A Comparison of Multi-Factor Term Structure Models for Interbank Rates

Abstract

In this paper, we present a robust predictive comparison of several continuous-time multi-factor models in the context of interbank rates. Recognizing the specific dynamics of the short-term segment of the yield curve, we examine the U.S. money market by extending two continuous-time frameworks with different factor structures, the Chan-Karolyi-Longstaff-Sanders (CKLS) model and the arbitrage-free dynamic Nelson-Siegel (AFDNS) model. A battery of formal forecasting accuracy tests is employed to select a subset of superior predictive models. Despite a better goodness-of-fit measure, additional factors improve the forecasting performance only for the CKLS family. With implications for monetary policy formulation, we found evidence of two separate maturity segments as the three-factor AFDNS and the five-factor CKLS models outperform parsimonious benchmarks in predicting the interbank rates for very short maturities. Our comparative forecasting results are re-confirmed with stronger out-of-sample performance for the five-factor CKLS model when the post global financial crisis sub-sample is analyzed.

Key words: interbank rates, continuous-time models, multi-factor term structure models, forecasting tests

JEL: G15, G17, C22, C13

1. Introduction

Research on interest-rate dynamics evolved in a rather continuous manner with an impressive list of models being developed over the last 40 years, from the earliest classic interest rate models such as Vasicek (1977), Cox, Ingersoll and Ross (1985), and Chan et al. (1992) to the most recent additions such as Ajello et al. (2021) and Vayanos and Vila (2021).

The focus of this study is driven by the important role that out-of-sample model validation plays in the modelling of the yield curve (Campbell and Shiller, 1991; Cochrane and Piazzesi, 2005; Coroneo and Caruso, 2022). Models that perform well empirically are of great relevance to portfolio allocation, risk management, asset pricing, and monetary policy formulation. In its early stages, the empirical literature on the term structure of interest rates concentrated on three-factor modelling (e.g., Chen, 1996; Balduzzi et al., 1996), given the findings by Litterman and Scheinkman (1991) that three factors (described as level, slope and curvature) can explain over 95% of the fluctuations in changes in the yield curve. Following the global financial crisis of 2007-2009 (hereafter GFC), the Basel II Committee on Banking Supervision (2010) recommended that banks should model the yield curve using more risk factors. In line with these recommendations, several studies (see Christensen et al., 2009; Adrian et al., 2013; van Deventer et al., 2013; Steeley, 2014a) considered a larger set of factors in pricing the time series and cross section of interest rates.

In this paper, we follow this line of recommendations, and examine the benefits of extending well-established multi-factor modelling frameworks in terms of their forecasting performance. From the vast range of models proposed in the literature for the term structure of interest rates, we focus on the extension of two multi-factor general frameworks. The first class of models we employ is a popular approach among central banks in estimating the yield curve, namely the arbitrage-free dynamic Nelson-Siegel (AFDNS) family developed by Christensen, Diebold and Rudebusch (2009, 2011and 2013). The second class of models belongs to the general multi-factor Chan-Karolyi-Longstaff-Sanders (Chan et al., 1992) modelling framework (CKLS).

The main objective of our research is to conduct an extensive and robust comparative forecasting analysis of various extensions of CKLS and AFDNS models along with more parsimonious benchmark methods. Within the CKLS family, following Nowman (2006), we examine the four- and five-factor extensions, while for AFDNS models we explore the three- and five-factor specifications (see Christensen et al., 2009).

The CKLS and AFDNS generalised modelling frameworks are both arbitrage-free, affine and retain a yield-based[1] factor structure (i.e., they do not involve macro-economic variables). Nevertheless, they are different by construction, through the nature (observable versus latent) of the factors that enter the model. The CKLS family directly employs interest rates of different maturities as correlated observable "factors", while the AFDNS models are curve-fitting models based on latent factors. Consequently, another objective of this study is to examine which factor structure yields superior out-of-sample performance (CKLS versus AFDNS) and to explore the benefits of such extensions inside each modelling framework (three-factor versus five-factor AFDNS and four-factor versus five-factor CKLS). The CKLS model can be interpreted under the general affine framework described in Duffie and Kan (1996), whereby the state variables are the spot rates themselves. The AFDNS models are derived from the dynamic Nelson-Siegel model (Diebold and Li, 2006) that initially generate a term structure that is not arbitrage-free. To address this problem, Christensen, Diebold, and Rudebusch (2007) (hereafter CDR) developed an arbitrage-free dynamic Nelson-Siegel (AFDNS).

Our study focuses on interbank LIBOR rates and is one of the very few studies concentrating only on the short-term segment of the yield curve (see Bali and Wu, 2006). Interbank rates are fundamental to money markets, and they are the reference rates in many other financial products from asset-backed securities to mortgage loans and structured products. According to State Street Global Advisors, almost $200 trillion of debt contracts use

[1] Joslin et al. (2013) differentiates between macro and yield-based models of the term structure of interest rates.

LIBOR as the reference rate.[2] When studying only the interbank rates we take advantage of the fact that they are deposit rates and therefore they can be conceptualised as the market factors themselves. This type of modelling can bring significant benefits for interest rate derivatives applications such as pricing structured products or performing cash-flow risk management simulations for securitizations.

We employ daily USD-LIBOR rates of various maturities to analyse the predictive power of seven model-specifications over the period January 1998 to July 2019. The models in competition are the continuous-time three- and five-factor AFDNS models, the four- and five-factor CKLS models on one side, and the more parsimonious univariate and vector autoregressive (AR and VAR) models, and the random walk process, on the other side. The out-of-sample model performances are evaluated using formal statistical tests including the equal predictability tests of Diebold-Mariano (1995) and Clark-West (2007), as well as the superior predictive ability (SPA) test of Hansen (2005) and the model confidence set of Hansen, Lunde, and Nason (2011) (hereafter, MCS).

Despite mixed results from various static measures of forecasting accuracy, based on the formal statistical tests mentioned above, we also find that including additional factors have the opposite effect on the forecasting performance of the two continuous-time frameworks, with an improvement in the accuracy of forecasting LIBOR rates only for the CKLS extended framework. In the case of the AFDNS family, the three-factor specification has superior out-of-sample performance than the extended five-factor model. In terms of maturity, the continuous-time models are better at predicting the shortest-term LIBOR rates (one-week, one-month, and three-month maturities) than more parsimonious models such as random walk and

[2] https://www.ssga.com/investment-topics/environmental-social-governance/2019/05/what-happens-when-libor-is-phased-out.pdf

autoregressive models. To check for the robustness of our results in terms of out-of-sample comparison, we conduct the same analysis over the post-crisis sub-sample and reach the same conclusions.

The structure of this paper is as follows. Section 2 provides a literature review. Section 3 outlines the methodology behind the two continuous-time multi-factor modelling frameworks. Following a description of the data in Section 4, we present and interpret the empirical results across all fitted models in Section 5. We conduct the forecasting analysis and discuss the results in Section 6. Concluding remarks and further research suggestions are presented in the last section.

2. Literature review

An overwhelming number of approaches have been proposed to accurately characterise and predict the evolution of interest rates over time. There are theoretical models imposing no-arbitrage restrictions and there are empirical models that fit the yield curve extremely well. Factor structure is pervasive in financial markets and financial economic theory, as financial asset returns display this structure (see Campbell et al., 1997). Early arbitrage-free factor models such as Vasicek (1977) and Cox, Ingersoll, and Ross (1985), CIR hereafter, involved only one factor, thereby failing to capture certain shapes observed empirically in the dynamics of the yield curve. Following the findings of the principal component analysis (PCA) reported by Litterman and Scheinkman (1991), modern empirical term structure models recognise a higher-dimensional factor structure, with three factors (level, slope, and curvature) being able to explain well over 95% of yield variation. Consequently, many three-factor term structure models were proposed in the literature. Duffie and Kan (1996) proposed the general affine no-arbitrage theory where pure discount bond prices are affine exponential combinations of the latent factors. Dai and Singleton (2000) classified the affine models into nine classes of equivalence, comprising numerous classical no-arbitrage factor models. Pursuing a different approach, Diebold and Li (2006) transformed the static model of Nelson and Siegel (1987) into a dynamic factor model (hereafter, DNS) by changing the parameters of the Nelson-Siegel model into time-variant factors, while keeping the same functional form for the factor loadings. Easy to implement and calibrate, the DNS model became a popular empirical model that produced superior fitting to the yield curve in comparison to the affine type models, despite its lower forecasting performance. Its empirical success prompted the development of further generalisations, including the four-factor extensions proposed by Svensson (1995) and Bjork and Christensen (1999).

In an earlier study, Steeley (1990) explored a four-factor extension of the NS model and concluded of a non-significant contribution of the additional factor in explaining the variation in the yield curve. However, more recently Steeley (2014a) found that the change in the volatility (as a fourth factor) was responsible for the undulations observed in the shape of the yield curve on the day quantitative easing policy was announced. Filipovic et al. (2014) suggested that a minimum of five factors (three term structure factors and two unspanned factors) seem to do well in capturing the dynamics of both the term structure and the volatility of interest rate changes over the period that followed the GFC. In a study with major implications for term structure modelling, Cochrane and Piazzesi (2005) identified a new return-forecasting factor − a tent-shaped linear combination of forward rates, which despite being poorly related to the level, slope and curvature factors, is able to forecast changes in short-term interest rates.

A controversial aspect in term structure modelling is the no-arbitrage consistency condition. In theory, a good yield curve model must ensure consistency between the time dynamics of the yields and the shape of the yield curve across maturities. Putting theory and practice together, imposing the affine no-arbitrage structure of Duffie and Kan (1996) on the DNS model, Christensen, Diebold, and Rudebusch (2007) proposed a new class of arbitrage-free NS models called the AFDNS models. The effect of imposing a no-arbitrage restriction on out-of-sample performance has resulted in mixed empirical results. Earlier studies, including Ang and Piazzesi (2003) and Monch (2008), found that the no-arbitrage restriction significantly improves the yield curve forecasts for VAR models; Almeida and Vicente (2008) reported similar evidence for polynomial models. Using different data sets, Gimeno and Marques (2009) and Christensen et al. (2011) conducted similar comparisons between DNS and AFDNS and reach the same conclusion. Nevertheless, other studies, including Coroneo et al. (2011), argue that there is little gain in imposing no-arbitrage. Using a parametric block-bootstrapping procedure, Coroneo et al. (2011) compare the estimated factor loadings from the Nelson-Siegel model with their counterparts in the no-arbitrage Nelson-Siegel affine version and conclude there is no statistical difference. Meanwhile, Duffee (2011) and Joslin et al. (2011, 2013), for example, are more pessimistic and argue that the predictive gains from dynamic restrictions are superior to those from arbitrage-free restrictions. Furthermore, Duffee (2011) found that the restriction on the first factor as a random walk produced superior forecasts to both affine and dynamic Nelson-Siegel type model specifications. At the same time, Joslin et al. (2011) show that forecasts of yield factors using an arbitrage-free Nelson-Siegel (AFDNS) model are equivalent to forecasts based on an unconstrained VAR(1) representation of yield factors. A similar conclusion has been obtained for canonical Gaussian macro-finance term structure models for which the estimated joint distribution is almost identical to the estimate from an economic-model-free factor vector autoregression, see Joslin et al. (2013). These advances are useful when the pricing factors determining bond prices (and indirectly yield rates) are known.

By comparison with the current best state-of-the-art models such as AFDNS (three- and five-factors), the multi-factor CKLS framework employs observed yields as factors and directly estimates the correlation structure among the interest rates. For particular values of certain parameters, the CKLS framework nests multivariate versions of classical models such as the Vasicek and CIR models.

Many studies (e.g. Duffee, 2002; Diebold and Li, 2006; Almeida and Vincente, 2008, Carriero, 2011, Carriero and Giacomini, 2011, Duffee, 2011; Matsumura et al., 2011, Steeley, 2014b) have evaluated the forecasting performance of competing term-structure models, employing a number of statistical and economic loss functions as measures of forecasting accuracy. The studies by Steeley (1990), Christensen et al. (2011), and Duffee (2011) presented evidence that a random walk process may produce similar results. Steeley (2014b) demonstrated that this may be true only for some maturities of the spot yields but not for all when short-term rates are near zero.

3. Modelling Frameworks

3.1. The CKLS framework

Following the approach of Nowman (2006), we estimate the four- and five-factor extensions of the CKLS multivariate model. We employ the Gaussian estimation methods of continuous-time dynamic systems developed by Bergstrom (1984) to obtain quasi-maximum likelihood (QML) parameter estimates. Since the four-factor specification is nested within the five-factor specification, we provide the explicit dynamics for the latter only.

