informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 000–000
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Q:1 Q:2
Tractable Term Structure Models

Bruno Feunou,[a] Jean-Sébastien Fontaine,[a] Anh Le,[b] Christian Lundblad[c]

Q:3 Q:4 [a] Bank of Canada; [b] Smeal College of Business, Pennsylvania State University, University Park, Pennsylvania; [c] Kenan–Flagler Business School, University of North Carolina, Chapel Hill, North Carolina 27599-3490

Contact: jsfontaine@bankofcanada.ca (BF); js.leon.fontaine@gmail.com (J-SF); anh.le.finance@gmail.com (AL); christian_lundblad@unc.edu, https://orcid.org/0000-0003-1937-5462 (CL)

Received: February 16, 2020
Q:5 **Revised:** April 6, 2021
Accepted: May 24, 2021
Published Online in Articles in Advance:
[redacted], 2021

https://doi.org/10.1287/mnsc.2021.4214

Abstract. We introduce a new framework that facilitates term structure modeling with both positive interest rates and flexible time series dynamics but that is also tractable, meaning amenable to quick and robust estimation. Using both simulations and U.S. historical data, we compare our approach with benchmark Gaussian and stochastic volatility models as well as a shadow rate model that enforces positive interest rates. Our approach, which remains arbitrarily close to arbitrage free, offers a more accurate characterization of bond Sharpe ratios because of a better fit of the volatility dynamics and a more efficient estimation of the return dynamics. Further, the shadow rate and stochastic volatility models exhibit important restrictions that are largely absent in our approach.

History: Accepted by Agostino Capponi, finance.
Supplemental Material: The data files and online appendix are available at https://doi.org/10.1287/mnsc.2021.4214.

Q:6 **Keywords:** term structure • lower bound • no arbitrage • no dominance

Q:7
1. Introduction

Q:8 Q:9 Q:10 As interest rates remain at or below zero in the face of persistent unconventional monetary policy in many industrialized countries, the need for a tractable term structure model—one that is amenable to quick and robust estimation—that can accommodate the near–lower-bound behavior of interest rates is stronger than ever before. Despite its practical importance even after years at the lower bound, characterizing bond risks and returns remains vexing. Unfortunately, existing bond pricing models that incorporate bounded interest rates (starting with Black 1995) suffer from intractable bond pricing, inflexible volatility dynamics, or both. Our contribution is to offer a framework to build potentially nonlinear models that are tractable yet arbitrarily close to arbitrage free.

Subjecting this framework to both realistic data simulations and U.S. historical data, we find significant improvements relative to benchmark models along several important dimensions. In a simulation environment, we show that our approach offers more accurate characterizations of bond Sharpe ratios. By combining information on risks and returns, Sharpe ratios are an important statistic to differentiate models on economic grounds. Second, in U.S. historical data, we find that our tractable models offer yield and return forecasts with the same or better accuracy than other benchmark models and improve yield volatility forecasts. These results occur both away from and near the lower bound, as well as in both in-sample and out-of-sample settings.

The elaborate simulation and out-of-sample evidence also provides valuable empirical insights that are relevant for researchers, policy makers, and portfolio managers alike. First, researchers have long struggled to obtain an accurate decomposition of risk premia in terms of prices and quantities of risk. It is well known that, even in normal times, many existing models face a tension in simultaneously capturing both components accurately; this is only more complicated during zero lower-bound (ZLB) episodes. In its ability to capture volatility dynamics without compromising the estimation of risk premia, our framework offers a tractable avenue to overcoming this tension. Second, policy makers are also interested in the ability of a model to provide reliable forecasts of these components over multiple horizons to better inform, for example, monetary policy and systemic risk analysis. Q:11 Q:12

Finally, portfolio managers, particularly those with a large fraction of their portfolios invested in U.S. Treasuries, will find value in a model that is able to provide reliable reward-to-risk metrics such as the Sharpe ratio.[1]

To guarantee tractability in pricing, we directly specify a recursive construction that connects bond prices at successive maturities to a small set of state variables, consistent with the pronounced evidence that bond yields are driven by a small number of factors. The direct specification of bond prices that we employ is reminiscent of the dynamic Nelson–Siegel approach (henceforth, DNS) (see Diebold and Li 2006). Although the DNS is nested in our framework, we allow for a high degree of flexibility in modeling the time series dynamics of the pricing factors, including stochastic volatility, along with nearly complete freedom in modeling the short rate, including nonlinear formulations that are consistent with a lower bound.

Although the pricing functions are explicitly specified, we recover prices that remain economically sensible. One reason put forth in Christensen et al. (2011) as to why DNS models are popular is that DNS bond prices appear to be *almost* arbitrage free even if they are not exactly free of arbitrage in a frictionless market (Bjork and Christensen 1999, Filipovic 1999). We pursue this intuition formally for our more general setting, showing that our broad family of models delivers bond prices that are free of dominant trading strategies in the sense of Rothschild and Stiglitz (1970) and Levy (1992). Intuitively speaking, the no-dominance (ND) property rules out almost all arbitrage opportunities, and we show that any remaining bond trading arbitrage opportunities cannot be self-financing and do not survive in the presence of nonzero short-selling costs. Overall, this analysis suggests that the practical relevance of DNS models generalizes to our broader, more flexible framework.

Although our framework precludes dominant bond trading strategies under far more general assumptions, we consider an example model that facilitates comparison with existing models. Specifically, in the spirit of Black (1995) and Wu and Xia (2016), we impose a "hockey stick"-like lower bound on interest rates. Second, we permit a multivariate conditional volatility process similar in spirit to the multivariate volatility models introduced by Noureldin et al. (2011). We show how to undo the nonlinearity of this example model, typically a major obstacle in estimation, and to take advantage of recent advances (see Joslin et al. 2011) such that estimation is highly convenient, fast, and robust.

To benchmark performance, we include in our analysis three popular alternatives. First, we consider the no-arbitrage (NA) affine Gaussian term structure model that does not allow for a lower bound and features constant yield volatility. Second, we consider the no-arbitrage shadow rate model in Black (1995) and Wu and Xia (2016) that enforces a lower bound but also features essentially constant yield volatility away from the lower bound.[2] Third, we consider a standard no-arbitrage affine stochastic volatility model (Dai and Singleton 2000) that does not facilitate a lower bound. Importantly, robust global estimates of all three models can be obtained in a straightforward manner.

To compare Sharpe ratio estimates across our model and these benchmarks, we set up a simulation environment where, in contrast to the case of historical data, the true conditional Sharpe ratios are known. To generate simulations, we use a data-generating process that incorporates the salient features of the U.S. bond market data—such as an upward sloping average yield curve, downward sloping average yield volatility, predictable bond excess returns, and time-varying volatility—that includes episodes at the lower bound toward the end of each sample. The results show that our approach offers a more accurate characterization of bond Sharpe ratios than the benchmark models.

One of the deeper insights from our results is that the Sharpe ratio estimates are more accurate because the flexible volatility dynamics in our framework readily translate into more accurate return forecasts (the *numerator* of the Sharpe ratio), which are because of the econometric efficiency gained by the much-improved volatility forecasts (the *denominator* of the Sharpe ratio). Another reason why our framework delivers more accurate Sharpe ratios is that the standard shadow rate model exhibits a pattern of biases in return and volatility forecasts that substantially aggravates the accuracy of bond and portfolio Sharpe ratio estimates. Put differently, we document that a full no-arbitrage implementation of the shadow rate model introduces a tension between returns and volatility forecasts. A related tension is well known for models relying on square root processes (Dai and Singleton 2000, Joslin and Le 2021). Hence, one surprising finding is that both stochastic volatility and Black shadow rate models suffer from tensions because of the no-arbitrage restrictions binding the bond return and volatility dynamics, even if the exact nature of these tensions differs across these two models.

Moving beyond the simulation setup, we fit the same models to the U.S. historical data in both full and out-of-sample contexts. Interestingly, we obtain pricing errors and yield forecasts that are very similar, in a statistical sense, across candidate models (both near and away from the ZLB). In contrast to the simulation exercise, estimates of conditional first moments may not allow for a clean separation of models in small samples, possibly because of the near-unit root nature of bond yields. In sharp contrast, we do find that our model offers a far better fit of yield volatilities, both in and out of sample (and near and away from the ZLB). Of course, including flexible volatility dynamics contributes to the improvement. For example, the benchmark no-arbitrage stochastic volatility model that we consider offers, in some subsamples at least, fairly competitive volatility forecasts. However, as in our simulations, easing the tension between fitting yields and volatilities also plays an important role. Despite some ability to forecast volatility, the benchmark no-arbitrage stochastic volatility model fails to capture bond empirical risk premia dynamics on a real-time basis; our preferred model does not suffer this limitation.

By way of comparison, it should also be noted that Engle et al. (2017) design an important extension of the dynamic Nelson–Siegel model tailored for long-term scenario generation and forecasts. Their design incorporates a lower bound, time-varying volatility, and correlation, as well as regime changes that increase persistence. Our goal is different. We provide a broad framework for tractable models precluding dominant bond trading strategies across a rich set of dynamic specifications. Instead of long-horizon forecasts, our empirical application focuses on conditional Sharpe ratio estimates.

Two other important papers combine time-varying volatility with a lower bound in no-arbitrage models. Monfort et al. (2017) allow for a combination of stochastic volatility and the lower bound for interest rates within the affine no-arbitrage framework. Filipovic et al. (2017) introduce the class of linear-rational no-arbitrage models featuring several realistic features. Each of these models requires a particular combination of state dynamics and the pricing kernel to deliver analytical bond prices. Importantly, they often rely on filtering and estimation procedures, including possibly using auxiliary financial market data, that can be computationally challenging to implement in practical day-to-day applications.

Finally, a substantial body of work uses no-arbitrage shadow rate models to study bond yields and bond risk premia.[3] Without exception, this stream of literature relies on constant variances and correlations for the risk factors. However, studying the risk premium and volatility jointly has become more important given how pervasive the lower bound has become. As Bauer and Rudebusch (2016) note, the boundary introduces an asymmetry in the relationship between bond yields and bond risk factors, implying that the bond volatility influences many important model forecasts. Recognizing this issue, Cieslak and Povala (2016) end their sample in 2010 before long-term yields, burdened by the tightness of the boundary, become less sensitive to news (Swanson and Williams 2014, Bauer and Rudebusch 2016). Similarly, Creal and Wu (2015) introduce time-varying macroeconomic uncertainty but end their sample in mid-2012. Both models do not prevent negative yields.

2. Tractable Dynamic Term Structure Models

This section introduces a new family of dynamic term structure models in which bond prices are specified directly, guaranteeing that bond yields are available in closed form with only minimal assumptions about the risk factor dynamics. In particular, we specify the relation between bond prices and the state vector, along with the associated state dynamics, in a way that bond yields are free of dominant trading strategies (which we will explain). Our framework admits as a special case the popular dynamic Nelson–Siegel approach of Diebold and Li (2006). However, our framework is significantly more general in its ability to accommodate realistic features of the data. In our empirical analysis, we employ a version of our proposed modeling framework that allows for the presence of a lower bound on yields as well as flexible volatility dynamics for yields.

2.1. A No-Dominance Model with a Lower Bound

Consider zero-coupon bonds with a face value of one dollar. Assumption 1 is a direct specification for the price of the n-period bond $P_n(X_t)$, where X_t is the state vector.

Assumption 1. *The n-period zero-coupon bond price $P_n(X_t)$ is given recursively by*

$$P_n(X_t) = P_{n-1}(g(X_t)) \times \exp(-m(X_t)), \tag{1}$$

$$P_0(X_t) \equiv 1, \tag{2}$$

where $X \in \underline{X} \subseteq \mathbf{R}^N$, where $m(X) \in \underline{M} \subseteq \mathbb{R}$ is a scalar, and $g(X) \in \underline{X}$.

Because zero-coupon bond prices are available in closed form, all yields and forward rates are also available. The forward rate $f_{n,t}$ applicable to a one-period loan n periods in the future is given by

$$f_{n,t} = m(g^{\circ n}(X_t)), \tag{3}$$

where $g^{\circ k}(X_t)$ is the function $g(\cdot)$ applied k times (see the appendix).

