Markov switching vector autoregressive models

A literature has recognized that allowing for regime switching dynamics may be of key importance to correctly capture the dynamics of riskless interest rates and thus obtain more accurate forecasts, see, among others, Ang and Bekaert (2002), Bansal and Zhou (2002), Chib and Kang (2013), Dai, Singleton, and Yang (2007), Gray (1996), Guidolin and Timmermann (2009), Hamilton (1988), Rudebusch and Wu (2007), Smith (2002), and Startz and Tsang (2010). Markov-switching vector autoregressions (MSVAR(p)) generalize the framework in (2)-(3) to the case in which a (latent) discrete state variable $S_{t+1} = 1, \dots, K$ may cause shifts in the vectors and matrices of parameters of the econometric model of the DNS factors:[8]

$$\boldsymbol{f}_{t+1} = \boldsymbol{\mu}_{S_{t+1}} + \sum_{j=1}^{p} \mathbf{\Phi}_{j,S_{t+1}} \boldsymbol{f}_{t+1-j} + \mathbf{\Sigma}_{S_{t+1}}^{1/2} \boldsymbol{z}_{t+1} \qquad \boldsymbol{z}_t \equiv \mathbf{\Sigma}_{S_{t+1}}^{-1/2} \boldsymbol{\eta}_{t+1} \sim IID\ N(0, \mathbf{I}_{\mathrm{N}}). \qquad (4)$$

Because in (4), the vector of intercepts, the matrix of vector autoregressive coefficients, and the (Choleski factorization of the) covariance matrix all depend on S_{t+1} (as in Levant and Ma, 2016), we call this framework a MSIAH(K, p) model, where the I stands for the fact that $\boldsymbol{\mu}_{S_{t+1}}$ is regime switching, A for the fact that $\mathbf{\Phi}_{j,S_{t+1}}$ is regime switching, and H refers to $\mathbf{\Sigma}_{S_{t+1}}^{1/2}$. Note that

[6] The one-step, state-space methods imply potentially challenging optimization problems and are characterized by a large number of parameters to be estimated which sometimes makes optimization unstable (see, among others, Diebold and Rudebush, 2013).

[7] As a result, we obtain that $Cov[\varepsilon_{t,\tau_n}, \eta_{(\cdot),t}] = 0$ for all factors and yields, but this is in any case one of the assumptions required for the optimality of the Kalman filter also in single-step, state space estimation.

[8] Hevia, Gutierrez, Sola, and Spagnolo (2015) and Levant and Ma (2016) have instead maintained a simpler VAR dynamics for the factors but instead jointly estimated a regime switching process for λ.

$\mathbf{\Sigma}_{S_{t+1}}^{1/2}(\mathbf{\Sigma}_{S_{t+1}}^{1/2})' = Var[\boldsymbol{\eta}_{t+1}|S_{t+1}]$. As it is well known (see Guidolin and Pedio, 2018), in a MSIAH(K, p) with $K > 1$ and $p \geq 1$, the potential of the conditional mean parameters ($\boldsymbol{\mu}_{S_{t+1}}$ and $\mathbf{\Phi}_{j,S_{t+1}}$) to move across regimes in ways that synchronize the direction of change of the conditional mean makes it possible to capture non-linear patters of association of the series of factors over time and beyond cross- and own-serial correlation (measured by $\mathbf{\Phi}_{j,S_{t+1}}$) and cross simultaneous correlation (estimated as $\mathbf{\Sigma}_{S_{t+1}}^{1/2}(\mathbf{\Sigma}_{S_{t+1}}^{1/2})'$), i.e., simple, classical linear association patterns. Moreover, time variation in the covariance matrix makes it possible for (4) to capture patterns of volatility clustering and thin/fat tails, which are all well-known features of time series of interest rates (see, e.g., Hess and Kamara, 2005; Koopman, Mallee, and van der Wel, 2010; Startz and Tsang, 2010).

In particular, in the time-homogeneous hidden Markov state specification of regime switching models, the latent regime variable S_{t+1} is generated by a discrete, time homogeneous, irreducible and ergodic first-order Markov chain such that $\Pr\left(S_t = j|\{S_j\}_{\tau=1}^{t-1}, \{\boldsymbol{f}_l\}_{l=1}^{t-1}\right) = \Pr(S_t = j|S_{t-1} = i) = p_{i,j}$, where $p_{i,j}$ is the generic [i, j] element of the $K \times K$ transition matrix, $\mathbf{P}$. The *irreducible* nature of the chain means that there is a non-zero-probability of transitioning from any state to any other over time and this implies that the vector of ergodic, unconditional probability implied by the process is positive element-by-element, $\bar{\boldsymbol{\xi}} > 0$. Finally, it is of the *first order* because the current state is only affected by the state of the previous period.

Even in their time-homogeneous Markov implementation (similarly to Xiang and Zhu, 2013), MSVAR models tend to be richly parameterized. If we let M be the number of variables collected in $\boldsymbol{f}_t$ (e.g., M = 3 but below we shall examine cases on which the vector is expanded to include four or five variables), then the general MISIAH(K, p) specification in (4) implies the estimation of a number of parameters equal to:

$$\mathrm{K}\left[\mathrm{M} + pM^2 + \frac{M(M+1)}{2} + (K-1)\right]. \tag{5}$$

Reducing the number of parameters that are regime-dependent allows not only to simplify the estimation but also to make it more reliable and tends to provide a better basis in forecasting applications. Especially when the number of parameters returned by (5) is large relative to the total number of available observations (MT), or equivalently when the *saturation ratio* given by MT divided by the number of parameters in (5) is inferior to 20 (or at least a two-digit threshold), besides avoiding specifications with K exceeds two or three regimes, imposing restrictions on the rich structure in (4) may pay off. One restricted version of (4) that turns out to be important in our empirical analysis is the MSIH(K)-VAR(p) model,

$$\boldsymbol{f}_{t+1} = \boldsymbol{\mu}_{S_{t+1}} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j \boldsymbol{f}_{t+1-j} + \boldsymbol{\Sigma}_{S_{t+1}}^{1/2} \boldsymbol{z}_{t+1} \qquad \boldsymbol{z}_t \equiv \boldsymbol{\Sigma}_{S_{t+1}}^{-1/2} \boldsymbol{\eta}_{t+1} \sim IID\ N(0, \mathbf{I}_{\mathrm{N}}). \tag{6}$$

which obtains when the VAR matrices are time-homogeneous, as in Xiang and Zhu (2013). In this case, the number of parameters to be estimated declines to $pM^2 + \mathrm{K}\left[\mathrm{M} + \frac{M(M+1)}{2} + (K-1)\right]$.

As customary in the time series literature, in the frequentist domain, the estimation of MSVAR models is performed by maximum likelihood, using the Expectation Maximization (EM) algorithm first applied to MS models by Hamilton (1990). As by a product, this technique also allows an iterative calculation of the one-step ahead forecast of the Kx1 state probabilities vector $\boldsymbol{\xi}_{t+1|t}$ given the entire information set at time t. These will be used in Sections 4 and 5 to follow to compute factor and hence yield predictions.

2.3. Extending Models to Monetary Policy Indicators

The third logical step of our methodology extends both the linear model in (3) and the non-linear frameworks encompassed by (4) and (6) to include one or more variables describing the stance of monetary policy over and beyond what the past history of the factors subsumes, to be collected in a Qx1 vector time series $\boldsymbol{m}_{t+1}$. In general, the variables in $\boldsymbol{m}_{t+1}$ are allowed to both affect the dynamics of the DNS factors and, on their turn, to be predicted by such factors, if the data so require. The literature abounds of indications of the fact that slope (see Levant and Ma, 2016; Wu, 2001) and curvature (see Bekaert, Cho, and Moreno, 2010; Dewachter and Lyrio, 2006) DNS factors are affected by monetary policy, not to mention the obvious linkages between the FFR controlled by the Fed and the level factor. For instance, in the case of a MSIAH(K, p) model, if we collect all the variables of interest in the Mx1 vector $\boldsymbol{x}_{t+1} \equiv [\boldsymbol{f}'_{t+1}\ \boldsymbol{m}'_{t+1}]'$ (where $M = Q + 3$) we obtain:

$$\boldsymbol{x}_{t+1} = \boldsymbol{c}_{S_{t+1}} + \sum_{j=1}^{p} \mathbf{A}_{j,S_{t+1}} \boldsymbol{x}_{t+1-j} + \boldsymbol{\Omega}_{S_{t+1}}^{1/2} \boldsymbol{v}_{t+1} \qquad \boldsymbol{v}_t \sim IID\ N(0, \mathbf{I}_{\mathrm{M}}). \tag{7}$$

When the VAR matrices, in some or all states and at some or all lags, display a block-diagonal structure, this means that the monetary policy variables constitute an autonomous sub-system inside the overall MSVAR framework; in this case, monetary conditions impact on the DNS factors only through simultaneous shocks (when $\boldsymbol{\Omega}_{S_{t+1}}$ is itself not block-diagonal) and/or through the information they reveal on the Markov chain variable governing regimes. As emphasized by Chada, Turner, and Zampolli (2013) with reference to the yield curve effects of the maturity composition of the net public debt held outside the Fed's balance sheet, a regime switching framework offers the key advantage of including pre-crisis data in the estimation at the same time allowing the novel, unconventional nature of the monetary policies pursued since 2008 to affect the relative predictive performance of different models.

When the VAR matrices, in some or all states and at some or all lags, display an *upper* block-triangular structure, this means that while the monetary policy variables predict future DNS factors, the opposite is not the case, i.e., monetary policy promptly incorporates all information in the factors, but monetary conditions gradually propagate over time, also through the factors, to the shape of the riskless yield curve, as in Ang and Piazzesi (2002). In this case, financial markets are unlikely to be (weakly) informationally efficient. When the VAR matrices, in some or all states and at some or all lags, display a *lower* block-triangular structure, this means that while the current DNS factors and hence the shape of the yield curve predict subsequent monetary policy reactions, the opposite does not hold, i.e., the US Treasury markets are efficient enough to account already of all available information on the stance of monetary policy, as featured in the seminal paper by Estrella and Hardouvelis (1991). This case is compatible with policymakers learning about the state of the economy from the features of the riskless yield curve.[9] Of course, when $\mathbf{A}_{j,S_{t+1}}$ fails to display any special structure (it is potentially full) in at least one regime, then we are facing the case of bi-direction feedbacks already uncovered by Diebold, Rudebusch, and Aruoba (2006) or by Kozicki and Tinsley (2005).

3. THE DATA

3.1. Treasury bond yields

To infer the DNS factors implicit in the Treasury yield curve using the DNS model, we summarize the Treasury term structure using a set of yields of different maturities: the 3-month, 6-month, 1-, 2-, 3-, 5-, 7-, 10-, and 30-year constant maturity yields for a January, 1982 - July, 2015 monthly sample, retrieved from FRED, at the Federal Reserve Bank of St. Louis.[10]. These yields are interpolated from daily yield curve data based on the closing market bid prices for actively traded Treasuries on the over-the-counter market.[11] Through such interpolation, we obtain yields in correspondence to fixed maturity points on the yield curve even when there are no outstanding bonds with that particular time-to-maturity, which is rather likely for tenors past 1-year.

[9] Of course, also in the block-triangular cases, it remains possible for monetary conditions to be related to the DNS factors simultaneously, via cross-correlation in the shocks and/or through the information they reveal on the Markov chain variable governing the regimes of the whole system.

[10] The 30-year yields are unavailable between February 2002 and August 2006. For this period, following Hamilton and Wu (2012), we use instead the 20-year rate minus 21 bps, which is the amount by which the 20-year rate exceeded the 30-year rate both immediately before and after the gap.

[11] For details on the interpolation of the Constant Maturity Treasury yields, see www.treasury.gov.

In Figure 1, we plot the constant maturity yields, whereas in Table 1 we present summary statistics. Both reveal features that are consistent with commonly known facts about the US term structure. For instance, a comparison of mean yields across different maturities confirms that the yield curve is on average increasing and concave; moreover, the sample standard deviations of the yields generally decrease with maturity. In addition, from these statistics we obtain pervasive evidence of non-normality in the data, mostly (with the exception of the 10-year and 30-year maturities) caused by a negative excess kurtosis (i.e., tails thinner than under a Gaussian distribution). In fact, a Jarque-Bera test confirms that the null hypothesis of a normal distribution can be rejected for all of the series at a 5% size level. As it can be seen from Panel B of Table 1 and from Figure 1, the correlations among the series are high and tend to be the highest for adjacent maturities, ranging between 0.90 and 0.99.

4.2. Monetary policy indicators

To test whether and how much the variables that capture the stance of monetary policy may add predictive power to the standard DNS factors, we select/construct five observable series to represent a range of features that may characterize policy, with particular emphasis to the unconventional measures adopted by the Fed during the GFC (i.e., balance sheet policies). The five variables are: two variables meant to proxy for the *size of the Fed's balance sheet* (the level of the Fed's total Assets and a Divisia money aggregate index); two variables that measure the composition of the Fed's balance sheet (the percentage of Treasuries on the Fed's total assets and the average maturity of the Fed's portfolio of Treasuries), and a measure of the *Fed's interest rate policy* (the Effective FFR). We describe each of the variables in detail and explain why these have been associated to specific aspects of monetary policy actions. Figure 2 shows the five series under investigation.

The level of the Fed's total Assets is available from the Federal Reserve's weekly H41 release. However, for the purpose of our analysis we consider the level in correspondence to the last week of each month.[12] Because the data are available only from July 1996, we are forced to perform our analysis on a shorter sample for this variable (i.e., July 1996-July 2015). This variable is selected to reflect the growth in the Fed's balance sheet as a consequence of the LSAP and QE (1

[12] Particularly, we use the natural logarithm of the level of total assets as suggested in Gambacorta, Hofmann, and Peersman (2014). Gambacorta and Hofmann (2012) have argued that, although other variables, mainly monetary base, seem to be affected by QE policies, the partial sterilization made by central banks on base money should make assets a better gauge to measure unconventional monetary policy during the GFC.

through 3) programs enacted between 2008 and 2014.[13] Figure 2 shows that the log of Assets increased dramatically between October 2008 and March 2009 as a consequence of the Fed's purchases of Treasuries and other longer term securities for an exceeding one trillion USD. In the following years, an additional series of large-scale asset purchases (LSAPs) led the Fed's total assets to grow even more (by about 3.5 trillion dollars in total).

As an alternative proxy for the total aggregate monetary base, and hence of the size of Fed's balance sheet, we follow the classical literature on the advantages of the Divisia Index.[14] In particular, we use the log level of the Divisia MZM Index published by the Federal Reserve Bank of St. Louis.[15] The sample for this series starts in January, 1982, as the constant maturity Treasury series. Figure 2 shows the natural log of the Divisia MZM aggregate (but our empirical estimates also use growth rates defined as differences in logs). The growth rate accelerated in 2007 and 2008 in response to the unconventional Fed policies; in particular, it picked up in December 2008, after the announcement of the LSAP program. Subsequently, it slowed during 2009 (this can be seen as a consequence of the substantial outflows experienced by money market funds) and strengthened again during 2010, in response to further quantitative easing policies (see Anderson and Jones, 2011, for details).

The percentage of Treasuries on the Fed's total assets is intended as a way to measure any variations in the composition of the Fed's balance sheet. This series is calculated by dividing the

[13] Between 2008 and 2010, the Fed has more than doubled the size of its balance sheet through asset purchases. In November 2008, the Fed announced the purchase of up to $100 billion agency debt and up to $500 billion agency mortgage-back securities (original LSAP program). In March 2009, to improve conditions in private credit markets, it announced the intention to purchase additional $850 billion agency debt and up to $300 billion of long-term Treasury securities (the so-called QE1). Furthermore, on November 2010, the Fed announced an additional expansion by purchasing a further $600 billion of long-term Treasuries (QE2). In September 2012, the Fed announced it would buy $40 billion *per month* in mortgage-backed securities until occupational figures would have improved (QE3). In December 2012, the FOMC announced an increase in the amount of the purchases from $40 to $85 billion per month.

[14] Belongia and Ireland (2012) have extended the classical Bernanke and Blinder (1988) to show that monetary aggregates still possess a significant explanatory power for monetary policy provided the aggregates are built not to be simple sum indexes because these cannot internalize pure substitution effects, like Divisia indices. The Divisia index of money weighs different assets by the value of the monetary services they provide, see, e.g., Barnett (1982). However, Gambacorta, Hofmann, and Peersman, 2014) have argued that in the US, in the aftermath of the GFC standard indicators of monetary growth fail to capture the full thrust of monetary policies.

[15] MZM stands for "money-zero maturity": it corresponds to M2, less small time deposits, plus Institutional Money Market Mutual Funds (MMMFs). It contains items that are immediately convertible, without penalty, to some form of medium of exchange. As emphasized by Belongia and Ireland (2012), the choice between different levels of aggregation (i.e., between M1, M2, and MZM) does not seem relevant. Their empirical analysis shows that the information content of the Divisia MZM and other Divisia Indexes is very similar, whereas important differences exist between Divisia and simple sum aggregates.

total amount of Treasuries on the Fed's portfolio by the Fed's total assets. Because with reference to the amount of Treasuries held the Federal Reserve's weekly H41 release is the only source of data and it is available from July 1996, we use a 1996 – 2015 also for this variable. The variable is meaningful because it has been generally observed that the size of the Fed balance sheet alone may fail to account for the impact of the credit policies implemented by the Fed.[16] Effectively, because an extensive use of credit policy took place especially during the initial (sometimes referred to sub-prime or credit crunch) stages of the GFC, it seems appropriate to also make use of such a variable. Borio and Disyatat (2010) emphasize that the main channel trough which balance sheet policies operate is by altering the composition of the portfolios in the private sector; for instance, through the purchase of less liquid or risky assets the Fed can reduce yields and ease financing constraints.[17] Figure 2 confirms that this variable is suitable to proxy for shocks in the composition of the Fed's portfolio resulting from the LSAP program and the policies undertaken between late 2007 and 2008 when—to lower borrowing costs and ease credit flows in the private sector—the Fed departed from its practice to only trade Treasury securities through open market operations and started instead to extend credit to the private sector and government agencies. Consequently, the Treasuries portfolio, that used to represent around 90% of the entire Fed's portfolio until May 2008, dropped dramatically to almost 20% during the following 6 months, to settle back to a steady level after December 2008. This pattern reflects the purchase of a large amount of mortgage-backed securities and agency debt during 2008.

The average maturity of the FED's portfolio of Treasuries is the variable selected to summarize the effects of MEP.[18] Because MEP is a sterilized operation in which the proceeds from the sales of short-term Treasuries are used to buy long-term Treasury securities, this operation is balance sheet neutral (i.e., it does not affect any of the variables listed so far).[19] Hence, the only way to

[16] According to the definition in Goodfriend (2011), we refer to credit policy as a policy that shifts the composition of the Fed's asset portfolio between Treasury securities and credit to the private sector (or to non-Treasury government entities), holding the total fixed.

[17] Such policies may produce real effects if the assets exchanged are not perfect substitutes, as in Modigliani and Sutch's preferred-habitat theory, recently revived by Vayanos and Vila (2009).

[18] Hamilton and Wu (2012) have stressed that the maturity structure of the government debt issued to the public may be expected to affect the pricing of Level, Slope, and Curvature risk. Because Treasuries of different maturities carry different risk, and this is priced by the market, replacing long-term debt with short-term debt in private sector portfolios would affect equilibrium yields. For instance, Kuttner (2006) finds that shifts in Federal Reserve holdings of government debt toward long maturities lower the risk premia of two-, three-, four-, and five-year bonds. Greenwood and Vayanos (2014) have reported a positive correlation between the maturity structure of the US government debt and the riskless term structure.

[19] In September 2011, the FOMC announced it would buy Treasuries with maturities between 72 and 360 months and sell an equal amount of securities with remaining maturities of 3 to 36 months, for an amount up to $667 billion. In contrast to the earlier Large-Scale Asset Purchase operations, at that time the MEP

capture the effect of this type of policy is to measure the average maturity of the Treasuries portfolio. The New York Fed database provides System Open Market Account (SOMA) holdings from 2003 to the present, divided by rough maturity breakdowns (less than 15 days, 16-90 days, 91 days to 1 year, over 1 year to 5 years, over 5 years to 10 years, and over10 years). We use these data to calculate the average maturity of the Fed's Treasuries portfolio for each end of month between January 2003 and July 2015. Moreover, to cover the same sample period as the constant maturity yields, we obtain data for the period January 1982 - December 2002 using the average maturity of the Fed's Treasury holdings as calculated in Kuttner (2006).[20] From inspection of Figure 2, we have confirmation that this variable has been strongly affected by the Fed's unconventional policies: the average maturity doubled in 2008, because of the QE1 program and it especially picked up after September 2011, after the MEP was launched.

Finally, the Effective FFR is meant to act as a proxy for the conventional, interest rate policy typically enacted by the Fed. We include this variable into our analysis to control whether and to what extent the standard monetary policy instrument may strengthen the predictability of the term structure, especially with MSVAR set ups that may endogenously account for sequences of monetary policy regimes. The data on the FFR are obtained from Bloomberg for the entire sample. To be consistent with the other variables, we download the effective rate in correspondence to the last trading day of each month. As shown in Figure 2, the variability in the FFR had been substantial before 2008. Then, after the lower bound of the target rate was cut to zero in December 2008 (i.e., the zero lower bound had been reached), the escalating recession forced the Fed to maintain its target rate at a zero level for a long time and to credibly communicate such an intention. Therefore, since then, no significant changes in the FFR have occurred for the rest of the sample. This suggests that, at least during the crisis and in its aftermath, the FFR may have turned insufficient to explain the monetary policy of the Fed and have lost any earlier predictive power.