The CKLS five-factor continuous-time model of the term structure (hereafter CKLS-5F) is represented by the general system of stochastic differential equations:

$$
\begin{aligned}
dr_1(t) &= [\alpha_1 + \beta_{11} r_1(t) + \beta_{12} r_2(t) + \beta_{13} r_3(t) + \beta_{14} r_4(t) + \beta_{15} r_5(t)]dt + \zeta_1(dt) \\
dr_2(t) &= [\alpha_2 + \beta_{21} r_1(t) + \beta_{22} r_2(t) + \beta_{23} r_3(t) + \beta_{24} r_4(t) + \beta_{25} r_5(t)]dt + \zeta_2(dt) \\
dr_3(t) &= [\alpha_3 + \beta_{31} r_1(t) + \beta_{32} r_2(t) + \beta_{33} r_3(t) + \beta_{34} r_4(t) + \beta_{35} r_5(t)]dt + \zeta_3(dt) \\
dr_4(t) &= [\alpha_4 + \beta_{41} r_1(t) + \beta_{42} r_2(t) + \beta_{43} r_3(t) + \beta_{44} r_4(t) + \beta_{45} r_5(t)]dt + \zeta_4(dt) \\
dr_5(t) &= [\alpha_5 + \beta_{51} r_1(t) + \beta_{52} r_2(t) + \beta_{53} r_3(t) + \beta_{54} r_4(t) + \beta_{55} r_5(t)]dt + \zeta_5(dt)
\end{aligned} \tag{1}
$$

where $r(t) = [r_1(t), r_2(t), ..., r_5(t)]'$ is the vector of the observable LIBOR rates variables, $\alpha = [\alpha_1, \alpha_2, ..., \alpha_5]'$ is the vector of the *drift* parameters, $\beta = \{\beta_{ij}\}_{1 \le i, j \le 5}$ is the *feedback* matrix, and the innovations $\{\zeta_i\}_{1 \le i \le 5}$ are correlated random measures defined on all subsets of the half line $0 < t < \infty$ with finite Lebesgue measure, such that $E[\zeta_i(dt)] = 0$ for all $i = 1, ..., 5$. In the context of interest rates, Nowman (1997) relaxed the assumption of constant volatility by allowing a special type of heteroskedasticity.[3] Volatility is considered as a step function, changing the value at the beginning of each unit observation period and then remaining constant over that time interval. For any $t > 1$, the unit period is denoted by the interval $[t'-1, t']$ where $t'-1$ is the largest integer less than t. Nowman (1997) adjusted the conditional volatility only and proposed an approximate continuous-time model to which Bergstrom's Gaussian methods can be applied, also reducing the temporal aggregation problem.

When multi-factor models are considered (see Nowman, 2006), the assumption of constant volatility during the unit period translates into a different variance-covariance matrix of the innovations with the following adjusted elements: $\Sigma^*(r,t) = \{\sigma^*_{ij}\}_{1 \le i, j \le n}$ where

[3] Some of the classical models nested in CKLS general framework, such as Cox, Ingersol, and Ross (1985), and Brennan and Schwartz (1980) are non-Gaussian. To estimate them, Nowman (1997) employs their Gaussian approximations to obtain quasi-ML estimates.

$\sigma_{ii}^* = \sigma_i^2 r_i^{2\gamma_i}(t'-1)$ and $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j r_i^{\gamma_i}(t'-1) r_j^{\gamma_j}(t'-1)$ involving the correlations $\rho_{ij}(1 \le i < j \le 5)$ and the leverage-effect parameters γ_i. The complete vector of structural parameters $\theta = (\alpha_i, \beta_{ij}, \sigma_i, \gamma_i, \rho_{ij})_{1\le i,j\le 5}$ comprises 50 parameters. By imposing specific restrictions on γ such as $\gamma = 0$, $\gamma = 0.5$ and $\gamma = 1$, one could obtain multi-factor versions of the Vasicek (1977), Cox, Ingersol, and Ross. (1985), and Brennan and Schwartz (1980) models, respectively. The inference is based on the unique "exact" discretization algorithm introduced by Bergstrom (1984, Theorem 2).

The flexibility of the multifactor CKLS framework is twofold: the drift component allows for feedback effects $(\beta_{ij})_{1\le i,j\le 5}$ and the diffusion part allows for correlated factors under a general formulation involving the leverage-effect parameters $\gamma_i(1 \le i \le 5)$. The multi-factor CKLS family of models takes into account the close relationship that exists among yields of different maturities. The empirical feature that interest rates for different maturities are highly correlated (Coroneo et al., 2011) is captured through the measurement of the time-variant covariance matrix $\Sigma^*(r,t)$.

To estimate the five-factor CKLS model, we use the following discrete-time analogue model for the Gaussian approximate model

$$r(t) = e^{\beta} r(t-1) + (e^{\beta} - I)\beta^{-1}\alpha + \varepsilon(t) \qquad t = 1,2,...,T \tag{2}$$

where $r(t) = [r_i(t)]'_{1\le i\le 5}$, $\varepsilon(t) = [\varepsilon_i(t)]'_{1\le i\le 5}$, $\alpha = (\alpha_i)'_{1\le i\le 5}$, $e^{\beta} = I + \sum_{k=1}^{\infty} \frac{1}{k!}\beta^k$ and

$$E[\varepsilon(t)\cdot\varepsilon'(t)] = \int_0^1 e^{r\beta}\Sigma^*(r,t)e^{r\beta'}dr \equiv \Omega(r,t)$$

The parameter estimates are the elements of the solution θ^* that minimizes the minus twice the logarithm of the Gaussian likelihood function LF:

$$L(\theta) = -2\log(LF(\theta)) = \sum_{t=1}^{T}\log(|\Omega(r,t)|) + \sum_{t=1}^{T}\varepsilon_t'\Omega^{-1}(r,t)\varepsilon(t) \tag{3}$$

3.2. The AFDNS framework

For our forecasting comparison, we will also investigate the class of arbitrage-free Nelson Siegel models (AFDNS) with three and five factors (hereafter, AFDNS-3F and AFDNS-5F, respectively). This new class of models combines the canonical affine structure of Duffie and Kan (1996) with the widely used Nelson-Siegel yield-curve dynamic specification (Diebold and Li, 2006). As a result, AFDNS models have several desirable features: theoretical consistency, tractability, and significant goodness of fit.

The derivation of the AFDNS yield curve model and its extension (generalization) to five factors is described in detail in Diebold and Rudebusch (2013). Given the limited space of exposition, here we only present the five-factor variant developed by Christensen et al. (2009). Typically, the state-space AFDNS models are defined by two equations: the measurement and the transition equations.

3.2.1. The measurement equation

Trying to approximate an arbitrary smooth yield curve across maturities, Nelson and Siegel (1987) proposed the following static model:

$$NS(1987): \qquad y(\tau) = \beta_0 + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)\beta_1 + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)\beta_2 \qquad (4)$$

where $\beta_0, \beta_1, \beta_2$ and $\lambda > 0$ are constant parameters.

Diebold and Li (2006) transformed the static NS model into a dynamic state-space model (DNS) by considering the parameters $\beta_0, \beta_1, \beta_2$ as time-variant autoregressive processes. Keeping the same functional form for the factor loadings, the new parameters can be interpreted as the classic latent factors explained by Litterman and Scheinkman (1991), namely the level (L_t), slope (S_t) and curvature (C_t) factors respectively:

$$DNS(2006): \qquad y(t,\tau) = L_t + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)S_t + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)C_t \qquad (5)$$

where τ is the time to maturity. The DNS model is still an empirical model lacking solid theoretical grounds such as the arbitrage-free restriction (see Filipovic, 1999; Diebold et al., 2009). Intuitively, Christensen et al. (2007) searched for the best approximation to DNS within the Duffie-Kan affine arbitrage free class that preserves the form of DNS factor loadings.

According to the general affine theory, affine models are tractable with the spot rates being affine combinations of the state variables:

$$y_t(t,\tau) = -\frac{B(t,\tau)'}{\tau} X_t - \frac{C(t,\tau)}{\tau} \tag{6}$$

Christensen (2009) shows that the arbitrage-free restriction has a minimal impact, leaving the dynamics under the real-world unchanged with only the addition of a constant yield-adjustment term in the measurement equation when compared to the respective DNS equation (5). Moreover, Steeley (2014c) shows how the yield-adjustment term can be also decomposed into shape-based components that bring extra contribution to the original level, slope, and curvature factors, explaining in this way the superior flexibility of AFDNS models over NS models.

To extend the AFDNS framework beyond three factors, Christensen et al. (2009) follow the approach used in the four-factor Svensson (1995) extension of the Nelson-Siegel model (DNSS), where a second curvature is introduced. They show that if the Nelson-Siegel analytical form for the factor loadings is to be preserved, then an arbitrage-free AFDNS approximation for four factors is not possible under Gaussian dynamics.[4] Moreover, they conclude that in order to impose arbitrage-free restrictions, both the slope and curvature factor have to come in pairs and with the same mean-reversion rate. Consequently, the AFDNS framework can be extended this way only to an odd number of factors that can be interpreted

[4] Sharef and Filipovic (2004) propose another four-factor arbitrage DNSS model by assuming square root processes as in the CIR model instead of Ornstein-Uhlenbeck processes for the state variables.

as level, slope(s) and curvature(s). Therefore, the next arbitrage-free AFDNS extension is the five-factor representation.

Allowing for measurement errors in the spot rates $y_t(\tau)$, the final measurement equation for the five-factor AFDNS model is given by:

$$y_t(t,\tau) = \Lambda X_t - \frac{C(t,\tau)}{\tau} + \varepsilon_t \tag{7}$$

where $X_t = (X_{1t}, X_{2t}, X_{3t}, X_{4t}, X_{5t}) = (L_t, S_{1t}, S_{2t}, C_{1t}, C_{2t})$ is the vector of the latent factors; the measurement errors $\varepsilon_t \in R^N$ are independent with the diagonal volatility matrix $H \in R^{N\times N}$ and $-\frac{C(t,\tau)}{\tau}$ is the yield adjustment term explicitly characterised in Christensen et al. (2009, Appendix B). The five-factor loadings matrix $\Lambda \in R^{N\times 5}$ is computed such that the DNS structure is maintained:

$$\Lambda = \begin{pmatrix} 1 & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} - e^{-\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} - e^{-\lambda_2\tau_1} \\ 1 & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2} & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} - e^{-\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_2}}{\lambda_2\tau_2} - e^{-\lambda_2\tau_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N} & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} - e^{-\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_N}}{\lambda_2\tau_N} - e^{-\lambda_2\tau_N} \end{pmatrix} \tag{8}$$

3.2.2. The transition equation

The transition equation for the AFDNS model is implied by the canonical affine form $A_0(3)$ of Dai and Singleton (2000), which can be naturally extended to a new theoretical $A_0(5)$ class of equivalence to accommodate five factors. Models from this affine class are state-space models for which the short rate r_t is a linear combination of the state variables $\{X_{it}\}_{1\le i\le 5}$ with specific dynamics:

$$r_t = \delta_0 + \delta' X_t \tag{9}$$

$$dX_t = K[\theta - X_t]dt + \Sigma dW_t \tag{10}$$

where $\delta_0 \in R$, $\delta \in R^5$, $\theta \in R^5$ and $K \in R^{5\times5}$ are the mean-reversion parameters, $W_t \in R^5$ are independent Wiener processes with the diagonal volatility matrix $\Sigma \in R^{5\times5}$.

The state-space (discrete-time) transition equation for the state-space AFDNS model is derived from the continuous-time equation (10):

$$X_t = (I - \exp(-K\Delta t))\theta + \exp(-K\Delta t)X_{t-1} + \eta_t \tag{11}$$

where Δt is the time between observations (1/252) and the disturbances η_t are uncorrelated with the volatility matrix $Q \in R^{5\times5}$ and iid distributed $\eta_t \sim N(0,Q)$.

The state-space AFDNS model, defined by the measurement error (7) and transition equation (11), is estimated[5] using the standard estimation approach of combining the maximum likelihood method with Kalman filter technique, which yields both efficient and consistent parameter estimates.

4. Data

We employ interbank LIBOR rates for just over 20 years, from 2 January 2, 1998 to July 16, 2019. In line with the practice of central banks and to avoid unwanted effects (Nymand-Andersen, 2018; Wahlstrom et al., 2022), we collect the daily one-week, one-, three-, six- and 12-month LIBOR rates for the U.S. dollar (USD), from Bloomberg Markets. The model estimations are based on the period from January 2, 1998 to September 29, 2017, with a total of 4,990 observations. The forecasting sample contains the following 452 daily observations (October 2, 2017 - July 16, 2019) which are used to produce the series of 200 forecasts for 1-, 3-, 6- and 12-month horizons.

[5] The Kalman filter estimation approach in the context of AFDNS is presented in detail in Christensen et al. (2007).

Figure 1 presents the time series plots of the 1-week, 3- and 12-month LIBOR rates over the analysed period. We try to interpret Figure 1 in the context of both continuous-time modelling frameworks, ADDNS and CKLS respectively. The time series look non-stationary with inconsistent mean and variance. Their time-paths are very similar to random walks, especially for the longer maturities of 3- and 12-month LIBOR rates, while for the 1-week time series there are several significant spikes and drops. In addition, the 1-week LIBOR time series exhibits a more pronounced zig-zag effect suggesting a higher volatility compared to the smoother dynamics of 1-year LIBOR curve. In addition, while these LIBOR rates seem to move relatively together, they are not solely moving in parallel shifts. We can observe that there are several periods where the 1-year LIBOR rate is lower than the 1-week LIBOR rate (1998, 2001, 2003, 2008 and 2019), implying that at these points in time there are changes in the slope of the LIBOR curve. However, this relative position between the 1-week and 12-month LIBOR rates is short lived, creating the so called "butterfly effect" (Martellini et al., 2003) which can be interpreted as the presence of a minimum of three latent factors (level, slope, and curvature) within a multi-factor model. When we turn to the CKLS framework, Figure 1 shows important information regarding the covariance-dynamics between 1-week and 12-month LIBOR rates. Not including this information in a multi-factor CKLS model might affect both the parameter estimation and the forecasting accuracy of the model.