The appendix shows that this general framework guarantees the absence of dominant trading strategies (Rothschild and Stiglitz 1970, Levy 1992). That is, prices generated by Assumption 1 always satisfies the ND condition, a concept closely related to the well-known NA condition. Specifically, the NA condition holds if any portfolio with positive cash flows for a strictly positive probability and zero cash flows otherwise must command a strictly positive price. The weaker ND condition holds if and only if portfolios with strictly positive payoffs in all states must have positive prices. Section 2.2 provides more detail on the ND condition with respect to tractable term structure models.

The functions $m(\cdot)$ and $g(\cdot)$ in Equation (1) are the key primitives in our construction of bond prices. Applying $n = 1$ to this equation leads to the one-period bond price of $P_1(X_t) = \exp(-m(X_t))$. Thus, the function $m(X_t)$ captures the one-period interest rate as a function of the state vector X_t. The function $g(\cdot)$ embodies how the price of the bond is discounted back by one period. Equation (2) corresponds to our assumption that bonds are redeemed at their face value of one dollar.

Our empirical results are based on a particularly relevant example of an ND model that imposes a lower bound on interest rates in the spirit of Black (1995) and Wu and Xia (2016). For the short rate function $m(\cdot)$, we let

$$r_t = m(X_t) = \theta\, w\left(\frac{s_t}{\theta}\right), \tag{4}$$

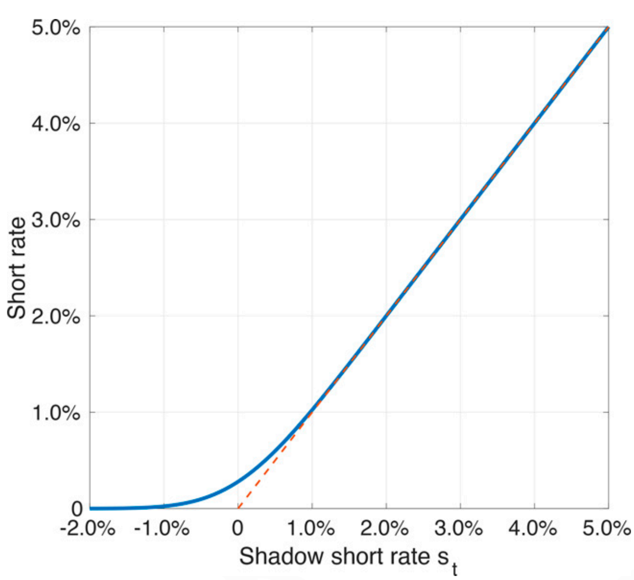
where $\theta > 0$ is a scalar and $s_t = \delta_0 + \delta_1' X_t$ is the shadow rate. The $w(x)$ function, closely related to the short rate function adopted by Wu and Xia (2016), is defined as

$$w(x) = x\Phi(x) + \phi(x), \tag{5}$$

where $\Phi(x)$ and $\phi(x)$ are the cumulative probability function and the density function of a standard normal distribution, respectively.

To illustrate Equations (4) and (5), Figure 1 displays the short rate r_t for values of the shadow rate s_t between −2% and 5% where we fix $\theta = 0.0070$, producing the familiar hockey stick pattern. It is visually apparent (and it can be shown analytically) that this choice for $w(x)$ implies the following properties: (i) r_t has a lower bound, (ii) r_t increases with the shadow rate s_t, and (iii) r_t converges to the shadow rate for large values of s_t. We choose to pin down the lower bound to zero for simplicity because the short maturity yields in our sample do not take negative values.

Q:16 **Figure 1.** (Color online) Hockey Stick



Note. The hockey stick transformation from the shadow rate s_t to the short rate r_t is given by $\theta w(s_t/\theta)$, where $w(x) = x\Phi(x) + \phi(x)$ and $\theta = 0.0070$. $\Phi(x)$ and $\phi(x)$ are the cumulative probability function and the density function of a standard normal distribution, respectively.

As our setting is quite general, other modeling choices are available. For instance, it is possible to set $w(x) = \log(1 + \exp(x))$, which also guarantees positive interest rates. In addition, the online appendix shows other examples of ND models where yields are given by linear or linear-quadratic functions of the pricing factors.

To obtain yields and forward rates, we adopt, for simplicity, a linear choice for $g(\cdot)$:

$$g(X_t) = KX_t, \tag{6}$$

where K is an $N \times N$ matrix. Although the choice of the function $w(x)$ embeds a reduced-form interpretation that s_t is the shadow short rate, Equations (4)–(6) together imply that the one-period forward rate n periods ahead is given by

$$f_{n,t} = \theta\, w\left(\frac{\delta_0 + \delta_1' K^n X_t}{\theta}\right), \tag{7}$$

which suggests a similar interpretation linking the forward rate $f_{n,t}$ to a shadow forward rate $s_{n,t} = \delta_0 + \delta_1{}' K^n X_t$. Like the short rate, the observed forward rate remains positive for all values of the shadow forward rate. Of course, the observed and shadow forward rates converge to each other for large values of the shadow rate $s_{n,t}$.

2.2. Properties

The solution for forward rates in Equation (7) is because of the recursive assumption for bond prices in Equation (1). By way of comparison, the online appendix shows that under a specific parameterization, our framework nests the Nelson–Siegel loadings. To this extent, our approach is reminiscent of the DNS framework of Diebold and Li (2006). This connection is important given the premium on tractability. Although the DNS framework does not guarantee positive yields, that class of models remains popular to a large extent because it is easy to use and the models provide good forecasting performance. One reason for the popularity of DNS models is that they seem *almost* arbitrage free, as put forth in Christensen et al. (2011).[4]

Therefore, an important question is whether more general choices of $m(\cdot)$ and $g(\cdot)$ within our framework produce bond prices that are also almost arbitrage free and that can be trusted in practical forecasting applications. We provide an answer to this question in the appendix. As mentioned, bond prices in our framework do not offer dominant trading strategy. We show, in addition, that any remaining arbitrage opportunity our framework could permit cannot be self-financing and in addition, that the opportunity does not survive in the presence of nonzero short-selling costs. This result holds for any choices of $m(\cdot)$

and $g(\cdot)$ and requires only mild conditions for the dynamics of the risk factors X_t. Indeed, a substantial literature has documented the costs to establish and maintain short Treasury bond positions, even for the most liquid issues (Duffie 1996, Krishnamurthy 2002, Vayanos and Weill 2008, Banerjee and Graveline 2013).

Therefore, it appears that the practical relevance of the DNS models may generalize to our broader, more flexible framework. By swapping the NA requirement for the weaker ND requirement, we hope to broaden the scope of tractable yet economically sensible term structure models. Choosing a model in the ND framework does not preclude that we recover a useful representation of the data. In addition, choosing a model in the NA framework does not guarantee that we recover a plausible representation of the data by all accounts. For instance, Duffee (2010) shows that estimates of fully flexible NA models produce maximum Sharpe ratios that are astronomically high.

Beyond flexibility and tractability, the ND framework offers another potential benefit. To see this, compare the ND pricing Equation (7) that one would obtain in a standard NA setting. First, the ND intercept δ_0 is invariant to the volatility specification. The intercepts in the NA model would contain Jensen terms that vary with maturity and depend on the variance parameters. This difference resembles the distinction between the constant term in the Nelson–Siegel model and the additional Jensen term in a standard NA Gaussian model. Second, the scaling parameter θ in the ND pricing Equation (7) is also invariant to the volatility specification. The scaling parameters in an NA model would vary with maturity and depend strongly on the volatility of yields. Therefore, it is plausible that the no-arbitrage restriction introduces a tension between the fit of yields and estimates of the volatility parameters, as in Joslin and Le (2021) for square root processes, that is relaxed in the ND setting.

2.3. State Dynamics

To facilitate comparison with existing models, we consider conditionally Gaussian dynamics for convenience. However, we reemphasize that our framework precludes dominant bond trading strategies under far more general dynamics. For example, future work could easily consider additional features including a shifting end point in the spirit of Kozicki and Tinsley (2001) and Bauer and Rudebusch (2020), a long-memory process as in Goliński and Zaffaroni (2016), or switching regimes as in Bansal et al. (2004).[5]

Every model that we consider builds on the following discrete time vector autoregressive VAR(1) dynamics:

$$X_{t+1} = K_0 + K_1 X_t + \Sigma_t^{1/2} \varepsilon_{t+1}, \tag{8}$$

where ε_{t+1} is i.i.d. standard normal. We also consider a general case where the conditional dynamics for volatility Σ_t are given by

$$\Sigma_t = \Sigma_0 + a\Sigma_{t-1} + b\Delta X_t \Delta X_t{}', \tag{9}$$

where a and b are both scalar and $\Delta X_t = X_t - X_{t-1}$, which is similar in spirit to the multivariate volatility models introduced by Noureldin et al. (2011). We label this model SV-B_{nd}, where SV indicates time-varying volatility, B indicates a Black-type shadow rate, and the subscript $_{nd}$ indicates the no-dominance property. For comparison with the existing shadow rate models, we also consider a case with a constant covariance matrix $\Sigma_t = \Sigma_0$, which we label B_{nd}. Both the SV-B_{nd} and B_{nd} specifications can be estimated quickly and robustly. The case with constant volatility is closely related, and hence, it can be compared with the widely used NA implementation of Black (1995) by Wu and Xia (2016).

2.4. Estimation Strategy

A central practical objective of our paper is to introduce a set of models that are both accurate and quick to estimate. An important insight in Joslin et al. (2011) is that implementation of affine models is easier if the candidate model is rotated to an observationally equivalent representation in which the state variables are the first principal components of yield or forward rates such that the state variables become observable.

The same insight is applicable to shadow rate models even if they do not belong to the affine class of models studied by Joslin et al. (2011). In the online appendix, we show that we can essentially undo the nonlinearity of our model and transform it back into a linear space for the purpose of estimation. In the case of shadow rate models, the transformation recovers a shadow short rate and shadow forward rates that we use as observable risk factors with linear dynamics. Then, based on this representation, we provide the closed-form conditional likelihood of the data. Finally, we analytically concentrate out parameters of the VAR(1) from the likelihood, which preserves another appealing feature of the approach in Joslin et al. (2011). The fact that our construction for Σ_t uses $\Delta X_t \Delta X_t'$ to update the volatility is an important building block for this approach. In this way, the conditional mean (K_0, K_1) parameters in Equation (8) do not enter the volatility dynamics in Equation (9) and thus, can be optimally obtained as a GLS estimate of the VAR(1).

Overall, we can achieve fast, convenient, and robust estimation of every model. For robustness, we also check that our key results are unchanged if we assume that the pricing portfolios are measured with error and use a Kalman filter to derive the likelihood. Note that a quick and robust estimation procedure is available more generally as long as the short rate function $m(.)$ is invertible and the $g(.)$ function remains affine. For even more general specifications of these functions, one could use the robust estimation method proposed by Andreasen and Christensen (2015) for nonlinear dynamic term structure models with potentially non-Gaussian latent factors.

3. Model Evaluation in a Simulated Environment

We first estimate and evaluate several bond pricing models using a large number of simulated yield data sets. Using simulated yields makes comparisons between models more accurate because all true predictive moments, such as yield forecasts, volatility forecasts, and conditional Sharpe ratios, are known at each point in time in every simulation. This is unlike using historical data (which we do in the next section), where predictive moments are not known and some are difficult to estimate precisely. This is especially true if we want to carry the assessments to periods over which yields are close to the lower bound.

The data-generating process for the simulations is the linear-rational term structure model of Filipovic et al. (2017) that satisfies the lower bound and allows for time-varying volatility. We select simulations with 30 years of data and that include an episode at the lower bound toward the end of the sample. The online appendix provides more details, and it shows that this sophisticated environment produces realistic simulated paths for yields and captures the upward sloping yield curve as well as the downward sloping yield volatility curve. This environment also produces realistic conditional Sharpe ratios that we can use to assess term structure models. Table 1 reports quantiles of the realized Sharpe ratios in U.S. data for investment horizons of 3, 6, and 12 months for bonds with 2, 5, and 10 years to maturity. For each bond and each investment horizon, we compute the realized Sharpe ratios by dividing the average excess return by its standard deviation over rolling windows of 36 months. Our rolling calculations attempt to capture, in a model-free way, the time variation in Sharpe ratios. Importantly, the same calculations can be applied to both historical yields and simulated yields to facilitate comparison.