4. EMPIRICAL RESULTS

In this Section, we first estimate the time series of the factors fitting the static NS model to the yield curve in each period; next, we capture the dynamic structure of the extracted Level, Slope and Curvature factors using the models described in Section 2. We start the analysis with a baseline single-state, VAR(1) model for the DNS factors only. Then, we introduce regimes in the

---

explicitly aimed to increase the average maturity of the Fed's Treasury holdings without increasing the overall size of the central bank's balance sheet.

[20] Data are available at http://econ.williams.edu/people/knk1/research.

model obtaining six MSVAR models (some of them will also generate a few variants described below): the first includes only the three yield curve factors, whereas the others are augmented by alternative variables, each representing the stance of monetary policy.

4.1. The DNS factors

As described earlier, DL's version of Nelson-Siegel's model has been estimated at each time t by standard OLS applied to the cross section of yields, holding λ fixed at 0.0609. Cross-sectional estimation has been recursively repeated over time, between January 1982 and July 2015 obtaining a set of 403 sequential estimates of the three unobservable factors $\hat{L}_t, \hat{S}_t, \hat{C}_t$. In Figure 3, we plot $\hat{L}_t, \hat{S}_t, \hat{C}_t$, along with the observable counterparts of level, slope and curvature of the yield curve. The empirical counterparts of the factors are proxies derived as in DL (2006): we define the observable Level as the 10-year yield, Slope as the difference between 10-year and 3-month yields, and Curvature as twice the 2-year yield minus the sum of the 3-month and 10-year yields. The correlations between the estimated factors and their empirical counterparts are very high, ranging between 0.98 and 0.99. These findings confirm our interpretation that the DNS factors largely correspond to the empirical level, slope, and curvature of the term structure and reassure us of the fact that we shall be working with factors with a similar nature and interpretation as in earlier literature.

Table 2 presents the key summary statistics of the estimated DNS factors. As we would expect, Level is on average positive and largely follows the path of the long term bond yields (10- and 30-year rates). On the contrary, Slope is on average negative: this is not surprising because the NS loading on Slope is designed in such a way that negative values of the Slope factor corresponds to the typical upward sloping yield curve. In turn, positive, null and even small negative values of the Curvature correspond to the typical concave yield curve; therefore, the slightly negative mean of the factor is compatible with the empirical fact that the term structure is on average concave. Moreover, from the standard deviations, it emerges that the Curvature is the most volatile among the factors, followed by Level; Slope is the least volatile factor. Finally, while Level and Slope are largely uncorrelated, Curvature is positively correlated with all other factors, meaning that when rates are relatively high and the term structure positively sloped, then concavity emerges.

4.2. $VAR(1)$ models

As a benchmark, we model the dynamics in the NS factors using a standard VAR(p) as in (3). We experiment with p = 1, 2, and 3, but we find that all information criteria (see Section 4.3 for definitions) as well as many of the adjusted R-square coefficients (on a equation-by-equation basis) decline when the number of lags is set at 2 or a higher value. To the contrary, and consistently with DL (2006), p = 1 is required although simple LR tests reject the null hypothesis that $\mathbf{\Phi}_1$ should be diagonal, as it is instead assumed by DL.

Table 3 reports the parameter estimates of the VAR(1) model applied to the factors. All factors are very persistent; however, Level is the most persistent ($\widehat{\mathbf{\Phi}}_1[1,1] = 0.989$) factor, followed by Slope ($\widehat{\mathbf{\Phi}}_1[2,2] = 0.934$), and Curvature ($\widehat{\mathbf{\Phi}}_1[3,3] = 0.916$). These results are in line, in spite of a longer sample that encompasses the GFC and the unconventional policy measures ensuing from it, with the literature. Moreover, our findings are consistent with the commonly reported pattern that *yield* dynamics (that in DNS framework is approximated by the Level factor) is more persistent than the *spread* dynamics (in our case, approximated by the Slope factor). The vast majority of the estimated coefficients is significant at least at a 5% level. In particular, all factors have significant forecasting power for Slope and Curvature, but not for Level; if anything, $\widehat{\mathbf{\Phi}}_1$ is upper triangular, not diagonal, which is consistent with the unconditional correlations in panel B of Table 2. As has already been documented in DL (2006) and Diebold, Rudebusch, and Aruoba (2006), the VAR(1), pure DNS factor model explains the cross-sectional variation of interest rates of different maturities very well over time, with R-squares exceeding 99% for intermediate maturities. Finally, in a single-state model, the residual correlations are small and not statistically significant, i.e., the covariance matrix of the shocks is close to diagonal, which is a case explored by Diebold, Rudebusch, and Aruoba (2006).[21]

4.3. Markov switching VAR models

On the basis of earlier literature on the presence of regimes in monetary policy and its connections to the yield curve, we have a prior that a single-state VAR(1) model may be insufficient to capture all features of the term structure. Therefore we proceed to estimate a range of K-state MSVAR models, both limited to the three DNS factors and extended to include one or two (simultaneously) additional variables to represent the stance of monetary policy (i.e., the first difference in the log of the assets in the Fed's balance sheet, the first difference in the log of the Divisia index, the percentage of the Fed's balance sheet assets represented by Treasuries, the average maturity of the Fed's Treasury portfolio, and the effective Fed funds rate). Of course, dealing with the case of Q = 1 offers the advantage of emphasizing the specific predictive role

[21] We have estimated restricted, block diagonal and upper triangular VAR(1) models in which the macro variables carry no cross-serial correlations with DNS factors or in which past macro variables forecast future DNS factors but not the opposite. All these restrictions were rejected by LR tests. We have also estimated a range of VAR(1) models for vectors of variables that include macroeconomic variables. These are not specifically commented, but they are used as benchmarks in Section 5.

played by each of the variables in isolation, and tends to maximize their forecasting power. The occasional inspection of a few cases in which two variables were simultaneously used (Q = 2) did allow us to test whether two different variables may contain different and complementary information on the stance of monetary policy. Because of data limitations, we only experiment with K = 2 and 3. Although other models (e.g., with constant covariance matrix of the shocks) were estimated, in the following we discuss the two types of models already emphasized, i.e., the MSIAH(K, 1) and MSIH(K)-VAR(1) frameworks.[22]

4.3.1. Model selection

As a first step, we conduct a specification search to select the model that best describes the underlying dynamics of the yield curve. First, we want to assess whether a multi-state framework is truly required by the data. Second, if that were to be the case, we aim at selecting the most appropriate MSVAR model among the various specifications presented in Section 2. In such a framework we define a model by selecting the VAR order, the appropriate number of regimes, and which are the parameters subject to regime switching.

To assess whether a multi-state framework is appropriate, we apply a corrected LR test as the standard version of the test is not valid in this case (it fails to have an asymptotically chi-square distribution) because of a nuisance parameter problem (see Garcia, 1998). Table 4 presents the LR linearity test—i.e., the null hypothesis is K = 1 vs. a composite alternative of K' > 1—along with Davies' (1977) adjusted p-values in parenthesis.[23] Visibly, for all sets of variables (as defined by the identity of the monetary variable added to the DNS factors, if any), all selections of K' = 2 and 3, and all MSVAR models, the LR test always rejects the null hypothesis of $K = 1$ at any confidence level. We conclude that our US interest rate data require regimes, consistently with earlier literature (see e.g., Ang and Bekaert, 2002; Bansal and Zhou, 2002; Dai, Singleton, and Yang, 2007; Rudebusch, and Wu, 2007; Startz and Tsang, 2010; Xiang and Zhu, 2013).

Once we establish that $K > 1$ is the appropriate number of states, we have to select the most suitable model within the MSVAR class. We analyze two possible specifications, MISIH(K)-

[22] Estimating MSIAH(K,p) models with p equal to or exceeding 2 turned out to be unfeasible. We limit ourselves to Q = 1 or 2 because richer MSVAR models imply a substantial loss of degrees of freedom, or equivalently, very low saturation ratios that either prevent outright convergence of the EM estimation routines, or in any event make us extremely skeptical about the reliability of the resulting estimates. In fact, Table 4 reports the estimation outputs for only one such model with Q = 2 (i.e., for a M= 5 variable MSVAR), for which we managed to obtain reliable convergence statistics, that is yet characterized by as many as 141 parameters and hence by a saturation ratio below 5.

[23] Davies (1977) provides an upper bound for the significance level of the LR test under nuisance parameters.

VAR(1) and a MISIAH(K, 1). In fact, it is known that when the data require a regime switching structure, especially when K ≥ 3 is considered, it is common to find that $p > 1$ becomes redundant, see e.g., Guidolin and Timmermann (2006). To select one of the four models among MISIH(2)-VAR(1), MISIH(3)-VAR(1), MISIAH(2,1), and MISIAH(3,1), we resort to three information criteria: the Akaike information criterion (AIC), the Hannan-Quinn criterion (HQ), and the Schwarz criterion (SIC). As can be seen from Table 4, none of the criteria suggests MSIH(2)-VAR(1) or MSIAH(2,1) as the best models, hence the selection is restricted to models with three regimes. Among them, for three out of seven models all the information criteria uniformly select a MISIH(3)-VAR(1). In three cases, while AIC advises a richly parameterized MSIAH(3,1) model, the more parsimonious H-Q and SIC criteria point towards a MSIH(3)-VAR(1) specification. As observed before in a variety of applications, it is typical of information criteria to provide heterogeneous indications, especially when the AIC is compared to H-Q and SIC (see, e.g., Fenton and Gallant, 1996). In only one case—a model in which the three DNS factors are augmented by the FFR—all criteria point towards a MSIAH(3,1). However, in all cases in which the information criteria selected a MSIAH(3,1) model, this was characterized by either 96 o 141 parameters to be estimated (with a total number of observations ranging between 684 and 1,608, depending on M) and, as a result, by saturation ratios well below a suggested, healthy threshold of 10-20:[24] 12.5 in three cases, 7.1 in one case, when the series on the size of the Fed's balance sheet is added to the DNS factors. Because in a majority of cases a more parsimonious MSIH(3)-VAR(1) had been selected by the information criteria, because this model leads to acceptable saturation ratios between 10.7 and 28.7, and because of a strong prior from earlier literature (see Xiang and Zhu, 2013), we settle on a uniform use of MSIH(3)-VAR(1) models for all sets of variables under analysis.[25]

4.3.2. *Three-state MSIH VAR models*

The selected MISH(3,1)-VAR(1) framework has been estimated for seven specifications: one for the three DNS factors only and other six for four-variable models built as described in Section 2.

---

[24] The saturation ratio is defined as the ratio between the total number of observations and the total number of parameters to be estimated in a model. Values of this ratio below 10 are considered deeply unsatisfactory because they may lead to inaccurate forecasts and unstable optimization, much dependent on starting conditions.

[25] In Table 4, we present also ML estimation statistics for a 5-variable Markov switching model including both the rate of growth of the size of the Fed's balance sheet and its average maturity. This was the only case in which we could achieve convergence and stable estimates for all four MSVAR models in the table. However, the models for 5 variables lead to alarming saturation ratios between 4.9 and 10.2 and has not been pursued further.

In Tables 5-7, we report the estimation outputs for three cases. All estimated processes imply point estimates of the VAR(1) matrix with eigenvalues outside the unit circle, i.e., they are covariance stationary in all regimes. The same holds for single-state VAR(1) models, even when occasionally a Reader might detect coefficients on the main diagonal of the $\mathbf{\Phi}_1$ matrices exceeding 1. Complete, full-sample estimation results for all seven cases (as well as for the corresponding, single-state VAR(1) models) are collected in an Appendix available upon request. Table 5 reports the estimates obtained for the factors in isolation. First, while most regime switching intercepts are not precisely estimated, the time invariant VAR(1) matrix collects only statistically significant (at a size of 10% or less) estimates and therefore stops displaying a triangular structure. When regimes are taken into account, not only there is evidence that the DNS factors are tightly interconnected because they jointly shift regimes, but also display a stronger, more accurately estimated cross-serial correlation structure. However, the point estimates of the coefficients are not drastically different (but the curvature factor becomes less persistent) and in fact they all retain the same signs as in Table 3: therefore a MSVAR framework just allows a more accurate estimation of the VAR components, even though the regime-specific intercepts are not always precisely estimated. Interestingly, also the regime-specific covariance matrix stops being approximately diagonal because in at least two regimes out of three, all factor shocks are positively and significantly correlated. Also in this case, the cross-sectional variation of interest rates of different maturities is explained almost entirely by the MSVAR model for the DNS factors, with all the R-squares exceeding 99%, and those for the intermediate tenors now approaching essentially 100%.

As far as their interpretation is concerned, an inspection of Table 5 and of the smoothed probabilities plotted in Figure 4 reveals that regime 1 is a state of stable, low rates characterized by a steeply upward sloping and concave term structure. The regime is the most persistent of the three, with a "stayer" ($p_{1,1} = \Pr(S_t = 1|S_{t-1} = 1)$) probability of 0.90 and an average duration of approximately 10 months. In this regime, shocks to level and slope have a strongly negative correlation, probably indicative of a period in which the efforts to reduce the short-term rates were aimed at supporting a steepening of the yield curve to express an improvement in the expectations concerning medium- and long-term growth prospects (hence, inflation). Figure 4 reveals that regime 1 tends to follow the start of recessionary periods (e.g., late 1993, 2002-2003, and 2009-2014) and especially corresponds to the great recession and the successful recovery of the US economy spurred by unconventional monetary policy measures; this is also reflected by the modest volatility of Level and Slope and therefore of the rates themselves.

Regime 2 is a state of high average rates, flat and approximately linear or even convex (when downward sloping) term structure, in which shocks to the factors are slightly more volatile than under regime 1 (with the exception of Curvature). This state is more persistent than regime 1 is, with a "stayer" probability of 0.94, which implies a duration in excess of 15 months. Figure 4 shows that the regime characterizes financial bull markets and the mature stages of expansionary periods, such as the long booms of 1996-1999 and 2004-2006, which were both characterized by high effective Fed fund rates.

Regime 3 is the least persistent of the states with an average duration of 4.4 months and a "stayer" probability of 0.77. This is a state characterizing the early stages of bear markets and, more generally, of economic downturns as revealed by the implied intermediate average level and slope of the yield curve, that are both lower vs. regime 1. In this regime, shocks to all factors are highly volatile, with an estimated volatility that may be twice the estimates for regime 1. Figure 4 shows that regime 3 coincides with NBER recession periods and occasionally spikes in correspondence to crisis episodes. For instance, this regime occurs in correspondence to the fallout from the Russian default in 1998, late 2000-2001 when the internet bubble burst, and finally the GFC of 2007-2008. Interestingly, the estimated transition matrix has a structure that reveals that regimes 1 and 3 "communicate" with relatively large and statistically significant probabilities: $\Pr(S_t = 3|S_{t-1} = 1) = 0.09$, $\Pr(S_t = 1|S_{t-1} = 3) = 0.16$;[26] on the contrary, regimes 1 and 2 and 2 and 3 are less likely to communicate. Importantly, this special structure justifies why regime 1, in spite of a lower duration and "stayer" probability carries a higher ergodic probability (0.485) than regime 2 does (0.254).This suggests that in the fixed income market, one typically switches from a state of turbulent markets in times of crisis, to more quiet markets, often under the influence of policy interventions, like in the final part of our sample, 2009-2014.

Even though our working hypothesis is that adding variables that capture the stance of monetary policy to a regime switching DNS set up improves forecasting power with special emphasis on the crisis sub-sample, Table 6—with reference to a M = 4 system expanded to include the log of the size of the Fed's balance sheet—shows results qualitatively similar to Table 5 as far as the parameter estimates and regime classification are concerned.[27] In fact, the fraction of accurately estimated parameters grows. Regime 1 is still interpretable as a regime of low rates, moderate

[26] Because of the "by-row" sum up constraint on the transition probabilities, in each row only the standard errors of K – 1 probabilities can be computed.

[27] In the case of the macro variables, the implied unconditional means reported in Tables 6 and 7 refer to the first difference of the log and hence to the implied monthly growth rate.

slope and concavity of the yield curve and in which the size of the Fed's balance sheet grows at a (relatively, compared to the average) rapid rate; shocks to both DNS factors and the log-size of the balance sheet are not very volatile and moderately correlated, to indicate that as measured by log-asset size, monetary policy is dominated by a very persistent dynamics (close to a stochastic trend) of the Fed's assets and not by a rapid response to market conditions. This regime is rather persistent (average duration is almost 8 months) and characterizes approximately 30% of a 1996-2015 sample. An unreported set of pictures of smoothed state probabilities similar to Figure 4 shows that regime 1 still captures an unconventional monetary policy regime, even though in this case between 2011 and 2015 there is a tendency of regime 1 to alternate with regime 2. This is sensible, as regime 2 describes a better state of higher average rates within upward sloping and approximately linear (if not slightly convex, with a hyperbolic shape) term structures, typical of the initial stages of economic expansions. However, this regime lasts on average 3.5 months, characterizes only 22% of our sample and in this respect is more "peculiar" and hence short-lived than regime 2 in Table 5, when only the DNS factors were considered. Regime 3 therefore plays a residual role in logical terms, describing "normal times", and it carries a high persistence of 2 years, and dominates our sample (its ergodic frequency of 48%); therefore, in this regime the yield curve is at intermediate levels, it is upward sloping and concave, as we would expect of a normal state. In this respect, regime 3 in Table 6 is very similar to regime 2 in Table 5.

Interestingly, the VAR(1) matrix in Table 6 turns out to be lower triangular, which indicates that the current size of the Fed's balance sheet does forecast subsequent DNS factors but not the opposite, similarly to the empirical results in Diebold, Rudebusch, and Aruoba (2006), who however use data for output growth and inflation. In fact, we have estimated a restricted, lower triangular MSVAR model in which the log size of the Fed balance sheet is not predicted by past DNS factors finding a maximized log-likelihood of 561.25; when compared to the unrestricted maximized log-likelihood of 566.32 in Table 4, under six restrictions, this gives an (asymptotically chi-square distributed) likelihood ratio test statistic of 10.134, which leads to a p-value of 0.119. Because the p-value seem indecisive and because the restriction was rejected for a number of other monetary policy indicators (as in Table 7), in the rest of the paper we work with unrestricted MSVAR models even though we are aware that in some cases this may worsen they realized, OOS forecasting performance. As such our results in Sections 5 and 6 ought to be taken as lower bounds of the potential ones.

Table 7 reports the ML estimates of a MSIH(3)-VAR(1) model similar to the one in Table 6, but when the size of the quantitative easing expansion engineered by the Fed in 2008-2015 is measured more indirectly, through its effects on the total quantity of money, as captured by the log-MZM Divisia index. Qualitatively, the estimation results, nature and persistence of the three regimes implied by the table are very similar to those commented with reference to Table 6; for instance, also in this case, most of the unconventional monetary policy effects on the shape and dynamics of the US risk-free yield curve are captured by a rather persistent regime 1, the smoothed probability of which raise towards 1 and remain close to that level during most of the period 2009-2014. In fact, the point of displaying Table 7 alongside Table 6 is to document the fact that, even though qualitatively the estimates and predictions from the two models may differ, in qualitative terms they have similar interpretation and implications. The same can be said with reference to the remaining three MSIH(3)-VAR(1) models that alternatively expand the set of variables composed of the three DNS factors to the time series of the average maturity of the asset-side of the Fed's balance sheet, its composition in terms of the percentage represented by Treasuries, and the effective Fed funds rate. In all cases, we recognize as a regime 1 a state that tends to characterize most of the final part of our sample and that is characterized by low volatility of the shocks, moderate correlations, low average rates, and intermediate levels of the Slope and Curvature factors.[28]

5. FORECASTING WITH MACRO-AUGMENTED REGIME SWITCHING DNS MODELS

To investigate the value created by the regime-switching framework and our conjecture that regimes exist in which DNS factors can be predicted not only from their own past but also using the stance of monetary policy, after a description of our pseudo OOS design, in this section we test three hypotheses. The first question is whether the MSVAR models (with and without macroeconomic factors) outperform a single state VAR in terms of their forecasting accuracy. This issue is related to the role of regime shifts in forecasting with DNS models and extends the work by Xiang and Zhou (2013). Secondly, we investigate whether the addition of macroeconomic variables to a baseline MSVAR model for the DNS factors may generate

[28] In the case of the MSVAR model including the FFR, regime 1 is characterized by a near-zero, basically constant rate. Moreover, the estimated VAR matrix shows that the FFR highly depends on the lagged values of Level and Slope although the opposite is less evident (i.e., FFR do not show high predictive ability for Level and Slope); on the other hand, the FFR has significant forecasting power for Curvature. In the case of the models including average maturity and the composition of the Fed's balance sheet, regime 1 implies an above-average maturity and a below-average weight assigned to Treasuries, as we would expect from Figure 2.

additional forecasting power and whether such a power changes across different periods, with particular emphasis on the GFC and the ensuing policy measures. Finally, we compare the forecasting performance of the baseline VAR and MSVAR models with that of a natural benchmark, much used in the literature, the Random Walk (henceforth, RW) that recently has been shown to have performed well in forecasting US spot rates since the GFC (see Eo and Kang, 2018).

5.1. The pseudo out-of-sample recursive forecasting experiment

We implement a typical recursive, expanding window, pseudo OOS design. We start by estimating all M-variables MSIH(K)-VAR(1) models—as defined by K = 1, 3 and by whether M = 3 or 4 (just DNS factors or factors expanded to include macro variables)— and the benchmarks (see Section 5.2) on a sample ending in December 2005 to forecast both the DNS factors and the associated yields one-month forward, as of the end of January 2006.[29] We then add one vector of estimated DNS factors, in this case corresponding to January 2006, and perform afresh estimation on this expanded data set to compute one-month ahead forecasts for February 2006. We proceed in this fashion until we exhaust the useful panel of estimated DNS factors, in June 2015, when we obtain the last one-step ahead prediction of factors and hence yields, for July 2015. Therefore, we obtain a sequence of 113 predicted DNS factors and the associated 113 predicted yield curves (in terms of the 9 maturities listed in Section 3).