[FIGURE 1 ABOUT HERE]

While the sample period proves to be quite rich in terms of unanticipated and policy-induced shocks and their complex consequences, the main shock corresponds to GFC. The volatility-clustering pattern in the first difference series indicate two distinct periods of higher volatility in daily rate changes. The first period, 2000-2002, corresponds to the impact of a multitude of factors such as the introduction of the euro, the creation and burst of the dot.com bubble, and the 9/11 terrorist attack. The second period, 2007-2009, relates to the GFC, triggered by the housing bubble and the collapse of the U.S. sub-prime mortgage market. The sharp decrease in the level of interbank rates appears in October 2008 after the collapse of Lehman Brothers.

Table 1 contains the summary statistics for all five LIBOR rates in this study. For these panel data set the means increase with maturity, however the LIBOR rate with the shortest maturity (1-week) has the highest standard deviation. Looking at the autocorrelation structure, the majority of the time series examined are highly persistent and non-stationary as reflected by the augmented Dickey-Fuller (ADF) test results. These results are confirmed by other formal statistical tests such as the Phillips-Perron (*PP)* and Kwiatkowski-Phillips-Schmidt (KPS) tests. We conclude that all LIBOR rates under consideration are integrated of order 1 (I1) as all the time series become stationary in the first difference.

[TABLE 1 ABOUT HERE]

For the U.S. money market data, we investigate a possible change in the dynamics of the USD-LIBOR rates because of the GFC. The initial estimation sample is divided into the pre-crisis sub-sample (January 2, 1998 – October 16, 2008) and post-crisis sub-sample (October 17, 2008– September 29, 2017) and all the models are re-examined. The summary statistics for the two subsamples are presented in Table 2. As these descriptive statistics indicate a possible structural break in the parameters because both the sample mean, and volatility are much lower in the post-crisis period than in the pre-crisis period across all the maturities. As a robustness check of our forecasting analysis, we also examine and compare the out-of-sample performance of the seven competing models based on the post-crisis sub-sample spanning from October 17, 2008 to September 29, 2017.

[TABLE 2 ABOUT HERE]

5. Estimation results

5.1. Estimation results for the four- and five-factor CKLS models

For the estimation of the four-factor CKLS extensions, we analysed the LIBOR rates with one-week, one-, three- and six-month maturities, while for the five-factor CKLS specification we considered the same maturities but added the one-year LIBOR rates.[6] We obtained the QMLE

[6] For the four-factor CKLS model, we used the first four maturities of the LIBOR curve that are available. The reason for not including the one-year LIBOR rate in the initial four-factor

estimates of θ, comprising 34 and 50 parameters for the four- and five-factor models, respectively.

The parameters of interest are the feedback matrix β_{ij}, the level effects γ_i, and the covariance matrix σ^*_{ij}. We present the drift parameters for both extensions in Table 3 and the diffusion parameters in Table 4. The drift parameters are almost all statistically significant indicating that there is a strong feedback effect among the interbank rates. For the CKLS model, the drift parameters are the five-dimensional intercept vector α with most of its components being significant and the feedback matrix β of 25 components whose estimates produce evidence of feedback in most directions.

[TABLE 3 ABOUT HERE]

[TABLE 4 ABOUT HERE]

As for the diffusion part, the gamma parameters are significant for the USD-LIBOR rates. It seems that the transition from four to five factors introduces a positive bias, as the highest values are realised at one-week maturity within the five-factor extensions ($\gamma_1(USD) = 1.087$). The volatility parameter estimates decrease in value with maturity. This suggests that the term-structure of interest rates is structurally similar in our multi-factor modelling and perhaps closer to a multi-factor Brennan-Schwartz type model.

The estimates for the correlation coefficients are all positive under the CKLS model. The correlation level decreases as the distance between maturities increases. For any two adjacent maturities, the correlation is strongest when compared to other more distant maturities. Moreover, this adjacent correlation increases along the cross-section of the yield curve (i.e., $\rho_{12} < \rho_{23} < \rho_{34} < \rho_{45}$). For example, in Table 4, for the CKLS-4F model, the

---

CKLS specification is based on the possibility of producing one-year maturity rates from other financial products such as Eurodollar futures and one-year swaps.

three-and six-month rates are most highly correlated $(\rho_{34} = 0.8315)$, while for the CKLS-5F model the six- and 12-month LIBOR rates are mostly correlated with $(\rho_{45} = 0.9007)$.

This inverse relationship that we observe between correlation and volatility as maturity increases from one week to one year, may be explained in two ways. First, for longer maturities, the relatively smaller volatilities will induce relatively larger correlations and vice versa (i.e., larger volatilities lead to smaller correlations). Second, LIBOR rates are an important indicator of liquidity patterns inside the money markets (Farooq and Christophersen, 2011). Trade volume for loans or bonds with maturities nearer one year are lower than, for example, one-month loans. Increased economic activity at shorter maturities combined with less transparency given the OTC nature of many money markets products, may result in a lower level of movement synchronization, therefore reducing the correlation between rates of shorter maturities.

Relative to the four-factor specifications, the five-factor models gain more explanatory power and the maximum values for the likelihood functions confirm that. To test the significance of this increase in explanatory power, we employ the likelihood-ratio (LR) test as the four-factor representations are nested in the five-factor models. Under the null hypothesis of the four-factor CKLS model, the test statistics $LR = 2[\log(LF_{5F}) - \log(LF_{4F})]$ are larger than the critical value $\chi^2(16df, 1\%) = 32$. Hence, we formally reject the restrictions in the four-factor specifications.

5.2. Estimation results for the three- and five-factor AFDNS models

In this subsection, we present the estimation results for the two versions of the AFDNS model, AFDNS-3F and AFDNS-5F, respectively. We choose to estimate the AFDNS models with an independent factor structure, as opposed to a correlated factor structure, due to the empirical findings by Christensen et al. (2007) that more parsimonious models (i.e.

independent factors) have superior out-of-sample predictive power over more flexible models that are otherwise better able to fit the models. The estimation results for the three- and five-factor AFDNS models are presented in Table 5. As expected, the in-sample performance improves when additional factors are considered. The likelihood-ratio (LR) test indicates a that the improvement in the model's goodness-of-fit by adding two extra factors is statistically significant, $LR = 2[LogL(5F) - LogL(3F)] = 3{,}183.40$ which exceeds the critical value $\chi^2(7df, 1\%) = 18.48$.

Adding two new factors (a second slope and curvature) may impact the estimation results in two ways. A possible impact is that the explanatory power will spread over more factors, and/or another impact is that more factors will partially account for the unexplained part of fewer factors. While the latter impact is difficult to measure, we observe clear signs of the former impact of additional factors. Indeed, if we focus on the level factor, its contribution/coefficients are lower in the extended five-factor model than in the three-factor model. This shift of explanatory power is clearly observed in the results in Table 5 when we compare the coefficients regarding the level factor. More specifically, the drift coefficients in the SDE dynamics of the level factor have decreased from the mean-reverting level of $k_{11} = 0.0627$ and speed-reversion coefficient $\theta_1 = 0.0893$ in the three-factor model to $k_{11} = 0.0072$ and $\theta_1 = 0.0397$ in the five-factor model, respectively. In addition, the volatility of the level factor is lower in the extended model, which is 0.0025 compared to 0.007 for the three-factor model. Therefore, we observe that all three parameters that characterize the dynamics of the level factor are lower in the three-factor model, implying that it takes longer for the level factor in the AFDNS-5F model to induce the same magnitude parallel shift in the yield curve that it produces within the ADNS3F model. In other words, over a given time interval, the contribution of the level factor to the changes in the yield is less in the AFDNS-5F model compared to the AFDNS-3F model. When we look at the other two types of latent factors, this effect is not easily identifiable as the contribution of the slope and curvature factors involves not only different dynamics but also different factor loading, as the decay parameter changes from $Lambda = 1.7175$ in the AFDNS-3F model to $Lambda1 = 2.7831$ in the AFDNS-5F model.

Following Diebold and Rudebusch (2013), we can give an economic interpretation of the estimation results from Table 5, where the K parameters represent the factor mean-reversion matrix, the Σ parameters represent the factor volatility matrix, and the θ parameters describe the factor means.

[TABLE 5 ABOUT HERE]

Since most studies using AFDNS models use monthly yields, we can convert the continuous-time K matrix into one-month conditional mean-reversion matrix with $\exp(-K^p \frac{1}{12})$ and the volatility matrix into one-month conditional covariance matrix with $Q = \int_0^{\frac{1}{12}} e^{-K^P s} \Sigma \Sigma' e^{-(K^P)' s} ds$.

The elements of the one-month conditional mean-reversion and covariance matrices for the AFDNS models are presented in Table 6. Our estimation results reflect important stylized facts concerning the yield curve such as high persistence in the behaviour of yields and much less persistence in spread dynamics (Diebold and Rudebusch, 2013). Indeed, we can observe that the level factor exhibits highest persistence being associated with an autoregressive coefficient near one. The stylized fact of less persistent spreads dynamics is reflected in less persistent dynamics of the slope factor(s) with higher volatility and lower mean-reversion estimates than the level factor. The curvature factor displays a much higher variation than the other two factors (level and slope) which is in line with other previous studies (e.g., Cochrane and Piazzesi, 2008; Afonso and Martins, 2012). This behaviour of the latent factors can be explained through their association with macro-economic variables. The level factor has been linked to target for inflation, while the slope factor to monetary policy shocks and market structure such as the demand for liquidity. The interaction between central banks and investors who follow announcements of the monetary policy committee, creates more volatility in the slope factor. The interpretation of the curvature factor is not straightforward, with no clear association to a specific macro-economic or policy variable. Nevertheless, Moench (2010) found that this factor has predictive information about both the future evolution of the yield curve and the macro-economy. He argues that unexpected increases of the curvature signal a flattening yield curve and a significant decline of output more than one year ahead. The connection of this third factor with macroeconomic variables leads to a much larger volatility of the curve estimate through time since news regarding macroeconomic variables emerge almost in a continuous manner. In a more recent study, Caruso and Coroneo (2022) consider a real-time joint model where macro-variables are also latent factors and find evidence that revised macroeconomic information has superior power in predicting the level interest rates when compared to real-time macro data.

[TABLE 6 ABOUT HERE]

6. Forecasting analysis

We conduct the forecasting analysis along three directions: across seven forecasting methods, across four horizons (one, three-, six-, and 12 months) and across multiple measures of forecasting accuracy. The more complex CKLS-4F, CKLS-5F, AFDNS-3F and AFDNS-5F models are compared in terms of forecasting power with parsimonious benchmark models such as the random-walk (RW), first-order vector-autoregressive VAR(1), and first-order autoregressive AR(1) models.

Several studies such as Steeley (1990), Christensen et al (2011) and Duffee (2011) presented evidence that a RW process may produce similar dynamics with the observed dynamics of the yield curve. Steeley (2014b) showed that this is only true for some maturities of the spot rates when short-term rates are near zero. Our extensive forecasting analysis brings important evidence in line with these findings. Our choice of considering another two discrete benchmark models, VAR(1) and AR(1), is consistent with the generalized specification of the discrete analogue model implied by Bergstrom's methodology, where for a $k-$ th order linear stochastic differential system the discrete analogue model is a $VARMA(k, k-1)$ model. The CKLS continuous-time models considered for estimation in this study correspond to the particular case of $k=1$, hence their discrete analogues are VAR(1). Moreover, AR(1) and the multivariate version VAR(1) are the natural departure from market efficiency extensively used as parsimonious benchmarks in numerous studies (e.g., Vasicek, 1977; Wu and Zhang, 1996).

6.1. Construction of out-of-sample forecasts

The forecasts for the CKLS models are produced by discretising the differential equation describing the dynamics assumed for the observable rates. Based on the estimated drift parameters, the forecast time $t+h$ given the value of the yield at time t, is given by

$$\hat{r}_{CKLS}(t+h) = \alpha[1+\beta\Delta t + \ldots \beta^{h-1}\Delta t^{h-1}]\Delta t + (\beta\Delta t)^h r(t) \,,$$

where α and β are the vector and matrix drift parameters.

For the AFDNS models, the computation of the forecasts involves the time series of the optimally filtered latent factors:

$$\hat{r}_{AFDNS}(t+h) = E_t\left[X_1(t+h)\right] + E_t\left[X_2(t+h)\right]\left(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau}\right) + E_t\left[X_3(t+h)\right]\left(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau}\right) + E_t\left[X_4(t+h)\right]\left(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau}\right) + E_t\left[X_5(t+h)\right]\left(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau}\right) - \frac{C(\tau)}{\tau}$$

Following Christensen (2007), we employ the the conditional expectations given by:

$$E_t[X(t+h)] = \left(\sum_{i=0}^{h-1} A^i\right)(I-A)\theta + A^h X_t \,,$$

where matrix $A = \exp(-K\Delta t)$ is the estimated one-day conditional mean-reversion matrix.