Overall, although the estimates in Filipovic et al. (2017) were not targeted to match moments of the historical Sharpe ratios, the simulation closely matches the realized moments. Because yields have declined and average returns were higher after 2008, the historical median estimates are higher than the comparable statistics reported by Duffee (2010). The ex post realized Sharpe ratios are higher than the conditional Sharpe ratios that actually prevailed in bond markets; this is likely because a substantial share of that decline in yields over the last few decades was unexpected. This may explain why the historical estimates tend to be somewhat higher than the estimates from simulations.

With simulations in place, we assess the performance of a number of models. Specifically, we consider an NA affine Gaussian model (G_{na}), an NA affine stochastic volatility model ($A1_{na}$), an NA Black shadow rate model (B_{na}) as in Wu and Xia (2016), a tractable ND Black shadow rate model (B_{nd}), and a tractable ND Black shadow rate model with stochastic volatility (SV-B_{nd}). For comparability, all models feature three pricing factors. The G_{na} and the $A1_{na}$ models correspond respectively to the well-known $A_0(3)$ and $A_1(3)$ models, respectively, of Dai and Singleton (2000). The difference between each of the considered models and our adopted data-generating process reflects the inescapable reality that the modeler will never know with certainty the true data dynamics. As an aside,

Table 1. Sharpe Ratios Using Historical and Simulated Data

		Simulation					U.S. data				
m, months	h, years	2.5	25	Median	75	97.5	2.5	25	Median	75	97.5
3	2	−0.51	−0.14	0.03	0.20	0.54	−0.64	−0.08	0.35	0.68	1.64
	5	−0.50	−0.12	0.07	0.24	0.59	−0.49	−0.06	0.17	0.41	0.80
	10	−0.49	−0.10	0.09	0.26	0.62	−0.50	−0.05	0.11	0.30	0.61
6	2	−0.76	−0.20	0.06	0.31	0.84	−0.97	−0.17	0.39	0.88	2.27
	5	−0.76	−0.17	0.12	0.38	0.93	−0.81	−0.09	0.24	0.56	1.16
	10	−0.74	−0.15	0.14	0.41	0.96	−0.88	−0.09	0.15	0.43	0.87
12	2	−1.13	−0.28	0.12	0.52	1.36	−1.38	−0.22	0.38	1.16	2.49
	5	−1.20	−0.24	0.21	0.63	1.51	−1.57	−0.12	0.32	0.75	2.07
	10	−1.18	−0.22	0.25	0.67	1.61	−1.66	−0.17	0.24	0.68	2.00

Notes. For each bond maturity m and each investment horizon h, we compute the realized Sharpe ratios by dividing the average excess return by its standard deviation over rolling windows of 36 months. The same calculations are applied to both historical yields and simulated yields. Columns (3)–(7) show quantiles of the Sharpe ratios across all simulated samples. Columns (8)–(12) show quantiles of the Sharpe ratios for the U.S. data.

although we focus on a linear-rational model to assess these models, we note that we find similar results using the Gaussian quadratic term structure model as the data-generating process.

3.1. Sharpe Ratios

We use estimates of the conditional Sharpe ratios to evaluate and rank models. For a given investment horizon, the ratio's numerator is the conditional expected return in excess of the risk-free rate. The denominator is the conditional volatility of the corresponding return. Intuitively, the Sharpe ratio measures the expected return per unit of risk—where risk is measured by total volatility. Our ranking presumes that term structure models that accurately estimate conditional Sharpe ratios are more useful than other models that do not. For completeness, the online appendix reports results based on the estimates of the conditional mean and volatility of yields across models.

To establish this ranking, we consider three investment horizons up to 1 year ahead and four bond maturities up to 10 years. We measure the accuracy with the root median squared errors (RMedSEs) between true and model-implied Sharpe ratios. We use the median both along the time dimension and across simulations. Using median statistics mitigates the risk that one extreme observation might skew the performance of any of the models.

Table 2 reports the RMedSE statistics implied by all models considered. The columns labeled μ and σ report the means and standard deviations of Sharpe ratio estimates based on estimates of bond returns and volatility computed using rolling windows of 36 months. The column labeled T reports the RMedSE based on the rolling-window Sharpe ratio estimate, thereby providing a practical and realistic baseline against which to gauge the other models. The left side of Table 2 reports statistics computed using the full sample in every simulation. The right side of the table reports statistics using only subsamples close to the lower bound.

Looking at the results, we ask two questions. First, how do the NA and ND implementations of the Black shadow rate model compare with each other? The answer is that the B_{nd} model delivers more accurate Sharpe ratio estimates than the B_{na} model, on average, for every investment horizon and bond maturity that we consider, without exception. In fact, the B_{na} model sometimes even underperforms the simpler Gaussian G_{na} model. This contrast is rather stark over the lower-bound sample. Whereas the B_{nd} model handily outperforms the model G_{na}, the B_{na} model performs consistently worse than the Gaussian G_{na} model, barring one exception. We discuss the causes underlying these differences in the next subsection.

The second question we ask is what the relative gain is from modeling both the lower bound and the time-varying volatility of yields. The answer to this question is that the ND Black model with stochastic volatility, SV-B_{nd}, significantly outperforms other models. In full samples, the SV-B_{nd} model delivers the most accurate Sharpe ratios, on average, for all considered horizons and bond maturities. The RMedSE gains are sizeable relative to the baseline rolling-window estimates and relative to the NA Gaussian model, ranging from 10% to 30%. Finally, the SV-B_{nd} gains are smaller relative to the B_{nd} model but larger

Table 2. Estimating Conditional Sharpe Ratios in Simulations

		Entire sample								ZLB sample							
h, months	m	μ	σ	T	G_{na}	$A1_{na}$	B_{na}	B_{nd}	SV-B_{nd}	μ	σ	T	G_{na}	$A1_{na}$	B_{na}	B_{nd}	SV-B_{nd}
3	6 months	0.02	0.26	0.27	0.22	0.22	0.36	0.27	0.18[a]	0.18	0.18	0.51	0.29	0.30	0.28	0.17[a]	0.21
	2 years	0.04	0.26	0.38	0.47	0.46	0.76	0.48	0.36[a]	0.22	0.17	1.00	0.64	0.69	0.68	0.33	0.33[a]
	5 years	0.07	0.27	0.57	0.83	0.78	1.15	0.75	0.58[a]	0.23	0.19	1.28	1.03	0.99	1.16	0.70	0.47[a]
	9 years	0.08	0.27	0.73	1.17	1.09	1.40	1.07	0.84[a]	0.23	0.21	1.41	1.26	1.32	1.27	1.03	0.77[a]
6	6 months	0.05	0.38	0.34	0.18	0.18	0.32	0.24	0.16[a]	0.29	0.27	0.53	0.20	0.20	0.23	0.14[a]	0.17
	2 years	0.07	0.40	0.45	0.41	0.40	0.68	0.39	0.32[a]	0.35	0.27	0.92	0.48	0.49	0.61	0.27	0.24[a]
	5 years	0.11	0.42	0.65	0.73	0.70	1.05	0.67	0.50[a]	0.38	0.29	1.23	0.75	0.83	0.97	0.58	0.38[a]
	9 years	0.13	0.42	0.78	1.07	0.99	1.26	0.97	0.75[a]	0.37	0.31	1.38	1.06	1.12	1.16	0.97	0.70[a]
12	6 months	0.10	0.59	0.46	0.16	0.14	0.25	0.19	0.14[a]	0.52	0.43	0.69	0.16	0.14	0.18	0.12[a]	0.13
	2 years	0.15	0.64	0.60	0.35	0.33	0.55	0.34	0.27[a]	0.62	0.45	1.00	0.35	0.32	0.47	0.25	0.19[a]
	5 years	0.21	0.67	0.80	0.64	0.60	0.93	0.61	0.42[a]	0.65	0.46	1.26	0.65	0.64	0.82	0.48	0.33[a]
	9 years	0.23	0.68	0.88	0.93	0.85	1.15	0.90	0.63[a]	0.65	0.49	1.44	0.91	0.85	1.02	0.86	0.55[a]

Notes. Precision of the conditional Sharpe ratio estimates at horizon h (column (1)) and maturity m (column (2)). μ and σ refer to the median across simulations of the mean and standard deviations of rolling-window Sharpe ratio estimates, respectively. The other columns report the RMedSEs between true and estimated conditional Sharpe ratios. T refers to the RMedSE of rolling-window Sharpe ratio estimates. G_{na}, $A1_{na}$, and B_{na} refer to the NA affine Gaussian model, the NA affine stochastic volatility model, and the NA shadow rate model, respectively. B_{nd} and SV-B_{nd} refer to the ND Black models with and without BEKK volatility, respectively. The lower-bound subsamples collect dates when the one-month rate is 25 basis points or less.

Q:18

[a]The best performance for a given horizon, maturity, and sample.

Proof Only

relative to the B_{na}. The $A1_{na}$ model also offers gains in full samples.

When focusing on the lower-bound samples, the baseline rolling-window estimates offer much poorer accuracy relative to the full-sample results. However, the SV-B_{nd} model still delivers the best performance for all horizon-maturity combinations, with the exception of one case where the B_{nd} is ranked first (six-month horizon and six-month terminal maturity). In all cases, the RMedSE gains are large relative to either of the NA models, ranging between 15% and 50%.

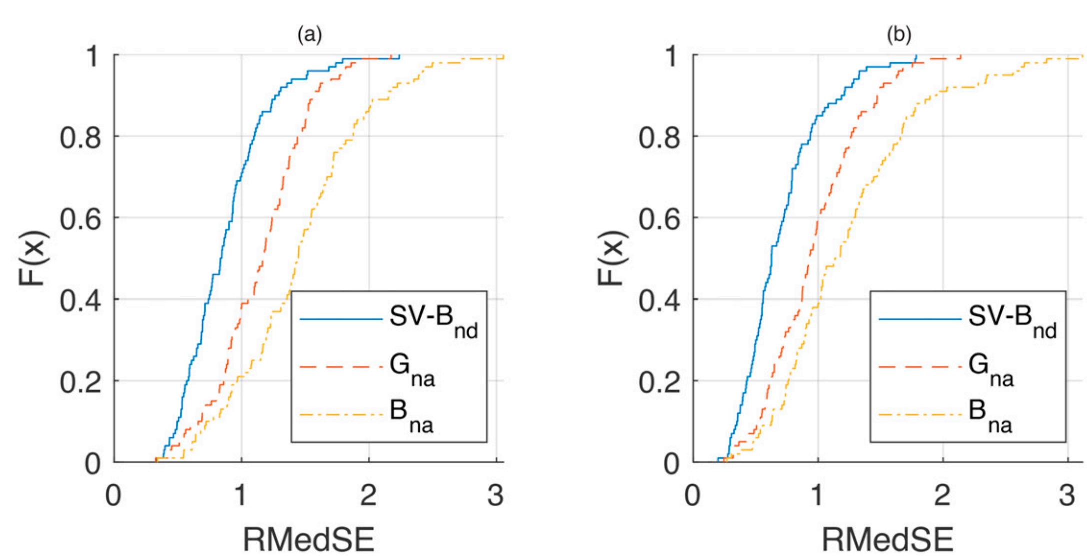
The winning model in Table 2 is always statistically significantly better than the G_{na} model. To see this clearly, we can look at the distribution of simulation results. Figure 2 reports the cumulative empirical distribution of RMedSEs across simulations. For clarity, we include results for the SV-B_{nd}, G_{na}, and B_{na} models (the distribution for the $A1_{na}$ model is close to the G_{na}, like their respective medians in Table 2). This distribution accounts for the sampling variability because of the sample size, the model estimation, and misspecification.

The ranking between the three models clearly emerges. The SV-B_{nd} exhibits first-order stochastic dominance over the benchmark G_{na}, which itself exhibits first-order stochastic dominance over the B_{na} model. This means that any quantile for RMedSe in the distribution from the SV-B_{nd} model is lower than in the distribution from the G_{na}. A test for equality also rejects the null hypothesis for arbitrary small critical values. As mentioned earlier, we find similar results (unreported) when using the Gaussian quadratic term structure model as the data-generating process.

We also evaluate the accuracy of Sharpe ratios for *portfolios* of bonds, which depends on the models' ability to forecast the correlations among future yields. We form three simple equal-weighted portfolios consisting of zero-coupon bonds with maturities of six months and one, two, … nine years. The first portfolio combines all 10 maturities, the second portfolio combines the first 5 short-term maturities, and the third portfolio combines the last 5 maturities. Table 3, sharing the same column structure as Table 2, reports the performance by the four models in forecasting the Sharpe ratios of the three portfolios. The key messages are essentially unchanged. The SV-B_{nd} model clearly outperforms all other models for every category. The performance gains obtained by the SV-B_{nd} model, relative to the Gaussian model, are similar to those observed for the case of the individual bonds reported in Table 2. In particular and as expected, the gains are especially large across samples near the lower bound.