We supplement the models described in Section 4 with the standard benchmarks commonly employed in the yield curve forecasting literature (see, e.g., Guidolind and Thornton, 2018):

- Random walk: $\hat{\boldsymbol{f}}_{t+1|t} = \boldsymbol{f}_t$;
- VAR(1) model: $\hat{\boldsymbol{f}}_{t+1|t} = \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{f}_t,$ which nests DL (2006).

However, differently from what is typical of the literature, such benchmarks are applied to the DNS factors to obtain $\hat{\boldsymbol{f}}_{t+1|t} \equiv [\hat{L}_{t+1|t} \;\; \hat{S}_{t+1|t} \; \hat{C}_{t+1|t}]'$ and then converted into yield predictions using:

[29] The estimation sample starts in January 1982 for four of the six choices of the variables included in $\boldsymbol{x}_t$ and in July 1996 for the models based on the log-size of the Fed's balance sheet and the proportion represented by Treasuries. We also compute forecasting performance for the factors themselves because the trading strategies in Section 6 are directly based on factor predictions. Of course, it is easy to extend our OOS design to multi-step ahead forecasts, but because we aim at analyzing the economic value of such forecasts, it seems more relevant to report results focusing on short-horizon predictions.

$$\begin{bmatrix} \hat{y}_{t+1,\tau_1} \\ \hat{y}_{t+1,\tau_2} \\ \vdots \\ \hat{y}_{t+1,\tau_N} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_n}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{bmatrix} \begin{bmatrix} \hat{L}_{t+1|t} \\ \hat{S}_{t+1|t} \\ \hat{C}_{t+1|t} \end{bmatrix}. \tag{8}$$

As in Diebold, Rudebusch, and Aruoba (2006), the yields only load on the predicted DNS factors and not on the macroeconomic factors.

5.2. Evaluating the forecasting performance

We define the time *t* forecast error from model *m*, for a target variables *j* as:

$$e^{j,m}_{t,t+1} \equiv y^{j}_{t+1} - \hat{y}^{\,j.m}_{\,t+1|t}\,, \tag{9}$$

where $\hat{y}^{\,j.m}_{\,t+1|t}$ comes from either one of the models described in Section 4 or from a benchmark, and it can be a DNS factor or a yield prediction. Once we have obtained the time series of forecast errors $\{e^{j,m}_{t,t+1}\}$, for *t* that goes from December 2005 through June 2015 and for each model *m*, we proceed to compute several measures of prediction accuracy.

The popular *Root Mean Squared Forecast Error (RMSFE)*, is defined as

$$RMSFE^{j,m} = \sqrt{\frac{1}{P}\sum_{t=1}^{P}\left(e^{j,m}_{t,t+1}\right)^2}\,, \tag{10}$$

where P = 112 is the total size of the pseudo OOS period. However, RMSFE is a point estimate of forecast accuracy and does not take into consideration sampling uncertainty; because we are interested in comparing the predictive power of alternative models (e.g., MSVAR models vs. the baseline VAR(1)), to test whether these models carry statistically significant differences in forecasting accuracy, we resort to the Diebold and Mariano (1995)-West (1996)-McCracken (2004) nonparametric tests for non-nested models, which also account for any incremental sample variation in forecast errors due to parameter uncertainty. The DMWM test for a pair of models indexed as m_1 and m_2 is based on the statistic

$$DMWM^{j}(m_1, m_2) \equiv \frac{\bar{d}}{\sqrt{\widehat{Var}(\bar{d})}}, \tag{11}$$

where $\bar{d}$ is an average over P observations of the values taken by some differential in loss functions, $d_t \equiv \ell(e^{j,m_1}_{t,t+1})$ - $\ell(e^{j,m_2}_{t,t+1})$ where $\ell(\cdot)$ is a generic loss function, and $\widehat{Var}(\bar{d})$ is an estimator of the variance of $\bar{d}$. In practice, in coherence with the RMSFE in (10), in the paper we use $\ell(e^{j,m}_{t,t+1})$

= $(e_{t,t+1}^{j,m})^2$. The DMWM statistic has an asymptotic standard normal distribution under the null hypothesis that E[d_t] = 0, which corresponds to a null of no differential predictive accuracy. Following standard practice, the variance of d_t is estimated using a heteroskedastic-autocorrelation consistent estimator,

$$\widehat{Var}(\bar{d}) = P^{-1}\left[\hat{\varphi}_0 + P^{-1}\sum_{j=1}^{B-1}(P-B)\hat{\varphi}_j\right] \quad (12)$$

where $\hat{\varphi}_j \equiv (P-j)^{-1}\sum_{t=j+1}^{P}(d_t-\bar{d})(d_{t-j}-\bar{d})$. West (1996) has shown that in general, when the loss functions depend on estimated parameters, (12) provides a valid estimate of the asymptotic variance of $\bar{d}$ only in special circumstances, e.g., when the models are estimated consistently by OLS and the loss function is a squared function (i.e., under MSFE). In general, however, the structure of $Var(\bar{d})$ is

$$Var(\bar{d}) = \widehat{Var}(\bar{d}) + 2\vartheta_{dm}(\boldsymbol{FU}Cov'(\boldsymbol{d},\boldsymbol{s})) + \vartheta_{mm}\boldsymbol{FU}Var(\boldsymbol{s})\boldsymbol{U}'\boldsymbol{F}' \quad (13)$$

Where, in the case of a recursive forecasting exercise, $\vartheta_{dm} \equiv$ 1-(R/P)ln(1+(P/R)), $\vartheta_{dm} \equiv$ 2[1-(R/P)ln(1+(P/R))], R is the length of the estimation sample from which the back-testing exercise is initialized (288 and 120, depending on the model considered). **F** and **U** are matrices that depend on the data used in estimation as well as on the derivatives of the loss functions with respect to unknown parameters to be computed in correspondence to the true but unknown population parameters (see McCracken, 2004). Finally, **s** denotes the time series of the scores generated by each model, when estimation occurs by QML.[30]

The *Relative RMSFE* represents another popular measure that allows us to compare the performances of different forecasting models. This is calculated as the ratio between the RMSFE of model *i* and the RMSFE of a given benchmark model. Thus, by construction, values of this statistic lower than one indicate that model *i* produces more accurate forecasts than the benchmark. Specifically, we shall compare RMSFEs of the MS models with two very successful benchmarks in the term structure literature: the Random Walk (RW), and the VAR(1)-DNS model.

The *Mean Absolute Percentage Error (MAPE)* is perhaps the most widely used unit-free measure of predictive accuracy (see, e.g., Armstrong and Collopy, 1992) and is defined as:

$$MAPE^{j,m} = \frac{1}{P}\sum_{t=1}^{P}\left|\frac{e_{t,t+1}^{j,m}}{y_{t+1}^{j}}\right|. \quad (14)$$

[30] McCracken (2004) proposes to estimate **F** without deriving the functional form for the derivatives of the loss function or making strong assumptions about the joint distribution of the observables. The idea is that unknown derivatives can be approximated numerically by using the finite difference method.

Because it measures the sample mean of absolute errors as a percentage of realized values, MAPE has the advantage of being scale-independent. However, because it penalizes positive errors more than negative errors, MAPE has been sometimes criticized (see, e.g., Hyndman and Koehler, 2006). Moreover, it may become unstable when y^j_{t+1} is zero or very close to it: however, in our sample and even with 3-month rates this was never the case, although these yields become very small between 2009 and 2014.

The *Hit Ratio* (HR) is the percentage of correct sign predictions offered by a given model *i*,

$$HR^{j,m} = \frac{1}{P}\sum_{t=1}^{P-1} I_{\left\{\Delta y^j_{t,t+1}\Delta\hat{y}\,^{j,m}_{t,t+1}>0\right\}}\,, \tag{15}$$

where $I_{\left\{\Delta y^j_{t,t+1}\Delta\hat{y}\,^{j,m}_{t,t+1}>0\right\}}$ is an indicator variable that takes a unit value when $\Delta y^j_{t,t+1}$ (the *actual* variation of the target variable *j* between *t* and *t+1*) and $\Delta\hat{y}\,^{j,m}_{t,t+1}$ (the *predicted* variation of variable *j* between *t* and *t+1*) have the same sign. As often demonstrated in finance applications, for many trading strategies (especially the directional ones) it is more relevant that the forecasting model produces predictions with a correct sign rather than predictions that are quantitatively (more) accurate (see, e.g., Christoffersen and Diebold, 2006; Leitch and Tanner, 1991). We shall return to this point in Section 6, when we assess the relative economic performance of our models.

5.3. Empirical results

Table 8 reports the realized RMSFE results for six MSVAR models and the two benchmarks. Panel A concerns DNS factor forecasting and panels B-D cover yield forecasting. In each panel and in correspondence to each combination of factor (or tenor)/sample, we boldface the model leading to the lowest realized RMSFE. Because we also aim at testing whether the inclusion of monetary variables my differentially improve forecasting performance in correspondence to the GFC, it is instructive to compute performance accuracy not only over the full 2006-2015 pseudo OOS period, but also to distinguish between Crisis (August 2007 – March 2009) and Non Crisis periods (the rest of the OOS period). Because January 2006 – July 2007 is relatively short, the Non Crisis period almost entirely corresponds to the late management and extension of unconventional monetary policies.

Visibly, with reference to the full sample, there is considerable heterogeneity concerning the identity of the best performing model; in fact, for long-term yields and the Level factor, it also occurs that a RW offers the best OOS performance. However, in the two sub-samples, it is always the case that richer MSVAR models that generally contain macro predictors lead to the most accurate performances. In particular, in the Crisis sample, out of a total of 12 cells (9 yield tenors and three DNS factors), the MSIH(2)-VAR(1) that includes the average maturity of the Treasuries in the Fed's balance sheet yields the lowest RMSFE six times and the model extended to the Divisia Index log differences scores best four times. During the Non-Crisis period, the regime switching model that includes the first difference in the log-size of the Fed's assets gives the lowest RMSFE six times and the one including the fraction of the Fed's balance sheet composed by Treasuries, three times. Even though it remains complex to grasp why certain variables may work best only sometimes, it is a fact that during the GFC the Fed reacted at first by changing the composition of its balance sheet (by selling Treasuries, also to support market liquidity and refinancing operations by banks, and purchasing MBS and agency debt) and subsequently by also exponentially expanding its size. Especially in the case of yields (the forecasts of the Slope and Curvature factors behave more erratically, while Level behaves coherently with long-term yields), the predictive performance of MSVAR models seem to reflect this, by selecting the predictive framework accordingly. For instance, while over the full-sample the best DNS multivariate model to forecast 5-year rates is a random walk for the factors with a RMSFE of 27 bps, during the GFC the lowest RMSFE (37 bps) is given by a MSVAR that includes the maturity variable; in the post-crisis period, the lowest RMSFE (only 14 bps, when rates are much lower) is given by a MSVAR that includes the log-size of the balance sheet, consistently with a QE-driven period.[31]

Figure 5 documents how sensible it can be to switch from including the average maturity of the Fed's balance sheet in favor of its size in correspondence to the end of GFC, during 2009. The top row of plots shows the time series of the two variables and emphasizes that average maturity "jumps" twice, in the second half of 2008 and then again between late 2011 and mid 2013; on the contrary, the log-size of the Fed's balance sheet records a sizeable "jump" after Lehman's bankruptcy, between October 2008 and March 2009. Such discrete changes are visible and represent signals that could be hardly ignored in real time, and yet one of them would have led to incorrect trades and a reduction in cumulative strategy returns. In the second row, to the left, we have plotted the evolution of the yield curve between June and December 2008: the term

[31] The fact that in overall terms, the RW outperforms all other models reminds us of typical results in the literature (see Duffee, 2002) but it is not a bizarre result: if better fine-tuned econometric models may give strong regime-dependent OOS performances but are unstable in such performance, it is possible that both maturity- and size-augmented MSVARs may be dominated by the RW. This means that if an investor were to acknowledge the existence of regimes but were unable to monitor their occurrence in real time (which is not as demanding as accurately forecasting them but requires to make inference on them), then a DNS RW benchmark may be a useful support to investment decisions. Section 6 further discusses this point.

structure shifts down by almost 150 bps, it becomes steeper and its concavity declines; indeed the estimated DNS factors reported in the third row, to the left, follow exactly this path, as shown by the arrows. These are precise trading signals on the three DNS factors that would have been predicted by the mid-2008 jump in the average maturity of the Fed's balance sheet. As a result, Table 8 correctly indicates that during 2008 a MSVAR model that includes this variable would have outperformed all other models. However, by 2010, it would have become unwise to forecast with this model: when in late 2011 a second jump occurs as a result of the MEP, the effect on the yield curve is ambiguous and while level declines and slope steepens at first, these outcomes subside in late 2012 (see rightmost plot in the second row), as also shown by the estimated DNS factors in the rightmost plot of the third row. In fact, after 2009, large jumps in monetary variables tend to imply much weaker trading signals, as shown in the fourth row of Figure 5 in that concerns the yield curve and the estimated DNS factors between October 2008 and March 2009: the yield curve at first shifts down and flattens but these shifts are reversed and fail to provide exploitable trading signals. A MSVAR model that were to include the average maturity of the Fed's balance sheet instead of its size would give signals that are unlikely to generate trading profits. In any event, Table 8 gives evidence that regimes are needed to support accurate predictions. In general, MSIH(3)-VAR(1) for the DNS factors forecasts better that a single-state VAR(1) and adding variables that capture the stance of monetary policy also tends to help, at least most of the time, which may be taken as an indication that past interest rates do not incorporate all information useful to predict future rates.

Table 9 confirms the results in Table 8 by displaying the Relative RMSFEs vs. the VAR(1) benchmarks with no regimes. We use the plural because for each model, as characterized by a selection of one macroeconomic variable (or none), we build a corresponding M-variable (M = 3 or 4) VAR(1) that includes the same variables. Moreover, in this case besides boldfacing all relative RMSFEs below 1, in the case of *the full sample* (to retain sufficient data) we also report (using stars next to the RMSFE statistics) indications concerning the outcomes of two-tailed DMWM tests of the null that the models in the tables have identical MSFE vs. the VAR(1). Strikingly, when used to forecast either yields in the full sample or the Level and Slope factors, all models lead to relative RMSFEs well below 1; most of them are also statistically significant, in the sense that the null of no differential predictive accuracy can be rejected with p-values of 1% or smaller in the case of yields, and for most models with p-values of 5% or smaller in the case of the DNS factors. This evidence is overwhelming for the Level factor and for short- to medium-term rates, in the sense that the relative RMSFE is below 0.9 (0.6 for the rates) for all models with regimes.

To give further support to these conclusions, in Table 10 we offer a variation of Table 9 in which the MSVAR models are pitted against the RW benchmark which, based on the evidence in Table 8, provides a much taller bar to clear and that can be seen as a highly parsimonious, restricted version of a VAR(1) model for the factors (one obtains the RW from a VAR(1) by assuming $\mathbf{\Phi}_1 = \mathbf{I}_3$). Given that it has been reported that for econometric models of interest rates, it is generally hard to outperform the RW, Table 10 is encouraging. On the one hand, *on average*, MSVAR models are not better than a RW either in terms of factor or of yield prediction: in the former case, the average relative RMSFE is 1.028; in the latter case, it is 1.001, even though Table 10 reveals that during the GFC many models with regimes for the DNS factors forecasted rates better than the RW. Of course, differences are hardly significant because they are on average small. However, at least with reference to the full sample, the (full-sample) DMWM tests offer indications of a few occasions in which selected MSVAR models can be established to lead to lower RMSFE than the RW does. In the case of factor predictions, this is the case of Slope, even though the rejections are obtained with p-values between 4% and 5%. The fact that Slope can be predicted very accurately will turn out to be important in Section 6 and this is consistent with the evidence in Wu (2001) that the steepness of the yield curve is heavily influenced by monetary policy. However, spot rates at 3-, 6-, and 12-month (essentially, T-bill rates) can surely be predicted more accurately with MSVAR DNS factors models than with a RW. Most recorded differences in MSFEs are in fact highly statistically significant, even when parameter uncertainty is (albeit imperfectly) taken into account. For these maturities, the MSVAR models including the size of the Fed's balance sheet, the average maturity of the Treasuries in the balance sheet, and the rate of growth of the Divisia aggregate, often take the relative RMSFE well below 0.9. It is instead harder to outperform the RW in OOS tests applied to maturities exceeding 2 years and very hard at the 5-, 7-, and 10-year tenors.

Nonetheless, Tables 8-10 are all based on the same loss function, the classical square loss, $\ell(e^{j,m}_{t,t+1}) = (e^{j,m}_{t,t+1})^2$. Even though we have not performed all DMWM tests afresh, to check the robustness of our earlier findings to using a different loss function, in Table 11 we report results obtained using the MAPE criterion, which is based on the absolute value of the errors scaled by the value of the target. Table 11 reports MAPE scores for our recursive OOS period. We fail to note any major qualitative differences between MAPE results and the RMSFE findings in Table 8:[32] in the Crisis sample, the MSIH(2)-VAR(1) including the average maturity of the Treasuries in

[32] In Table 11, MAPE declines with the yield maturity because in (11) we scale by the level of the predicted yield, which is on average increasing with maturity.

the Fed's balance sheet yields the lowest MAPE ten times (out of 12 possible); during the Non-Crisis period, the MSVAR that includes the first difference in the log-size of Fed's balance sheet assets gives the lowest MAPE seven times and the one including the fraction of the Fed's balance sheet composed by Treasuries, four times. Although the overall predictive performance is less impressive for the full-sample period, when the RW often returns the lowest MAPE (especially for medium- and long-term bonds), regimes are needed to support accurate predictions and adding variables that capture the stance of monetary policy reduces MAPE, an indication of partial inefficiency of the Treasury market that would not always incorporate all available information on the stance of monetary policy.

Because the ability of nonlinear time series models to forecast the sign of the changes in target variables has been proven to be essential to their economic value in financial applications, in Table 12 we report the realized Hit ratios from the models under consideration. Note that we omit the ratios of the RW from factor forecasts because, since the RW forecasts are always of "no change" (i.e., the factor at $t + 1$ is expected to be the same as in t), the ratio will always be zero or close to zero. In the table, we draw red squares around the maximum Hit ratio across models and boldface all ratios that exceed 50%, for added visibility. Although the general feeling for the results is similar to Tables 8-11, Table 12 marks a shift of the balance of the evidence on predictive power in favor of MSVAR models expanded to include the log-size of the Fed's balance sheet. In fact, out of a total number of best model cells equal to 36 across factors/yields/sub-periods, 11 are occupied by this model; no other model stands out. While with reference to the non-crisis period, the OOS outperformance of the model including size is to be expected in the light of earlier comments, the finding that also during the crisis sub-sample this variable improves accuracy is interesting. In the case of the full sample, where the length of the OOS period makes our remarks more reliable, we note that while a top Hit ratio of 66% for the Slope factor is rather impressive, in the case of yields, we observe top performance of around 55% in the case of short rates that increases to exceed 60% for longer rates.[33]

In the light of the previous results, we also experimented with a M = 5-variable MSIH(3)-VAR(1) model, the one reported at the bottom of Table 4, that includes, besides the DNS factors, both the log-asset and the Fed's budget maturity series. Our goal is to assess whether the forecasting

[33] Interestingly, although the length of the OOS is rather limited and any inference ought to be taken with caution, the percentage of correct sign prediction is generally higher over the Crisis sub-sample than in the full sample. Panel A of Table 12 traces back this result to the fact that during the GFC it becomes relatively easier, through recursive estimation, to predict the most volatile factors implicit in the yield curve (Slope and Curvature) vs the least variable one (Level).

power of MSVAR models further improves when, in essence, we try to account for both the QE and MEP programs simultaneously. Yet, such a model achieves OOS predictive results—both in terms of realized OOS RMSFE and relative RMSFE—similar to four-variable MSVARs in which either the log-size or Treasury maturity composition of the Fed balance sheet are investigated individually. This holds with references to both sub-periods, when the model fails to outperform the simpler four-variable MSVARs with very limited exceptions. Yet, in terms of Hit ratio, this extended model obtains superior results, and in particular the highest Hit ratio for both the Level and Slope factors (66% and 62%, respectively). However, the advantage obtained in the Slope forecasts do not translate into better sign predictions of the short-term yields, although for the 10- and 30-year yields this five-variable model reaches the highest Hit ratios. On the one hand, these results strengthen our feeling that variables measuring the stance of monetary policy through the size and maturity of the Fed's balance sheet carry forecasting power for Level and Slope, respectively: whereas the size of the Fed balance sheet mainly influences the general level of the interest rate, the maturity composition of the Fed's portfolio of Treasuries affects the spread between long-, and short-term yields. On the other hand, the low saturation ratio of this model and its disappointing predictive accuracy for short-term rates advises us not to pursue further tests of this model.

6. ECONOMIC VALUE

In Section 5, we have shown that MSVAR models, and especially some of the models extended to include macroeconomic variables, can be successfully used to predict the movements of the yield curve. Therefore, at least in some regimes, the shape of the riskless yield curve—as summarized by the DNS factors—fails to summarize all available information on the current and future stance of monetary policy. In this Section, we dig deeper in earlier results and investigate whether and to what extent such evidence of predictability can be used to implement a systematic trading strategy using bond butterfly strategies. We first explain how to build butterfly strategies and the trading rule used to set them up. Subsequently, we illustrate how to measure their economic value, and discuss their empirical performance.

