Tractable No-Dominance Term-Structure Models and the Zero Lower Bound

Bruno Feunou\textsuperscript{1}, Jean-Sébastien Fontaine\textsuperscript{1}, Anh Le\textsuperscript{2} and Christian Lundblad\textsuperscript{3}

\textsuperscript{1}Bank of Canada
\textsuperscript{2}Smeal College of Business, Pennsylvania State University
\textsuperscript{3}Kenan–Flagler Business School, University of North Carolina, Chapel Hill

This draft: November 21, 2018

Abstract

We greatly expand the space of tractable dynamic term-structure models. Key to our approach is a direct specification of closed-form bond pricing functions, which is a class of generalized Nelson-Siegel (NS) models. Models in this class guarantee the absence of dominant trading strategies, consistent with the perception that NS models are “nearly” free of arbitrage. Term structure forecasting remains vexing for researchers, investment managers, and central banks in an environment characterized by time-varying volatility and a zero lower bound (ZLB). We design a special case of our model adapted to these two critical empirical features. We show, through simulations, that a case with time-varying volatilities and correlations significantly improves forecast accuracy for bond returns, volatilities, and Sharpe ratios. The case with constant variance performs comparably to existing ZLB models.

Keywords: Term-Structure, Zero Lower Bound, No-Arbitrage, No-Dominance
JEL codes: G12

We thank for comments and suggestions Alison Arnott, Peter Christoffersen, Greg Duffee, Antonio Diez de los Rios, Peter Feldhutter, Jing-zhi Huang, Kris Jacobs, Richard Luger, and seminar and conference 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, and the 2014 Tremblant Risk Conference, 2015 CIRANO finance workshop, 2015 Bank of Canada-FRBSB Fixed Income Conference, 2017 MFA meetings. Corresponding author: jsfontaine@bankofcanada.ca.
1 Introduction

The absence of arbitrage is a cornerstone of realistic bond-pricing models. Managers of bond portfolios should discard risk and return estimates whenever the underlying yield forecasts offer feasible arbitrage opportunities. Similarly, its risk manager should attribute zero probability to configurations of yield forecasts that offer feasible arbitrage opportunities. Unfortunately, have been long acknowledged that realistic term structure models incorporating flexible volatility dynamics can be challenging to implement. Further, in the face of persistent unconventional monetary policy holding interest rates close to zero or negative in the post-crisis years, realistic bond pricing models have become particularly challenging to implement.

We take a new look at the construction of bond prices, taking a step back from the standard paradigm. Our approach constructs bond prices directly, guaranteeing tractability. It uses a recursive construction similar to Nelson and Siegel (1987) tying together bond prices (and forward rates) at successive maturities. It is general enough to arm researchers with the freedom to explore the realm of realistic yield dynamics and to improve the fit of important stylized facts. Yet, the specifications that we propose remain robust and convenient to implement.

Our construction builds on and expands the dynamic Nelson-Siegel (DNS) model in Diebold and Li (2006), which, strictly speaking, does not guarantee the no arbitrage (NA) condition. The connection with the DNS model, a special case of our model, is one important thread to understand our results. Just like the DNS model, our construction does not guarantee the absence of arbitrage. But, despite lacking the NA properties, the DNS model performs as well as its NA counterparts along important dimensions and thus remains popular among a remarkable range of users, including practitioners, researchers, and foreign reserves managers.

One reason why the DNS model is popular, put forth in Christensen, Diebold, and Rudebusch (2010), is that the DNS model is “almost” arbitrage-free. We refine this argument and show that our entire family of models, including the DNS model, are free of dominant trading strategies (Rothschild and Stiglitz, 1970; Levy, 1992). To highlight this important property, we refer to our models as no-dominance (ND) models. We show that there always

---

1Bjork and Christensen (1999) and Filipovic (1999) show the DNS model is not consistent with NA in a frictionless world.

2The IMF reports that foreign reserves managers oversee $6,280 billion in USD assets. This underlines the practical importance of term structure forecasting models that are realistic and easy to implement on a day-to-day basis to manage their portfolios.
exists a sequence of NA models that converge in the limit to each member of our model family. In this sense, our models lie right in the outer border of the closed set of NA models. In a more practical note, we show that any remaining arbitrage opportunity that our framework could permit must reduce to a self-financing strategy. This means that this type of opportunity is fragile to short-selling costs and may have little practical implications. Overall, the insight that DNS models can be used with confidence in practical applications generalizes to our new family of tractable term structure models. This is especially important and reassuring given the popularity and the broad range of DNS models used in practice (see e.g., the review in Diebold and Rudebusch (2012)).