3.1.1. Discussion of the Simulation Results. The combination of time-varying volatility and satisfying the lower bound appears important in driving the performance in these simulations. In the online appendix, we show that the flexible volatility dynamics help improve forecasts of the denominator in the Sharpe ratio (which is expected) as well as the numerator (which is

Figure 2. (Color online) RMedSE Distributions



Notes. (a) m = 9 years and h = 3 months. (b) m = 9 years and h = 12 months. The cumulative distribution function of simulated RMedSEs from Sharpe ratio estimates is shown. G_{na} and B_{na} refer to the NA affine Gaussian and Black models, respectively. SV-B_{nd} refers to the ND Black models with BEKK volatility. Both panels report results for holding a bond with nine years to maturity at the end of the investment horizon. Panel (a) shows results for the three-month holding horizon. Panel (b) shows results for the 12-month holding horizon.

Q:17

Table 3. Bond Portfolio Sharpe Ratios

		Entire sample					ZLB sample				
h, months	m	G_{na}	$A1_{na}$	B_{na}	B_{nd}	SV-B_{nd}	G_{na}	$A1_{na}$	B_{na}	B_{nd}	SV-B_{nd}
3	6 months to 9 years	0.5	0.97	1.51	0.94	0.72[a]	0.7	0.99	1.04	0.49	0.45[a]
	6 months to 4 years	0.9	0.92	1.39	0.95	0.68[a]	1.0	0.99	1.10	0.73	0.49[a]
	5–9 years	1.0	0.91	1.30	0.92	0.68[a]	1.2	1.04	1.04	0.76	0.52[a]
6	6 months to 9 years	0.5	0.96	1.60	0.98	0.77[a]	0.6	0.99	1.20	0.58	0.46[a]
	6 months to 4 years	0.8	0.95	1.43	0.94	0.69[a]	0.8	1.06	1.22	0.81	0.53[a]
	5–9 years	0.9	0.95	1.30	0.94	0.68[a]	0.9	1.06	1.15	0.86	0.58[a]
12	6 months to 9 years	0.4	0.96	1.61	1.03	0.77[a]	0.4	0.91	1.39	0.72	0.52[a]
	6 months to 4 years	0.7	0.94	1.45	1.00	0.66[a]	0.7	0.96	1.24	0.79	0.53[a]
	5–9 years	0.8	0.94	1.31	1.01	0.66[a]	0.8	0.99	1.27	0.95	0.58[a]

Notes. G_{na}, $A1_{na}$, and B_{na} refer to the NA affine Gaussian model, the NA affine stochastic volatility model, and the NA shadow rate model, respectively. B_{nd} and SV-B_{nd} refer to the ND Black models with and without BEKK volatility, respectively. For the G_{na} model, we report the RMedSEs between true and model-implied estimates. For other models, we report the RMedSE statistics relative to the G_{na} model. The first two columns give the forecast horizon h and the maturity m at the end of the investment periods, respectively. The lower-bound subsamples collect dates for which the one-month rate is 25 basis points or less.

[a]The best performance for a given forecast horizon, bond maturity, and sample.

more surprising). The latter gains are because of the increased econometric efficiency.

It is easy to understand why the SV-B_{nd} model does a better job, given its flexible volatility dynamics. However, what can explain the poor performance of the NA Black model B_{na} relative to the ND Black model B_{nd}? Fortunately, the simulation exercise can provide further insights into the differences.

Conceptually, the only difference between these models can be found in the pricing equations, which are where the answer must lie. To help facilitate comparison, the pricing equation in the B_{na} model is given by

$$f_{n,t} \approx S_n\, w\left(\frac{A_n + B_n' X_t}{S_n}\right),$$

where A_n and B_n are restricted by the absence of arbitrage (the coefficients S_n, A_n, and B_n are given in the online appendix). Some of the parameters that constitute S_n play a dual role in the B_{na} models. They govern bond pricing and contribute to fitting the volatility of yields. This dual role may create tension between fitting yields and volatilities. Notably, this tension is largely absent in ND models (see Equation (7)).[6]

Any trade-off between yield and volatility forecasts will affect the Sharpe ratio estimates. However, poorer yield and volatility estimates do not always lead to worse Sharpe ratio estimates. To illustrate this possibility, consider an investment with an expected excess return and expected volatility of 0.3 and 0.1, respectively, where the true Sharpe ratio is three. An error of +0.1 in the forecasts of returns and volatilities implies a conditional Sharpe ratio of $0.4/0.2 = 2$. This is a difference of one relative to the true ratio. By contrast, an error of -0.05 in the forecasts of the expected returns and return volatility implies a conditional Sharpe ratio of $0.35/0.05 = 5$. This is now a difference of two relative to the true ratio. The smaller absolute forecast errors produce a larger Sharpe ratio error.

Following this illustration and to be more precise, we can express the mean squared error of Sharpe ratios analytically. Define the forecast error $fe_j(x)$ and the mean squared forecast error $msfe_j(x)$ for some forecast x from model j:

$$\begin{aligned} fe_j(x) &\equiv (x_j - x_0) \\ msfe_j(x) &\equiv E[fe_j(x)^2 \mid x_0] = E[(x_j - x_0)^2 \mid x_0], \end{aligned}$$

where we suppress the horizon and maturity index for clarity, x_j is the model forecast from model j, and x_0 is the true forecast in the simulation. For instance, $msfe_j(xr^{(m)}_{t,t+h})$ gives the mean squared return forecast error for the yield with maturity m at the horizon h, and $msfe_j(\sigma^{(m)}_{t,t+h})$ gives the mean squared volatility forecast error. Analytically, it can be shown that $msfe_j(SR^{(m)}_{t,t+h})$ is given by

$$\begin{aligned} msfe_j(SR) \approx{}& msfe_j(xr) + b_1 msfe_j(\sigma) \\ &+ b_2 E[\, fe_j(xr)] \times E[\, fe(\sigma)] \\ &- b_3 COV(\, fe_j(xr); fe_j(\sigma)), \end{aligned} \tag{10}$$

where the coefficients are positive $b_1 > 0$, $b_2 > 0$, and $b_3 > 0$, and they depend on the true Sharpe ratio (see the online appendix). This expression is useful to confirm the intuition. The first term indicates that the accuracy of Sharpe ratio estimates moves one for one with the accuracy of returns forecasts. The second term indicates that the impact from the accuracy of the volatility forecasts depends on the level of the Sharpe ratio. The third term indicates that $msfe_j(SR)$ also increases if the product of the biases is positive, which is caused by the nonlinearity in the ratio. Finally, the last term shows that the $msfe_j(SR)$ increases with the correlation between the return forecast error and the volatility forecast errors $fe_j(xr)$ and $fe_j(\sigma)$.

In the online appendix, we show that the forecast biases are the key drivers for the relatively poor performance of the NA models in the simulation exercise. Both the B_{na} and G_{na} models exhibit large biases in their yield and volatility forecasts. In the case of the B_{na} model, both biases are negative and reduce the accuracy of the Sharpe ratio estimates. For the G_{na} model, the biases have opposite signs, which mitigates the effect on the accuracy of the Sharpe ratio estimates. We also show that the root mean squared errors (RMSEs) of yield and volatility forecasts do not explain the relatively poor performance of NA models in estimating conditional Sharpe ratios. Unsurprisingly, the SV-B_{nd} model delivers the most accurate forecasts with the smallest biases across the simulation results.

4. Empirical Illustrations

In this section, we complement our simulation evidence and examine the performance of the G_{na}, $A1_{na}$, B_{na}, B_{nd}, and SV-B_{nd} models based on historical data.[7] We use rates, measured monthly from January 1970 through August 2020, on one-month forward loans with 13 different maturities: one month, three months, six months, one year, two years, and annually until year 10. Specifically, we use the data provided by Gurkanyak et al. (2007) for maturities longer than six months. For maturities less than or equal to six months, we bootstrap the forward rates from Treasury bond prices provided by the Center for Research in Security Prices.

Based on these historical data, we report pricing errors, yield and volatility forecasts, and shadow rate estimates relative to the G_{na} model. Following Wu and Xia (2016) and for comparability with existing results, we use the Kalman filter estimate of the B_{na}, B_{nd}, and G_{na} models in our analysis. Consistent with the evidence in Joslin et al. (2013), our results are unchanged if we estimate every model with shadow forward portfolios priced without error. For the $A1_{na}$ and SV-B_{nd} models, we use the estimate obtained by assuming the portfolios of shadow forwards are priced perfectly. This approach is easy to implement and converges quickly and robustly to global parameter estimates.

4.1. Pricing Errors

Table 4 reports the RMSEs in fitting yields for selected maturities in the full sample and in the subsample where the short rate is at the lower bound. First, the results are very similar across models. This is not surprising because every model that we consider has three factors, which is well known to be enough to fit the cross-section of yields.

Second, we contrast the results between the NA and ND shadow rate models with constant volatility, B_{na} and B_{nd}. We find that the pricing errors are very close, within a fraction of one basis point (bp) over the full sample. This result parallels existing work finding negligible differences between the pricing errors of the Gaussian G_{na} models and those of the DNS models. Table 4 extends this result to models with a Black truncated short rate. Note that both models offer a small improvement over the standard Gaussian G_{na} model. As expected, this difference is concentrated in the period close to the lower bound.

Table 4. Pricing Errors

	Entire sample					ZLB sample				
m	G_{na}	$A1_{na}$	B_{na}	B_{nd}	SV-B_{nd}	G_{na}	$A1_{na}$	B_{na}	B_{nd}	SV-B_{nd}
6 months	22.2	21.1	18.2	17.7	21.3	15.1	14.8	13.7	12.1	16.1
1 year	18.7	18.5	15.2	15.3	18.4	11.9	11.1	7.2	7.6	10.9
2 years	12.4	12.7	10.4	10.6	12.1	9.8	10.3	4.5	5.1	6.7
5 years	5.2	3.2	4.0	3.9	3.5	4.5	2.0	3.0	2.7	3.3
10 years	1.9	1.3	1.4	1.4	1.4	1.4	0.9	0.9	0.7	1.1

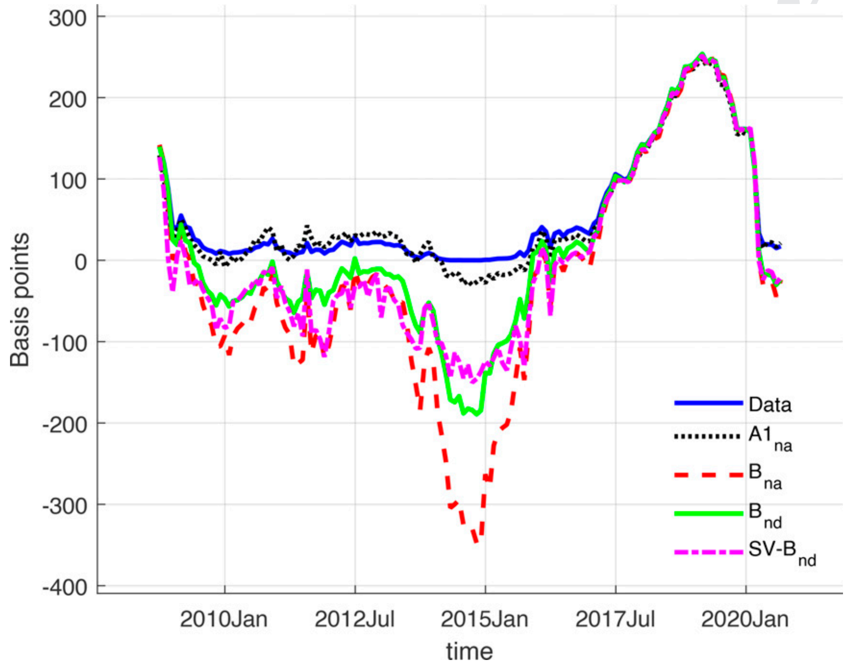
Notes. Pricing error RMSEs for NA and ND models (in basis points) are shown. G_{na}, $A1_{na}$, and B_{na} refer to the NA affine Gaussian model, the NA affine stochastic volatility model, and the NA shadow rate model, respectively. B_{nd} and SV-B_{nd} refer to the ND Black models with and without BEKK volatility, respectively. The sample period is 1970:January to 2020:August.