6.1. Trading strategies

Bond butterflies are among the most common active strategies used by trading desks to exploit views on spot interest rate changes. Here we follow Fabozzi, Martellini, and Priaulet (2007) who have shown how to use them within a NS framework. To bet on specific views on changes of the shape of the term structure, one can implement at least two different types of trading strategies:

one that bets on the changes of the individual factors of the yield curve; another one that directly bets on the implied yield changes. In the former case, we construct a set of semi-hedged, long-short, zero outlay (self-financing) strategies in which the portfolio is exposed to movements of one factor at the time, remaining neutral to all others. We describe such strategies below. For instance, in the case of Level bets, we have to neutralize the exposure of the portfolio to Curvature and Slope movements and as a result, when we forecast a downward movement of the yield curve we cannot buy all the maturities in the same amount (like we do in case of yield bets) because otherwise we would be exposed to the risk of changes in the Slope or Curvature. Instead, we have to combine long and short positions in order to be hedged from non-parallel shifts of the yield curve. In the latter case, using a given model, we derive time t predictions of yields for time t + 1: we then proceed to sell (buy) all Treasuries for which the rate is predicted to increase (decrease), according to an equally weighted scheme (these are sets of bullet trades); the strategy is turned into a self-financing one by adjusting all short (long) positions to cover the sum of all long (short) positions when the number of long (short) positions exceeds the number of the short (long) ones.[34]

As far the butterfly bets on DNS factors are concerned, for instance, suppose that a trader predicts a change in the Curvature of the term structure between t and t + 1, but she has no view on Level and Slope movements. Therefore, she needs to build a portfolio that is exposed to Curvature changes but insensitive to Level and Slope changes. To this purpose, she may design a duration-weighted hedged and self-financing butterfly strategy using three bonds with different maturities (a short-, a medium-, and a long-term bond). The portfolio weights of such a strategy are obtained as the solution to the following system of equations:

$$\begin{cases} q_s\, \mathrm{MD}_s \mathrm{L}_s + q_l\, \mathrm{MD}_l \mathrm{L}_l + q_m\, \mathrm{MD}_m \mathrm{L}_m = 0 \\ q_s\, \mathrm{MD}_s \mathrm{S}_s + q_l\, \mathrm{MD}_l \mathrm{S}_l + q_m\, \mathrm{MD}_m \mathrm{S}_m = 0 \\ q_s + q_l + q_m = 1 \end{cases}, \tag{16}$$

where q_s , q_m , and q_l are the amounts of the short-, medium-, and long-term bonds (interpretable as weights thanks to the third constraint), MD_s, MD_m, and MD_l are the modified durations of the short-, medium-, and long-term bonds;[35] L_s, L_m, and L_l are the sensitivities of the short-, medium-, and long-term yields to the Level factor as calculated in Nelson-Siegel's model;

[34] For instance, if the predictions from a given model imply that 7 tenors ought to be purchased and 2 of them sold, given an equal weighting scheme for the 7 long tenors, each of the 2 tenors short must then be leveraged according to a 3.5 = 7/2 factor. In case all tenors should be purchased, we arbitrarily finance the position by shorting 3-month T-bills; in case all tenors should be shorted, we arbitrarily invest the proceeds from going long in Treasuries between 1-month and 10-year tenors, in 30-year Treasuries.

[35] The modified duration is a measure of the price sensitivity, hence the risk exposure, of a bond, defined as the percentage change in price of a bond for a unit change in yield.

finally, S_s, S_m, and S_l are the exposures of the short-, medium-, and long-term yields to the Slope factor (i.e., the loadings of Slope). Moreover, since $\mathrm{L}_s = \mathrm{L}_m = \mathrm{L}_l = 1$ by construction, hedging against the Level factor is equivalent to a duration-neutral condition. Therefore the problem in (16) simplifies to:

$$\begin{cases} q_s\,\mathrm{MD}_s + q_l\,\mathrm{MD}_l + q_m\mathrm{MD}_m = 0 \\ q_s\,\mathrm{MD}_s = -q_m\mathrm{MD}_m\gamma_C \\ q_s + q_l + q_m = 1 \end{cases}, \tag{17}$$

where: $\gamma_C \equiv \left(\frac{\mathrm{S}_l - \mathrm{S}_m}{\mathrm{S}_l - \mathrm{S}_s}\right)$. Similarly, another type of Nelson-Siegel-weighted butterfly can be designed to be hedged against movements in the Level and Curvature factors, while retaining exposure to changes in the Slope factor. This can be done as in equation (17), except for using a different γ coefficient, $\gamma_S \equiv \left(\frac{\mathrm{C}_l - \mathrm{C}_m}{\mathrm{C}_l - \mathrm{C}_s}\right)$, where C_s, C_m, and C_l are the sensitivities of the short-, medium-, and long-term yields to the Curvature factor (i.e., the loadings of Curvature). Of course, a third kind of strategy can be easily implemented to bet only on the Level factor and that consists of simply modifying equation (16) in order to neutralize the Slope and Curvature exposures.

Following this procedure, we build three kind of butterfly strategies, each betting on a different factor. In particular, our butterflies are based on of 2-, 5-, and 10-year bonds, set up and unwounded on a monthly basis.[36] To decide the direction of the trades (i.e., which maturities to buy and which to sell) we rely on the forecasts generated by the seven models described in Sections 4 and 5 and therefore, the direction only depends on the predicted direction of the change in the factor. Just to continue with the earlier example on Curvature bets, on the one hand, when we predict a sharply humped yield curve, we sell the “body”, the center of the butterfly (i.e., the 5-year maturity) and buy the “wings” (i.e., the 2- and 10-year maturities); on the other hand, when we predict a decline in Curvature, we flip around the direction of the trades, i.e., we buy the body and sell the wings. Following the same logic, we can apply this rule also to Level and Slope bets.

6.2. Measuring the economic value of the strategies

The first step towards measuring the performance of a strategy is to calculate its total return. Given the definition of modified duration, we can approximate the total return between *t* and *t+1* of a bond with maturity *j* with the well-known formula,

[36] This combination of maturities is selected because it is typical of what practitioners do. To check the robustness of our results, we have also used another popular combination of maturities (i.e., 2-, 5-, 7-year maturities) finding qualitatively similar and often quantitatively indistinguishable empirical results.

$$R_j \equiv \frac{\Delta P_j}{\mathrm{P}_j} \cong -\mathrm{MD}_j \Delta y_j, \qquad\qquad j = s, m, l \qquad\qquad (18)$$

where ΔP_j is the price change of the bond between two adjacent dates, s, m, and l denote the tenors of the short-, medium-, and long-term Treasuries that are traded, and Δy_j is the change in the yield between t and $t + 1$.[37] At this point, we take the weighted average of the short-, medium-, and long-maturities bond returns to obtain the total return of the butterfly strategy based on model μ, between t and $t + 1$, $Total\ Return_t^{\mu} \equiv \sum_{j=s,m,l} q_{j,t}\, R_{j,t}$. We perform this calculation at end of each month in the pseudo OOS period (January 2006 - June 2015), thus obtaining for each strategy a vector of 112 monthly returns. Then, we can calculate the average monthly performance of each butterfly strategy as

$$Average\ Monthly\ Return_t^{\mu} \equiv \prod_{t=1}^{P} \left(1 + Total\ Return_t^{\mu}\right)^{1/P} - 1, \qquad (19)$$

which is a geometric average.

To further test whether the performance of a strategy is statistically significant (or significantly different from some other strategy), we use again DMWM tests. Here, the null hypothesis of the test is that mean returns are zero or that difference between the total returns of two strategies is zero, which are the natural yardsticks for a zero-outlay long-short strategy. In particular, in the case of the MSIH(3)-VAR(1) model for the DNS factors, the null is that of equal performance vs. a single state VAR(1) model; in all other cases, the null hypothesis is that of equal performance vs. the MSIH(3)-VAR(1) model that excludes the monetary variables. The former hypothesis is useful to test whether regimes may generate economic value, while the latter concerns whether adding additional information on the stance of monetary policy—interacted with regimes—may be exploited.

One additional way to assess the benefits of active strategies consists of examining the risk-return score they generate. For this reason, we calculate the most popular measure of realized risk-adjusted performance, the Sharpe Ratio (SR),

$$SR_P^{\mu} \equiv \frac{\sum_{t=1}^{P} Total\ Return_t^{\mu}}{\sqrt{P^{-1} \sum_{t=1}^{P} (Total\ Return_t^{\mu} - P^{-1} \sum_{t=1}^{P} Total\ Return_t^{\mu})^2}}, \qquad (20)$$

[37] For additional precision, we could have included a convexity term, which is omitted here because positions are just held for a period of one month, so that small changes in rates prevail and make first-order Taylor expansions viable. In robustness checks, we confirm that the factor forecast-driven results in Tables 13 and 14 are unaffected to the third decimal digit when the convexity term is accounted for.

in which the numerator does not include a risk-free rate adjustment when the strategy is a zero net-outlay one (any funding costs will be deducted from the total return). Because the Sharpe ratio is subject to estimation error as much as Diebold-Mariano's type statistics, and such errors derive from the fact that the recorded total returns depend on estimated model parameters which are sample statistics, we want to quantify the uncertainty surrounding SR_P^{μ} and test hypotheses concerning it.[38] Under some conditions (see Lo, 2002), the asymptotic variance of the Sharpe Ratio estimator is $V_{SR}(\mu, P) = P^{-1}[1 + \frac{1}{2}\left(SR_P^{\mu}\right)^2]$ so that the standard error of the SR estimator is:

$$\mathrm{SE}_{SR}(\mu, P) = \sqrt{P^{-1}\left[1 + \frac{1}{2}(SR_P^{\mu})^2\right]}. \tag{21}$$

In the calculation of total strategy returns, we also entertain two distinct cases. At first, we do not consider the funding costs on margin positions. This applies to traders that may be already hold the Treasuries being shorted and that therefore end up only paying a fractional haircut (that may be even zero or negative in some market states) that is in general modest in the case of Treasuries, especially at benchmark maturities such as 5-, 7-, and 10-year.[39] Because in any event retail traders and most other desks do pay some funding costs when Treasuries are shorted, we perform the same calculations (of realized OOS return, DMWM tests, and SRs) afresh accounting for the presence of these costs. According to US regulations (Regulation T established by the Federal Reserve Board), to sell a security short, a trader must establish a margin account with a brokerage firm and deposit margins in cash and/or securities as a collateral. Under Regulation T there are no restrictions on government securities, hence the margin requirements for Treasuries are independently determined by the broker. Although there are different practices that vary according to the brokerage firm, a thorough search of standard practices reveals that a margin of 10% is reasonably conservative.[40] Once a margin is established, it is necessary to calculate the cost of short selling and this will depend on the quality and liquidity of the collateral posted. Given that we assume that Treasury securities will be posted instead of cash, the repo

---

[38] In general, this is not straightforward, but under the assumption that the realized portfolio returns are approximately IID, due to the Central Limit Theorem, as P diverges, the divergence of the numerator (denominator) from its expectation multiplied by $\sqrt{P}$ converges to a normal distribution with mean zero and variance $P^{-1}\sum_{t=1}^{P}(Total\ Return_t^{\mu} - P^{-1}\sum_{t=1}^{P} Total\ Return_t^{\mu})^2$.

[39] Moreover, it is typical of hedge funds to benefit of discounts or to buy services in bundles from prime brokers, presumably in exchange of trading commissions or even sheer trading volume.

[40] See Rule 4210 by the Financial Industry Regulatory Authority (FINRA) for a detailed discussion on margin requirements in the US market. The margins are assumed to be constant for the entire period in which the strategies are implemented because our search showed that margins have not been subject to changes between 2006 and 2015.

rate should be a good proxy of the funding cost of short positions.[41] In particular, on each month of our OOS period, we use the corresponding average repo rate obtained from the Federal Reserve Bank of New York data base. As a result, we calculate the net return subtracting to the return in (18) the costs on margin positions, set to equal 0.1 × repo rate.

6.3. Empirical results

Table 13 and 14 report results concerning the OOS performance of trading strategies and hence the economic value of alternative models and variable selections. Differently from earlier forecasting evaluations, here the comparisons involve two different levels of analysis. One concerns the performance of alternative models (i.e., single-state VAR benchmark vs. the MSVAR with and without macroeconomic factors) for each trading strategy; the other concerns the performance of different strategies (i.e., bets on the yields as well as specific bets on Level, Slope, or Curvature) regardless of the model used to produce the forecasts. In other words, in addition to assessing whether using a forecasting framework may increase the realized profits of the same strategy, we also ask whether there are any differences in the risk-return profile of the strategies common to all models.

Table 13 concerns the case in which funding costs are disregarded. A bird's eye view focused on the boldfaced DMWM statistics that indicate rejection of the null of no difference vs. simpler models, reveals that significant trading profits can only be derived by either trading Slope forecasts during the Crisis period (see panel A)—and this actually obtains using MSVAR models that account for both regimes and the predictive power of macroeconomic variables that represent monetary policy—or by trading all yields without implementing butterfly strategies, again mostly during the GFC (see panel D). In light of Table 12 and of the literature on the links between accurate sign predictions and trading performance, this may have been expected, because the highest Hit ratios were found for the Slope factor, especially with reference to the crisis sub-sample.

In the non-crisis OOS and when either Level or Curvature are traded, profits are harder to seize, in the sense that even why they are positive, the Sharpe ratio are seldom significantly so.[42] In particular, exploiting the forecasts from a MSIH(3)-VAR(1) that includes DNS factors as well as

---

[41] A Repurchase Agreement (repo) represents a collateralized loan in which a party lends to a borrower and receives securities as collateral until the loan is repaid. In particular, the *repo rate* is the interest paid by the borrower to the cash lender.

[42] However, trading all Treasuries under consideration in the full-sample on the basis of forecasts from MSVAR models that include monetary aggregates, the size of the Fed's balance sheet, or its average maturity, tends to lead to positive and statistically significant Sharpe ratios (see panel D).

the percentage of Treasuries in the Fed's balance sheet, during the GFC would have led to an average monthly return of 38.6 bps, a statistically significant Sharpe ratio of 0.524 (the standard error is 0.250) when used to give signals to a butterfly strategy; the model would have outperformed a MSIH(3)-VAR(1) for the DNS factors only with a DMWM p-value of essentially zero, i.e., conditioning on the stance of monetary policy would have generated economic value. The same MSVAR would have scored an average monthly return of 68.9 bps, a statistically significant Sharpe ratio of 0.502 (the standard error is 0.222) when used to support equal-weighted trades on the entire yield curve; in this case, a DMWM test of equality vs. a MSIH(3)-VAR(1) for the DNS factors only, rejects with a p-value of less than 0.01. Yet, the same model, when used to trade the Slope factor over the full OOS period, yields a much smaller average monthly return of 13.5 bps, to imply an insignificant Sharpe ratio of 0.051 (standard error is 0.104), leading to a DMWM p-value in excess of 0.28 when the model is compared to the MSVAR that drops the macro variable. In fact, over the full OOS period, the same model leads to negative returns and Sharpe ratios when the entire yield curve is traded. Even if we limit the analysis to the GFC, this model leads to disappointing results when applied to butterfly strategies that trade either Level or Curvature predictions: in the former case, the average monthly return is -5.5 bps which implies a negative (but at least not significant) Sharpe ratio; in the latter case, panel C of Table 13 shows large average returns of 87.5 bps that however turn out to be highly volatile, to the point that the 0.372 Sharpe ratio fails to be statistically significant (the standard error is 0.234), while the null of no difference vs. a model that omits the stance of monetary policy cannot be rejected at conventional significance levels.

Table 14 has the same structure as Table 13 but now reports trading strategy results when realistic funding costs to a retail investor are taken into account. Note that such funding costs are considered on an ex-ante basis, i.e., they are charged on all short positions *before* the decision to implement the strategy has been taken. This explains the fact that in some cases the raw, monthly returns in Table 14 exceed those in Table 13: some trades that are executed in Table 13 on a thin forecasting "margin" and that end up losing money, are not executed in Table 14 because of the implicit filter represented by the funding cost on the short positions. Despite this, Table 14 shows that—with very few exceptions—transaction costs may completely wipe out the ability of both regimes in DNS models and of adding variables descriptive of the monetary policy stance to generate positive economic value. For instance, in panel B, with reference to the GFC period, when Slope predictions are traded while the average returns from models that exclude monetary proxies are generally low (5-9 bps per month), MSVAR models that include such proxies lead to stronger performance, between 15 and 35 bps per month. However, such returns are highly uncertain and as a result the Sharpe ratios are modest and hardly ever significant; for the same reason, most comparisons of realized returns across models with different structure in terms of including regimes or macroeconomic variables, fail to lead to rejection of the null of no difference. As in Table 13, trading the entire yield curve tends to give less extreme and erratic mean performances vs. trading the factors. For instance, in the case of Level bets, the constraints deriving from the need to hedge induce a trader to buy and sell larger quantities of bonds vs. unconstrained trades on all yields and this results in a more risky strategy. However, DMWM tests performed against a zero return strategy show that, with few exceptions, both strategies achieve gains statistically different from zero for all the models. For the majority of the models, accounting for regime switching implies better performances than single-state VARs do. However, these results should be treated with caution because they are hardly ever significant according to DMWM tests. Therefore, even though in qualitative terms the results in Table 14 are not different from those in Table 13, when realistic transaction costs are considered, the number of trades decreases and the volatility of the resulting outcomes increases enough to prevent us from drawing precise and statistically robust inferences on the economic value of the models entertained in this paper.

7. CONCLUSIONS

In this paper, in the spirit of Diebold and Li (2006) and Xiang and Zhu (2013), we have used a range of regime switching models to test whether standard DNS factors derived from the current and past shape of the yield curve carry all the information necessary to forecast the dynamics of the very yield curve or—as an alternative—whether the direct modelling of macroeconomic variables able to proxy for the current stance of monetary policy expand the needed information set. To deal in an effective way with the risk of overfitting presented by relatively rich models with Markov regimes extended to include macroeconomic variables besides the DNS factors, we have also investigated the economic value of the predictions generated by the models. We have done so both by adopting simple strategies that buy (sell) all Treasuries whose yield is forecast to decrease (increase) between t and $t + 1$, as well as hedged butterfly strategies that allow us to place pure bets on specific DNS factors when these are predicted to change over time in reliable ways. The interaction between introducing Markov regimes in multivariate time series models for DNS factors and extending the vector of DNS factors to include variables that proxy for monetary policy seems natural in light of the drastic shift in the tools adopted by the Fed during the GFC and its aftermath, based for instance on quantitative easing objectives and on the attempt to control the slope of the yield curve through maturity extension programs. Moreover, such unconventional policy measures may have affected (hence, forecast) different features of the DNS factors, that currently the literature has been increasingly using to summarize the shape and dynamics of the yield curve.

We find that—even when regimes are taken into account—the standard DNS factors hardly contain all the information required to forecast the yield curve. As conjectured, such a lack of predictive power becomes stronger during (but it is not limited to) the GFC, when monetary policy shifted tools and objectives. This means that, at least when considered in isolation and conditioning on the imperfect (but very accurate) fit to the term structure provided by the DNS factor, the US Treasury market may be considered weakly informationally efficient but not semi-strong efficient, as appropriate indicators of the stance of monetary policy provide tradeable signals that improve the forecasting power otherwise obtainable. Such signals lead in fact to positive and sometimes accurately estimated trading profits, especially when the strategies are based on placing bets on Slope factor predictions and during the GFC, when the changes in policy tools were more drastic.

Several extensions of this paper appear to be natural. First of all, we have not tried to incorporate all the macroeconomic variables into a single eight-variable MSVAR system to investigate if and to what extent the forecasting performance would change. The numerical and inferential problems we would encounter, due to the lack of a sufficient number of observations on the very monetary variables, would be probably unsurmountable. Yet, with longer time series or resorting to higher frequency (weekly) data, such an effort might be possible. Another natural extension would consist of making our MSVAR models arbitrage-free in the tradition of Christensen, Diebold, and Rudebusch (2011) to control whether enforcing the theoretical rigor of the model would deteriorate or improve the forecasting performance. Moreover, because our results partially suggest that also yields have forecasting power for macroeconomic variables, thus along the lines of Estrella and Hardouvelis (1991), this route could be further investigated in a Markov switching framework. Third, we have limited our forecast horizon to one month because this seems natural when trading strategies are implemented and the resulting economic value estimated, but it could be interesting to understand our framework to multi-step forecasts, as in Moench (2008) or Eo and Kang (2018); the latter paper, in fact, also experiments with regime-switching, predictable forecast combination schemes, that we have ignored in our work. Fourth, while in our paper we have turned forecast accuracy in economic value using trading strategies, in the tradition of Leitch and Tanner (1992), another, increasingly popular metrics also when applied to bond portfolio is risk measurement and prediction. For instance, Tu and Chen (2018) have recently studied whether expanding the DNS framework (as in Caldeira, Moura, and Santos,

2015) to include non-yield curve related financial and macroeconomic variables may lead to more accurate Value-at-Risk estimates and have reported that plain-vanilla, three-factor NS models are rarely sufficient to provide reasonably accurate VaR forecasts. It would be interesting to assess the robustness of our findings to using risk management-related loss functions. Finally, in the paper we have used simple trading strategies that just exploit forecasts for either the DNS factors or for the implied yields. Recently, some researchers recognized the benefits of exploiting predictability in the shape of the yield curve also for portfolio choice. Campbell and Viceira (2001) and Korn and Koziol (2006), pioneered this literature by employing the Vasicek (1977) model to perform bond portfolio selection, in the former case using intertemporal optimization for a long-term investor. Caldeira, Moura, and Santos (2016) propose a novel method that explicitly uses the DNS model (extended to a general class of dynamic heteroskedastic factor models) to build optimal mean-variance bond portfolios. It could be interesting to extend the empirical work in our paper to an explicit mean-variance set up.