A robust forecasting comparison is conducted where the daily optimal forecasts are computed recursively for all spot rates under each model and all horizons *h* length such as 1-month, 3-month, 6-month and 1-year. The forecast horizon (*h*) is a vital component in the forecasting analysis, as the conclusions regarding the forecasting accuracy may vary across different horizons and/or different loss functions (Diebold and Lopez, 1996). In this regard, we employ a horizon of *h* equal to 22, 66, 126 and 252 days, corresponding to 1-month, 3-month, 6-month, and 1-year, respectively. Modelling and forecasting interbank rates are of great interest to policy makers and money market participants who prefer to operate within this segment of the financial markets. Close monitoring of the volatility of short-term interest rates is crucial for a well-functioning money market, as any disruption in this segment can have repercussions in the credit supply and banks operations/services to firms, government entities, and households. While model forecasts tend to lose their predictive power as the forecasting horizon is increased, we chose four forecasting horizons based on investors preferences when operating within money markets. Assuming that investors' preferences depend on the structure of assets and liabilities of their portfolios (market segmentation theory), we span our forecasting analysis across horizons characteristic only to money markets, namely one, three, six, and 12 months. Shifting to a longer maturity segment changes the nature of the structure of assets and liabilities, which then must be actively priced over longer periods.

To gauge the forecasting performance of the models, we repeat our forecasting exercise based on a daily-extended estimation window for the next 200 days, for all models and for four horizons. To determine the forecasting accuracy of the models, the forecast errors are aggregated using various statistical and economic forecasting metrics. Over the last two decades, the literature on measures of forecast error still portrays a controversial picture documenting their various limitations (Hyndman and Koehler, 2005). We employ a range of stylized statistical and economic metrics calculated over 200 forecasts at one, three-, six- and 12-month horizons: the mean error (ME), the square root of variance error (STDE), the root mean squared error (RMSE) and the directional forecasting measure – the correct direction change percentage prediction (CDCP). Choosing the right loss function is relative to the particular purpose at hand, as forecasts can be used in various decision environments either by trading desks or by government officials (Diebold, 1993). Despite their sensitivity to the presence of outliers (Armstrong, 2001), the loss functions ME and RMSE are still the most commonly used scale-dependent measures and have been used in forecasting analysis mainly due to their relevance in statistical modelling. However, it is important to consider other types of loss functions such as direction of change measures to provide additional information useful in the decision-making process. It is still possible for forecasts produced by models declared inferior based on statistical measures, to yield a profit when used in market trading strategies (Gerlow, et al., 1993). We employ the financial/economic loss function proposed by Refenes (1995) which is an indicator of a model's ability to correctly predict directional changes. In contrast with most statistical measures of forecasting accuracy, where penalties are applied differently to large errors relative to small errors, the proportion of correct sign prediction measure penalizes both types of errors equally.

6.2. Forecasting analysis based on forecasting accuracy measures

In this section we discuss the forecasting results over the long sample period based on three forecasting accuracy metrics ME, STDE and RMSE (Table 7, panels a, b, c) and one directional forecasting measure CDCP (Table 8). The evaluation of the forecasting metrics is presented for all USD-LIBOR rate time series across four forecasting horizons. The results are rather mixed across maturities and horizons and across the forecasting measures

Based on the ME measure (Table 7 panel a) the continuous-time model CKSL-5F and AFDNS-3F have superior predictive accuracy over the other models for shorter maturities (1-week, 1-, and 3months) and shorter (1- and 3-month) horizons, while the RW is the dominant model for longer (6- and 12-month) maturities and longer (6- and 12-month) horizons. All models seem to under-forecast as most forecasting errors (forecast minus actual) are negative, resulting in all the ME estimates being negative. This can be due to a long-term impact of the GFC on the parameter estimates and therefore on the forecasting ability of the models.[7] According to the results for the STDE measure (Table 7, panel b) the superior model is AR1 with clear dominance for the longer (6- and 12-month) maturities across all horizons; CKLS-5F and AFDNS-3F perform better for shorter maturities. As we move to the RMSE measure (Table 7, panel c), the forecasting results are like those for ME, as the continuous-time models CKLS-5F and AFDNS have higher forecasting accuracy for shorter maturity LIBOR rates, while the parsimonious RW model is superior for the remaining longer maturities. We should interpret these results with caution as we designate as the best model the one with the smallest ME or RMSE value, and very often the second-best model is very close. Nevertheless, there seems to be a separation within the maturity spectrum into two smaller segments: first segment from 1-week to 3-month maturity and the second segment from 3-month to 12-month maturity. It is interesting to note is that the CKLS-4F and AFDNS-5F models do not forecast well at all in comparison with the other models, suggesting that the extension does benefit the predictive performance only for the CKLS modelling framework.

[7] To avoid the long-term impact of the GFC on the forecasting performance of the models, one could employ different techniques based on non-parametric models such as weighted historical simulation where different weights are given to different age data, more weight is given to more recent data, and less weight to more distant data.

Regarding the extensions of the two continuous-time models, the results in Table 7 show that increasing the number of factors has a different impact on the forecasting ability of these models. While adding new factors improves the goodness-of-fit for both models, the accuracy of forecasting the LIBOR rates increases only for the CKLS extended model. This result may be just one particular sample realization, as the empirical results from interest rate modelling are highly sensitive to the data sample (country, period, size, and frequency) and estimation method (Episcopos, 2000; Lo, 2005, Diaz et al., 2016). To check the robustness of our results, we repeat the entire forecasting analysis over a shorter post-GFC period of our initial sample. At the same time, each model is extended in a different way, with the additional factors having a different interpretation. For the CKLS model an extra factor requires more cross-sectional data (i.e., another maturity of LIBOR rates). Adding the 1-year LIBOR rate as the fifth factor better captures the volatility and correlation structure across the LIBOR curve through an extended variance-covariance matrix and this may contribute to a statistically significant improvement in the forecasting performance of the CKLS model. For the AFDNS model, the extension includes two new latent factors (a second slope and a second curvature) which are independent of the data sample. We find that while five factors fit the data better than three factors, it is the three-factor model that has an overall higher forecasting power. This is in line with similar studies (see Gurkaynak et al., 2007) as it is unlikely for a second hump to be observed inside the LIBOR segment of the yield curve. In addition, the curvature factors could be subject to confounding effects (Wahlstrom et al., 2022), making the second curvature factor superfluous.

[TABLE 7 ABOUT HERE]

The results for the economic loss function CDCP are more consistent (Table 8), suggesting that the RW forecasts are the best (16 out of 20 forecasts) in predicting the directional change in the daily USD-LIBOR rates. For 1-, 3- and 6-month forecasting horizons, the RW model outperforms all the other six models in competition across all the maturities. These findings do not necessarily indicate that interbank rates behave like a random walk. Theoretically, interest rates are bounded and hence, they cannot be modelled as a pure random walk. However, there is some empirical evidence (e.g., Ait-Sahalia, 1996) suggesting that the mean reversion characteristic is present only at extreme (low and high) levels of interest rates, while for the middle range values (4%- 17%) the dynamics of interest rates follow a random walk. Additionally, the CDCP forecasting accuracy measure takes into account only the direction of the daily interest rate changes, while it disregards the magnitude of the forecasting errors.

Whether to rely or not on the CDCP metric depends on the purpose at hand. The CDCP measure can be used in improving a decision-making process by reducing costs or increasing utility. For the one-, three- and six-month horizons, the RW forecasts produce the highest CDCP. The highest value corresponds to the 6-month USD LIBOR for the 3-month horizon, where a CDCP of 0.705 indicates that based on the RW forecasts, there is a high chance (70.5%) to correctly predict the change of direction in the movements of the 6-month LIBOR occurring in three months from now. Speculators can make most use of these directional predictions to speculate on the future movements of interest rates. Reliable directional forecasts can be used by financial analysts to design more complex short- and long-term strategies for risk management.

[TABLE 8 ABOUT HERE]

6.3. Forecasting performance tests

Forecasting accuracy measures may not precisely indicate the superiority or not of a particular forecasting model. Sometimes forecasting measures are very close in value, meaning that a model may be considered better just due to sampling noise. Hence, we need to formally test different models regarding their forecasting ability (West, 2006).

We test the statistical significance of the out-of-sample forecasts using more formal procedures such as the Diebold-Mariano (DM) test (Diebold and Mariano, 1995) for any two sets of forecasts and the Clark-West (2007) (CW) test for nested models. The CW test provided an adjustment for the DM test such that the test statistic had an approximately zero mean under the null hypothesis. Using the following two separate sequences of nested models $(RW \subset AR1 \subset VAR1 \subset CKLS4F \subset CKLS5F)$ and $(AFDNS4F \subset AFDNS5F)$, we employ the CW test for 11 pairs of models and present their results in 11 small windows in Table 9. One small panel for a particular pair of two models $M_1 \subset M_2$ presents 20 p-values corresponding to the five maturities of LIBOR rates and all four forecasting horizons. For the first nesting sequence, for one-week, three-months, one-year and occasionally six-month maturities, we reject the null of equal predictive power at 5% level of significance and conclude that the more flexible CKLS models have superior predictive ability over all the other three parsimonious models RW, AR(1) and VAR(1). As the nesting model in the CW test, the CKLS-5F model performs in general better than CKLS-4F model for 9/16 times. Also, the CKLS-5Fperforms better than VAR(1) model for 13/20 times, better than AR(1) for 16/20 times and better than RW model for 12/20 times, Interesting to note, for the one-month USD-LIBOR rate, we consistently fail to reject the null, implying that there is no difference between the CKLS family and the benchmark parsimonious models. However, for the one-month USD-LIBOR rate, we consistently fail to reject the null, implying that there is no difference between the CKLS family and the benchmark parsimonious models. The results of the CW test for the second nesting sequence, $AFDNS-3F \subset AFDNS-5F$, conclude (see the last panel in the Table 9) that there is no significant difference between the forecasting ability of the AFDNS-3F and AFDNS-5F models. Over all twenty outcomes (five maturities and four forecasting horizon), the AFDN-5F is better than the nested AFDNS-3F model only 3/20 times, while for 17/20 times we conclude that the two models have equal forecasting performance.

The DM test is applied under the quadratic error loss, following the approach outlined in Diebold (2015), where the forecasts produced by the various models are compared instead of the models themselves. The results for the DM test are reported in Table 10. We assess the forecasts across 10 model-pairs M1 versus M2, where M1 is one of the following models RW, AR(1), VAR(1), CKLS 4F or CKLS-5F), and M2 can be AFDNS-3F or AFDNS-5F. The testing has three possible outcomes: for p-values higher than 0.05 we fail to reject the null and conclude that M1 = M2 (i.e., the two models have equal predictive power); for p-values under 0.05, the decision is based on a comparison of the accuracy measure ME of each model. In Table 10, we indicate by a plus sign ("+") when M1 is superior to M2 and by a minus sign ("−") when M1 is inferior to M2. Using a 5% level of significance, we found that the M2 (AFDNS) specifications are at least as good as or better than the other models. The comparison between the RW and the AFDNS models, shows that there are six minus signs between RW and ANDNS-3F and three between RW and ANDNS-5F, while the remaining signs are all equal, with no positive signs at all. Therefore, the AFDNS specifications are better or equal to the RW model, with the simplerAFDNS-3F model being relatively more superior to the RW than the more general AFDNS-5F model. Only the CKLS-5F and CKLS-4F models seem to occasionally outperform the AFDNS models. Moreover, we observe that for one-month USD-LIBOR rates, the AFDNS family achieves the best prediction. The overall conclusion from Table 10 is that both AFDNS-5F and AFDNS-3F models are either equal or superior to the benchmark models across all maturities and forecasting horizons. When we compare directly the AFDNS models with the CKLS models, the AFDNS models seem overall superior to the CKLS specifications. If we analyze by horizon, for the one-month horizon only one sign (1/18) is positive implying that the CKLS-5F is superior to the AFDNS-5F for the one-year LIBOR rate; the remaining 17/18 signs are distributed as 11 negative and 6 equal signs, meaning that AFDNS models are either better or equal in forecasting performance. For the three-month horizon, there are 1/18 positive sign, 4/18 negative and 13/18 equal signs; for the six-month horizon, there are 2/18 positive signs, 6/18 negative and 10/18 equal signs; and finally for one-year horizon, there is no statistical difference in the forecasting performance between the two modelling frameworks as there are 18/18 equal signs.