4.2. Shadow Short Rates

Figure 3 plots the one-month rate in the $A1_{na}$ and the one-month shadow rates implied by the Black models B_{na}, B_{nd}, and SV-B_{nd} along with the observed one-month short rate, focusing on the period after 2008. For much of the first lower-bound period, it is rather interesting that the shadow rates implied by the two ND models, despite their different volatility specifications, remain close to one another. Overall, the NA model B_{na} shadow rates remain within a range of 50 bps with one clear exception; for much of 2014, the shadow rate implied by the NA model B_{na} falls much deeper into the negative region, often by more than 100 bps, relative to the two ND shadow rates. In the more recent lower-bound period, starting in March 2020, these three models produce similar shadow rate estimates. The estimated shadow rate for the B_{na} model that we report is broadly similar to results reported by the Federal Reserve Bank of Atlanta based on the Wu and Xia (2016) model. In particular, both also report a shadow rate at −3% in 2014.

The pattern in Figure 3 suggests that although there has been considerable interest in studying the economic content of shadow rates, this must be done with caution given that shadow rates can be highly model dependent. This is also an important observation documented in Bauer and Rudebusch (2016).

Figure 3. (Color online) Shadow Rates



Notes. One-month short rate and shadow rate (1970:January to 2020:August) are shown. $A1_{na}$ and B_{na} refer to the NA affine stochastic volatility and the NA shadow rate models, respectively. B_{nd} and SV-B_{nd} refer to the ND Black models with and without BEKK volatility, respectively.

4.3. Short Rate Forecasts

Figure 4 shows each model's forecast of the one-month rate, focusing on the period close to the lower bound after 2008. We report results for the 3- and 12-month forecast horizons. Results for other maturities lead to similar observations (not reported).

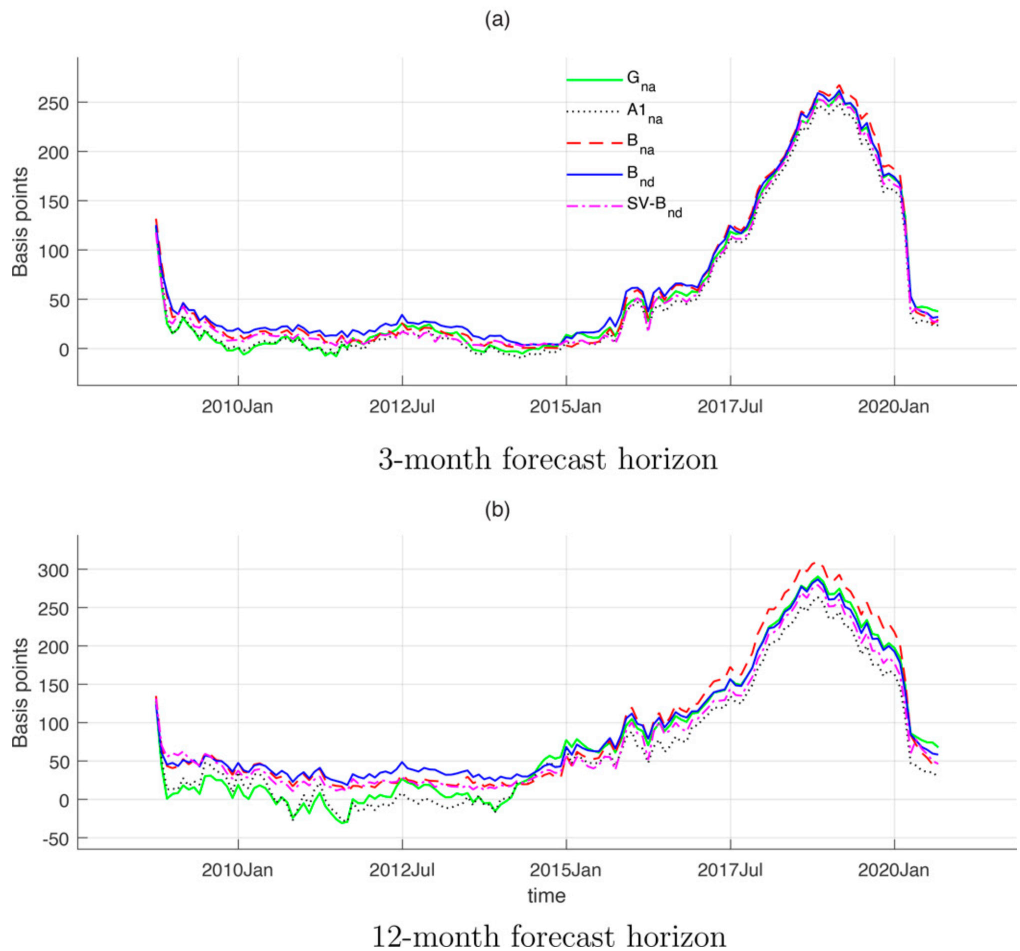
Two observations are worth nothing. First, the forecasts implied by the NA affine Gaussian model (green lines) and the NA stochastic volatility model (dotted lines) can be visibly distinct from the forecasts implied by the shadow rate models. In particular, their forecasts significantly violate the lower bound (at all horizons) in the prolonged period near the lower bound after 2008. Second, regardless of the model type—NA, ND, constant variance, or stochastic volatility—the forecasts by the shadow rate models are close to one another. Their proximity suggests that it can be challenging to accurately compare the different models in one short subsample close to the lower bound. This observation strengthens the case for the exercise in Section 3, where the models are evaluated in a controlled simulation environment. Additionally, this observation serves as a reminder that even large differences in the model-implied shadow rates, as documented in the previous subsection, should be interpreted with caution. As we can see here, large differences in shadow rates do not translate to empirically meaningful differences in short rate forecasts.

4.4. Volatility Forecasts

It might be challenging to compare models based on the forecast of yields in one short sample close to the lower bound, as we have argued. However, this argument does not carry fully to comparison based on the models' implied yield volatility. Figure 5 plots the volatility forecasts of the 24-month yield implied by the models over two forecast horizons: 3 and 12 months. As a benchmark, we compute the realized variances of the 24-month yield using 3 months of daily data leading up to each point in time t. It is known that realized volatility can be estimated precisely using relatively high-frequency data, even in a short window.

Every shadow rate model appears to do a good job during the lower-bound episodes starting in 2008 and 2020. Away from the lower bound, the SV-B_{nd} model performs well matching the large fluctuations of the RV series (ranging from as low as 40 bps to as high as 300 bps). This justifies the simple construction for the volatility dynamics in Equation (9). The volatility forecasts from the $A1_{na}$ model follow the broad pattern displayed by the realized volatility. Specifically, the $A1_{na}$ model tracks rather closely the realized volatility pattern in the recent episode near the lower bound and continues to do so when yields are lifted off the lower bound. However, the $A1_{na}$ volatility forecasts significantly miss the short period of very high volatility in the early 1980s. These results are consistent with the conclusion by Jacobs and Karoui (2009) that the

Figure 4. (Color online) Short Rate Forecasts



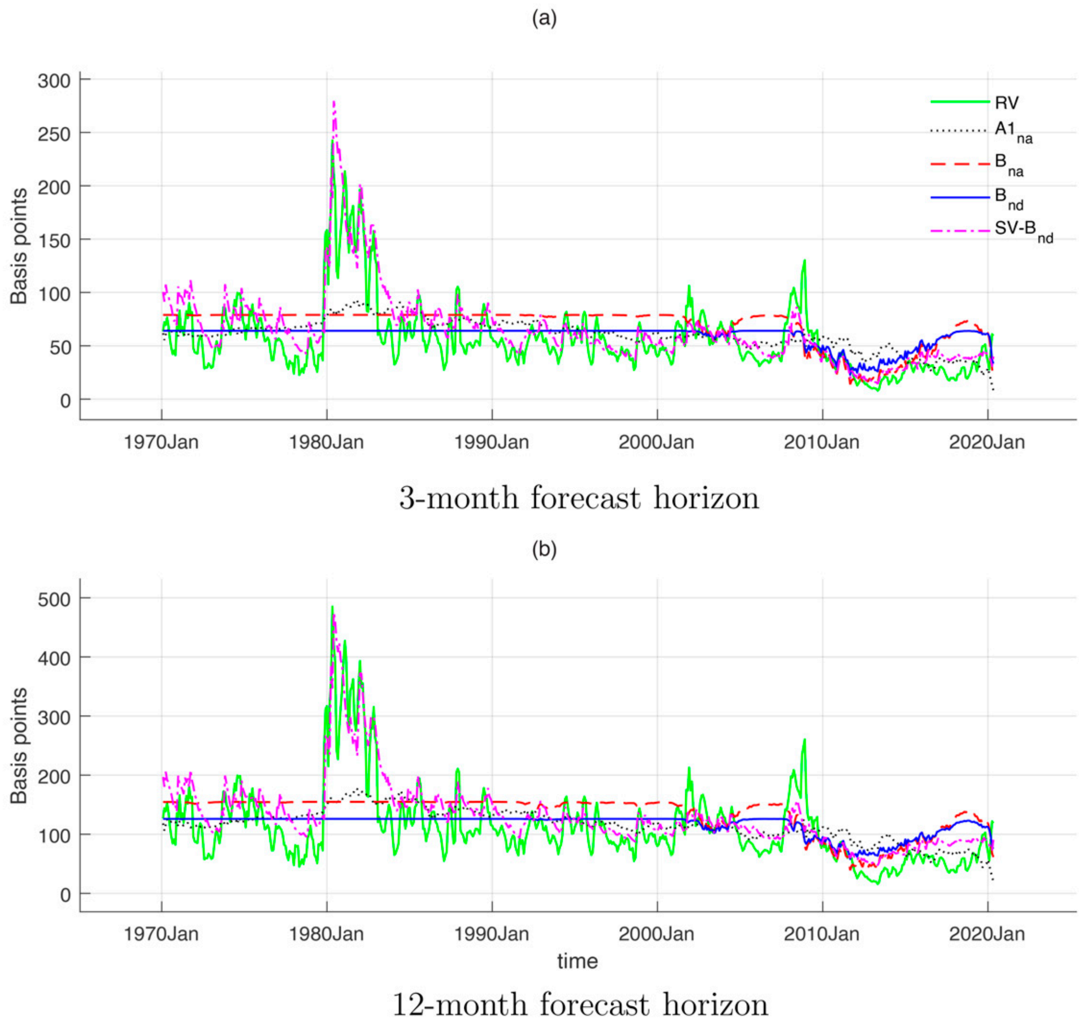
Notes. (a) Three-month forecast horizon. (b) Twelve-month forecast horizon. Forecasts of the one-month rate (1970:January to 2020:August) are shown. G_{na}, $A1_{na}$, and B_{na} refer to the NA affine Gaussian model, the affine stochastic volatility model, and the shadow rate model, respectively. B_{nd} and SV-B_{nd} refer to the ND Black models with and without BEKK volatility, respectively.

ability of this class of models to capture conditional volatility depends on the sample period.

In contrast and as expected, it is clear that the yield volatilities implied by the constant-variance Black models are essentially a flat line for periods away from the lower bound. Loosely speaking, the flat lines correspond to the average volatility levels away from the lower bound implied by the B_{na} and B_{nd} models. It is notable that they are not at the same level. Specifically, the average volatility level implied by the NA model B_{na} is higher than the level implied by the ND model B_{nd} and higher than the corresponding RV series for much of the sample, with the exception of the early 1980s. The fact that the B_{na} model seems to miss the average volatility level indicated by the RV series constitutes evidence that the B_{na} model might be constrained. As we argue before, the scaling parameter S_n in the pricing equation of the B_{na} model is directly linked to the volatility parameters of the model. As such, this dual role of the scaling parameter might create a tension in the ability of the model to fit the multiple dimensions of the data.

These results are confirmed by the RMedSEs reported in Table 5. Across maturities and horizons, the SV-B_{nd} model is able to reduce volatility forecast errors by 10%–30% relative to the $A1_{na}$ model during the full sample. The improvement can be even larger in specific cases or in subsamples. The reduction is particularly large for short-term yields near the lower bound, where the improvement ranges from 70% to 85%. This is precisely where we expect the SV-B_{nd} model to provide the largest improvement. Across other maturities and in the period near the lower bound, the improvement ranges from 10% to 50%. Comparing the B_{na} and B_{nd} models during the period near the lower bound, where these models exhibit meaningful volatility dynamics, does not produce a clear winner.