References

Anderson, R. G., and Jones, B. (2011). A comprehensive revision of the US monetary services (Divisia) indexes. *Federal Reserve Bank of St. Louis Review*, 93(September/October 2011).

Ang, A., and Bekaert, G. (2002). Regime switches in interest rates. *Journal of Business and Economic Statistics* 20, 163–182

Ang, A., Dong, S., and Piazzesi, M. (2007). No-arbitrage Taylor rules. National Bureau of Economic Research, Working paper No. 13448.

Ang, A., and Piazzesi M.. (2003)."A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables." *Journal of Monetary Economics* 50, 745-787.

Armstrong, J. S., and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting* 8, 69-80.

Balfoussia, H., and Wickens, M. (2007). Macroeconomic sources of risk in the term structure. *Journal of Money, Credit and Banking* 39, 205-236.

Bansal, R., and Zhou, H. (2002). Term structure of interest rates with regime shifts. *Journal of Finance* 57, 1997-2043.

Barnett, W. A. (1982). The optimal level of monetary aggregation. *Journal of Money, Credit and Banking* 14, 687-710.

Bank for International Settlements (2005). Zero-coupon yield curves: technical documentation. Technical Report.

Belongia, M. T., and Ireland, P. N. (2012). Quantitative easing: interest rates and money in the measurement of monetary policy. *Scandinavian Journal of Economics* 16, 635-668.

Bekaert, G., Cho, S., and Moreno, A. (2010). New Keynesian macroeconomics and the term structure. *Journal of Money, Credit and Banking* 42, 33-62.

Bernanke, B. S. and Blinder A. B. (1988) Credit, money, and aggregate demand. NBER Working Paper No. 2534.

Byrne, J. P., Cao, S., and Korobilis, D. (2017). Forecasting the term structure of government bond yields in unstable environments. *Journal of Empirical Finance* 44, 209-225.

Borio, C., and Disyatat, P. (2010). Unconventional monetary policies: an appraisal. *The Manchester School* 78, 53-89.

Caldeira, J. F., Moura, G. V., and Santos, A. A. (2015). Measuring risk in fixed income portfolios using yield curve models. *Computational Economics* 46, 65-82.

Caldeira, J. F., Moura, G. V., and Santos, A. A. (2016). Bond portfolio optimization using dynamic factor models. *Journal of Empirical Finance* 37, 128-158.

Campbell, J. Y., and Viceira, L. (2001). Who should buy long-term bonds? *American Economic Review* 91, 99-127.

Carriero, A., and Giacomini, R. (2011). How useful are no-arbitrage restrictions for forecasting the term structure of interest rates? *Journal of Econometrics* 164, 21-34.

Chadha, J. S., Turner, P., and Zampolli, F. (2013). The interest rate effects of government debt maturity. BIS Working paper No. 415.

Chib, S., and Kang, K. H. (2013). Change-points in affine arbitrage-free term structure models. *Journal of Financial Econometrics* 11, 302-334.

Christensen, J.H.E., F.X. Diebold, and G.D. Rudebusch. (2011). The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models. *Journal of Econometrics* 164, 4-20.

Christoffersen, P. F., and Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52, 1273-1287.

Coroneo, L., Nyholmy, K. and Vidova-Kolevaz, R. (2011). How Arbitrage-Free is the Nelson-Siegel Model? *Journal of Empirical Finance* 18, 393-407.

Coroneo, L., Giannone, D., and Modugno, M. (2016). Unspanned macroeconomic factors in the yield curve. *Journal of Business and Economic Statistics* 34, 472-485.

Dai, Q., Singleton, K. J., and Yang, W. (2007). Regime shifts in a dynamic term structure model of US Treasury bond yields. *Review of Financial Studies* 20, 1669-1706.

Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247-254.

Dewachter, H., and Lyrio, M., 2006. Macro factors and the term structure of interest rates. *Journal of Money, Credit, and Banking* 38, 119–140.

Diebold, F.X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337-364.

Diebold, F.X., and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253-263.

Diebold, F., and Rudebush G.D. (2013). *Yield Curve Modeling and Forecasting. The Dynamic Nelson-Siegel Approach*. Princeton University Press.

Diebold, F.X., Rudebush G.D., and Arouba S.B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics* 131, 309–338.

Doshi, H., Jacobs, K., and Liu, R. (2018). Macroeconomic determinants of the term structure: Long-run and short-run dynamics. *Journal of Empirical Finance* 48, 98-122.

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

Duffee, G.R. (2008). Forecasting with the term structure: the role of no-arbitrage. Working Paper, University of California, Berkeley.

Duffee, G. R. (2011). Information in (and not in) the term structure. *Review of Financial Studies* 24, 2895-2934.

Duffee, G.R. (2013). Bond pricing and the macroeconomy. In: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), *Handbook of the Economics of Finance*.

Eo, Y., and Kang, K., H. (2018). Forecasting the Term Structure of Interest Rates with Potentially Misspecified Models. Working paper, University of Syndey.

Estrella, A., and Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *Journal of Finance*, 46, 555-576.

European Central Bank (2008), The New Euro Area Yield Curves (February, Monthly Bulletin).

Evans, C. L., and Marshall, D. A. (2007). Economic determinants of the nominal treasury yield curve. *Journal of Monetary Economics* 54, 1986-2003.

Fabozzi, F.J., Martellini L., and Priaulet P. (2005). Predictability in the shape of the term structure of interest rates. *Journal of Fixed Income* 15, 40–50.

Fenton, V. M., and Gallant, A. R. (1996). Qualitative and asymptotic performance of SNP density estimators. *Journal of Econometrics* 74, 77-118.

Gambacorta, L., and Hofmann, B. (2012). The economic effects of central bank bond purchase programmes. *Bancaria* 3, 38-50.

Gambacorta, L., Hofmann, B., and Peersman, G. (2014). The effectiveness of unconventional monetary policy at the zero lower bound: a cross-country analysis. *Journal of Money, Credit and Banking* 46, 615-642.

Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review* 39, 763-788.

Goodfriend, M. (2011). Central banking in the credit turmoil: an assessment of Federal Reserve practice. *Journal of Monetary Economics* 58, 1-12.

Gray, S. (1996). Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process. *Journal of Financial Economics* 42, 27–62.

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

Guidolin, M., and Pedio, M. (2018). *Essentials of Time Series for Financial Applications*. Academic Press.

Guidolin, M., and Thornton, D. L. (2018). Predictions of short-term rates and the expectations hypothesis. *International Journal of Forecasting* 34, 636-664.

Guidolin, M., and Timmermann, A. (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics* 21, 1–22.

Guidolin, M., and Timmermann, A. (2009). Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics* 150, 297-311.

Gürkaynak, R.S. and Wright, J.H. (2012). Macroeconomics and the term structure. *Journal of Economic Literature* 50, 331–367.

Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control* 12, 385-423.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39-70.

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

Hess, A. C., and Kamara, A. (2005). Conditional time-varying interest rate risk premium: Evidence from the Treasury bill futures market. *Journal of Money, Credit and Banking* 37, 679-698.

Hevia, C., Gonzalez-Rozada, M., Sola, M., and Spagnolo, F. (2015). Estimating and forecasting the yield curve using a Markov switching dynamic Nelson and Siegel model. *Journal of Applied Econometrics* 30, 987-1009.

Hördahl, P., Tristani, O., and Vestin, D. (2006). A joint econometric model of macroeconomic and term-structure dynamics. *Journal of Econometrics* 131, 405-444.

Hyndman, R. J., and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, 679-688.

Iania, L., and Dewachter, H. (2012). An extended macro-finance model with financial factors. *Journal of Financial and Quantitative Analysis* 46, 1893-1916.

Kim, H., and Park, H. (2013). Term structure dynamics with macro-factors using high frequency data. *Journal of Empirical Finance* 22, 78-93.

Koopman, S. J., Mallee, M. I., and Van der Wel, M. (2010). Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters. *Journal of Business and Economic Statistics* 28, 329-343.

Korn, O. and Koziol, C. (2006). Bond portfolio optimization. a risk-return approach. *Journal of Fixed Income,* 15: 48–60.

Kozicki, S., and Tinsley, P. A. (2005). What do you expect? Imperfect policy credibility and tests of the expectations hypothesis. *Journal of Monetary Economics* 52, 421-447.

Kuttner, K. (2006). Can central banks target bond prices? National Bureau of Economic Research, Working paper No. 12454.

Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review* 81, 580-590.

Levant, J., and Ma, J. (2016). Investigating United Kingdom's monetary policy with macro-factor augmented dynamic Nelson–Siegel models. *Journal of Empirical Finance* 37, 117-127.

Levant, J., and Ma, J. (2017). A dynamic Nelson-Siegel yield curve model with Markov switching. *Economic Modelling* 67, 73-87.

Lo, A.W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal* 58, 36–52.

McCracken, M. W. (2004). Parameter estimation and tests of equal forecast accuracy between non-nested models. *International Journal of Forecasting* 20, 503-514.

Moench, E. (2008). Forecasting the yield curve in a data-rich environment: a no-arbitrage factor-augmented VAR approach. *Journal of Econometrics* 146, 26-43.

Moench, E. (2012). Term structure surprises: the predictive content of curvature, level, and slope. *Journal of Applied Econometrics* 27, 574-602.

Nelson, C. and Siegel, A. (1987). Parsimonious Modeling of Yield Curves. *Journal of Business* 60, 473–489.

Rudebusch, G. D., and Wu, T. (2007). Accounting for a shift in term structure behavior with no-arbitrage and macro-finance models. *Journal of Money, Credit and Banking* 39, 395-422.

Smith, D. R. (2002). Markov-switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business and Economic Statistics* 20, 183-197.

Startz, R., and Tsang, K. P. (2010). An unobserved components model of the yield curve. *Journal of Money, Credit and Banking* 42, 1613-1640.

Tu, A. H., and Chen, C. Y. H. (2018). A factor-based approach of bond portfolio value-at-risk: the informational roles of macroeconomic and financial stress factors. *Journal of Empirical Finance* 45, 243-268.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.

Van Dijk, D., Koopman, S. J., Van der Wel, M., and Wright, J. H. (2014). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics* 29, 693-712.

Vayanos, D., and J. Vila. (2009). A preferred-habitat model of the term structure of interest rates. NBER Working Paper No. 15487.

Ullah, W., Tsukuda, Y., and Matsuda, Y. (2013). Term structure forecasting of government bond yields with latent and macroeconomic factors: do macroeconomic factors imply better out-of-sample forecasts? *Journal of Forecasting* 32, 702-723.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067-1084.

Wu, T. (2002). Monetary policy and the slope factor in empirical term structure estimations. Federal Reserve Bank of San Francisco, Working paper No. 2002-07.A

Xiang, J., and Zhu, X. (2013). A Regime-switching Nelson–Siegel term structure model and interest rate forecasts. *Journal of Financial Econometrics* 11, 522–555.

Yu, W. C., and Zivot, E. (2011). Forecasting the term structures of Treasury and corporate yields using dynamic Nelson-Siegel models. *International Journal of Forecasting* 27, 579-591.

Table 1

Summary Statistics of Treasury Bond Yields

Panel A shows summary statistics for Constant Maturity Treasury yields, at various maturities. The data are expressed in annualized percentage terms. We use monthly data for the sample January 1982-February 2015. The last row contains the Jarque-Bera test applied to the null hypothesis of normal distribution for each of the series (the 5% critical level is 5.991). Panel B shows the correlations between pairs of Treasury yields.

Panel A

	3 months	6 months	1 year	2 years	3 years	5 years	7 years	10 years	30 years
Mean	4.387	4.582	4.761	5.142	5.363	5.742	6.036	6.246	6.634
Median	4.870	5.040	5.000	5.120	5.330	5.600	5.720	5.810	6.060
Maximum	14.280	14.810	14.730	14.820	14.730	14.650	14.670	14.590	14.220
Minimum	0.010	0.040	0.100	0.210	0.330	0.620	0.980	1.530	2.590
Std. Dev.	3.095	3.200	3.259	3.329	3.278	3.129	3.010	2.859	2.577
Skewness	0.318	0.363	0.388	0.404	0.434	0.521	0.595	0.705	0.832
Kurtosis	2.693	2.798	2.804	2.783	2.819	2.930	2.983	3.073	3.149
Jarque-Bera	8.133	9.271	10.434	11.428	12.780	17.749	23.069	32.470	45.477

Panel B

Correlation	3 months	6 months	1 year	2 years	3 years	5 years	7 years	10 years	30 years
3 months	1.000								
6 months	0.999	1.000							
1 year	0.995	0.998	1.000						
2 years	0.985	0.991	0.996	1.000					
3 years	0.977	0.983	0.991	0.999	1.000				
5 years	0.959	0.967	0.977	0.991	0.996	1.000			
7 years	0.945	0.954	0.965	0.982	0.990	0.998	1.000		
10 years	0.930	0.939	0.952	0.972	0.982	0.994	0.998	1.000	
30 years	0.899	0.909	0.924	0.948	0.961	0.979	0.988	0.995	1.000

Table 2

Summary Statistics for Recursively Estimated Dynamic Nelson-Siegel Factors

Panel A shows summary statistics for the estimated DNS factors: Level, Slope, and Curvature. The data are reported in annualized percentage terms. We use monthly data from January 1982 to February 2014. The last row contains the Jarque-Bera test applied to the null hypothesis of normal distribution for each of the series (the 5% critical level is 5.991). Panel B shows the correlations between the DNS factors.

Panel A

	Level	Slope	Curvature
Mean	6.771	-2.428	-1.366
Median	6.095	-2.660	-0.777
Maximum	14.107	0.775	6.980
Minimum	2.587	-5.070	-7.423
Std. Dev.	2.536	1.536	2.741
Skewness	0.834	0.229	-0.261
Kurtosis	3.128	1.972	2.492
Jarque-Bera	45.556	20.648	8.636

Panel B

Correlation	Level	Slope	Curvature
Level	1.000		
Slope	-0.002	1.000	
Curvature	0.712	0.445	1.000

Table 3

ML Estimates of a Single-State VAR(1) Model for the DNS Factors

	Level	Slope	Curvature
Intercept term			
	0.069	**0.171*****	**-0.395*****
	(0.058)	(0.075)	(0.150)
VAR(1) matrix			
Level (t-1)	**0.989*****	**-0.039*****	**0.058*****
	(0.008)	(0.011)	(0.022)
Slope (t-1)	0.012	**0.934*****	**0.055****
	(0.011)	(0.014)	(0.028)
Curvature (t-1)	-0.005	**0.053*****	**0.916*****
	(0.009)	(0.011)	(0.022)
Covariance matrix			
Level	0.063***		
Slope	-0.040	0.107***	
Curvature	0.032	-0.046	0.430***

*significat at 10% level, **significant at 5% level, ***significant at 1% level

Table 4

Model Specification Search for Expanded Markov Switching VAR Models

The table shows the result of a model search based on the maximization of the optimized log-likelihood and the minimization of three alternative information criteria (Akaike's, Hannan-Quinn's, and the Bayes-Schwartz's) that penalized according to alternative functions the number of parameters implied by each model. The saturation ratio is the ratio between the total number of observations available for estimation and the total number of parameters implied by each model. The structure of the four MSVAR models estimated is reported in the main text. The LR linearity test is a LR test of the null hypothesis that a linear VAR model with the same number of lags as the MSVAR indicated in the row is adequate to fit the data; the p-values unreported in parentheses are computed using Davies' (1977) upper bound that adjusts for the presence of nuisance parameters in the MSVAR model under the null.

MSVAR(k, p)	Number of parameters	Maximized Log-Likelihood	LR Linearity Test	AIC	H-Q	SIC	Saturation ratio
Dynamic Nelson-Siegel Factors (Jan. 1982 - July 2015)							
MSIH(2,1)	29	-381.55	135.96 (0.000)	2.100	2.220	2.400	41.59
MSIH(3,1)	42	**-167.13**	564.80 (0.000)	**1.073**	**1.242**	**1.500**	28.71
MSIAH(2,1)	38	-281.50	336.08 (0.000)	1.638	1.792	2.025	31.74
MSIAH(3,1)	60	-170.90	557.26 (0.000)	1.800	1.420	1.790	20.10
DNS Factors + Federal Funds Rate (Jan. 1982 - July 2015)							
MSIH(2,1)	46	60.063	660.55 (0.000)	-0.0721	0.1133	0.3957	26.22
MSIH(3,1)	64	170.15	880.73 (0.000)	-0.5444	-0.2864	0.1065	18.84
MSIAH(2,1)	62	76.352	693.13 (0.000)	-0.0736	0.1763	0.5569	19.45
MSIAH(3,1)	96	**266.68**	1073.78 (0.000)	**-0.8753**	**-0.4883**	**0.1010**	12.56
DNS Factors + Natural Log of Divisia Money (Jan. 1982 - July 2015)							
MSIH(2,1)	46	1296.7	433.10 (0.000)	-6.5313	-6.3432	-6.0571	25.83
MSIH(3,1)	64	1380.3	600.23 (0.000)	-6.8736	**-6.6119**	**-6.2139**	18.56
MSIAH(2,1)	62	1197.9	235.49 (0.000)	-5.9318	-5.6782	-5.2927	19.16
MSIAH(3,1)	96	**1433.4**	706.42 (0.000)	**-6.9838**	-6.5912	-5.9942	12.56
DNS Factors + Average Maturity of FED's Balance Sheet (Jan. 1982 - July 2015)							
MSIH(2,1)	46	-988.84	178.04 (0.000)	5.307	5.492	5.775	26.22
MSIH(3,1)	64	-722.04	711.66 (0.000)	4.031	**4.289**	**4.682**	18.84
MSIAH(2,1)	62	-953.06	249.61 (0.000)	5.205	5.455	5.836	19.45
MSIAH(3,1)	96	**-680.06**	795.60 (0.000)	**3.980**	4.367	4.956	12.56
DNS Factors + Natural Log of Size of Assets on FED's Balance Sheet (July 1996 - July 2015)							
MSIH(2,1)	46	344.11	163.57 (0.000)	-2.7603	-2.4699	-2.0414	14.87
MSIH(3,1)	64	566.32	607.99 (0.000)	-4.6511	**-4.2471**	**-3.6510**	10.69
MSIAH(2,1)	62	434.95	345.24 (0.000)	-3.4532	-3.0618	-2.4844	11.03
MSIAH(3,1)	96	**611.39**	698.13 (0.000)	**-4.7721**	-4.1661	-3.2720	7.13
DNS Factors + Percentage of FED's Balance Sheet Made Up by US Treasuries (July 1996 - July 2015)							
MSIH(2,1)	46	564.85	367.96 (0.000)	-4.800	-4.510	-4.085	14.87
MSIH(3,1)	64	627.37	492.97 (0.000)	**-5.216**	**-4.812**	**-4.216**	10.69
MSIAH(2,1)	62	481.61	201.44 (0.000)	-3.885	-3.494	-2.916	11.03
MSIAH(3,1)	96	**638.09**	506.40 (0.000)	-4.982	-4.376	-3.482	7.13
DNS Factors + Average Maturity and Log Size of Assets on FED's Balance Sheet (July 1996 - July 2015)							
MSIH(2,1)	67	-826.63	435.88 (0.000)	4.4865	4.6723	4.9552	10.21
MSIH(3,1)	91	**-705.23**	678.68 (0.000)	**3.9549**	**4.2134**	**4.6070**	7.52
MSIAH(2,1)	92	-944.63	199.87 (0.000)	5.1755	5.4259	5.8072	7.43
MSIAH(3,1)	141	-736.15	616.84 (0.000)	4.2784	4.6662	5.2566	4.85

Table 5

ML Estimates of a Three-State VAR(1) Model for the DNS Factors

The table shows the estimates of the 42 parameters implied by a MSIH(2)-VAR(1) model for the dynamically estimated Nelson-Siegel factors from US Treasury yields over a 1982:01-2015:06 sample. For conditional mean coefficients, OGP-derived p-values are reported underneath the corresponding coefficient. Coefficients that are significantly different from zero are boldfaced. In the case of residual standard errors, significant means that the estimated coefficient >0.0001.

	Level(t)	*Slope(t)*	*Curvature(t)*
Intercept term			
Regime 1	0.0601	**0.1066**	-0.1943
	(0.222)	(0.077)	(0.127)
Regime 2	0.0069	0.1073	**-0.8636**
	(0.911)	(0.149)	(0.000)
Regime 3	0.0390	-0.1214	-0.2014
	(0.593)	(0.171)	(0.264)
VAR(1) matrix			
Level(t-1)	**0.9939**	**-0.0170**	**0.0402**
	(0.000)	(0.072)	(0.038)
Slope(t-1)	**0.0195**	**0.9534**	**0.0644**
	(0.034)	(0.000)	(0.004)
Curvature(t-1)	**-0.0163**	**0.0375**	**0.8434**
	(0.049)	(0.000)	(0.000)
Unconditional Means			
Regime 1	7.8869	0.0615	0.8092
Regime 2	4.2035	-4.1854	-6.1569
Regime 3	-0.7556	-5.2689	-3.6391
Covariance matrix of shocks	(standard errors on main diagonal; correlation outside)		
Regime 1			
Level(t)	**0.1894*****	**-0.7273*****	**0.4268*****
Slope(t)		**0.2083*****	**-0.2545****
Curvature(t)			**0.5047*****
Regime 2			
Level(t)	**0.2291*****	**-0.9645*****	-0.1479
Slope(t)		**0.2679*****	-0.0939
Curvature(t)			**0.4236*****
Regime 3			
Level(t)	**0.3524*****	**-0.2858****	**0.1532***
Slope(t)		**0.4730*****	**-0.2053****
Curvature(t)			**0.9384*****
Transition matrix	*Regime 1*	*Regime 2*	*Regime 3*
Regime 1	**0.9004*****	0.0064	0.0932
Regime 2	0.0097	**0.9350*****	0.0553
Regime 3	**0.1758****	0.0514	**0.7728*****
	Observations	*Ergodic Probability*	*Duration*
Regime 1	195	0.4853	10.04
Regime 2	102	0.2538	15.39
Regime 3	105	0.2609	4.40

* significant at 10% size; ** significant at 10% size; *** significant at 1% size

Table 6

ML Estimates of a Three-State VAR(1) Model for the DNS Factors and the Log-Size of the FED's Balance Sheet

The table shows the estimates of a MSIH(2)-VAR(1) model for the Nelson-Siegel factors from US Treasury yields over a 1996:07-2015:06 sample. For conditional mean coefficients, OGP- p-values are reported underneath the coefficients, with boldfacing to indicate significance. In the case of residual standard errors, significant means that the estimated coefficient >0.0001.