[TABLE 9 ABOUT HERE]

[TABLE 10 ABOUT HERE]

We also considered tests for superior predictive ability (SPA), where an alternative forecast is tested against a given forecast. This procedure employs Hansen's (2005) test as an improvement of the reality check test proposed by White (2000).[8] The null hypothesis asserts that the average performance of the benchmark is as small as the minimum average performance across the rest of the models, while the alternative is that the minimum average loss across the models is smaller than the average performance of the benchmark. Any p-value above 0.05 implies that the reference model is superior to the other alternatives, while the model with the most p-values under 0.05 is consider inferior. Across all the 20 p-values available for each model, the RW is the most underperforming with only 6/20 p-values above 0.05, while both AFDNS-5F and AFDNS-3F do very well for all maturities LIBOR rates and across all horizons, being superior to the set of remaining models with 20/20 p-values above 5%. At the same time, we observe that the CKLS-5F model forecasts well against the set of all the other alternative models only for shorter maturities, namely the 1W-LIBOR and 3M-LIBOR. Hence, based on the SPA test, we can conclude that the AFDNS framework is overall superior to the other models we have considered in this study.

[TABLE 11 ABOUT HERE]

---

[8] Because of space constraints, we only present the results for the SPA test while the results for the White (2000) reality check are available from the authors upon request.

Finally, we use the model confidence set (MCS) test of Hansen et al. (2011) for the simultaneous comparison of multiple forecasts and the identification of the best model set. The results presented in Table 12 provide additional support for the superiority of the CKLS5F, AFDNS5F and AFDNS3F models over traditional forecasting models that are hard to surpass in general. The value "1" indicates that the respective model belongs to the best set of models forecasting a particular LIBOR for a given forecasting horizon. Thus, by counting the number of 1s for a model, we can see how well the model predicts LIBOR rates in general. We can analyse the results in Table 12, across the maturities of the LIBOR rates or across the forecasting horizons. Across maturities, for the one-week LIBOR rate, the confidence set includes the CKLS-5F solely for longer horizons (3, 6 and 12 month), while for the 1M horizon the confidence set can be extended to three models, namely CKLS-5F, AFDNS- 3F and VAR(1). Moving to the one-month LIBOR rate, the confidence set is the same across all horizons containing four models, namely AFDNS-5F, AFDNS-3F, AR(1), and VAR(1). These results are very similar to those from the CW test for one-month maturity where we could not single out a best predictive model and VAR1 specification is equal to the CKLS family. For the three-month LIBOR rate, the confidence is the same across all horizons and it includes only two models, CKLS-5F and AFDNS-3F. For the six-month LIBOR rate, the confidence set contains the AFDNS-3F, VAR(1), and WR(1) across all horizons, with the addition of the CKLS-5F model for 3M and 12M horizon. Finally, for the 12-month LIBOR rate, the two main models in the confidence set are AFDNS-5F and AFDNS-3 across all the horizons, with the addition to another three models for the 1Y horizon. Based on a general score, the model with the highest frequency of being included in the confidence set is the AFDNS-3F model with17/20 entries, followed by the CKLS-5F and VAR(1) models with 10/20 entries. Consistent with the findings from the previous formal tests, we conclude that the more flexible continuous-time models (AFDNS and CKLS) outperform overall the discrete-time parsimonious models.

[TABLE 12 ABOUT HERE]

6.4. Post-crisis Forecasting Analysis

To check the robustness of our comparative results even further, we repeated the entire forecasting analysis for the USD-LIBOR rates based on the post-crisis estimation period from 17 October 2008 to 29 September 2017. The subprime mortgage crisis had a significant impact on the general levels of interest rates as can be seen from Figure 1 and the summary statistics in Table 2. The objective is to investigate whether the new post-crisis parameter estimates will affect the overall results in our comparative forecasting analysis. All the empirical results for the post-crisis forecasting analysis are presented in the provided Online Appendix. Based on the statistical forecasting accuracy measures, we obtained similar results to those from the analysis of the entire sample, with the superiority of the CKLS and the AFDNS families slightly enhanced. More specifically, the performance in terms of the ME and RMSE measures of the CKLS family improved from the all-period analysis, especially for three- and twelve-month maturity rates. At the same time, the RW benchmark is losing terrain in front of the continuous-time models in the post-crisis analysis. The results for the CDCP measure remain unchanged, with the RW being the clear winning method (i.e. highest percentage) in correctly predicting the change in the direction of the level of interbank rates level. When the formal tests for forecasting comparison are employed, the general "forecasting performance" picture is similar to the one obtained using the entire data sample. That is, the CKLS and the AFDNS models remain "better" forecasting models with the only difference being that models from the CKLS family seem to forecast even better relative to AFDNS models.

Our findings show that the effect of the GFC of 2007-2009 fails to materialize in the comparative forecasting performance of the models studied, despite its clear effect on the estimation results. The confirmation of the same hierarchy among the models in terms of the performance of their forecasts over two different data samples only reinforces that the comparative results are reliable and hence the insights obtained from this extensive forecasting analysis may be inferred for other countries interbank curves over different periods.

7. Conclusions

In this empirical study, we investigate to what extent enhancing model flexibility by increasing the number of factors influences model forecasting ability. The superiority of multi-factor term structure models relative to single factor formulations is well-established in the literature and events such as the GFC of 2007-2009 highlighted the necessity of richer models by considering more risk factors. We also investigate another modelling question from a forecasting perspective, assessing the empirical trade-off between a parsimonious and more complex models.

We conduct an extensive forecasting comparison among various nested versions of two continuous-time models, namely CKLS and DNS models for the US inter-bank market. The comparison is made across many criteria: among seven models, across five maturity LIBOR rates, and across four forecasting horizons. The results based on the static forecasting accuracy and directional accuracy metrics are inconclusive, indicating a separation of the up-to-one year maturity spectrum into two smaller segments. Once we make that segmentation, a clearer general picture emerges. More specifically, for the one-week to three-month maturity segment the continuous-time models outperform the parsimonious models, while for longer (six to twelve months) maturity segment, the simpler models such as RW and AR(1) have higher predictive power.

It is well known that the results from the analysis of the static metrics for forecasting accuracy (here ME, STDE, RMSE, and CDPC) are difficult to interpret due to the particular statistical properties of these measures (Davydenko and Fildes, 2016). As a result, the forecasting literature has not yet found a consensus regarding the selection of the most suitable measure. To test the statistical significance of the forecasting accuracy across the models, we have employed more advanced and well-established techniques such as the Diebold-Mariano (1995), Clark-West (2007), the superior predictive ability test of Hansen (2005), and the model confidence set of Hansen, Lunde, and Nason (2011). Our empirical results from these four formal statistical tests significantly contribute to a robust comparative picture, confirming previous mixed evidence of predictive superiority between different models across the maturity spectrum of interbank yield curves (Steeley, 2014b). In addition, we found evidence that an improvement in the accuracy of forecasting LIBOR rates only for the CKLS extended framework. In the case of the AFDNS family, the three-factor specification has superior out-of-sample performance when compared the extended five-factor model, a finding that is in line with a recent study (Wahlstrom et al., 2022) which recommends that central banks should use less flexible yield curve models when formulating their monetary policy. The estimation results for the interbank yield curve reveal significantly higher decay parameters than those proposed in the literature, where the estimation window covers the whole maturity spectrum. This may be explained by the unique dynamics of money markets and how their specific instruments operate. This finding opens a question for future investigation of comparing the forecasts of interest rates from models supported by restricted versus larger maturity range.

We assess the robustness of our comparative results by repeating the entire forecasting analysis for post-crisis period 2008 -2017. Our findings show that the effect of the sub-prime crisis fails to materialize in the comparative forecasting performance of the studied models, despite its clear impact of changing the level of interest rates. The patterns observed in the model selection with best predictive power are similar with those inferred from the comparison over the entire period 1998 -2017. The only substantial change observed was that based on the post-crisis sample, the CKLS multi-factor structure (especially the five-factor model) gains more forecasting ability over the AFDNS models.

Our comprehensive and robust forecasting analysis highlights the complexity reported in the term-structure literature regarding the modelling and forecasting of interest rates. Overall, we conclude that model flexibility is still an important attribute for improving forecasting performance; interest rates of different maturities may follow particular dynamics and hence the best model choice for that particular maturity segment of the yield curve could be different. As a result, financial analysts should employ extensively formal testing given the drawbacks of various forecasting accuracy measures.

REFERENCES

Adrian, T., Crump, R.K., Moench, E. (2013) Pricing the term structure with linear regressions. Journal of Financial Economics 110, 110–138.

Afonso, A., Martins, M.M.F. (2012) Level, slope, curvature of the sovereign yield curve, and fiscal behaviour. Journal of Banking and Finance 36, 1789–1807.

Ait-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. The Review of Financial Studies 9, 385-432.

Ajello, A., Benzoni, L., Chyruk, O. (2020) Core and ‘Crust’: Consumer Prices and the Term Structure of Interest Rates. The Review of Financial Studies 33, 3719–3765.

Almeida, C., Vicente J. (2008) The role of no-arbitrage on forecasting: Lessons from a parametric term structure model. Journal of Banking and Finance 32, 2695–1705.

Ang, A., Piazzesi, M. (2003) A no arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–87.

Balduzzi, P., Das, S. R., Foresi, S., Sundaram, R. (1996) A simple approach to three factor affine term structure models. Journal of Fixed Income 6(3), 43-53.

Bali, T., Wu, L. (2006) A comprehensive analysis of the short-term interest-rate dynamics. Journal of Banking and Finance 30, 1269-1290.

Bergstrom, A. R. (1984) Continuous time stochastic models and issues of aggregation over time. In Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, Vol. 1. North-Holland: Amsterdam.

Bergstrom, A. R. (1989) Optimal forecasting of discrete stock and flow data generated by a higher order continuous time system. Computers and Mathematics with Applications 17, 1203-1214.

Bjork, T. Christensen, B.J. (1999) Interest rate dynamics and consistent forward rate curves. Mathematical Finance 9, 323 – 348.

Brennan, M. J. Schwartz, E. S. (1980) Analyzing convertible bonds. Journal of Financial and Quantitative Analysis 15, 907-929.

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

Campbell, J., R. Shiller. (1991) Yield Spreads and Interest Rate Movements: A Bird's Eye View. Review of Economics Studies 58 (3), 495-514.

Carriero, A. (2011) Forecasting the yield curve using priors from no-arbitrage affine term structure models. International Economic Review 52**,** 425–459.

Carriero. A., Giacomini R. (2011) How useful are no-arbitrage restrictions for forecasting the yield curve? Journal of Econometrics 164, 21–34.

Caruso, A., Coroneo, L. (2022) Does real-time macroeconomic information help to predict interest rates? Journal of Money Credit and Banking. Forthcoming.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., Sanders, A. B. (1992) An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209–1227.

Chen, L. (1996) Stochastic mean and stochastic volatility – A three-factor model of the term structure of interest rates and its application to the pricing of interest rate derivatives. Financial Markets, Institutions and Instruments 5, 1–88.

Christensen, J.H.E., Diebold, F.X., Rudebusch, G.D. (2007) The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models. NBER Working Paper 13611

Christensen, J.H.E., Diebold, F.X., Rudebusch, G.D. (2009) An arbitrage-free generalized Nelson-Siegel term structure model. Econometrics Journal 12, C33-C64.

Christensen, J. H. E., Diebold, F. X., Rudebusch, G. D. (2011) The affine arbitrage-free class of Nelson–Siegel term structure models. Journal of Econometrics 164, 4–20.

Clark, T.E.West, K.D. (2007) Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291-311.

Cochrane, J., Piazzesi M. (2005) Bond risk premia. American Economic Review 95**,** 138–160.

Cox, J. C., Ingersoll, J. E., Ross, S. A. (1985) A theory of the term structure of interest rates. Econometrica 53, 385-408.

Coroneo, L., Nyholm, K. and Vidova-Koleva, R. (2011) How arbitrage-free is the Nelson–Siegel model? Journal of Empirical Finance 18(3), 393–407.

Dai, Q., Singleton, K. J. (2000) Specification analysis of affine term structure models. Journal of Finance 55, 1943-78.

Davydenko, A., Fildes, R. (2016) Forecast Error Measures: Critical Review and Practical Recommendations. In Business Forecasting: Practical Problems and Solutions. John Wiley & Sons Inc.

Diaz, D., Theodoulidis, B., Dupouy, C. (2016) Modelling and forecasting interest rates during stages of the economic cycle: A knowledge-discovery approach. Expert Systems with Applications 44, 245–264.

Diebold, F. X. (1993) On the limitations of comparing mean square forecasts errors: Comment. Journal of Forecasting 12, 641-642.

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

Diebold, F. X, Lopez, J. A. (1996) Forecast evaluation and combination. Handbook of Statistics 14, 241-268.

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

Diebold, F. X., Rudebusch, G. D. (2013) Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. Princeton University Press, Princeton, N.J.

Doshi, H., Jacobs, K., Liu, R. (2021) Information in the term structure: A forecasting perspective. Management Science 67(8), 5255-5277.

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

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

Duffie, D., Kan, R. (1996) A yield-factor model of interest rates. Mathematical Finance 6, 379–406.

Episcopos, A. (2000) Further Evidence on Alternative Continuous Time Models of the ShortTerm Interest Rate. Journal of International Financial Markets, Institutions, and Money 10, 199-212.

Farooq, A.Q., Christophersen, C. (2010) Interbank overnight interest rates - Gains from systemic importance. Working Paper, No. 2010/11, ISBN978-82-7553-562-5, Norges Bank, Oslo.