Figure 5. (Color online) Volatility Forecasts



Notes. (a) Three-month forecast horizon. (b) Twelve-month forecast horizon. Forecasts of the two-year bond yield volatility (1970:January to 2020:August) are shown. G_{na}, $A1_{na}$, and B_{na} refer to the NA affine Gaussian model, the affine stochastic volatility model, and the shadow model, respectively. B_{nd} and SV-B_{nd} refer to the ND Black models with and without BEKK volatility, respectively.

5. Model Evaluation with Out-of-Sample Forecasts

This section continues the evaluation of the G_{na}, $A1_{na}$, B_{na}, B_{nd}, and SV-B_{nd} models in a real-time out-of-sample forecast exercise based on the same historical data employed in the previous section. To start, each model is first estimated based on monthly historical data from January 1970 through December 1995. We use rates on one-month forward loans with 13 different maturities: one month, three months, six months, one year, two years, and annually until year 10. We then evaluate three-month-ahead yield and volatility forecasts for the six-month, 2-year, 5-year, and 10-year bonds, and we record the associated forecast errors. The results are qualitatively similar if we look at 6- or 12-month forecasts. We then reestimate each model expanding the sample to January 1996 data and repeat the evaluation. We iterate this procedure, expanding the sample one month at a time, until August 2020. Figure 6 reports out-of-sample root median squared forecast errors for yields and volatilities across all time periods.

5.1. Yield and Volatility Forecasts

We start with an evaluation of yield forecast errors. Panels (a) and (b) of Figure 6 report the results in subsamples before 2008 and after 2008 (away or near the zero lower bound), respectively. Despite the heterogeneity that we observe in the simulation environment, we do not observe significant differences across the models in terms of their ability to characterize yield expectations out of sample. Regardless of whether we

Table 5. Volatility Forecasts

		Entire sample				ZLB sample			
h, months	m	$A1_{na}$	B_{na}	B_{nd}	SV-B_{nd}	$A1_{na}$	B_{na}	B_{nd}	SV-B_{nd}
3	1 month	35.85	1.15	0.87	0.76	34.49	0.61	0.50	0.16
	2 years	16.52	1.48	1.04	0.69	15.28	0.73	1.13	0.50
	5 years	12.00	1.43	1.18	0.83	8.77	1.23	1.40	0.79
	10 years	12.34	1.25	1.09	0.71	13.65	0.96	0.80	0.47
	Average	19.17	1.33	1.05	0.75	18.05	0.88	0.96	0.48
6	1 month	49.73	1.10	0.85	0.77	46.07	0.68	0.61	0.20
	2 years	22.61	1.50	1.08	0.73	20.89	0.79	1.21	0.66
	5 years	16.87	1.39	1.17	0.86	13.24	1.13	1.24	0.84
	10 years	17.64	1.22	1.07	0.74	19.60	0.92	0.82	0.51
	Average	26.71	1.30	1.04	0.77	24.95	0.88	0.97	0.55
12	1 month	67.85	1.05	0.86	0.71	60.66	0.88	0.73	0.31
	2 years	30.34	1.52	1.15	0.89	28.00	0.99	1.35	0.92
	5 years	23.05	1.39	1.17	0.92	19.78	1.04	1.18	0.88
	10 years	25.35	1.17	1.05	0.76	28.33	0.90	0.76	0.57
	Average	36.65	1.28	1.06	0.82	34.19	0.95	1.00	0.67

Notes. Root median squared differences between volatility forecasts and our realized volatility forecasts for a bond with maturity m and horizon h are shown. $A1_{na}$ and B_{na} refer to the NA affine stochastic volatility model and the NA shadow rate model, respectively. B_{nd} and SV-B_{nd} refer to the ND models with and without BEKK volatility, respectively. For the B_{na} model, we report the RMedSEs, in basis points, between RV and model-implied volatility forecasts. For the B_{nd} and SV-B_{nd} models, we report the ratio of RMedSE relative to the $A1_{na}$ model. Forecasts horizons are 3, 6, and 12 months.

are away from or near the lower bound, out-of-sample yield forecasts with an expanding estimation window do not seem to be a metric for which model delineation is particularly useful.

We next turn to an evaluation of volatility forecasts, where the realized volatility is the square root of the sum of daily squared returns. Panels (c) and (d) of Figure 6 also report the results in subsamples before 2008 and after 2008, respectively. Regardless of the proximity to the zero lower bound, the volatility forecasting performance of the Gaussian model, G_{na}, is, not surprisingly, relatively poor. The two best models in terms of volatility forecasting are, also not surprisingly, the $A1_{na}$ and SV-B_{nd} models. The SV-B_{nd} model produces volatility forecast errors that are consistently small, on average, across maturities, and its performance relative to other models is particularly notable for shorter maturity bonds. In contrast, the performance of the $A1_{na}$ model is overall but deteriorates for shorter maturity bonds.

When we compare the performance of the B_{na} and B_{nd} models, we find that the ND model performs better before 2008, particularly for shorter-maturity bonds. This difference can only be attributed to the bias introduced by the no-arbitrage restriction. However, the performance of both improves after 2008, and the wedge between the two models essentially closes (panel (d) of Figure 6). For long maturities, their performance is close to the results based on the SV-B_{nd} model. This latter result suggests that the volatility compression that arises near the lower bound is capturing most of the variation in volatility during these periods (see also Kim and Priebsch 2013, Christensen and Rudebusch 2014).

Overall, a real-time out-of-sample exercise based on yield forecasting proves less useful for delineating models. This is perhaps not so surprising given the extremely persistent, near-unit root nature of bond yields. This was also part of the motivation for the simulation exercise in Section 3. However, the exercise is able to delineate models based on volatility dimensions, highlighting the importance of the models that directly facilitate stochastic volatility. Taken together, our preferred SV-B_{nd} model fares favorably in an out-of-sample evaluation. From this perspective, the closest competitor (particularly for some longer maturities) is the $A1_{na}$ stochastic volatility model. Hence, the results in the historical sample offer a more positive message about this model than the results in simulated samples.

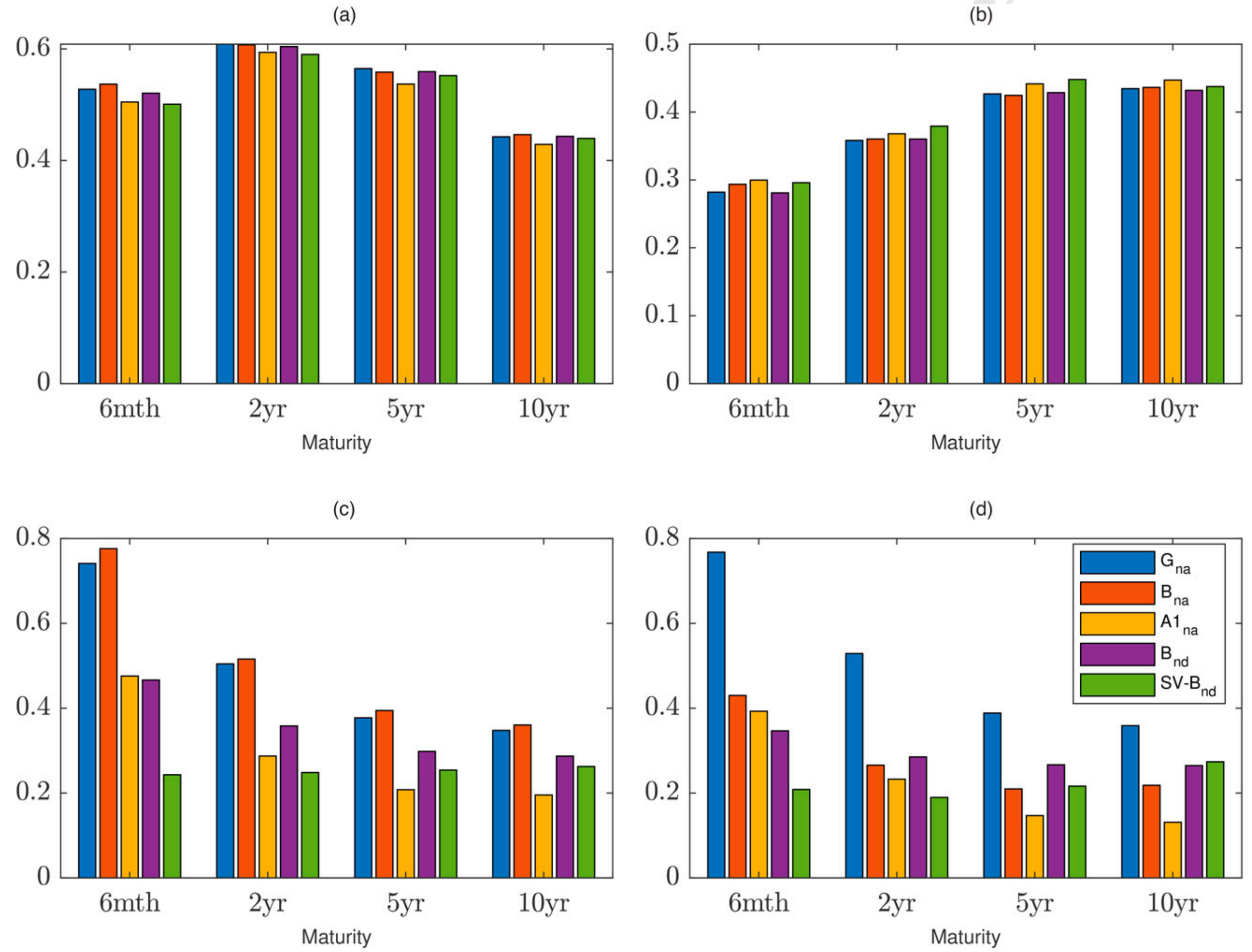
5.2. Linear Projection of Yield Changes

Beyond RMSE statistics, Dai and Singleton (2002) propose a modified version of the popular Campbell and Shiller (1991) expectations hypothesis regression as an alternative metric to assess the predictive power of a term structure model. The original Campbell and Shiller (1991) regression is often referred to as LPY(i), and the modified version is referred to as LPY(ii). In using essentially detrended variables such as yield changes and the slope of the yield curve, this approach affords an evaluation of each model's ability to predict yields while largely abstracting from the near-unit root behavior of bond data.

As is well known, LPY(i) involves regressing a series of yield changes on the corresponding slopes of the yield curve. The regression is designed such that if the expectation hypothesis holds, the coefficient of the regression should be one, regardless of yield maturity. What Campbell and Shiller (1991) find, however, is that these coefficients are significantly different from one. As a matter of fact, they are mostly negative and increasingly so as bond maturity increases. This pronounced failure of the expectations hypothesis translates to evidence for strong time variation of bond risk premia. LPY(ii) adds a *model-implied* risk premium component to the yield changes used in LPY(i). The premise is that if the risk premium adjustment is in agreement with the data, the regression coefficients will be corrected and moved closer to one.

We revisit the exercise to evaluate the G_{na}, $A1_{na}$, B_{na}, B_{nd}, and SV-B_{nd} models. As in Dai and Singleton (2002), we ask whether the LPY(ii) adjustment induced by each model can move the Campbell and Shiller (1991) regression coefficients closer to unity. It should be noted that in the out-of-sample spirit, we

Figure 6. (Color online) Out-of-sample Forecast Error



Notes. (a) Yield forecast RMSEs—before 2008. (b) Yield forecast RMSEs—after 2008. (c) Volatility forecast RMSEs—before 2008. (d) Volatility forecast RMSEs—after 2008. RMSEs in three-month-ahead out-of-sample forecasts of yields and yield volatilities are shown. Forecasts are based on estimation from January 1970 through December 1995 and adding one observation every month until August 2020. G_{na}, $A1_{na}$, and B_{na} refer to the NA affine Gaussian model, the NA affine stochastic volatility model, and the NA shadow rate model, respectively. B_{nd} and SV-B_{nd} refer to the ND Black models with and without BEKK volatility, respectively.

conduct this analysis on a *real-time* basis, whereas Dai and Singleton (2002) employ model estimates based on the entire sample.