	Level(t)	*Slope(t)*	*Curvature(t)*	*Size of Assets(t)*
Intercept term				
Regime 1	**1.6282**	-1.4715	0.9853	-0.0632
	(0.039)	(0.135)	(0.594)	(0.355)
Regime 2	**1.7830**	**-1.8075**	0.5958	-0.0450
	(0.040)	(0.067)	(0.747)	(0.518)
Regime 3	**1.7253**	-1.5834	-0.1162	-0.0654
	(0.045)	(0.119)	(0.951)	(0.351)
VAR(1) matrix				
Level(t-1)	**0.9119**	0.0433	0.0020	0.0015
	(0.000)	(0.229)	(0.978)	(0.568)
Slope(t-1)	0.0048	**0.9500**	-0.0116	-0.0006
	(0.734)	(0.000)	(0.753)	(0.603)
Curvature(t-1)	-0.0045	**0.0319**	**0.7841**	-0.0010
	(0.7258)	(0.044)	(0.000)	(0.603)
Size of Assets(t-1)	**-0.0945**	0.0966	-0.0780	**1.0043**
	(0.080)	(0.130)	(0.503)	(0.000)
Unconditional Means				
Regime 1	4.4452	-0.3901	-0.0962	0.3249
Regime 2	16.0975	-13.9899	2.4110	-0.6363
Regime 3	8.1290	-6.7922	-3.8984	0.0698
Covariance matrix of shocks	(standard errors on main diagonal; correlation outside)			
Regime 1				
Level(t)	**0.1580*****	**-0.7132****	0.0149	-0.0579
Slope(t)		**0.1893*****	**-0.2839****	**-0.1780***
Curvature(t)			**0.4703*****	0.0159
Size of Assets(t)				**0.0130*****
Regime 2				
Level(t)	**0.2190*****	**-0.7128*****	0.0480	-0.0696
Slope(t)		**0.3531*****	**0.1664***	-0.0846
Curvature(t)			**0.8831*****	**-0.1426***
Size of Assets(t)				**0.0910*****
Regime 3				
Level(t)	**0.2962*****	**-0.9581*****	**-0.1578***	0.1540
Slope(t)		**0.2561*****	-0.1135	**-0.1798***
Curvature(t)			**0.4452*****	0.0041
Size of Assets(t)				**0.0164*****
Transition matrix	*Regime 1*	*Regime 2*	*Regime 3*	
Regime 1	**0.8715*****	**0.1285***	0.0000	
Regime 2	**0.1884****	**0.7153*****	0.0963	
Regime 3	0.0000	**0.0420***	**0.9580*****	
	Observations	*Ergodic Probability*	*Duration*	
Regime 1	82	0.3081	7.78	
Regime 2	51	0.2102	3.51	
Regime 3	94	0.4817	23.8	

* significant at 10% size; ** significant at 10% size; *** significant at 1% size

Table 7

ML Estimates of a Three-State VAR(1) Model for the DNS Factors and the Natural Logarithm of the Divisia Index

The table shows the estimates of a MSIH(2)-VAR(1) model for the Nelson-Siegel factors from US Treasury yields over a 1982:01-2015:06 sample. For conditional mean coefficients, OGP- p-values are reported underneath the coefficients, with boldfacing to indicate significance. In the case of residual standard errors, significant means that the estimated coefficient >0.0001.

	Level(t)	*Slope(t)*	*Curvature(t)*	*Log of Divisia(t)*
Intercept term				
Regime 1	**2.2610**	**-2.8900**	0.2857	**0.0323**
	(0.001)	(0.002)	(0.874)	(0.000)
Regime 2	**2.3995**	**-3.0677**	0.8319	**0.0887**
	(0.001)	(0.001)	(0.643)	(0.000)
Regime 3	**2.3497**	**-2.8095**	1.0204	**0.0824**
	(0.001)	(0.02)	(0.561)	(0.000)
VAR(1) matrix				
Level(t-1)	**0.9370**	**0.0545**	0.0278	**-0.0020**
	(0.000)	(0.022)	(0.559)	(0.000)
Slope(t-1)	**0.0211**	**0.9526**	**0.0537**	0.0002
	(0.030)	(0.000)	(0.029)	(0.195)
Curvature(t-1)	**-0.0149**	**0.0382**	**0.8294**	-0.0006
	(0.077)	(0.044)	(0.000)	(0.195)
Log of Divisia(t-1)	**-0.2358**	**0.3025**	-0.1349	**0.9919**
	(0.002)	(0.001)	(0.461)	(0.000)
Unconditional Means				
Regime 1	6.1275	-3.2325	-3.4379	0.2823
Regime 2	7.4942	-2.9082	-1.3172	-0.1278
Regime 3	6.7049	-1.6482	-0.0296	-0.0752
Covariance matrix of shocks	(standard errors on main diagonal; correlation outside)			
Regime 1				
Level(t)	**0.2284***	**-0.9651***	**-0.1364***	-0.1116
Slope(t)		**0.2687***	-0.1022	**0.1641***
Curvature(t)			**0.4232***	**-0.3102***
Log of Divisia(t)				**0.0043***
Regime 2				
Level(t)	**0.3166***	**-0.5180***	**0.3325****	**-0.1956****
Slope(t)		**0.3589***	**-0.1654***	0.0100
Curvature(t)			**0.7927***	-0.1052
Log of Divisia(t)				**0.0049***
Regime 3				
Level(t)	**0.1931***	**-0.7314***	**0.3104****	-0.0729
Slope(t)		**0.1927***	**-0.1890***	-0.0045
Curvature(t)			**0.5439***	-0.0320
Log of Divisia(t)				**0.0028***
Transition matrix	*Regime 1*	*Regime 2*	*Regime 3*	
Regime 1	**0.9273***	**0.0589****	0.0132	
Regime 2	0.0399	**0.9074***	0.0527	
Regime 3	0.0141	0.0341	**0.9518***	
	Observations	*Ergodic Probability*	*Duration*	
Regime 1	102	0.2574	13.75	
Regime 2	126	0.3195	10.8	
Regime 3	168	0.4230	20.74	

* significant at 10% size; ** significant at 10% size; *** significant at 1% size

Table 8

Root Mean Squared Forecast Error Performance

The table reports the RMSFE for the six MSVAR models and the two benchmarks described in the text. Panel A reports the OOS performance for the factors, whereas Panels B-D for the yields. The results are provided for the total OOS sample (January 2006-July 2015) as well as for the Crisis (August 2007-March 2009) and Non-Crisis (January 2006-July 2007;April 2009-July 2015) sub-samples. We boldface the lowest RMSFE across models.

Panel A

	Full Sample			**Crisis Sample**			**Non-Crisis Sample**		
	Level	Slope	Curvature	Level	Slope	Curvature	Level	Slope	Curvature
VAR(1) for DNS factors	0.2977	0.5013	0.6030	0.4549	0.7350	0.8762	0.1501	0.3287	0.5928
Random walk	**0.2524**	0.3016	0.5182	0.3384	0.4094	0.7897	0.1553	0.2095	0.4421
MSIH(3)-VAR(1) for DNS factors	0.2544	0.2958	0.5339	0.3245	0.3468	0.7822	0.2028	0.2660	0.4757
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.2608	**0.2903**	0.5400	0.3618	0.3566	0.7694	0.1652	0.2232	0.4658
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.2535	0.2917	**0.5162**	0.3291	0.3569	**0.7596**	0.1657	0.2336	0.4667
MSIH(3)-VAR(1) DNS factors + Maturity	0.2531	0.2923	0.5606	**0.3015**	**0.3460**	0.8344	0.1564	0.2329	**0.4246**
MSIH(3)-VAR(1) DNS factors + Log Assets	0.2604	0.2966	0.6128	0.3180	0.3610	1.0309	**0.1409**	**0.1988**	0.4457
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.2769	0.2959	0.6402	0.3435	0.3804	1.1091	0.1712	0.2122	0.4869

Panel B

	Full Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	0.5916	0.4098	0.3625	0.3960	0.3889	0.3726	0.3572	0.2870	0.3402
Random walk	0.2729	0.1605	0.1917	0.2320	0.2352	**0.2694**	**0.2676**	**0.2219**	0.2823
MSIH(3)-VAR(1) for DNS factors	0.2563	0.1470	0.1848	0.2335	0.2505	0.2903	0.2842	0.2320	0.2764
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.2532	0.1461	**0.1810**	0.2274	0.2511	0.2952	0.2925	0.2360	0.2755
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.2377	0.1423	0.1877	0.2257	0.2310	0.2733	0.2751	0.2245	0.2760
MSIH(3)-VAR(1) DNS factors + Maturity	**0.2295**	**0.1395**	0.1819	**0.2180**	**0.2303**	0.2750	0.2719	0.2252	**0.2736**
MSIH(3)-VAR(1) DNS factors + Log Assets	0.2389	0.1413	0.1909	0.2412	0.2465	0.2770	0.2762	0.2268	0.2921
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.2477	0.1602	0.1947	0.2454	0.2684	0.3061	0.3012	0.2503	0.2966

Panel C

	Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	1.1231	0.8245	0.7009	0.7349	0.7540	0.6721	0.6237	0.4666	0.4163
Random walk	0.5302	0.3196	0.2977	0.3865	0.4214	0.3761	0.3530	0.2839	0.3264
MSIH(3)-VAR(1) for DNS factors	0.4996	0.2945	0.2873	0.3986	0.4434	0.4066	0.3807	0.2923	0.3217
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.4985	0.2713	0.2553	0.3769	0.4422	0.4188	0.4062	0.3122	0.3307
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.4366	0.2631	0.2630	0.3545	0.4009	0.3704	**0.3504**	**0.2725**	**0.3171**
MSIH(3)-VAR(1) DNS factors + Maturity	0.4359	**0.2521**	**0.2546**	**0.3504**	**0.3924**	**0.3680**	0.3516	0.2873	0.3240
MSIH(3)-VAR(1) DNS factors + Log Assets	**0.4312**	0.2747	0.2945	0.4051	0.4490	0.3959	0.3563	0.2728	0.3192
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.4617	0.2889	0.2876	0.4188	0.4838	0.4410	0.4146	0.3430	0.3934

Panel D

	Non-Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	0.3830	0.2236	0.2215	0.2886	0.2865	0.2558	0.2260	0.1875	**0.1443**
Random walk	0.1347	0.1416	0.1364	0.1519	0.1585	0.1592	0.1584	0.1591	0.1713
MSIH(3)-VAR(1) for DNS factors	0.1224	0.1199	0.1236	0.1542	0.1701	0.1807	0.1861	0.1910	0.1961
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	**0.1033**	0.1514	0.1526	0.1680	0.1747	0.1740	0.1735	0.1748	0.1810
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.1736	0.1539	0.1676	0.1864	0.1837	0.1721	0.1656	0.1695	0.1922
MSIH(3)-VAR(1) DNS factors + Maturity	0.0958	0.1668	0.1663	0.1544	0.1540	0.1536	0.1559	0.1596	0.1626
MSIH(3)-VAR(1) DNS factors + Log Assets	0.1538	**0.0945**	**0.1035**	**0.1417**	**0.1463**	**0.1428**	**0.1420**	0.1514	0.1916
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.1299	0.1376	0.1479	0.1589	0.1596	0.1541	0.1513	**0.1493**	0.1507

Table 9

Relative Root Mean Squared Forecast Error Performance vs. a VAR(1) Model

The table reports the Relative RMSFEs for the six MSVAR models described in the text against a single state VAR(1) that includes the same set of variables. Panel A reports the OOS relative performance for the factors, whereas Panels B-D for the yields. The results are provided for the total OOS sample (January 2006-July 2015) as well as for the Crisis (August 2007-March 2009) and Non-Crisis (January 2006-July 2007;April 2009-July 2015) sub-samples. We boldface relative RMSFEs below 1. Only in the case of the full sample statistics, the asterisks refer to the outcomes of a West-McCraken test of the null of no superior predictive accuracy against a VAR(1). * and ** indicate rejection of the null of equal predictive accuracy at 5 and 1 percent size, respectively.

Panel A

	Full Sample			Crisis Sample			Non-Crisis Sample		
	Level	Slope	Curvature	Level	Slope	Curvature	Level	Slope	Curvature
MSIH(3)-VAR(1) for DNS factors	0.8546*	0.5901*	0.8855	0.7134	0.4718	0.8928	1.3508	0.8094	0.8025
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.8761*	0.5791**	0.8956	0.7954	0.4852	0.8096	1.1002	0.6793	0.7858
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.8516*	0.582**	0.8561*	0.7234	0.4855	0.8669	1.1037	0.7108	0.7873
MSIH(3)-VAR(1) DNS factors + Maturity	0.8502	0.5831*	0.9297	0.7551	0.4843	0.9523	1.0418	0.7086	0.7164
MSIH(3)-VAR(1) DNS factors + Log Assets	0.8747*	0.5918**	1.0164	0.6991	0.4911	1.1767	1.1403	0.6458	0.8214
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.9301	0.5903*	1.0617	0.8827	0.5284	1.2659	0.9386	0.6048	0.7519

Panel B

	Full Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
MSIH(3)-VAR(1) for DNS factors	0.4333**	0.3588**	0.5099**	0.5895**	0.6442**	0.7790**	0.7955**	0.8083*	0.8125*
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.4280**	0.3566**	0.4995**	0.5743**	0.6457**	0.7924**	0.8187**	0.8223*	0.8099**
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.4017**	0.3473**	0.5177**	0.5699**	0.5940**	0.7335**	0.7702**	0.7822**	0.8112**
MSIH(3)-VAR(1) DNS factors + Maturity	0.3879**	0.3403**	0.5017**	0.5503**	0.5923**	0.7380**	0.7611**	0.7845*	0.8042**
MSIH(3)-VAR(1) DNS factors + Log Assets	0.4038**	0.3447**	0.5267**	0.6090**	0.6339**	0.7434**	0.7731**	0.7900*	0.8585*
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.4187**	0.3908**	0.5372**	0.6198**	0.6900**	0.8215**	0.8431*	0.8719	0.8717

Panel C

	Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
MSIH(3)-VAR(1) for DNS factors	0.4448	0.3572	0.4099	0.5423	0.5880	0.6049	0.6105	0.6265	0.7729
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.4439	0.3291	0.3642	0.5128	0.5865	0.6231	0.6514	0.6690	0.7944
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.3887	0.3191	0.3752	0.4823	0.5316	0.5511	0.5619	0.5839	0.7618
MSIH(3)-VAR(1) DNS factors + Maturity	0.3882	0.3058	0.3675	0.4769	0.5203	0.5476	0.5638	0.6158	0.7783
MSIH(3)-VAR(1) DNS factors + Log Assets	0.3839	0.3332	0.4201	0.5513	0.5954	0.5890	0.5712	0.5845	0.7667
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.4111	0.3503	0.4103	0.5699	0.6417	0.6562	0.6648	0.7351	0.9451

Panel D

	Non-Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
MSIH(3)-VAR(1) for DNS factors	0.3197	0.5361	0.5579	0.5341	0.5938	0.7064	0.8235	1.0186	1.3588
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.2697	0.6771	0.6892	0.5821	0.6098	0.6802	0.7678	0.9319	1.2538
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.4532	0.6880	0.7568	0.6457	0.6413	0.6729	0.7328	0.9040	1.3320
MSIH(3)-VAR(1) DNS factors + Maturity	0.2502	0.7458	0.7509	0.5350	0.5375	0.6006	0.6898	0.8512	1.1270
MSIH(3)-VAR(1) DNS factors + Log Assets	0.4016	0.4225	0.4675	0.4908	0.5108	0.5583	0.6286	0.8074	1.3279
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.3390	0.6152	0.6677	0.5503	0.5573	0.6026	0.6697	0.7964	1.0445

Table 10

Relative Root Mean Squared Forecast Error Performance vs. a Random Walk Model

The table reports the Relative RMSFEs for the six MSVAR models described in the text against a random walk for the DNS factors. Panel A reports the OOS relative performance for the factors, whereas Panels B-D for the yields. The results are provided for the total OOS sample (January 2006-July 2015) as well as for the Crisis (August 2007-March 2009) and Non-Crisis (January 2006-July 2007;April 2009-July 2015) sub-samples. We boldface relative RMSFEs below 1. In the case of the full sample statistics, the asterisks refer to the outcomes of a West-McCraken test of the null of no superior predictive accuracy against a random walk for the DNS factors. * and ** indicate rejection of the null of equal predictive accuracy at 5 and 1 percent size, respectively.

Panel A

	Full Sample			**Crisis Sample**			**Non-Crisis Sample**		
	Level	Slope	Curvature	Level	Slope	Curvature	Level	Slope	Curvature
MSIH(3)-VAR(1) for DNS factors	1.0081	**0.9807**	1.0303	**0.9589**	**0.8471**	**0.9905**	1.3059	1.2695	1.0761
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	1.0334	**0.9625***	1.0421	1.0691	**0.8711**	**0.9743**	1.0638	1.0652	1.0537
MSIH(3)-VAR(1) DNS factors + Log Divisia	1.0045	**0.9672***	**0.9961**	**0.9725**	**0.8718**	**0.9619**	1.0670	1.1148	1.0557
MSIH(3)-VAR(1) DNS factors + Maturity	1.0029	**0.9691***	1.0818	**0.8909**	**0.8452**	1.0566	1.0071	1.1115	**0.9605**
MSIH(3)-VAR(1) DNS factors + Log Assets	1.0318	**0.9833**	1.1825	**0.9397**	**0.8818**	1.3055	**0.9073**	**0.9487**	1.0082
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	1.0972	**0.9810**	1.2354	1.0150	**0.9292**	1.4045	1.1024	1.0127	1.1014

Panel B

	Full Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
MSIH(3)-VAR(1) for DNS factors	**0.9394****	**0.9163****	**0.9642**	1.0061	1.0653	1.0774	1.0618	1.0456	**0.9792**
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	**0.9279****	**0.9108****	**0.9446***	**0.9802**	1.0677	1.0959	1.0929	1.0637	**0.9759**
MSIH(3)-VAR(1) DNS factors + Log Divisia	**0.8710****	**0.8870****	**0.9791**	**0.9726**	**0.9823**	1.0144	1.0281	1.0119	**0.9776**
MSIH(3)-VAR(1) DNS factors + Maturity	**0.8410****	**0.8692****	**0.9489***	**0.9393***	**0.9795**	1.0208	1.0159	1.0148	**0.9692***
MSIH(3)-VAR(1) DNS factors + Log Assets	**0.8756****	**0.8804****	**0.9960**	1.0394	1.0483	1.0282	1.0319	1.0220	1.0347
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	**0.9078****	**0.9981**	1.0160	1.0578	1.1411	1.1362	1.1254	1.1279	1.0505

Panel C

	Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
MSIH(3)-VAR(1) for DNS factors	**0.9422**	**0.9215**	**0.9651**	1.0312	1.0521	1.0809	1.0786	1.0297	**0.9856**
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	**0.9402**	**0.8490**	**0.8576**	**0.9751**	1.0494	1.1136	1.1508	1.0996	1.0130
MSIH(3)-VAR(1) DNS factors + Log Divisia	**0.8234**	**0.8233**	**0.8836**	**0.9171**	**0.9513**	**0.9849**	**0.9928**	**0.9598**	**0.9715**
MSIH(3)-VAR(1) DNS factors + Maturity	**0.8222**	**0.7890**	**0.8554**	**0.9067**	**0.9311**	**0.9785**	**0.9962**	1.0122	0.9925
MSIH(3)-VAR(1) DNS factors + Log Assets	**0.8132**	**0.8597**	**0.9892**	1.0482	1.0654	1.0525	1.0093	**0.9608**	**0.9777**
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	**0.8707**	**0.9039**	**0.9663**	1.0836	1.1482	1.1725	1.1746	1.2083	1.2052

Panel D

	Non-Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
MSIH(3)-VAR(1) for DNS factors	**0.9087**	**0.8465**	**0.9058**	1.0147	1.0736	1.1350	1.1749	1.2007	1.1449
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	**0.7667**	1.0692	1.1191	1.1058	1.1025	1.0929	1.0955	1.0986	1.0565
MSIH(3)-VAR(1) DNS factors + Log Divisia	1.2883	1.0863	1.2289	1.2266	1.1595	1.0811	1.0455	1.0657	1.1224
MSIH(3)-VAR(1) DNS factors + Maturity	**0.7113**	1.1776	1.2192	1.0164	**0.9718**	**0.9651**	**0.9841**	1.0034	**0.9496**
MSIH(3)-VAR(1) DNS factors + Log Assets	1.1417	**0.6671**	0.7591	**0.9323**	**0.9236**	**0.8971**	**0.8968**	**0.9518**	1.1189
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	**0.9637**	**0.9715**	**1.0842**	1.0455	1.0076	**0.9682**	**0.9555**	**0.9388**	**0.8801**

Table 11

Mean Absolute Percentage Forecast Error Performance

The table reports the MAPFE for the six MSVAR models and the two benchmarks described in the text. Panel A reports the OOS performance for the factors, whereas Panels B-D for the yields. The results are provided for the total OOS sample (January 2006-July 2015) as well as for the Crisis (August 2007-March 2009) and Non-Crisis (January 2006-July 2007;April 2009-July 2015) sub-samples. We boldface the lowest MAPFE across models.