We specialize our ND framework to a class of models with realistic features. To match yield dynamics, we specify a VAR(1) for the yield factors, where the associated innovations follow a conditional Gaussian distribution. To match yield volatility dynamics, we specify a simple but flexible BEKK model for the time-varying covariances among these innovations. To keep yields positive, we choose a short-rate equation in the spirit of Black (1995). These specification choices make it easy to compare our results with existing research incorporating some (though not all) of these features in NA models.

Our first set of results focus on the comparison between NA and ND implementations of positive yield models in the spirit of Black (1995). For this comparison, we focus on the construction with the same constant variance for the states dynamics, so that the no-arbitrage restriction is the only distinction between the NA and ND models. We ask if the observation that the linear DNS model is “almost” arbitrage-free generalizes to the popular non-linear Black model? Pricing errors are essentially identical between the NA and ND Black models across bond maturities, in the whole sample and in sub-samples. Both models offer the same improvements relative to a benchmark NA Gaussian model in the sub-sample when the short rate stays at the lower bound.

However, we do find differences in the estimated state dynamics. In a realistic simulation environment in which the lower bound is occasionally binding and with time-varying yield volatility, we find that the NA model under-performs when matching conditional Sharpe ratios. The Sharpe ratio is an important statistic to differentiate models on an economic ground. The ratio combines expected returns and conditional volatilities in such a way that is useful to investors and researchers. We find that the NA Black model does not match the yields and volatility dynamics as well as the the ND model, especially in samples when yields are close to the lower bound. Our results suggest that the no-arbitrage restrictions introduce

---

3The lower bound can be any constant, and not necessarily zero.
a tension between fitting yields and fitting volatility in the case of the popular no-arbitrage Black models.

Our second set of results focus on the comparison between ND implementations of the Black model with constant state volatility and its counter-part with time-varying volatility. Again, there is little difference to separate these models in terms of pricing performance. If anything, time-varying volatility improves the pricing for yields with the shortest maturities. Again, we find differences between the estimated state dynamics. However, the time-varying volatility model offers a better fit of the Sharpe ratios than the constant volatility model in samples away from the lower bounds. The gains are significant and uniform across investment horizons and across bond maturities. Of course, part of the gains are due to the much improved fit of the yield volatility dynamics. But we also find significant gains in fitting yield forecasts. The source of this gain is more subtle. The inclusion of time-varying volatility improves the estimation efficiency for the conditional mean parameters. In samples close to the lower bound, the constant volatility model offers a better fit of the Sharpe ratio for short maturities and the time-varying volatility model offers a better fit for other maturities. In spite of these differences, both models are better than the NA Black model.

The simulation environment we use in our analysis is a realistic set-up that is useful to compare bond pricing models. The data generating process features an upward sloping average yield curve, downward sloping average yield volatility, predictable bond excess returns, time-varying volatility, and it includes episodes at the lower bound toward the end of each sample. The main advantage of our simulation exercise is that we can use true predictive moments of the simulated yields to assess and compare the moments implied from the models that we estimate. This is unlike evaluation based on historical data where true moments are unknown and conditional moments, such as conditional Sharpe ratios and forecasts are estimated very imprecisely. In addition, the assessment and comparison are still valid in specific sub-samples around the lower bound, without regard to the statistical uncertainties that arise in small historical sub-samples.

We pay great attention to the speed and robustness of the estimation procedure for Black models under consideration (both NA and ND). Making estimation fast and robust is essential for tractability and a necessary step in using a rich simulation environment for model evaluation. To achieve this, we combine several techniques known in the literature. First, following Joslin, Singleton, and Zhu (2011), we invert the pricing equation to recover observable pricing factors. In the case of Black models, the inversion recovers a shadow short rate and shadow forward rates that we use to construct observable portfolios as risk
factors with linear dynamics. Second, given these portfolios of shadow forwards, we derive the conditional likelihood of the data. Finally, we also preserve from JSZ the appealing feature that VAR(1) conditional mean parameters can be compartmentalized in the likelihood and, therefore, can be estimated using least squares in a separate step. We show how to analytically concentrate out these parameters from the likelihood, which is feasible in our case because the mean and volatility parameters are not mingled together in the specifications that we consider.