Filipovic, D., Larsson, M., Trolle, A. (2014) Linear-rational term structure model. Journal of Finance 72, 655–704.

Gerlow, M.E., Irwin S. H., Liu T. (1993) Economic evaluation of commodity price forecasting models. International Journal of Forecasting 9, 387 – 397.

Gimeno, R., Marques, J. M. (2009) Extraction of financial market expectations about inflation and interest rates from a liquid market. Banco de Espana, Working Paper 0906.

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

Hansen, P.R., Lunde, A., Nason, J.M. (2011) The model confidence set. Econometrica 79, 453-497.

Hansen, P.R. (2005) A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365-380.

Hyndman, R.J., Koehler, A.B. (2005) Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679-688.

Jarrow, R. A. (2013) The zero-lower bound on interest rates: Myth or reality? Finance Research Letters 10, 151–156.

Joslin, S., Le A., Singleton, K. (2013) Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs. Journal of Financial Economics 109**,** 604–622.

Joslin, S., Singleton K., Zhou H. (2011). A new perspective on Gaussian dynamic term structure models. Review of Financial Studies 24, 926–970.

Litterman, R., Scheinkman, J. (1991) Common factors affecting bond returns. Journal of Fixed Income 1, 54-61.

Lo, K. M. (2005) An Evaluation of MLE in a Model of the Nonlinear Continuous-Time Short-Term Interest Rate. Working Paper Number 45, Bank of Canada.

Matsumura, M., Moreira, A. Vicente, J. (2011) Forecasting the yield curve with linear factor models. International Review of Financial Analysis 20, 237–243.

Monech, 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. (2010) Term structure surprises: The predictive content of curvature, level, and slope. Journal of applied econometrics, John Wiley & Sons, Ltd., 27(4), 574-602.

Nelson, C. R., Siegel, A. F. (1987) Parsimonious modelling of yield curve. Journal of Business, 473 -89.

Nowman, K. B. (1997) Gaussian estimation of single-factor continuous time models of the term structure of interest rates. Journal of Finance 52, 1695-1706.

Nowman, K. B. (2006) Japanese Fixed Income Markets: Money, Bond and Interest Rate Derivatives, Elsevier, Amsterdam.

Patton, A.J. (2011) Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246-256.

Piazzesi, M., Cochrane, J. (2009) Decomposing the Yield Curve, Meeting Papers 18, Society for Economic Dynamics.

Refenes, A. P. (1995) Neural networks within the capital markets. In: Refenes A (ed) Wiley, Chichester.

Sharef, E., Filipovic, D. (2004) Conditions for consistent exponential-polynomial forward rate processes with multiple nontrivial factors. International Journal of Theoretical and Applied Finance 7, 685-700.

Steeley, J. (2014a) Yield curve dimensionality when short rates are near the zero lower bound. In Chadha, J.S., Duree, A., Joyce, M.S., Sarno, L. (Eds.), Development in Macro-Finance Yield Curve Modelling. Cambridge: Cambridge University Press.

Steeley, J. (2014b) Forecasting the term structure when short-term rates are near zero. Journal of Forecasting 33, 350-363.

Steeley, J. (2014c) A shape-based decomposition of the yield adjustment term in the arbitrage-free Nelson and Siegel (AFDNS) model of the yield curve. Applied Financial Economics 24, 661-669.

Steeley, J. (1990) Modelling the dynamics of the term structure of interest rates. Economic and Social Review - Symposium on Finance 21, 337-362.

Svensson, L. E. O. (1995) Estimating forward interest rates with the extended Nelson-Siegel method. Sveriges Riskbank Quarterly Review 3, 13-26.

van Deventer, Donald R., Imai, K., Mesler, M. (2013) Advanced Financial Risk Management: Tools and Techniques for Integrated Credit Risk and Interest Rate Risk Management, John Wiley & Sons, Singapore.

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

Wahlstrom, R. R. Paraschiv, F., Shurle, M. (2022) Comparative Analysis of Parsimonious Yield Curve Models with Focus on the Nelson-Siegel, Svensson and Bliss Versions. Computational Economics 59, 967 – 1004.

West, K.D. (2006) Forecast evaluation. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Elsevier.

White, H. (2000) A reality check for data snooping. Econometrica 68, 1097-1196.

Wu, Y., Zhang, H. (1996) Mean reversion in interest rates: New evidence from a panel of OECD countries. Journal of Money, Credit and Banking 28, 604-621.

Figure 1: Time series of USD-LIBOR rates covering the period January 1998 to July 2019 for 1-week, 3-month and 1-year maturities.

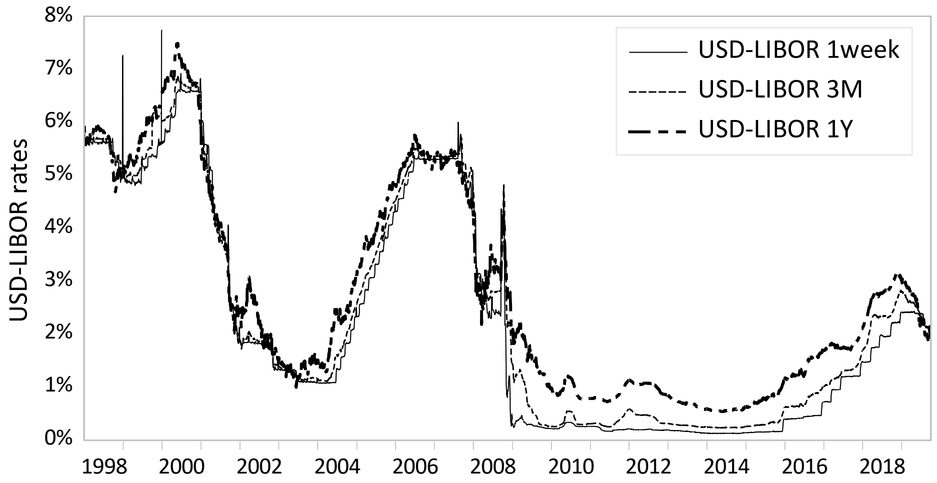

Table 1: Summary Statistics for USD-LIBOR rates (1998 -2019)

Note: This table contains the summary statistics for daily observations of USD-LIBOR rates (in percentage) across five maturities (1week, 1-, 3-, 6- and 12 months) obtained from Bloomberg. These statistics are computed over the whole sample period January 2, 1998 to Julyn16, 2019 for USD-LIBOR AC(x) is the autocorrelation coefficient at lag x, ADF is the augmented Dickey–Fuller test for a unit root, and ADFp is the associated probability value. These tests include a constant term but no time trend and select the lag length using SIC criterion.

Maturity	Obs.	Mean	Std. Dev.	Max.	Min.	AC(1)	AC(12)	AC(24)	ADF	ADFp
USD1W	5,441	2.208	2.122	7.756	0.116	0.999	0.992	0.984	-1.667	0.448
USD1M	5,441	2.255	2.12	6.821	0.148	1	0.993	0.985	-1.605	0.48
USD3M	5,441	2.378	2.094	6.869	0.223	1	0.994	0.987	-1.665	0.449
USD6M	5,441	2.515	2.037	7.109	0.319	1	0.995	0.988	-1.562	0.502
USD12M	5,441	2.738	1.946	7.501	0.534	1	0.994	0.987	-1.449	0.559

Table 2: Summary statistics for USD-LIBOR rates for the pre-crisis and post-crisis periods

Note: This table presents the descriptive statistics for two sub-samples: the pre-crisis period (January 2, 1998 – October 16, 2008) and post-crisis period (October 17, 2008 –September 29, 2017).

USD-LIBOR			**Pre-Crisis**		
Maturity	Obs.	Mean	Std. Dev.	Max.	Min.
USD1W	2,728	3.824	1.810	7.756	1.025
USD1M	2,728	3.855	1.816	6.821	1.020
USD3M	2,728	3.928	1.822	6.869	1.000
USD6M	2,728	3.998	1.806	7.109	0.980
USD12M	2,728	4.135	1.755	7.501	0.990
			Post-Crisis		
Maturity	Obs.	Mean	Std. Dev.	Max.	Min.
USD1W	2,262	0.309	0.286	3.156	0.116
USD1M	2,262	0.364	0.356	4.181	0.148
USD3M	2,262	0.530	0.437	4.419	0.223
USD6M	2,262	0.750	0.494	4.130	0.319
USD12M	2,262	1.075	0.489	3.973	0.534

Table 3. Estimation results for the extended CKLS models: Drift parameters

Note: The maximum likelihood estimation was performed using daily data for five interbank rates between January 2, 1998 and September 29, 2017.

Five-factor			Four-Factor		
Parameter	USD	S. E	Paramete r	USD	S.E.
alpha1	8.50E-06	0.00E+00	alpha1	1.17E-05	0.00E+00
alpha2	2.72E-05	0.00E+00	alpha2	2.46E-05	0.00E+00
alpha3	1.23E-05	5.90E-08	alpha3	2.02E-05	0.00E+00
alpha4	1.57E-05	1.24E-07	alpha4	2.39E-05	0.00E+00
alpha5	4.14E-05	2.47E-06	beta11	4.31E-02	2.06E-03
beta11	4.29E-02	2.56E-03	beta12	-5.19E-02	2.47E-03
beta12	-5.30E-02	3.22E-03	beta13	7.33E-03	1.28E-03
beta13	1.02E-02	2.07E-03	beta14	-4.33E-04	6.26E-04
beta14	-3.61E-03	1.76E-03	beta21	9.38E-02	2.35E-03
beta15	1.53E-03	5.88E-04	beta22	-1.05E-01	2.75E-03
beta21	9.31E-02	3.91E-03	beta23	1.15E-02	1.20E-03
beta22	-1.04E-01	4.73E-03	beta24	-1.32E-03	5.76E-04
beta23	1.01E-02	1.76E-03	beta31	4.74E-02	1.40E-03
beta24	1.50E-03	1.39E-03	beta32	-5.10E-02	1.28E-03
beta25	-1.58E-03	4.61E-04	beta33	7.10E-03	1.04E-03
beta31	4.74E-02	2.75E-03	beta34	-4.26E-03	5.46E-04
beta32	-5.56E-02	3.28E-03	beta41	4.53E-02	1.87E-03
beta33	1.71E-02	1.33E-03	beta42	-5.63E-02	1.99E-03
beta34	-1.35E-02	1.22E-03	beta43	2.16E-02	2.01E-03
beta35	3.80E-03	4.16E-04	beta44	-1.16E-02	9.56E-04
beta41	4.45E-02	2.20E-03			
beta42	-6.19E-02	1.98E-03			
beta43	3.62E-02	3.24E-03			
beta44	-2.39E-02	3.69E-03			
beta45	4.30E-03	1.22E-03			
beta51	4.17E-02	4.09E-03			
beta52	-5.72E-02	5.73E-03			
beta53	2.71E-02	4.28E-03			
beta54	-5.40E-03	3.82E-03			
beta55	-7.62E-03	1.82E-03			

Table 4. Estimation results for the extended CKLS models: Diffusion parameters

Note: This table reports the estimated diffusion parameters for the four- and five-factor CKLS models over the estimation period January 2, 1998 to September 29, 2017. The five factors considered are one week, one-, three- and six- and twelve-month USD-LIBOR rates. The estimated standard errors of the parameter estimates are included in the adjacent S.E. column.

CKLS	Five-factor		**CKLS**	Four-factor	
Parameter	USDS	S.E.	Parameter	USD	S.E.
Gamma1	1.0872	0.0017	Gamma1	1.0876	0.0016
Gamma2	0.9028	0.0041	Gamma2	0.9089	0.0029
Gamma3	0.8619	0.0022	Gamma3	0.8876	0.0012
Gamma4	0.8243	0.0022	Gamma4	0.9425	0.0026
Gamma5	0.793	0.005			
Sigma1	0.0319	0.0007	Sigma1	0.0319	0.0006
Sigma2	0.0092	0.0004	Sigma2	0.0095	0.0003
Sigma3	0.0056	0.0002	Sigma3	0.0063	0.0001
Sigma4	0.0055	0.0001	Sigma4	0.0088	0.0002
Sigma5	0.0066	0.0003			
Corr12	0.5211	0.0075	Corr12	0.5194	0.0073
Corr13	0.4097	0.0085	Corr13	0.4081	0.0084
Corr14	0.2924	0.0096	Corr14	0.2906	0.0094
Corr15	0.1742	0.01	Corr23	0.7381	0.0047
Corr23	0.7412	0.0054	Corr24	0.5982	0.0066
Corr24	0.6039	0.0071	Corr34	0.8315	0.0073
Corr25	0.4191	0.0092			
Corr34	0.8349	0.0032			
Corr35	0.6429	0.0087			
Corr45	0.9007	0.0027			
LL Function		424,084.90			338,389.31

Table 5. Estimation results for the three- and five-factor AFDNS models.

Note: This table reports the estimated K matrix and θ vector along with the estimated parameters of the Q matrix in the three- and five factor AFDNS models with independent factors for the sample period from January 2, 1998 to September 29, 2017. The estimated standard errors of the parameter estimates are included in the next S.E. column.