Panel A

	Full Sample			Crisis Sample			Non-Crisis Sample		
	Level	Slope	Curvature	Level	Slope	Curvature	Level	Slope	Curvature
VAR(1) for DNS factors	0.0541	0.9522	0.4711	0.0819	0.4038	0.2543	0.0264	4.2626	1.9987
Random walk	**0.0446**	0.5635	0.4821	0.0536	0.1850	0.2162	0.0247	2.5712	2.1213
MSIH(3)-VAR(1) for DNS factors	0.0461	1.0720	0.4851	0.0509	0.1478	0.2068	0.0366	5.2917	2.1432
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.0474	0.8067	0.4705	0.0623	0.1540	0.1978	0.0275	3.8841	2.0506
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.0449	0.9921	0.5893	0.0571	0.1402	0.2287	0.0252	4.8847	2.5179
MSIH(3)-VAR(1) DNS factors + Maturity	0.0456	0.8921	0.4880	**0.0487**	**0.1340**	**0.1900**	0.0272	4.3532	2.1186
MSIH(3)-VAR(1) DNS factors + Log Assets	0.0475	0.6848	**0.4047**	0.0538	0.1805	0.2858	0.0263	3.2065	**1.6321**
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.0497	**0.4530**	0.5059	0.0715	0.1667	0.3007	**0.0229**	**2.0049**	2.1363

Panel B

	Full Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	3.7815	0.6885	0.5678	0.3750	0.1809	0.1589	0.1170	0.0705	0.0702
Random walk	1.9164	**0.1546**	0.5277	0.2906	**0.1333**	**0.1433**	**0.0992**	**0.0586**	0.0584
MSIH(3)-VAR(1) for DNS factors	1.7793	0.1727	0.5099	0.2856	0.1451	0.1511	0.1049	0.0608	0.0566
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	1.7292	0.2223	0.5194	0.2752	0.1409	0.1564	0.1097	0.0634	0.0563
MSIH(3)-VAR(1) DNS factors + Log Divisia	1.7380	0.2091	0.5400	0.2904	0.1352	0.1484	0.1049	0.0604	**0.0557**
MSIH(3)-VAR(1) DNS factors + Maturity	**1.6792**	0.1668	**0.4989**	0.2833	0.1435	0.1438	0.1000	0.0590	0.0563
MSIH(3)-VAR(1) DNS factors + Log Assets	2.0184	0.2339	0.5284	0.2872	0.1414	0.1494	0.1054	0.0621	0.0601
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	2.0434	0.3420	0.5009	**0.2561**	0.1608	0.1650	0.1151	0.0677	0.0595

Panel C

	Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	3.4874	0.7684	0.4728	0.3693	0.3247	0.2346	0.1865	0.1145	0.0760
Random walk	1.2522	0.2191	0.1601	0.1746	0.1722	0.1237	0.0933	0.0633	0.0583
MSIH(3)-VAR(1) for DNS factors	1.2622	0.2031	0.1519	0.1785	0.1814	0.1318	0.1008	0.0645	0.0559
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	1.3099	0.1924	0.1396	0.1646	0.1709	0.1294	0.1085	0.0694	0.0606
MSIH(3)-VAR(1) DNS factors + Log Divisia	1.1307	0.1796	0.1350	0.1604	0.1670	0.1230	0.0934	**0.0578**	**0.0534**
MSIH(3)-VAR(1) DNS factors + Maturity	**0.8057**	**0.1486**	**0.1154**	**0.1424**	**0.1498**	**0.1115**	**0.0886**	0.0662	0.0583
MSIH(3)-VAR(1) DNS factors + Log Assets	1.1028	0.1866	0.1450	0.1867	0.1885	0.1338	0.0983	0.0624	0.0594
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.8900	0.2261	0.1796	0.2040	0.1995	0.1402	0.1150	0.0810	0.0745

Panel D

	Non-Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	0.0738	0.0398	0.0407	0.0522	0.0518	0.0456	0.0398	0.0334	0.0242
Random walk	0.0210	0.0225	0.0238	0.0267	0.0284	0.0286	0.0281	0.0276	0.0270
MSIH(3)-VAR(1) for DNS factors	0.0199	0.0189	0.0208	0.0275	0.0303	0.0329	0.0342	0.0338	0.0309
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.0171	0.0243	0.0250	0.0302	0.0309	0.0307	0.0307	0.0313	0.0295
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.0280	0.0235	0.0268	0.0298	0.0306	0.0296	0.0287	0.0288	0.0300
MSIH(3)-VAR(1) DNS factors + Maturity	**0.0161**	0.0292	0.0293	0.0264	0.0271	0.0277	0.0282	0.0281	0.0250
MSIH(3)-VAR(1) DNS factors + Log Assets	0.0261	**0.0161**	**0.0174**	**0.0234**	**0.0244**	**0.0247**	**0.0249**	0.0266	0.0299
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.0209	0.0216	0.0248	0.0280	0.0293	0.0284	0.0273	**0.0265**	**0.0233**

Table 12

Hit Ratio Forecast Performance

The table reports the hit ratio (defined as the percentage of the OOS observations for which a model estimated with data up to *t*-1 returns a correct prediction of the sign of the change between *t*-1 and *t*), for the six MSVAR models and the two benchmarks described in the text. Panel A reports the OOS performance for the factors, whereas Panels B-D for the yields. The results are provided for the total OOS sample (January 2006-July 2015) as well as for the Crisis (August 2007-March 2009) and Non-Crisis (January 2006-July 2007;April 2009-July 2015) sub-samples. For each sub-sample, target variable/maturity, we emphasize with a square the model returning the highest OOS hit ratio.

Panel A

	Full Sample			**Crisis Sample**			**Non-Crisis Sample**		
	Level	Slope	Curvature	Level	Slope	Curvature	Level	Slope	Curvature
VAR(1) for DNS factors	**55.8%**	**60.0%**	49.5%	45.0%	**65.0%**	**70.0%**	50.0%	**61.1%**	33.3%
MSIH(3)-VAR(1) for DNS factors	**58.5%**	**60.6%**	48.9%	45.0%	**65.0%**	**70.0%**	**61.1%**	**61.1%**	38.9%
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	**59.6%**	**58.5%**	**51.1%**	45.0%	**60.0%**	**70.0%**	**61.1%**	**61.1%**	50.0%
MSIH(3)-VAR(1) DNS factors + Log Divisia	**59.6%**	**59.6%**	47.9%	50.0%	**65.0%**	**70.0%**	**55.6%**	**55.6%**	27.8%
MSIH(3)-VAR(1) DNS factors + Maturity	**56.4%**	**62.8%**	47.9%	**75.5%**	48.4%	**95.0%**	**55.6%**	**61.1%**	33.3%
MSIH(3)-VAR(1) DNS factors + Log Assets	**56.4%**	**62.8%**	47.9%	45.0%	**70.0%**	**65.0%**	**55.6%**	**61.1%**	33.3%
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	**57.5%**	**66.0%**	48.9%	45.0%	**80.0%**	**65.0%**	**55.6%**	**66.7%**	33.3%

Panel B

	Full Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	46.8%	**53.2%**	47.9%	**52.1%**	**55.3%**	**56.4%**	**60.6%**	**55.3%**	**58.5%**
Random walk	48.9%	**57.4%**	**52.1%**	**52.1%**	**60.6%**	**61.7%**	**64.9%**	**57.4%**	**61.7%**
MSIH(3)-VAR(1) for DNS factors	50.0%	**55.3%**	**51.1%**	**52.1%**	**59.6%**	**62.8%**	**66.0%**	**57.4%**	**63.8%**
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	47.9%	**52.1%**	**52.1%**	48.9%	**56.4%**	**60.6%**	**62.8%**	**56.4%**	**60.6%**
MSIH(3)-VAR(1) DNS factors + Log Divisia	48.9%	**56.4%**	**53.2%**	50.0%	**58.5%**	**59.6%**	**63.8%**	**58.5%**	**61.7%**
MSIH(3)-VAR(1) DNS factors + Maturity	48.9%	**55.3%**	**51.1%**	**51.1%**	**55.3%**	**59.6%**	**61.7%**	**56.4%**	**60.6%**
MSIH(3)-VAR(1) DNS factors + Log Assets	50.0%	47.9%	**53.2%**	**53.2%**	**55.3%**	**61.7%**	**63.8%**	**58.5%**	**62.8%**
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	46.8%	**55.3%**	**51.1%**	**51.1%**	**59.6%**	**61.7%**	**62.8%**	**56.4%**	**58.5%**

Panel C

	Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	50.0%	**70.0%**	**80.0%**	**80.0%**	**75.0%**	**70.0%**	**75.0%**	**60.0%**	**55.0%**
Random walk	**55.0%**	**70.0%**	**90.0%**	**80.0%**	**80.0%**	**75.0%**	**80.0%**	**60.0%**	**60.0%**
MSIH(3)-VAR(1) for DNS factors	**60.0%**	**70.0%**	**80.0%**	**75.0%**	**75.0%**	**70.0%**	**75.0%**	**55.0%**	**55.0%**
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	**60.0%**	**70.0%**	**85.0%**	**80.0%**	**80.0%**	**75.0%**	**80.0%**	**65.0%**	**55.0%**
MSIH(3)-VAR(1) DNS factors + Log Divisia	**60.0%**	**75.0%**	**85.0%**	**80.0%**	**80.0%**	**75.0%**	**80.0%**	**65.0%**	**60.0%**
MSIH(3)-VAR(1) DNS factors + Maturity	**60.0%**	**70.0%**	**85.0%**	**75.0%**	**70.0%**	**65.0%**	**70.0%**	**55.0%**	**55.0%**
MSIH(3)-VAR(1) DNS factors + Log Assets	**60.0%**	**70.0%**	**85.0%**	**80.0%**	**75.0%**	**70.0%**	**75.0%**	**60.0%**	**65.0%**
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	**60.0%**	**75.0%**	**80.0%**	**75.0%**	**75.0%**	**70.0%**	**75.0%**	**55.0%**	50.0%

Panel D

	Non-Crisis Sample								
	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year	30-year
VAR(1) for DNS factors	**66.7%**	**61.1%**	44.4%	50.0%	**55.6%**	**55.6%**	**55.6%**	**55.6%**	**55.6%**
Random walk	**66.7%**	**61.1%**	44.4%	50.0%	**55.6%**	**55.6%**	**55.6%**	50.0%	**55.6%**
MSIH(3)-VAR(1) for DNS factors	**66.7%**	**61.1%**	44.4%	**55.6%**	**55.6%**	**61.1%**	**61.1%**	**55.6%**	**61.1%**
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	**61.1%**	**61.1%**	44.4%	44.4%	44.4%	50.0%	50.0%	44.4%	**61.1%**
MSIH(3)-VAR(1) DNS factors + Log Divisia	**66.7%**	**61.1%**	**55.6%**	44.4%	50.0%	50.0%	**55.6%**	50.0%	**55.6%**
MSIH(3)-VAR(1) DNS factors + Maturity	**61.1%**	**61.1%**	44.4%	**55.6%**	**55.6%**	**55.6%**	**55.6%**	50.0%	**55.6%**
MSIH(3)-VAR(1) DNS factors + Log Assets	**66.7%**	**61.1%**	50.0%	50.0%	**55.6%**	**61.1%**	**61.1%**	**55.6%**	**55.6%**
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	**66.7%**	**61.1%**	44.4%	**55.6%**	**55.6%**	**55.6%**	**55.6%**	50.0%	**55.6%**

Table 13

Realized Performance of Butterfly Trading Strategies

The table shows monthly percentage returns from the recursive implementation of butterfly strategies based on bets on either the individual factor forecasts under a variety of models or on trading the selected Treasury maturities. Sharpe ratios and the associated asymptotic standard errors are also reported. In order to implement the strategies, we trade the 2-, 5-, 10-year maturities on a monthly basis betting on the predicted changes either in the factors or in the yields implied by a range of models. The p-values in the W-McK column refer to West (1996)-McCracken's (2004) nonparametric tests for non-nested models of the null hypothesis that the return differentials are not different from zero. In particular, in the case of the MSIH(3)-VAR(1) model for DNS factors, the null is that of equal performance vs. a single state VAR(1) model; in all other cases, the null hypothesis is that of equal performance vs. the MSIH(3)-VAR(1) model. Next to the p-value, the symbols ↑ and ↓ indicate that a model as a superior (resp. inferior) performance vs. its benchmark and that the null of no difference is rejected with a p-value of 5% or lower. Throughout the table, we have boldfaced all p-values inferior or equal to a 5% size. The results are tabulated for the full OOS sample (January 2006-July 2015) as well as for the Crisis (August 2007-March 2009) and Non-Crisis (January 2006-July 2007;April 2009-July 2015) sub-samples. Panels A, B, C, and D show the results for recursive bets on Level, Spread, Curvature, and the yields, respectively.

Panel A: Trading Level Forecasts under (Partial) Hedging

	Full Sample				**Crisis Sample**				**Non-Crisis Sample**			
	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.
VAR(1) for DNS factors	-0.383	__	-0.071	0.103	0.541	__	0.014	0.224	-0.717	__	-0.273	0.240
MSIH(3)-VAR(1) for DNS factors	-0.842	0.256	-0.139	0.123	-2.838	**0.000↓**	-0.367	0.214	0.967	**0.000↑**	0.246	0.222
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	-0.223	0.793	-0.048	0.103	0.957	**0.031↑**	0.087	0.228	0.398	0.436	-0.003	0.249
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.192	0.435	0.012	0.093	-0.332	0.719	-0.052	0.232	0.662	0.317	0.075	0.225
MSIH(3)-VAR(1) DNS factors + Maturity	-0.641	0.434	-0.109	0.087	-2.440	0.329	-0.289	0.232	0.025	0.137	-0.114	0.234
MSIH(3)-VAR(1) DNS factors + Log Assets	0.166	0.904	0.008	0.085	-0.457	0.628	-0.065	0.216	1.539	0.434	0.357	0.229
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	-0.060	0.123	-0.024	0.122	-0.055	0.093	-0.022	0.210	-0.717	0.545	0.354	0.230

Panel B: Trading Slope Forecasts under (Partial) Hedging

	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.
VAR(1) for DNS factors	0.024	__	-0.199	0.226	0.053	__	**0.805**	0.257	0.017	__	-0.013	0.236
MSIH(3)-VAR(1) for DNS factors	0.149	0.098	0.076	0.083	0.098	0.541	**0.637**	0.247	0.663	**0.000↑**	**1.330**	0.324
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.155	0.122	0.093	0.096	0.331	0.062	**0.468**	0.225	0.482	0.159	0.162	0.216
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.061	0.266	-0.105	0.126	0.248	**0.000↑**	0.219	0.229	0.099	0.486	**-0.969**	0.286
MSIH(3)-VAR(1) DNS factors + Maturity	0.170	0.195	0.116	0.091	0.198	0.151	0.105	0.203	0.751	0.958	**1.142**	0.311
MSIH(3)-VAR(1) DNS factors + Log Assets	-0.015	**0.001↓**	**-0.273**	0.120	0.301	**0.000↑**	**0.503**	0.211	0.113	0.725	**-0.887**	0.254
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.135	0.283	0.051	0.104	0.386	**0.021↑**	**0.524**	0.250	0.512	0.445	0.232	0.230

Table 13 (continued)

Realized Performance of Butterfly Trading Strategies

Panel C: Trading Convexity Forecasts under (Partial) Hedging

	Full Sample				**Crisis Sample**				**Non-Crisis Sample**			
	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.
VAR(1) for DNS factors	-0.096	___	-0.142	0.225	-0.507	___	**-0.448**	0.235	0.003	___	-0.017	0.236
MSIH(3)-VAR(1) for DNS factors	-0.032	0.854	-0.096	0.110	-0.068	**0.041↑**	-0.225	0.212	0.268	**0.000↑**	0.137	0.220
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.130	0.762	0.015	0.084	0.608	0.206	0.223	0.233	0.276	0.270	0.181	0.227
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.138	0.065	0.020	0.119	0.730	**0.006↑**	0.282	0.233	0.515	**0.050↑**	0.136	0.224
MSIH(3)-VAR(1) DNS factors + Maturity	0.352	**0.006↑**	0.165	0.093	0.805	**0.009↑**	**0.501**	0.228	0.895	0.237	**0.656**	0.251
MSIH(3)-VAR(1) DNS factors + Log Assets	-0.210	0.589	**-0.220**	0.092	0.047	0.872	-0.049	0.226	0.384	0.292	-0.031	0.238
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.157	0.718	0.033	0.091	0.875	0.459	0.372	0.234	0.394	0.333	-0.023	0.244

Panel D: Trading Implied Yield Forecasts under (Partial) Hedging

	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.
VAR(1) for DNS factors	-0.140	___	-0.185	0.104	-1.228	___	**-1.269**	0.300	0.171	___	-0.051	0.236
MSIH(3)-VAR(1) for DNS factors	0.211	0.208	0.168	0.088	-1.564	0.641	**-1.602**	0.349	1.079	**0.000↑**	**0.858**	0.274
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	-0.083	0.718	-0.128	0.088	-0.898	0.963	**-0.938**	0.284	0.289	0.883	0.067	0.249
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.251	0.464	**0.205**	0.100	0.156	**0.008↑**	0.116	0.247	0.124	0.472	-0.101	0.250
MSIH(3)-VAR(1) DNS factors + Maturity	0.453	0.952	**0.408**	0.123	1.044	**0.009↑**	**1.003**	0.272	0.562	0.808	0.285	0.261
MSIH(3)-VAR(1) DNS factors + Log Assets	0.258	0.558	**0.211**	0.101	0.547	**0.033↑**	**0.960**	0.204	1.191	0.320	**0.965**	0.263
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	-0.102	0.862	-0.148	0.104	0.689	**0.007↑**	**0.502**	0.222	-1.029	**0.004↓**	**-1.068**	0.264

Table 14

Realized Performance of Butterfly Trading Strategies: The Effects of Accounting for Funding Costs on Margin Positions

The table shows monthly percentage returns from the recursive implementation of butterfly strategies based on bets on either the individual factor forecasts under a variety of models or on trading the selected Treasury maturities. Sharpe ratios and the associated asymptotic standard errors are also reported. In order to implement the strategies, we trade the 2-, 5-, 10-year maturities on a monthly basis betting on the predicted changes either in the factors or in the yields implied by a range of models. The p-values in the W-McK column refer to West (1996)-McCracken's (2004) nonparametric tests for non-nested models of the null hypothesis that the return differentials are not different from zero. In particular, in the case of the MSIH(3)-VAR(1) model for DNS factors, the null is that of equal performance vs. a single state VAR(1) model; in all other cases, the null hypothesis is that of equal performance vs. the MSIH(3)-VAR(1) model. Next to the p-value, the symbols ⇧ and ⇩ indicate that a model as a superior (resp. inferior) performance vs. its benchmark and that the null of no difference is rejected with a p-value of 5% or lower. Throughout the table, we have boldfaced all p-values inferior or equal to a 5% size. The results are tabulated for the full OOS sample (January 2006-July 2015) as well as for the Crisis (August 2007-March 2009) and Non-Crisis (January 2006-July 2007;April 2009-July 2015) sub-samples. Panels A, B, C, and D show the results for recursive bets on Level, Spread, Curvature, and the yields, respectively.