A substantial body of existing work use NA Black style models to study bond yields and bond risk premium. Without exception, this research stream relies on a specification with constant variances and correlations for the risk factors. But studying the risk premium and volatility jointly has become even more important given how pervasive the bond yield boundary has become. As Bauer and Rudebusch (2013) noted, the boundary implies that bond volatility plays a central role in many important model forecasts, because the boundary introduces an asymmetry in the relationship between bond yields and bond risk factors (the state variables). Forecasts of future yields increase with the risk factor volatilities even if the mean dynamics remain unchanged. Some recent work incorporates time-varying volatility but ignore the yield boundary. Cieslak and Povala (2015) recognize the issue and end their sample in 2010, before long-term yields become less sensitive to news, burdened by the tightness of the boundary (See e.g., Bauer and Rudebusch 2013; Swanson and Williams 2014). Similarly Creal and Wu (2015) end their sample in mid-2012.

Two important contributions combine time-varying volatility in a model consistent with the boundary. Monfort, Pegoraro, Renne, and Roussellet (2017) extend the affine no-arbitrage framework, allowing for the combination of stochastic volatility and the lowe bound for interest rates. Filipovic, Larsson, and Trolle (2017) introduce the class of linear-rational NA model featuring several realistic features. Each of these models relies on filtering and estimation procedures that can be challenging to implement in practical day-to-day applications. We propose a family of tractable models that can be easily implemented for a rich set of dynamics specifications.

Section 2 provides a generic construction of our family of tractable dynamic term structure models. In this section, we show theoretically that our bond prices can be made arbitrarily close to the absence of arbitrage. Section 3 develops a special case with time-varying volatilities, in which yields are consistent with the existence of a zero lower bound. Section 4

---

proposes a robust and computationally efficient estimation method for this class of models. Section 5 evaluates, in a simulated environment, the performance of ZLB models over a number of dimensions. Section 6 provides an empirical illustration of ZLB models using U.S. Treasury yields. Section 7 concludes.

2 Tractable Dynamic Term-Structure Models

This section introduces a new family of dynamic term structure models in which bond prices are specified directly. This direct approach, while flexible and tractable, guarantees that bond yields are available in closed form with only minimal assumptions about the risk factor dynamics. The proposed family of models includes as one of its members the yield curve model of Nelson and Siegel (1987) and the dynamic DNS version of Diebold and Li (2006). There is ample evidence in practical applications that models in the spirit of DNS are essentially indistinguishable from their no arbitrage counterparts (see, for example, Christensen, Diebold, and Rudebusch 2010). We analyze on theoretical grounds why this perceived feature is likely to generalize to our proposed family of dynamic term structure models.

2.1 Bond Prices

Consider $J$ zero-coupon bonds maturing in $n = 1, 2, \ldots, J$ periods with a face value of one dollar. Let $P_n(X_t)$ denote the price of the $n$-period bond, where $X_t$ is a state vector with support $X \subseteq \mathbb{R}^K$. Assumption 1 provides a direct specification for $P_n(X_t)$.

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

\[
P_0(X_t) \equiv 1, \tag{1}
\]

\[
P_n(X_t) = P_{n-1}(g(X_t)) \times \exp(-m(X_t)), \tag{2}
\]

for functions $m(\cdot)$ and $g(\cdot)$ such that $m(X) \in M \subseteq \mathbb{R}$ and $g(X) \in X$ for every $X_t \in X$.

The recursive structure of bond prices is the distinctive feature of Assumption 1 allowing us to derive the theoretical properties of this family of models below. For now, we note that, since zero-coupon bond prices are available in closed form for all maturities, all forward rates and bond yields are available in closed form.
Proposition 1. Assumption 1 implies that the one-period forward rate applicable to a one-period loan \( n \) periods in the future is given by

\[
f_{n,t} = m(g^n(X_t)),
\]

where \( g^n(X_t) = g(g(\ldots g(X_t))) \) and \( g^0(X_t) = X_t \) by convention.

Proof. See Appendix A.1.1. \( \square \)

Equation (1) corresponds to our assumption that bonds are redeemed at their face value of one dollar. The functions \( m(\cdot) \) and \( g(\cdot) \) in Equation (2) are the key primitives in our construction of bond prices. The case \( n = 1 \) in Equation (2) leads to the one-period yield \( y_{1,t} = -\log(P_t(X_t)) = m(X_t) \), which tells us that the \( m(X_t) \) is the one-period interest rate as a function of the states \( X_t \). The case \( n = 2 \) in Equation (2) leads to \( P_2(X_t) = P_1(g(X_t)) \times \exp(-m(X_t)) = \exp(-m(g(X_t)) - m(X_t)) \), which tells us that \( g(\cdot) \) embodies how the price of one bond is discounted back to its present value.