AFDNS	Five-factor		**AFDNS**	Three-factor	
Parameter	Estimate	S.E.	Parameters	Estimate	S.E.
Kappa11	0.0072	0.0003	Kappa11	0.0627	0.0001
Kappa22	1.7955	0.0007	Kappa22	0.4402	0.0000
Kappa33	0.1672	0.0007	Kappa33	1.0219	0.0000
Kappa44	0.3595	0.002	Sigma11	0.007	0.0292
Kappa55	0.3189	0.0005	Sigma22	0.0069	0.0268
Sigma11	0.0025	0.0142	Sigma33	0.0196	0.0168
Sigma22	0.0077	0.0026	Theta1	0.0893	0.0002
Sigma33	0.0086	0.0094	Theta2	0.0104	0.004
Sigma44	0.024	0.0023	Theta3	0.037	0.0024
Sigma55	0.0166	0.0013	Lambda	1.7175	0.0005
Theta1	0.0397	0.001			
Theta2	0.0046	0.0268			
Theta3	-0.0284	0.001			
Theta4	-0.0052	0.0002			
Theta5	-0.0072	0.0007			
Lambda1	2.7831	0.0012			
Lambda2	0.1662	0.0008			
LL Function		156,454.90			154,863.20

Table 6. One-month conditional mean-reversion parameters and volatilities for AFDNS-3F and AFDNS-5F Models.

Note: This table reports the off-diagonal elements of the conditional mean-reversion and volatility matrices corresponding to the level, slope, and curvature factors.

Model	AFDNS-3F		AFDNS-5F	
Factors	Mean-Reversion	Volatility	Mean-Reversion	Volatility
Level	0.995	4.06E-06	0.999	5.21E-07
Slope(s)	0.964	3.83E-06	0.861	4.27E-06
			0.986	6.08E-06
Curvature(s)	0.918	2.94E-05	0.970	4.66E-05
			0.974	2.24E-05

Table 7. Forecasting Accuracy Measures across all models

Panel a: The results for the ME forecasting accuracy measure computed at 1-, 3-, 6- and 12-month horizons for all models and all LIBOR rates.

Note: The results in bold font indicate the models of best forecasting accuracy. According to this measure, the RW model has an overall best performance (8/20), especially for longer maturities and longer horizons, while the continuous-time models CKLS-5F and AFDNS-3F share the best position for shorter maturities.

1W LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
ME 1M	-4.05E-04	-3.25E-04	**-2.27E-04**	-1.35E-03	-8.45E-04	-2.28E-04	-8.23E-04
ME 3M	-1.33E-03	**-1.10E-03**	-1.47E-03	-2.46E-03	-2.39E-03	-1.54E-03	-2.58E-03
ME 6M	-2.73E-03	**-2.26E-03**	-3.36E-03	-4.23E-03	-4.38E-03	-3.74E-03	-5.20E-03
ME-12M	-3.98E-03	**-3.06E-03**	-5.69E-03	-6.50E-03	-5.40E-03	-6.96E-03	-8.82E-03
1M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
ME 1M	-1.23E-03	-1.36E-03	**-6.68E-04**	-7.05E-04	-8.57E-04	-8.92E-04	-9.81E-04
ME 3M	-3.69E-03	-4.08E-03	**-2.04E-03**	-2.15E-03	-2.45E-03	-2.44E-03	-2.96E-03
ME 6M	-7.16E-03	-7.93E-03	**-3.90E-03**	-4.12E-03	-4.22E-03	-4.60E-03	-5.69E-03
ME-12M	-1.21E-02	-1.36E-02	-5.77E-03	-6.18E-03	**-4.68E-03**	-7.27E-03	-9.16E-03
3M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
ME 1M	-9.02E-04	**-8.06E-04**	-1.37E-03	-9.55E-04	-1.01E-03	-1.43E-03	-1.14E-03
ME 3M	-2.54E-03	**-2.25E-03**	-2.95E-03	-2.67E-03	-2.85E-03	-3.44E-03	-3.25E-03
ME 6M	-4.72E-03	-4.14E-03	-5.03E-03	-4.93E-03	**-4.03E-03**	-5.90E-03	-6.12E-03
ME-12M	-5.68E-03	-4.51E-03	-5.85E-03	-5.99E-03	**-3.45E-03**	-7.50E-03	-8.43E-03
6M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
ME 1M	-1.01E-03	-1.11E-03	-1.00E-03	**-9.67E-04**	-1.09E-03	-3.22E-03	-3.68E-03
ME 3M	-2.93E-03	-3.22E-03	**-2.76E-03**	-2.79E-03	-2.82E-03	-9.90E-03	-9.54E-03
ME 6M	-3.19E-02	-5.94E-03	-5.02E-03	-5.08E-03	**-4.23E-03**	-1.41E-02	-1.55E-02
ME-12M	-5.92E-03	-7.10E-03	-5.29E-03	-5.38E-03	**-3.17E-03**	-1.73E-02	-1.90E-02
12M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
ME 1M	-	**-9.79E-04**	-1.03E-03	-1.02E-03	-1.06E-03	-2.87E-03	-3.21E-03
ME 3M	-	-2.88E-03	-2.88E-03	-3.00E-03	**-2.66E-03**	-8.76E-03	-8.59E-03
ME 6M	-	-5.02E-03	-5.03E-03	-5.25E-03	**-3.83E-03**	-1.27E-02	-1.42E-02
ME 12M	-	-4.19E-03	-4.24E-03	-4.60E-03	**-1.53E-03**	-1.48E-02	-1.68E-02

Panel b. The results for the STDE forecasting accuracy measure computed at 1-, 3-, 6- and 12-month horizons for all models and all LIBOR rates.

Note: The results in bold font indicate the model of best forecasting accuracy. According to this measure, the AR1 model has an overall best performance (11/20), especially for longer maturities across all horizons, while the continuous-time models CKLS-5F and AFDNS-3F share the best position for shorter maturities.

The results are not clear-cut. However, we observe that the STDE measure produces

1W LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
STDE 1M	1.09E-03	1.09E-03	**8.55E-04**	1.47E-03	1.09E-03	8.66E-04	1.07E-03
STDE 3M	8.61E-04	8.81E-04	8.61E-04	1.25E-03	**6.98E-04**	9.93E-04	7.11E-04
STDE 6M	1.01E-03	1.08E-03	1.01E-03	1.02E-03	1.01E-03	1.37E-03	**5.55E-04**
STDE 12M	2.74E-03	2.86E-03	2.32E-03	1.82E-03	4.26E-03	2.62E-03	**1.64E-03**
1M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
STDE 1M	7.42E-04	7.39E-04	7.61E-04	**7.19E-04**	7.49E-04	7.85E-04	7.55E-04
STDE 3M	7.74E-04	**7.33E-04**	9.24E-04	8.35E-04	8.18E-04	1.04E-03	8.87E-04
STDE 6M	8.39E-04	**7.31E-04**	1.26E-03	1.08E-03	1.25E-03	1.58E-03	1.11E-03
STDE 12M	1.89E-03	**1.69E-03**	2.69E-03	2.41E-03	4.49E-03	2.94E-03	2.43E-03
3M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
STDE 1M	1.05E-03	1.06E-03	**9.47E-04**	9.74E-04	1.05E-03	9.69E-04	1.04E-03
STDE 3M	2.46E-03	2.50E-03	**1.97E-03**	2.16E-03	2.24E-03	1.99E-03	2.37E-03
STDE 6M	3.01E-03	3.10E-03	**2.28E-03**	2.30E-03	2.38E-03	2.44E-03	2.82E-03
STDE 12M	4.72E-03	4.91E-03	3.91E-03	**3.84E-03**	5.58E-03	4.07E-03	4.31E-03
6M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
STDE 1M	8.71E-04	8.63E-04	9.01E-04	8.66E-04	8.58E-04	1.01E-03	**8.05E-04**
STDE 3M	2.21E-03	2.16E-03	2.16E-03	2.18E-03	2.06E-03	1.45E-03	**1.24E-03**
STDE 6M	2.69E-03	2.53E-03	2.53E-03	2.47E-03	3.08E-03	1.16E-03	**9.40E-04**
STDE 12M	5.27E-03	5.08E-03	5.13E-03	5.07E-03	6.74E-03	2.71E-03	**2.24E-03**
12M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
STDE 1M	-	7.57E-04	7.76E-04	7.47E-04	7.23E-04	8.57E-04	**6.87E-04**
STDE 3M	-	1.94E-03	1.87E-03	1.87E-03	1.93E-03	1.23E-03	**1.06E-03**
STDE 6M	-	2.49E-03	2.36E-03	2.21E-03	3.47E-03	9.47E-04	**6.34E-04**
STDE 12M	-	6.20E-03	5.98E-03	5.78E-03	7.89E-03	3.81E-03	**3.24E-03**

Panel c. The results for the RMSE forecasting accuracy measure computed at 1-, 3-, 6- and 12-month horizons for all models and all LIBOR rates.

Note: The results in bold font indicate the model of best forecasting accuracy. According to this measure, the overall picture is that for longer maturities and horizons the RW1 has superior prediction power, while the continuous time models CKLS-5F and AFDNS-3F share the best position for shorter maturities.

1W LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
RMSE 1M	1.16E-03	1.13E-03	**8.83E-04**	1.99E-03	1.38E-03	8.94E-04	1.35E-03
RMSE 3M	1.59E-03	**1.41E-03**	1.70E-03	2.76E-03	2.49E-03	1.83E-03	2.68E-03
RMSE 6M	2.91E-03	**2.50E-03**	3.50E-03	4.35E-03	4.50E-03	3.99E-03	5.23E-03
RMSE12M	4.82E-03	**4.18E-03**	6.14E-03	6.75E-03	6.87E-03	7.43E-03	8.97E-03
1M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
RMSE 1M	1.43E-03	1.54E-03	**1.01E-03**	**1.01E-03**	1.14E-03	1.19E-03	1.24E-03
RMSE 3M	3.77E-03	4.15E-03	**2.24E-03**	2.30E-03	2.58E-03	2.65E-03	3.09E-03
RMSE 6M	7.20E-03	7.96E-03	**4.10E-03**	4.26E-03	4.40E-03	4.86E-03	5.80E-03
RMSE-12M	1.22E-02	1.37E-02	**6.37E-03**	6.63E-03	6.48E-03	7.84E-03	9.48E-03
3M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
RMSE 1M	1.38E-03	**1.33E-03**	1.67E-03	1.36E-03	1.46E-03	1.72E-03	1.54E-03
RMSE 3M	3.54E-03	**3.36E-03**	3.54E-03	3.43E-03	3.62E-03	3.97E-03	4.02E-03
RMSE 6M	5.60E-03	5.17E-03	5.53E-03	5.43E-03	**4.68E-03**	6.38E-03	6.74E-03
RMSE-12M	7.37E-03	6.66E-03	7.03E-03	7.11E-03	**6.55E-03**	8.53E-03	9.46E-03
6M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
RMSE 1M	1.33E-03	1.41E-03	1.35E-03	**1.30E-03**	1.38E-03	3.38E-03	3.77E-03
RMSE 3M	3.66E-03	3.88E-03	3.50E-03	3.54E-03	**3.49E-03**	1.00E-02	9.62E-03
RMSE 6M	3.21E-02	6.46E-03	5.62E-03	5.65E-03	**5.23E-03**	1.42E-02	1.56E-02
RMSE-12M	7.92E-03	8.72E-03	**7.36E-03**	7.39E-03	7.44E-03	1.75E-02	1.91E-02
12M LIBOR	CKLS-4F	CKLS-5F	AFDNS-3F	AFDNS-5F	RW	VAR1	AR1
RMSE 1M	-	**1.24E-03**	1.29E-03	1.26E-03	1.28E-03	2.99E-03	3.28E-03
RMSE 3M	-	3.47E-03	3.43E-03	3.53E-03	**3.28E-03**	8.84E-03	8.65E-03
RMSE 6M	-	5.60E-03	5.56E-03	5.69E-03	**5.16E-03**	1.27E-02	1.42E-02
RMSE-12M	-	7.47E-03	**7.32E-03**	7.37E-03	8.02E-03	1.52E-02	1.71E-02

Table 8. Correct Direction Change Percentage (CDCP) Measure

Note: In this table, we present the results from the CDCP measure across all models for all five maturities USD-LIBOR rates. This measure is computed at different horizons of one-, three-, six- and 12-month horizons. For 16 out of 20 combinations, the RW process seems to be superior in correctly predicting the change of direction in the USD-LIBOR rates. The results in bold font indicate the models of best out-of-sample performance in terms of this measure.