Specifically, LPY(ii) is derived from the decomposition of the bond return into a risk premium component and an expectations component. Based on this decomposition, one can use the model-implied premium component as a correction term for LPY(i):

$$y_{n-1,t+1} - y_{n,t} + \frac{e_t^n}{n-1} = \alpha_n + \phi_n\left(\frac{y_{n,t} - y_{1,t}}{n-1}\right) + \epsilon_{n,t+1}, \tag{11}$$

where e_t^n is the relevant risk adjustment:

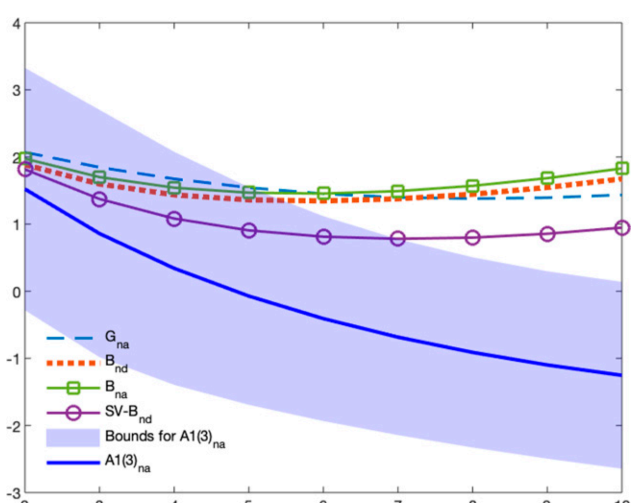
$$e_t^n = -(n-1)\hat{E}_t[y_{n-1,t+1} - y_{n,t}] + \hat{y}_{n,t} - \hat{y}_{1,t}. \tag{12}$$

The correction term e_t^n can be computed for every model. Equation (12) tells us that the ability of a candidate model to capture the risk adjustment in the context of LPY(ii) depends upon its ability to fit bond yields contemporaneously and to capture the time series yield dynamics. If that is the case, then the theoretical value of the ϕ_n coefficient is unity. Note that models with similar forecast RMSEs may offer different performance based on the LPY(ii) criteria—the estimates for ϕ_n can differ—especially if the yield forecast errors are correlated with the spread $y_{n,t} - y_{1,t}$.

To operationalize LPY(ii), we require yield forecasts as well as fitted yields from every model. The typical way to compute e_t^n is to estimate each model using the entire sample of data. However, given our focus on out-of-sample performance, we estimate each model every month in the expanding window framework discussed. We then compute e_t^n every month, collect the results from January 1995 until the end of the sample, and run the LPY(ii) in Equation (11). This means that the adjustment term is a real-time estimate without look-ahead bias.

Figure 7 plots the LPY(ii) coefficients ϕ_n across maturities n for every model. We consider an out-of-sample forecast horizon of 12 months. There are two key takeaways. First, similar to Dai and Singleton (2002), we find that the coefficients implied by the $A1_{na}$ model remain far from unity (confidence bounds are provided in the blue shaded area). The premium adjustments have a relatively small effect, and these coefficients are much closer to the standard Campbell–Shiller regression coefficients (i.e., ignoring the adjustment term). Despite the fact that the $A1_{na}$ performance in out-of-sample volatility forecasts is competitive, we infer that the well-known tension between yields, returns, and volatility dynamics in this no-arbitrage model has important negative consequences for the implications of out-of-sample forecasts in correctly capturing premium dynamics. Second, although there is relatively little separating the other models along this particular evaluation dimension, the coefficients implied by the SV-B_{nd} model are closer to unity across maturities than the $A1_{na}$ model. Taken together, from the perspective of yield and volatility forecasts along with LPY(ii) premium dynamics, the SV-B_{nd} model fares favorably out of sample in a manner consistent with its strong performance in the simulation environment.

Figure 7. (Color online) LPY(ii) Regression Coefficients



Notes. Estimates of coefficients ϕ_n in the LPY(ii) regression for annual yield changes based on Equations (11) and (12), where the adjustment e_t^n is computed based on the real-time model estimation with an expanding sample, are shown. G_{na}, $A1_{na}$, and B_{na} refer to the NA affine Gaussian model, the affine stochastic volatility model, and the shadow rate model, respectively. B_{nd} and SV-B_{nd} refer to the ND Black models with and without BEKK volatility, respectively. Confidence bounds for the LPY(ii) regression coefficients from the NA affine stochastic volatility model are provided in the blue shaded area.

6. Conclusion

We introduce a family of tractable no-dominance term structure models where bond prices are analytically available and very nearly arbitrage free for essentially any specification of the time series dynamics. Our results based on some special cases show how this new class of models can do a reasonable job, relative to existing models, in capturing the dynamics of conditional bond Sharpe ratios before and after 2008, when yields reach the lower bound.

Variation in the volatility term structure remains a challenge for existing models. Our family of no-dominance models is large and permits flexible specifications of the dynamic interactions between yield and macrovariables, including time-varying volatility and the interaction with the lower bound. This should allow future research to revisit several results involving the trade-off between the risk premium and yield volatility faced by investors, the influence of conventional and unconventional policy actions on this trade-off (including quantitative easing and forward guidance), and the correlations among international term structures (when far from or near to their respective lower bounds).

Acknowledgments

The authors thank Peter Christoffersen, Greg Duffee, Antonio Diez de los Rios, Peter Feldhutter, Jing-zhi Huang, Kris Jacobs, Richard Luger, and seminar and conference Q:13
participants at the Banque de France, the Federal Reserve Bank of New York, McGill University, the National Bank of Belgium, Penn State Smeal College of Business, Tsinghua University PBC School of Finance, the 2014 Tremblant Risk Conference, the 2015 CIRANO finance workshop, the 2015 Bank of Canada-FRBSB Fixed Income Conference, and the 2017 MFA meetings for comments and suggestions. They also thank Alison Arnot and Meredith Fraser-Ohman for editorial support.

Appendix. No-Dominance Term Structure Models

In the following, we study dominant and arbitrage trading strategies between bonds. For this purpose, note that, starting with $P_0(\cdot) \equiv 1$ and expanding the recursion (1), we get

$$P_n(X_t) = \exp\left(-\sum_{i=0}^{n-1} m(g^{\circ i}(X_t))\right). \qquad \text{(A.1)}$$

Equation (3) follows from the definition of the n-period yield and forward rate, $y_{n,t} \equiv -\log(P_n(X_t))/n$ and $f_{n,t} \equiv (n+1)y_{n+1,t} - ny_{n,t}$, respectively. In the following results, we rely on Assumption A.1, which is essential to clarify the set of admissible time series dynamics.

Assumption A.1. *The time series dynamics of X_t admit $\underline{X}$ as support.*

Assumption A.1 is analogous to the requirement in NA models that the historical and risk-neutral measures be equivalent measures. This is a mild requirement because this affords researchers the flexibility to consider rich dynamic and distributional assumptions. First, Theorem A.1 checks dominant bond trading strategies: portfolios of bonds with strictly positive payoffs.

Theorem A.1 (No Dominant Strategies). *Assumptions 1 and A.1 guarantee the absence of strictly dominant trading strategies in the bond markets.*

Proof. Let w_n denote the amount (in face value) invested in each n-period bond. Suppose that this portfolio guarantees positive payoffs: $\sum_n w_n P_{n-1}(X_{t+1}) > 0 \ \forall \ X_{t+1} \in \underline{X}$. From the pricing recursions in Equation (1), the price of this portfolio is given by

$$\sum_n w_n P_n(X_t) = \exp\left(-m(X_t)\right) \times \sum_n w_n P_{n-1}(g(X_t)). \tag{A.2}$$

Because $g(X_t) \in \underline{X}$ and because $\sum_n w_n P_{n-1}(X_{t+1}) > 0$ for all $X_{t+1} \in \underline{X}$, it follows that the price of this portfolio is strictly positive. Thus, a dominant trading strategy does not exist. □

Next, we consider portfolios that have a zero payoff with strictly positive probability but positive payoffs otherwise. The NA condition requires that the price of this portfolio be strictly positive. Theorem A.1 establishes that portfolios like this one cannot admit strictly negative prices in our framework.

Theorem A.2 (Nonnegative Payoffs). *Assumptions 1 and A.1 ensure that bond portfolios with strictly nonnegative payoffs cannot admit strictly negative prices.*

Proof. Let w_n denote the amount (in face value) invested in each n-period bond. Consider a portfolio with strictly nonnegative payoffs: $\sum_n w_n P_{n-1}(X_{t+1}) \geq 0 \ \forall \ X_{t+1} \in \underline{X}$. From the pricing recursion in Equation (1), the price of this portfolio is given by

$$\sum_n w_n P_n(X_t) = \exp\left(-m(X_t)\right) \times \sum_n w_n P_{n-1}(g(X_t)). \tag{A.3}$$

The price of this portfolio cannot be negative for it requires $\sum_n w_n P_{n-1}(g(X_t)) < 0$, but this would contradict $g(X_t) \in \underline{X}$ and $\sum_n w_n P_{n-1}(X_{t+1}) \geq 0$ for all $X_{t+1} \in \underline{X}$. □

Figure A.1 illustrates Theorem A.2 and clarifies that what is setting NA and ND apart is the set of zero-cost portfolios with zero or positive payoffs (both with positive probability). For portfolios with strictly nonnegative payoffs, the absence of arbitrage restricts prices to be on the positive half of the real line, *excluding* the origin. Assumption 1 allows for bond prices on the positive half of the real line, *including* the origin. The difference reduces to one point on the real line, the origin, graphically illustrated by Figure A.1.

Figure A.1 suggests that the distance between our models and the no-arbitrage paradigm is small. Does this difference matter? Note that Figure A.1 also tells us that any remaining arbitrage opportunities must involve a self-financing portfolio with strictly nonnegative payoffs in the future and with strictly zero setup costs today. However, a self-financing portfolio must involve short-selling bonds, and because short positions imply nonzero costs to set up and to maintain, the specific portfolio in Figure A.1 may not represent an arbitrage opportunity in a practical sense for sophisticated long-short investors. The presence of small transaction costs preventing self-financing strategies is *economically* plausible.

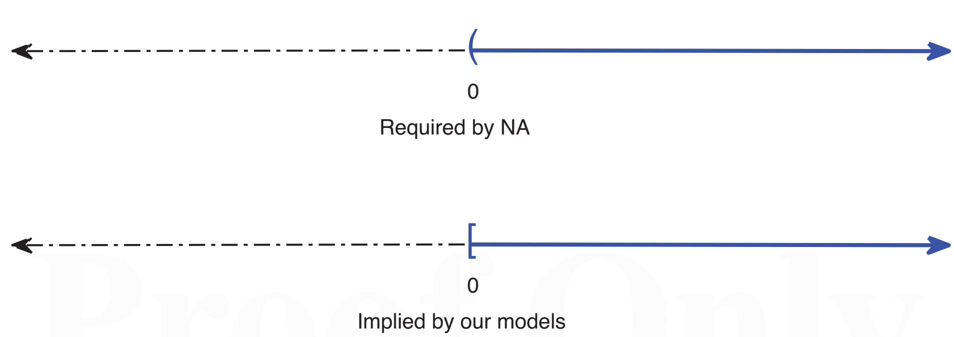
Theorem A.3 (Transaction Costs). *Assumptions 1 and A.1 rule out all arbitrage opportunities in bond prices if transaction costs for short positions are positive.*

Proof. Consider again a portfolio in which the amount (in face value) invested in each n-period bond is given by w_n. To account for transaction costs, let us assume that the setup cost at time t of the portfolio is C_0, and to realize the cash flows at time $t+1$, the transaction cost is given by C_1. The existence of transaction costs means that at least one of C_0 and C_1 is strictly positive. The one-period-ahead cash flows net of transaction costs are given by

$$CF(X_{t+1}) = \sum_n w_n P_{n-1}(X_{t+1}) - C_1.$$

Recall that the absence of arbitrage is *equivalent* to requiring that any portfolio with nonnegative payoffs must command a positive price. Thus, the key question here is as follows. If $CF(X_{t+1}) \geq 0$ for every $X_{t+1} \in \underline{X}$, then can we show that the price of the portfolio must be strictly positive, net of transaction costs? The price of this portfolios is given by

Figure A.1. (Color online) Prices of Portfolios with Strictly Nonnegative Payoffs



$$
\begin{aligned}
Price(X_t) &= \sum_n w_n P_n(X_t) + C_0 \\
&= \exp(-m(X_t)) \sum_n w_n P_{n-1}(g(X_t)) + C_0, \\
&= \exp(-m(X_t)) \times (CF(g(X_t)) + C_1) + C_0, \\
&= \exp(-m(X_t)) \times CF(g(X_t)) + \exp(-m(X)) C_1 + C_0,
\end{aligned} \tag{A.4}
$$

where the second line follows from Equation (1) and the third line follows from the definition of the cash flows $CF(X_{t+1})$. The first term on the right-hand side of (A.4) is nonnegative because $g(X_t) \in \underline{X} \Rightarrow CF(X_t) \geq 0$. Additionally, the last two terms must add up to a strictly positive number because C_0 and C_1 cannot be jointly zero. Thus, $Price(X_t) > 0$ as needed. □

Endnotes

[1] To provide an important example, we highlight the "official" foreign holdings of U.S. federal debt; roughly $4 trillion in 2020 represent, in part, the foreign currency reserve portfolios of various central banks (U.S. Treasury Bulletin). As these reserve portfolio managers face tight risk management restrictions that often preclude credit risk or derivatives, they tend to rely on easy-to-implement term structure models to manage their high-quality government debt duration risk exposures.