2.2 State Dynamics

Our family of models places minimal restrictions on the state dynamics and affords researchers the flexibility to study a rich set of time series dynamics. Assumption 2 clarifies the set of time series dynamics admissible to complete the construction of our models.

Assumption 2. The time-series dynamics of \( X_t \) admit \( X \) as support.

It is essential that the support of \( X_t \) under the time-series measure coincides with the support \( X \) for bond prices. This requirement is analogous to the requirement for no-arbitrage models that the time-series and risk-neutral measures must be equivalent measures. Roughly speaking, this requires that any event that is priced can happen with a non-zero probability.

Assumption 2 implies very mild restrictions on state dynamics, affording researchers the flexibility to study rather rich time series dynamics. For instance, 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 Joslin, Priebsch, and Singleton (2013), Duffee (2011) and Feunou and Fontaine (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); it can depend on the history \( \{X_t, X_{t-1}, \ldots\} \), in the spirit of the ARCH literature pioneered by Engle (1982); or it 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 (2014).

Nesting the Dynamic Nelson-Siegel Model

The direct approach to constructing bond prices in Assumptions 1-2 nests a long-standing practical tradition using DNS model to fit bond yields.

Proposition 2. Following Diebold and Li (2006), suppose that \( X_t \in \mathbb{R}^3 \) gaussian autonomous AR(1) specifications. Take \( m(\cdot) \) and \( g(\cdot) \) given by:

\[
\begin{align*}
  m(X_t) &= \begin{bmatrix} 1 & \frac{1-e^{-\lambda}}{\lambda} & \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{bmatrix} X_t, \\
  g(X_t) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda} & \lambda e^{-\lambda} \\ 0 & 0 & e^{-\lambda} \end{bmatrix} X_t,
\end{align*}
\]

then bond yields follow the DNS models in Diebold and Li (2006). In particular, Assumption 1 generates yields to maturity with NS loadings.

Proof. Direct computation of \( y_{t,n} = \frac{1}{n} \sum_{i=0}^{n-1} m(g^{\circ i}) \) yields the result. \( \square \)

Bjork and Christensen (1999) and Filipovic (1999) show theoretically that the DNS model, does not preclude all arbitrage opportunities. Despite this, DNS models remain popular among a remarkable range of users, including practitioners, researchers, and regulators. To a large extent, this popularity is due to the ease of using DNS models and the well-known performances in forecasting (see Diebold and Li (2006)). Another reason for this popularity, as put forth in Christensen, Diebold, and Rudebusch (2010), is that the DNS model is “almost” arbitrage-free. Indeed, Krippner (2013) shows that the DNS model can be seen as a low-order Taylor approximation of certain no-arbitrage Gaussian affine term-structure models. Empirically, Coroneo, Nyholm, and Vidova-Koleva (2011) find that estimated parameters of DNS models show no significant statistical differences compared to parameters of corresponding no-arbitrage term structure models.\(^5\)

\(^5\)See Diebold and Rudebusch (2012) for an excellent evaluation of the DNS model.
2.3 No-Dominance Properties of Bond Prices

While dynamic bond pricing models with NS loading arise as a special case in the class of models based on Assumption 1, the choices of \( m(\cdot) \) and \( g(\cdot) \) to construct bond prices as in Assumption 1 are much more general. One important question, then, is whether our proposed models also produce bond prices that are “almost” arbitrage-free. This section addresses this question and clarifies theoretically the economic sense in which our models are considered “almost” free of arbitrage.

Recall that the no-arbitrage (NA) condition holds if and only if any portfolio with strictly non-negative payoffs admits a strictly positive price.\(^6\) A related, but weaker, requirement is the no-dominance (ND) condition. The ND condition holds if and only if any portfolio with strictly positive payoffs in all states admits a strictly positive price. In other words, for any given trading strategy, there must not exist a strictly dominant trading strategy which costs the same or less to set up but delivers strictly better payoffs in every state of nature.

Checking for the absence of arbitrage is usually a difficult problem. However, building on the simple structure in Assumption 1, we show that our proposed bond prices rule out strictly dominant trading strategies.