Panel A: Trading Level Forecasts under (Partial) Hedging

	Full Sample				**Crisis Sample**				**Non-Crisis Sample**			
	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.
VAR(1) for DNS factors	-0.383	__	-0.062	0.103	0.512	__	0.013	0.224	-0.621	__	-0.249	0.240
MSIH(3)-VAR(1) for DNS factors	-0.740	0.256	-0.130	0.090	-2.712	**0.000⇩**	-0.362	0.243	0.909	**0.000⇧**	0.233	0.239
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	-0.206	0.793	-0.043	0.124	0.968	**0.033⇧**	0.080	0.243	0.380	0.436	-0.003	0.236
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.032	0.638	-0.011	0.117	0.053	0.834	-0.011	0.216	0.791	0.426	0.114	0.236
MSIH(3)-VAR(1) DNS factors + Maturity	-1.484	0.316	**-0.236**	0.083	-3.149	0.403	-0.408	0.214	0.385	0.577	0.070	0.236
MSIH(3)-VAR(1) DNS factors + Log Assets	0.055	0.774	-0.008	0.128	0.233	0.884	-0.041	0.210	0.845	0.533	0.131	0.237
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.329	0.695	0.032	0.121	-0.028	0.304	-0.019	0.228	0.829	0.684	0.126	0.237

Panel B: Trading Slope Forecasts under (Partial) Hedging

	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.
VAR(1) for DNS factors	0.025	__	-0.172	0.226	0.052	__	**0.818**	0.257	0.016	__	-0.012	0.236
MSIH(3)-VAR(1) for DNS factors	0.139	0.098	0.071	0.103	0.087	0.541	**0.634**	0.250	0.571	**0.004⇧**	**1.375**	0.307
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.151	0.122	0.096	0.121	0.347	0.062	0.448	0.243	0.428	0.159	0.145	0.235
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.119	0.069	0.020	0.118	0.266	0.062	0.285	0.205	0.482	0.708	0.126	0.228
MSIH(3)-VAR(1) DNS factors + Maturity	0.123	0.098	0.027	0.116	0.213	0.060	-0.433	0.246	0.482	0.804	0.126	0.258
MSIH(3)-VAR(1) DNS factors + Log Assets	0.095	0.199	-0.027	0.084	0.153	0.135	0.005	0.246	0.530	0.251	0.272	0.235
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.159	0.958	0.096	0.105	0.225	0.753	0.163	0.226	0.647	0.516	**0.606**	0.263

Table 14 (continued)

Realized Performance of Butterfly Trading Strategies: The Effects of Accounting for Funding Costs on Margin Positions

Panel C: Trading Convexity Forecasts under (Partial) Hedging

	Full Sample				Crisis Sample				Non-Crisis Sample			
	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.
VAR(1) for DNS factors	-0.100	__	-0.130	0.225	-0.449	__	-0.385	0.235	0.003	__	-0.016	0.236
MSIH(3)-VAR(1) for DNS factors	-0.031	0.854	-0.087	0.090	-0.071	**0.047**⬆	-0.200	0.226	0.234	**0.013**⬆	0.127	0.237
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	0.118	0.762	0.015	0.110	0.558	0.206	0.211	0.226	0.275	0.270	0.154	0.238
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.071	0.092	-0.026	0.110	0.499	0.210	0.166	0.225	0.740	0.221	0.473	0.249
MSIH(3)-VAR(1) DNS factors + Maturity	0.159	0.943	0.034	0.087	0.213	0.299	-0.433	0.234	0.482	0.084	0.126	0.237
MSIH(3)-VAR(1) DNS factors + Log Assets	-0.120	0.780	-0.157	0.094	0.130	0.214	-0.009	0.224	0.139	0.090	-0.385	0.244
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	0.092	0.588	-0.012	0.078	0.041	0.967	-0.051	0.224	0.484	0.345	0.101	0.237

Panel D: Trading Implied Yield Forecasts under (Partial) Hedging

	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.	Return	W-McK p-value	Sharpe ratio	SR Std. Err.
VAR(1) for DNS factors	-0.141	__	-0.185	0.104	-1.271	__	-1.102	0.300	0.149	__	-0.044	0.236
MSIH(3)-VAR(1) for DNS factors	0.220	0.208	0.160	0.119	-1.497	0.641	-1.666	0.338	0.977	**0.000**⬆	0.846	0.276
MSIH(3)-VAR(1) DNS factors + Fed Funds Rate	-0.074	0.718	-0.113	0.126	-0.881	0.963	-0.872	0.268	0.262	0.883	0.059	0.273
MSIH(3)-VAR(1) DNS factors + Log Divisia	0.118	0.949	0.073	0.127	-0.382	0.820	-0.419	0.233	0.470	0.396	0.245	0.239
MSIH(3)-VAR(1) DNS factors + Maturity	0.455	0.599	**0.411**	0.109	0.919	0.792	**0.882**	0.264	0.794	0.219	**0.542**	0.252
MSIH(3)-VAR(1) DNS factors + Log Assets	0.059	0.714	0.015	0.104	-0.434	0.973	-0.473	0.236	0.400	0.818	0.173	0.238
MSIH(3)-VAR(1) DNS factors + Treasuries Pct.	-0.127	0.722	-0.172	0.124	-1.328	0.201	**-1.370**	0.311	0.399	0.880	0.192	0.238

Figure 1

Treasury Bond Yields

The plot reports the 3-month, 6-month, 1-, 2-, 3-, 5-, 7-, 10-, and 30-year Constant Maturity Treasury yields. The sample covers the period from January 1982 to February 2015. The data are reported in annualized percentage terms.

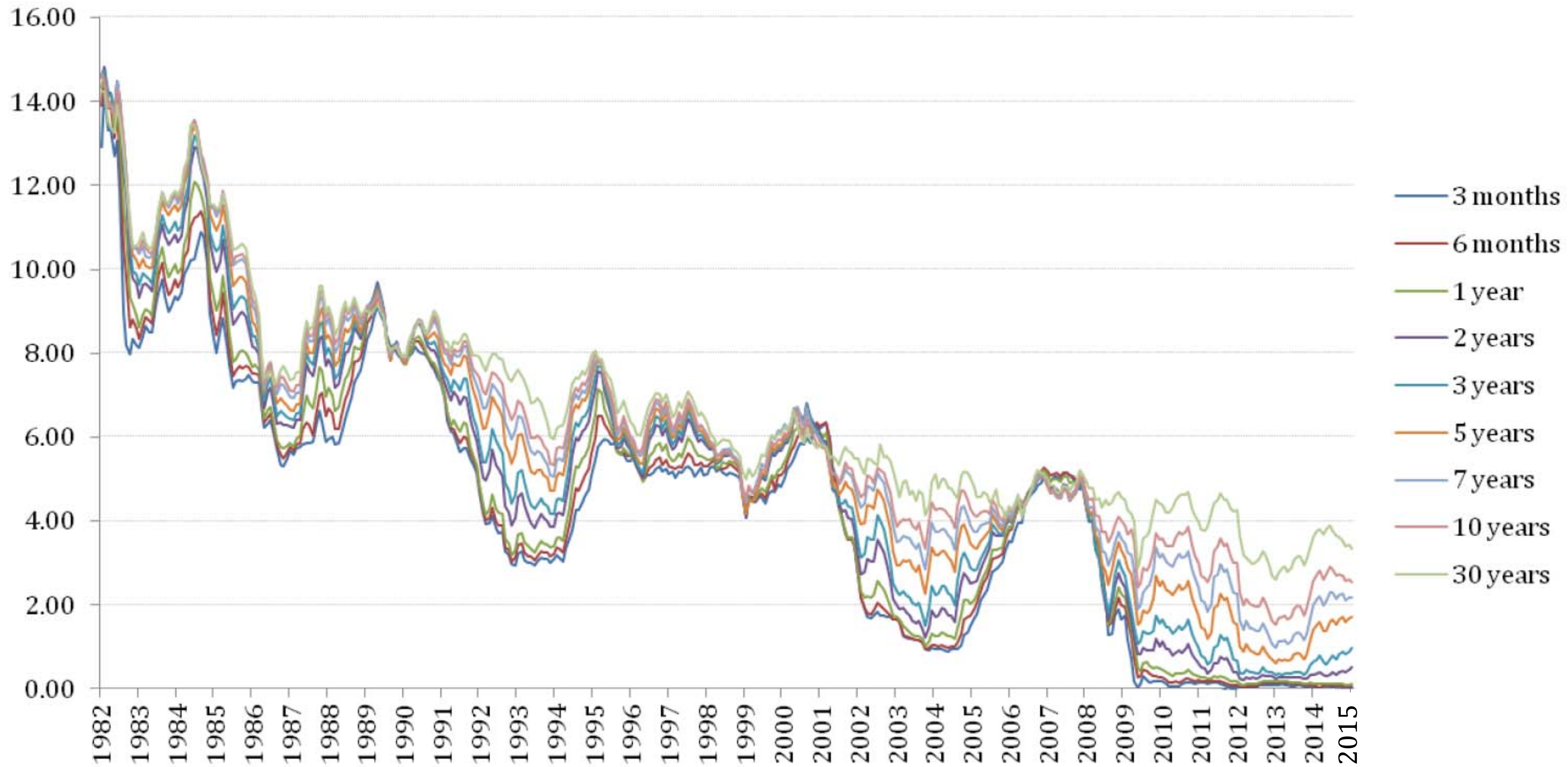


Figure 2

Plots of the Series Used to Capture the Stance of Monetary Policy

The effective Federal Funds Rate is from FRED, at the Federal Reserve Bank of St. Louis. As a proxy for total aggregate monetary base, we use (in natural logs) the Divisia Index MZM ("money-zero maturity", i.e., M2 less small time deposits, plus Institutional money market mutual funds) published by the Federal Reserve Bank of St. Louis. The total assets in the Fed's balance sheet (in natural logs) is from the Federal Reserve's weekly H41 release sampled in correspondence to the last week of each month. The amount of Treasuries in Fed's balance sheet is also sampled at a monthly frequency from the weekly H41 release. The average maturity of the FED's portfolio of Treasuries refers to the New York Fed data of System Open Market Account holdings computed from rough maturity breakdowns (less than 15 days, 16-90 days, 91 days to 1 year, over 1 year to 5 years, over 5 years to 10 years, and over10 years), by imputing the mid-point of each bracket, for the 2003-2015 sample. Between 1982 and 2002 we use the average maturity of the Fed's Treasury holdings calculated by Kuttner (2006).

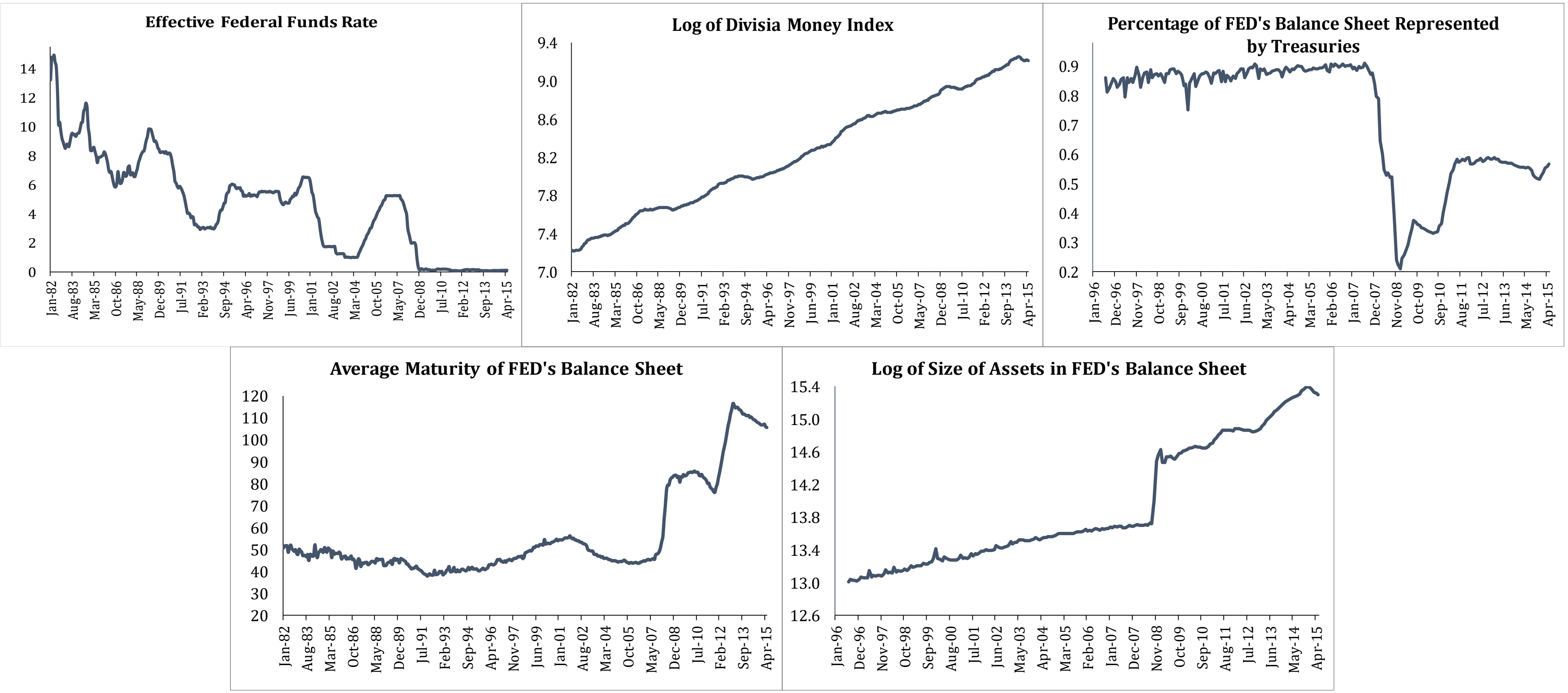
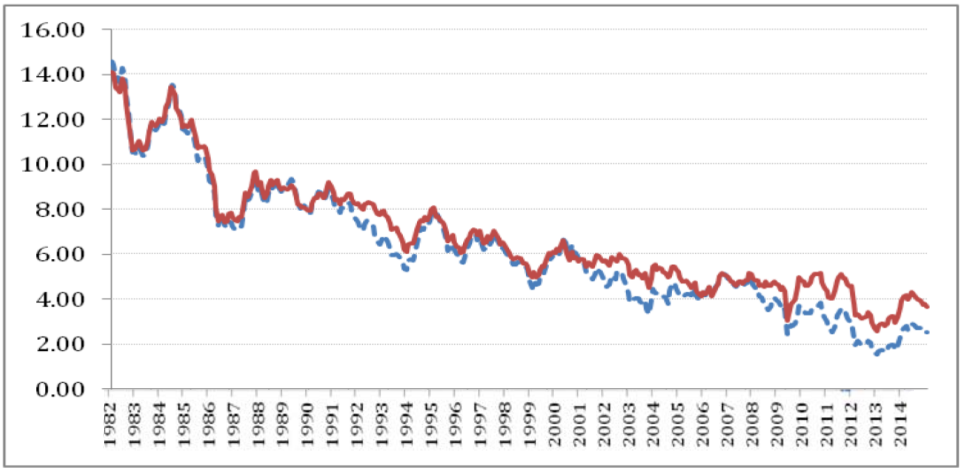


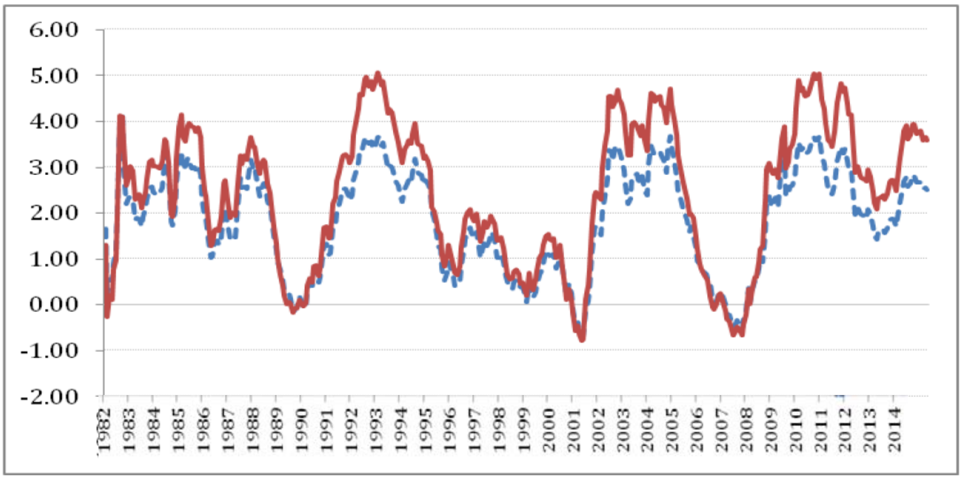
Figure 3

Estimated vs. Empirical Yield Curve Factors

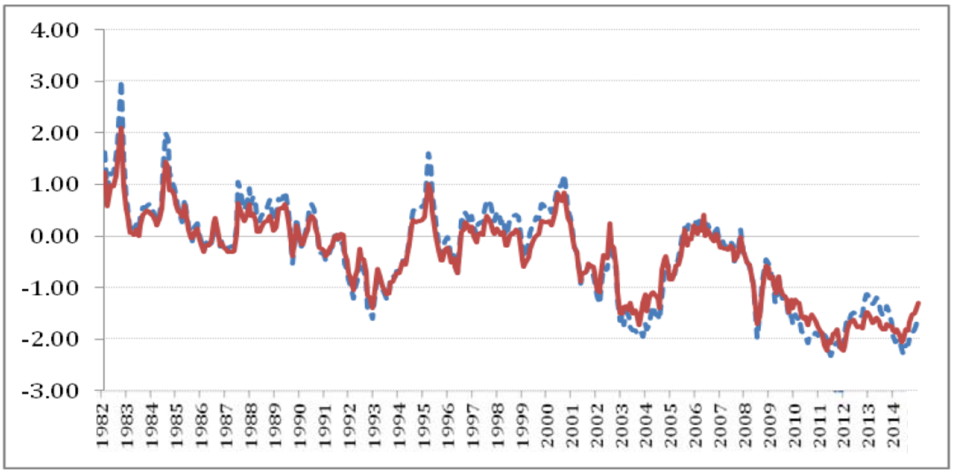
The plots show recursive, cross-sectional estimates obtained between Jan. 1982 and July 2015. We also show empirical proxies for level (the 10-year yield), slope (10-year minus 3-month yield), and curvature (twice the 2-year minus the sum of the 3-month and 10-year yields).



Red line: $\hat{L}_t$ Blue line: Empirical level



Red line: - $\hat{S}_t$ Blue line: Empirical Slope



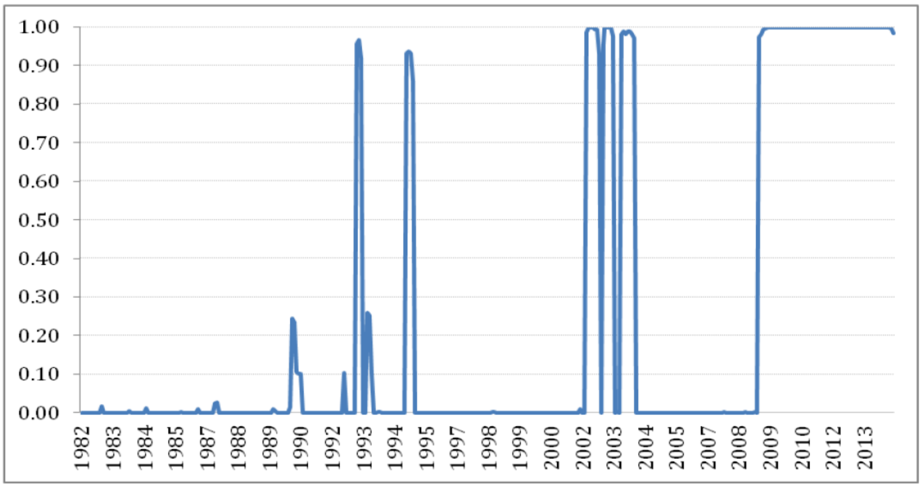
Red line: 0.3 $\hat{C}_t$ Blue line: Empirical Curvature

Figure 4

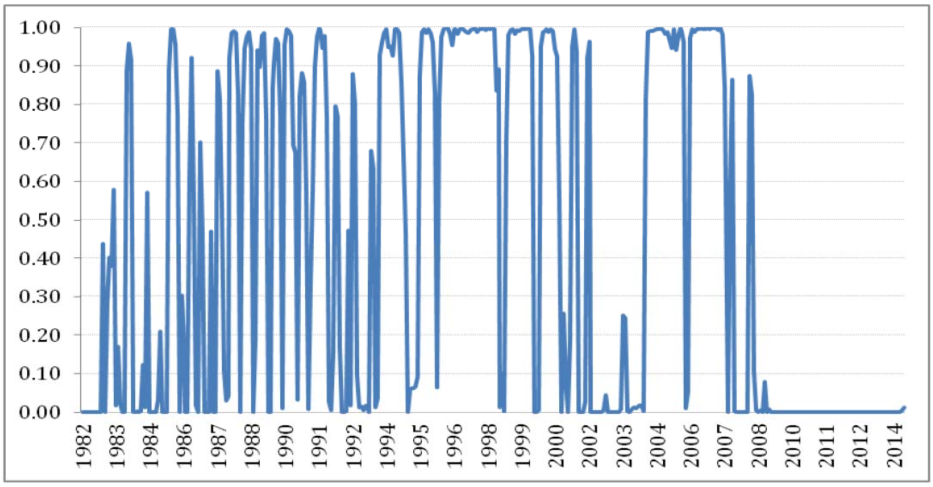
Smoothed Regime Probabilities from a Three-State VAR(1) Model for the DNS Factors

The plots show the full-sample, exp-post smoothed state probabilities implied by a MSIH(2)-VAR(1) model for the dynamically estimated Nelson-Siegel factors from US Treasury yields over a 1982:01-2015:06 sample.

Regime 1



Regime 2



Regime 3

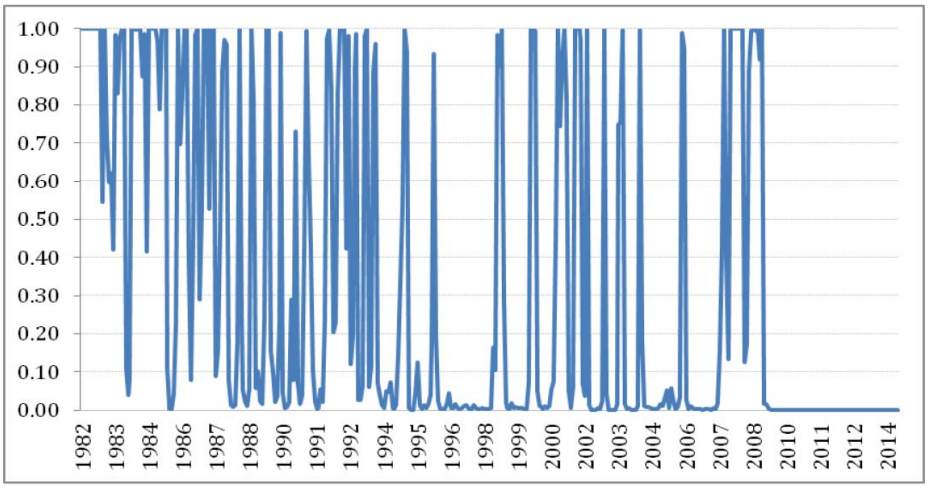


Figure 5

Changes in the Shape of the Yield Curve and Structure of DNS Factors on Selected Dates

The plots show how the yield curve and the estimated DNS factors change over the samples June 2008 – April 2009 and October 2011 – January 2013 when Markov switching VAR(1) models augmented with the log of the size of the FED's balance sheet and the average maturity of the FED's balance sheet lead to diverging forecasting performances.