CDCP	CKLS-5F					CKLS-4F				
horizon	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.500	0.245	0.245	0.220	0.265	0.495	0.245	0.280	0.240	-
3M	0.445	0.280	0.335	0.265	0.280	0.430	0.275	0.320	0.260	-
6M	0.425	0.345	0.455	0.345	0.370	0.410	0.345	0.370	0.350	-
1Y	0.505	0.490	0.420	0.490	0.470	**0.510**	0.495	0.505	0.480	-
CDCP	AR1					VAR1				
horizon	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.400	0.265	0.240	0.220	0.275	0.465	0.255	0.245	0.220	0.270
3M	0.375	0.280	0.300	0.265	0.280	0.415	0.280	0.300	0.265	0.280
6M	0.395	0.345	0.365	0.345	0.370	0.395	0.345	0.365	0.345	0.370
1Y	0.500	0.490	0.505	**0.515**	0.505	0.500	0.490	0.505	0.495	**0.510**
CDCP	AFDNS5F					AFDNS3F				
horizon	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.390	0.285	0.275	0.220	0.265	0.475	0.295	0.245	0.240	0.275
3M	0.390	0.280	0.300	0.265	0.280	0.420	0.275	0.300	0.265	0.280
6M	0.395	0.345	0.365	0.345	0.370	0.400	0.340	0.365	0.345	0.370
1Y	0.500	0.490	**0.515**	0.475	0.480	0.505	0.495	0.505	0.470	0.465
CDCP	RW									
horizon	1W	1M	3M	6M	1Y					
1M	**0.560**	**0.715**	**0.740**	**0.745**	**0.715**					
3M	**0.550**	**0.685**	**0.685**	**0.705**	**0.700**					
6M	**0.555**	**0.645**	**0.620**	**0.630**	**0.615**					
1Y	0.485	**0.500**	0.495	0.495	0.470					

Table 9. The results of the Clark-West test for USD-LIBOR rates computed at various horizons (1, 3, 6 and 12 months).

Note: This table presents the p-values of the CW test. The Clark-West test for the nested models is a one-tail test. The null hypothesis states that, based on the mean square prediction error (MSPE) statistic, two models M1 (the nested model) and M2 (the nesting model) have equal predictive ability, while the alternative hypothesis states that the more general M2 outperforms M1. We considered the following two sequences of model nesting $RW \subset AR1 \subset VAR1 \subset CKLS4F \subset CKLS5F$ and $AFDNS3F \subset AFDNS5F$. For p-values smaller than 0.05, we reject the null and conclude that M2 has a better forecasting ability than M1. We display the p-value results across the five maturities (horizontally) and across the four horizons (vertically).

M1 is nested in M2	M2:AR1					M2: VAR1					M2: CKLS-4F					M2: CKLS-5F				
M1: RW	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.37	1.00	1.00	1.00	1.00	0.00	0.50	0.95	0.98	1.00	0.00	1.00	0.00	0.00	-	0.00	1.00	0.00	0.10	0.00
3M	1.00	1.00	1.00	1.00	1.00	0.00	0.10	1.00	1.00	1.00	0.00	1.00	0.00	0.81	-	0.00	1.00	0.00	1.00	0.00
6M	1.00	1.00	1.00	1.00	1.00	0.00	0.07	1.00	1.00	1.00	0.00	1.00	0.00	1.00	-	0.00	1.00	0.00	1.00	0.00
1Y	1.00	1.00	1.00	1.00	1.00	0.00	0.22	1.00	1.00	0.99	0.00	1.00	0.00	1.00	-	0.00	1.00	0.00	1.00	0.01
M1: AR1	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M						0.00	0.02	0.97	0.96	0.99	0.00	1.00	0.00	0.00	-	0.00	1.00	0.00	0.00	0.00
3M						0.00	0.00	0.08	0.02	0.00	0.00	1.00	0.00	0.00	-	0.00	1.00	0.00	0.00	0.00
6M						0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	-	0.00	1.00	0.00	0.00	0.00
1Y						0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	-	0.00	1.00	0.00	0.00	0.00
M1: VAR1	1W	1M	3M	6M	1Y	1W	1W	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M											0.99	1.00	0.00	0.00	-	0.99	1.00	0.00	0.00	0.00
3M											0.02	1.00	0.00	0.00	-	0.00	1.00	0.00	0.02	0.00
6M											0.00	1.00	0.00	1.00	-	0.00	1.00	0.00	0.33	0.00
1Y											0.00	1.00	0.00	0.00	-	0.00	1.00	0.00	0.93	0.00
M1: CKLS-4F	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M																0.03	1.00	0.00	1.00	-
3M																0.00	1.00	0.00	1.00	-
6M																0.00	1.00	0.00	0.00	-
1Y																0.00	1.00	0.00	1.00	-
M1 is nested in M2																M2: AFDNS-5F				
M1: AFDNS-3F	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M																0.84	0.07	0.00	0.62	0.08
3M																0.98	0.50	0.00	0.83	0.16
6M																1.00	0.63	0.07	0.57	0.04
1Y																0.94	0.59	0.30	0.32	0.07

Table 10. The results of the Diebold-Mariano test for USD-LIBOR rates computed at various horizons (1, 3, 6 and 12 months).

Note: This table contains the p-values for the DM test for all pairs of non-nested models in our analysis. We compare the following pairs (M1, M2) of models: (RW, AFDNS-3F), (RW, AFDNS-5F), (AR1, AFDNS-3F), (AR1, AFDNS-5F), (VAR1, AFDNS-3F), (VAR1, AFDNS-5F), (CKL-4F, AFDNS-3F), (CKLS-4F, AFDNS-5F) (CKLS-5F, AFDNS-3F), (CKLS-5F, AFDNS-5F). For p-values smaller than 0.05 we reject the null hypothesis of equal predictive power between M1 and M2. The decision on the better model is based on the ME accuracy measures. We indicate by "-" sign when the model M1 on the left column is inferior to M2 (AFDNS-3F or AFDNS-5F) and by "+", if otherwise. The "=" sign indicates that at 5% level of significance we reject the null and conclude that the forecasts of the models that we compare have similar forecasting performance.

M1 \ M2	M2: AFDNS-3F					M2: AFDNS-5F				
M1: RW	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.03-	0.02-	0.32=	0.05=	0.04-	0.32=	0.05=	0.10=	0.04-	0.02-
3M	0.00-	0.03-	0.58=	0.28=	0.30=	0.74=	0.04-	0.35=	0.35=	0.28=
6M	0.00-	0.09=	0.72=	0.56=	0.62=	0.69=	0.43=	0.72=	0.69=	0.61=
1Y	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=
M1: AR1	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.03-	0.00-	0.43=	0.01-	0.01-	0.32=	0.01-	0.04-	0.01-	0.00-
3M	0.00-	0.00-	0.33=	0.16=	0.11=	0.90=	0.00-	0.21=	0.18=	0.10=
6M	0.00-	0.00-	0.40=	0.28=	0.30=	0.20=	0.11=	0.43=	0.39=	0.31=
1Y	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=
M1: VAR1	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.90=	0.02-	0.31=	0.00-	0.00-	0.16=	0.05=	0.02-	0.01-	0.00-
3M	0.76=	0.23=	0.11=	0.08=	0.01-	0.43=	0.36=	0.04-	0.08=	0.01-
6M	0.70=	0.50=	0.34=	0.22=	0.15=	0.88=	0.69=	0.39=	0.38=	0.18=
1Y	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=
M1: CKLS-4F	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.03-	0.00-	0.14=	0.04-	n/a	0.23=	0.00-	0.59=	0.03-	n/a
3M	0.09=	0.00-	0.98=	0.24=	n/a	0.17=	0.00-	0.51=	0.30=	n/a
6M	0.02+	0.00-	0.95=	0.01-	n/a	0.33=	0.00-	0.90=	0.01-	n/a
1Y	1.00=	1.00=	1.00=	1.00=	n/a	1.00=	1.00=	1.00=	1.00=	n/a
M1: CKLS-5F	1W	1M	3M	6M	1Y	1W	1M	3M	6M	1Y
1M	0.02-	0.00-	0.11=	0.02-	0.06=	0.22=	0.00-	0.52=	0.02-	0.02+
3M	0.01+	0.00-	0.14=	0.20=	0.30=	0.11=	0.00-	0.50=	0.24=	0.27=
6M	0.00+	0.00-	0.68=	0.39=	0.63=	0.18=	0.00-	0.82=	0.52=	0.62=
1Y	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=	1.00=

Table 11. SPA test results across all seven model specifications

Note: This table reports the p-values of the superior predictive ability (SPA) test proposed by Hansen (2005) for the 1-week, 1-month, 3-month, 6-month and 1-year USD-LIBOR rates at forecasting horizons of one-, three-, six- and 12 months. The forecasting analysis is based on the parameter estimates over the period 2nd of January 1998 to 29th of September 2017. Failure to reject the null hypothesis (at 5%) indicates that the reference model (name marked in bold at the top of the each column) has superior predictive ability over the set of all the other models. The results for this SPA test show that both AFDNS models have all the p-values above 0.05 across all maturities and also all horizons, implying their superiority against the group of the remaining six models. The next superior models in terms of forecasting power are the CKLS-5F model for 1-week and 3-month maturities and the AR(1) model for 1-month maturity.

Maturity	**1W USD LIBOR**							**1M USD LIBOR**						
H\M	**CKLS 5F**	**CKLS 4F**	**AR(1)**	**VAR(1)**	**AFDNS 5F**	**AFDNS 3F**	**RW**	**CKLS 5F**	**CKLS 4F**	**AR(1)**	**VAR(1)**	**AFDNS 5F**	**AFDNS 3F**	**RW**
1M	0.42	0.384	0.482	0.404	0.083	1	0	0	0	1	0.004	1	1	0.007
3M	1	0.207	0	0.319	0.239	1	0	0	0	1	0.002	1	1	0
6M	1	0	0	0.005	1	1	0	0	0	1	0.004	1	1	0
1Y	1	0	0	0	1	1	0	0	0	1	0	1	1	0
Maturity	**3M USD LIBOR**							**6M USD LIBOR**						
H\M	**CKLS 5F**	**CKLS 4F**	**AR(1)**	**VAR(1)**	**AFDNS 5F**	**AFDNS 3F**	**RW**	**CKLS 5F**	**CKLS 4F**	**AR(1)**	**VAR(1)**	**AFDNS 5F**	**AFDNS 3F**	**RW**
1M	1	0.048	1	0.007	1	0.098	0.09	0	0	0	0	1	1	0.001
3M	1	0.026	0.298	0	1	1	0	0	0	0	0	1	1	0.001
6M	1	0.052	0.001	0	1	1	0	0	0	0	0	1	1	0.009
1Y	1	0.024	0	0	1	1	0	0	0	0	0	1	1	0.004
Maturity	**12M USD LIBOR**													
H\M	**CKLS 5F**	**CKLS 4F**	**AR(1)**	**VAR(1)**	**AFDNS 5F**	**AFDNS 3F**	**RW**							
1M	0.001	-	0	0	1	0.264	0							
3M	0	-	0	0	1	0.46	0.002							
6M	0.011	-	0	0	1	0.198	0.007							
1Y	0.026	-	0.003	0.013	1	0.351	0.015							

Table 12. The results of the model confidence set (MCS) test (Hansen et al., 2011)

Note: This table presents the indicators for best predictor set of models proposed by Hansen et al. (2011), with a 5% critical level, 10,000 bootstrap replications and an average block size of 20, for the 1-week, 1-month, 3-month, 6-month and 1-year USD-LIBOR at forecasting horizons of one-, three-, six- and 12-months. The value 1 indicates that the model at the top of each column belongs to the model confidence set, while 0 says otherwise. The model with the most inclusions in the confidence set is AFDNS-3F (17/20), followed by CKLS-5F and VAR1 with 10/20 inclusions, while the RW model belongs consistently to the confidence set only for 6-month LIBOR rates. In addition, for 1-month LIBOR the confidence set comprises four models: AFDNS-3F, AFDNS-5F, VAR1 and AR1.

Maturity	**1W USD-LIBOR**							**1M USD-LIBOR**						
H/Model	CKLS 5F	CKLS 4F	AFDNS 5F	AFDNS 3F	AR1	VAR1	RW	CKLS 5F	CKLS 4F	AFDNS 5F	AFDNS 3F	AR1	VAR1	RW
1M	0	0	0	1	0	1	0	0	0	1	1	1	1	0
3M	1	0	0	0	0	0	0	0	0	1	1	1	1	0
6M	1	0	0	0	0	0	0	0	0	1	1	1	1	0
1Y	1	0	0	0	0	0	0	0	0	1	1	1	1	0
Maturity	**3M USD-LIBOR**							**6M USD-LIBOR**						
H/Model	CKLS 5F	CKLS 4F	AFDNS 5F	AFDNS 3F	AR1	VAR1	RW	CKLS 5F	CKLS 4F	AFDNS 5F	AFDNS 3F	AR1	VAR1	RW
1M	1	0	0	1	0	0	0	0	0	0	1	0	1	1
3M	1	0	0	1	0	0	0	1	0	0	1	0	1	1
6M	1	0	0	1	0	0	0	0	0	0	1	0	1	1
1Y	1	0	0	1	0	0	0	1	0	0	1	0	1	1
Maturity	**12M USD-LIBOR**													
H/Model	CKLS 5F	CKLS 4F	AFDNS 5F	AFDNS 3F	AR1	VAR1	RW							
1M	0	-	1	1	0	0	0							
3M	0	-	1	1	0	0	0							
6M	0	-	1	1	0	0	0							
1Y	1	-	1	1	0	1	1							