[2] A rich literature provides accurate and tractable approximation schemes to easily estimate shadow rate models with constant volatility. In particular, Krippner (2011), Kim and Priebsch (2013), Priebsch (2013), and Wu and Xia (2016) provided early practical approximation schemes for the model in Black (1995) and analyzed the accuracy.

[3] Important examples include Kim and Singleton (2012), Krippner (2013), Priebsch (2013), Christensen and Rudebusch (2014), Bauer and Rudebusch (2016), and Wu and Xia (2016).

[4] See Diebold and Rudebusch (2012) for an excellent survey of DNS models. Bjork and Christensen (1999) and Filipovic (1999) show theoretically that the DNS model does not preclude all arbitrage opportunities. Krippner (2013) shows it can be seen as a low-order Taylor approximation of certain no-arbitrage Gaussian affine term structure models. Coroneo et al. (2011) find that estimated parameters of DNS models are not statistically different from estimates of corresponding no-arbitrage models.

[5] Additionally, the conditional mean $E_t[X_{t+1}]$ at time t may not be completely spanned by X_t. This allows for notions of unspanned risks introduced by Duffee (2011), Feunou and Fontaine (2014), and Joslin et al. (2014). Likewise, the conditional variances $V_t[X_{t+1}]$ can be constant, as in standard Gaussian DTSMs; can depend on X_t itself, as in the $A_M(N)$ models of Dai and Singleton (2000); can depend on the history $\{X_t, X_{t-1}, \ldots\}$, in the spirit of the ARCH literature pioneered by Engle (1982); or can depend on the history of other risk factors, capturing the notion of unspanned stochastic volatility in Collin-Dufresne and Goldstein (2002), Li and Zhao (2006), and Joslin (2018).

[6] The scaling parameter θ of the B_{nd} model does not play a role in shaping the volatility dynamics of states. However, we require that $\theta w(\frac{0-\sigma_s}{\theta}) \geq 0.0005$ and $\theta w(\frac{0+\sigma_s}{\theta}) \leq 0.0050$. These inequalities constrain the variation of the shadow rate only locally, near the lower bound. The NA restrictions tie S_n at every maturity with the underlying volatility parameters.

[7] The linear-rational model of Filipovic et al. (2017), used as the data-generating process in Section 3, is capable of capturing many desirable features of the data. However, its estimation is significantly more involved and requires auxiliary information from derivatives markets, which they use in their paper. Hence, given our focus on tractability, we do not report empirical results based on the linear-rational model. To maintain a level playing field, we include only those models for which standard estimation is based on yield data only and whose global convergence can be obtained in a straightforward and robust manner.

References

Andreasen MM, Christensen BJ (2015) The SR approach: A new estimation procedure for non-linear and non-Gaussian dynamic term structure models. *J. Econometrics* 184:420–451.

Banerjee S, Graveline JJ (2013) The cost of short-selling liquid securities. *J. Finance* 68:637–664. Q:14

Bansal R, Tauchen G, Zhou H (2004) Regime shifts, risk premiums in the term structure, and the business cycle. *J. Bus. Econom. Statist.* 22:396–409.

Bauer M, Rudebusch G (2016) Monetary policy expectations at the zero lower bound. *J. Money Credit Banking* 48:1439–1465.

Bauer MD, Rudebusch GD (2020) Interest rates under falling stars. *Amer. Econom. Rev.* 110:1316–1354.

Bjork T, Christensen BJ (1999) Interest rate dynamics and consistent forward rates curves. *Math. Finance* 9:323–348.

Black F (1995) Interest rates as options. *J. Finance* 50:1371–1376.

Campbell J, Shiller R (1991) Yield spreads and interest rate movements: A bird's eye view. *Rev. Econom. Stud.* 58:495–514.

Christensen JHE, Rudebusch GD (2014) Estimating shadow-rate term structure models with near-zero yields. *J. Financial Econom.* 13:226–259.

Christensen JH, Diebold FX, Rudebusch GD (2011) The affine arbitrage-free class of Nelson-Siegel term structure models. *J. Econometrics* 164:4–20.

Cieslak A, Povala P (2016) Information in the term structure of yield curve volatility. *J. Finance* 71:1393–1436.

Collin-Dufresne P, Goldstein RS (2002) Do bonds span the fixed income markets? Theory and evidence for 'unspanned' stochastic volatility. *J. Finance* 57:1685–1730.

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

Creal DD, Wu JC (2015) Estimation of affine term structure models with spanned or unspanned stochastic volatility. *J. Econometrics* 185:60–81.

Dai Q, Singleton K (2000) Specification analysis of affine term structure models. *J. Finance* 55:1943–1978.

Dai Q, Singleton K (2002) Expectations puzzles, time-varying risk premia, and affine models of the term structure. *J. Financial Econom.* 63:415–441.

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

Diebold FX, Rudebusch GD (2012) *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach* (Princeton University Press, Princeton, NJ).

Duffee G (2010) Sharpe ratios in term structure models. Working paper, Johns Hopkins University, Baltimore. Q:15

Duffee G (2011) Information in (and not in) the term structure. *Rev. Financial Stud.* 24:2895–2934.

Duffie D (1996) Special repo rates. *J. Finance* 51:493–526.

Engle R (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50:987–1008.

Engle R, Rousselet G, Siriwardane E (2017) Scenario generation for long run interest rate risk assessment. *J. Econometrics* 201:333–347.

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

Filipovic D (1999) A note on the Nelson-Siegel family. *Math. Finance* 9:349–359.

Filipovic D, Larsson M, Trolle AB (2017) Linear-rational term structure models. *J. Finance* 72:655–704.

Goliński A, Zaffaroni P (2016) Long memory affine term structure models. *J. Econometrics* 191:33–56.

Gurkanyak R, Sack B, Wright J (2007) The U.S. Treasury yield curve: 1961 to the present. *J. Monetary Econom.* 54:2291–2304.

Jacobs K, Karoui L (2009) Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets. *J. Financial Econom.* 91:288–318.

Joslin S (2018) Can unspanned stochastic volatility models explain the cross section of bond volatilities? *Management Sci.* 64:1707–1726.

Joslin S, Le A (2021) Interest rate volatility and no-arbitrage affine term structure models. *Management Sci.*, ePub ahead of print February 23, https://doi.org/10.1287/mnsc.2020.3858.

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

Joslin S, Priebsch M, Singleton KJ (2014) Risk premiums in dynamic term structure models with unspanned macro risks. *J. Finance* 69:1197–1233.

Joslin S, Singleton K, Zhu H (2011) A new perspective on Gaussian dynamic term structure models. *Rev. Financial Stud.* 24:926–970.

Kim D, Priebsch MA (2013) The U.S. yield curve at the zero lower bound: Gaussian shadow-rate vs. affine-Gaussian term structure models. Working paper, Federal Reserve Bank of New York, New York.

Kim DH, Singleton KJ (2012) Term structure models and the zero bound: An empirical investigation of Japanese yields. *J. Econometrics* 170:32–49.

Kozicki S, Tinsley PA (2001) Shifting endpoints in the term structure of interest rates. *J. Monetary Econom.* 47:613–652.

Krippner L (2011) Modifying Gaussian term structure models when interest rates are near the zero lower bound. Working paper, Reserve Bank of New Zealand, Wellington, New Zealand.

Krippner L (2013) A theoretical foundation for the Nelson-Siegel class of yield curve models. *J. Appl. Econometrics* 30:97–118.

Krishnamurthy A (2002) The bond/old-bond spread. *J. Financial Econom.* 66:463–506.

Levy H (1992) Stochastic dominance and expected utility: Survey and analysis. *Management Sci.* 38:555–593.

Li H, Zhao F (2006) Unspanned stochastic volatility: Evidence from hedging interest rate derivatives. *J. Finance* 61:341–378.

Monfort A, Pegoraro F, Renne J-P, Roussellet G (2017) Staying at zero with affine processes: An application to term structure modelling. *J. Econometrics* 201:348–366.

Noureldin D, Shephard N, Sheppard K (2011) Multivariate high-frequency-based volatility (HEAVY) models. *J. Appl. Econometrics* 27:907–933.

Priebsch M (2013) Computing arbitrage-free yields in multi-factor Gaussian shadow-rate term structure models. Technical report, Finance and Economics Discussion Series, The Federal Reserve Board, Washington, DC.

Rothschild M, Stiglitz J (1970) Increasing risk. I. A definition. *J. Econom. Theory* 2:225–243.

Swanson ET, Williams JC (2014) Measuring the effect of the zero lower bound on medium- and longer-term interest rates. *Amer. Econom. Rev.* 104:3154–3185.

Vayanos D, Weill P-O (2008) A search-based theory of the on-the-run phenomenon. *J. Finance* 63:1361–1398.

Wu JC, Xia FD (2016) Measuring the macroeconomic impact of monetary policy at the zero lower bound. *J. Money Credit Banking* 48:253–291.

AUTHOR QUERIES

DATE 11/26/2021

JOB NAME MNSC

ARTICLE 20214214

QUERIES FOR AUTHOR Feunou et al.

THIS QUERY FORM MUST BE RETURNED WITH ALL PROOFS FOR CORRECTIONS

Q: 1_Please provide all funding information, if applicable, including the names of the institutions as well as any grant numbers associated with the funding, or confirm that no funding was received.

Q: 2_Your color figures appear in color in your page proof to simulate their presentation in color online, but they will be converted to grayscale in the print version of your article because you have not requested color processing for print. Please (a) check your figures to confirm that they will be sufficiently clear in grayscale presentation in the print version of your article and (b) modify your figure legends as necessary to remove any references to specific colors in the figure.

Q: 3_Please confirm that the article title, author names, affiliations, and email addresses are set correctly. If applicable, please provide author ORCID numbers.

Q: 4_Please provide complete details for affiliation “a”.

Q: 5_Please provide department or section in affiliation a; city, state, and country in affiliation a; and postal code in affiliations a and b.

Q: 6_Please confirm that keywords are correct as set.

Q: 7_Please confirm that heading levels are correct as set.

Q: 8_Please verify that all displayed equations and in-text math notations are set correctly.

Q: 9_Per INFORMS style, built-up fractions included within paragraph text should be converted to their in-line form, i.e., with a solidus rule, as in $[a + (b/x)]^{1/2}$, whenever possible. Long or complex in-line formulas (within paragraph text) should be set as display equations. Please review all in-text equations and modify as necessary to be represented accurately as either display or in-line equations.

Q: 10_Per INFORMS style, all variables should be italic and all vectors should be bold. Please confirm that all terms have been formatted properly throughout.

Q: 11_Per style, sentences that begin with mathematical notation should be avoided. Please rewrite sentences that begin with mathematical notation throughout the text.

Q: 12_Please define the abbreviations RV, i.i.d., GLS, VAR(1), LPY, DTSM, and ARCH in the text.

Q: 13_Please spell out PBC, CIRANO, FRBSB, and MFA in the acknowledgments.

Q: 14_Please provide issue numbers for all journal references.

Q: 15_Please update all working paper references.

Q: 16_Please check that all tables and figures, including their titles and notes (which may contain edits) are set correctly.

Q: 17_Please spell out BEKK in Figures 2-7 legends.

Q: 18_Please spell out BEKK in Tables 2-5 legends.