**Theorem 1.** *Assumption 1 guarantees the absence of strictly dominant trading strategies in the bond markets.*

**Proof.** See Appendix A.1.3. □

Dominant trading strategies consider bond portfolios with strictly positive payoffs. Consider now a portfolio that has zero payoff with (strictly) positive probability, but positive cash flows otherwise. Note that the NA condition requires that the price of this portfolio is strictly positive. We show in Theorem 2 that our proposed construction of bond prices in Assumption 1 guarantees that portfolios like this one cannot admit strictly negative prices.

**Theorem 2.** *Assumption 1 ensures that bond portfolios with strictly non-negative payoffs cannot admit strictly negative prices.*

**Proof.** See Appendix A.1.4. □

Theorem 2 clarifies the sense in which our bond price construction is “almost” arbitrage free. What is setting NA and ND apart is the set of zero-cost portfolios with zero or positive payoffs.

\(^6\)To be precise, any portfolio with positive cash flows for a strictly positive probability and zero cash flows otherwise must command a strictly positive price.
payoffs (both with positive probability). For portfolios with strictly non-negative payoffs, the absence of arbitrage restricts prices to be on the positive half of the real line, excluding the origin. The construction of bond prices in Assumption 1 allows for 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 1.

Figure 1: Prices of portfolios with strictly non-negative payoffs.

Figure 1 suggests that the distance between our models and the no-arbitrage paradigm is infinitesimal. In this sense, our construction of bond prices is almost free of arbitrage. Does this difference matter? We offer a theoretical and a practical answer to this question. From a theoretical perspective, Theorem 3 formalizes the intuition that this distance is small, showing that one can always find no-arbitrage models that are arbitrarily close to the model of bond prices in Assumption 1.

**Theorem 3.** There exists a sequence of no-arbitrage models that converge to the model of bond prices constructed in Assumption 1.

*Proof.* See Appendix A.1.5.

Figure 1 also tells us that any remaining arbitrage opportunities must involve a self-financing portfolio with strictly non-negative payoffs in the future and with strictly zero set-up costs. From a practical perspective, a self-financing portfolio must involve short-selling bonds and, since short positions imply some non-zero costs to set up and to maintain, the specific portfolio in Figure 1 may not represent an arbitrage opportunity in a practical sense for sophisticated long-short investors.\(^7\) Summing up, the results in Theorems 1-3 offer a constructive perspective to understand the empirical success and robustness of the DNS

---

\(^7\) A substantial literature has documented the costs to establish and maintain short Treasury bond positions, even for the most liquid issues. See e.g., Duffie, 1996; Krishnamurthy, 2002; Vayanos and Weill, 2008 and Banerjee and Graveline, 2013.
model. This perspective is general enough, suggesting that more flexible models may offer the same success as well as expand our ability to jointly fit time-varying volatility and the ZLB stylized facts.

2.4 Discussion

One common implementation of the no-arbitrage approach to bond pricing requires a pricing kernel $M_{t+1}/M_t$. The $n$-period bond prices are then computed based on the expectation $\int (M_{t+n}/M_t)d\mathbb{P}$ conditional on the relevant information set at time-$t$ and given the dynamics for $M_{t+n}$ under the historical measure $\mathbb{P}$. However, computing the integral in this conditional expectation is not analytically tractable for many interesting choices of $(M_t, \mathbb{P})$, in part because the kernel must be strictly positive almost surely. This contrasts with our approach for which bond prices can be computed in closed-form for the extremely wide range of historical dynamics consistent with Assumption 2.

The alternative implementation to no-arbitrage bond pricing specifies the dynamics for the short rate under the risk-neutral measure $\mathbb{Q}$. The $n$-period bond prices are then computed based on the conditional expectation $\int \exp(-\sum_{j=0}^{n-1} r_{t+j})d\mathbb{Q}$, given the dynamics for the one-period rate $r_t$. Again, this integral is not analytically tractable for many interesting choices of $(r_t, \mathbb{Q})$, including many cases where $r$ is strictly positive (or is bounded below). Again, this contrasts with our approach for which bond prices can be computed in closed-form for any short-rate function $m(\cdot)$.

As an example, Black (1995) uses $\max(0, r)$ to truncate the short rate and guarantee positive interest rates. However, the integral $\int \exp(-\sum_{j=0}^{n-1} r_{t+j})d\mathbb{Q}$ becomes analytically intractable because of the nonlinearity induced by the truncation. Various analytical approximation schemes have been proposed, but these schemes are limited to cases in which the state dynamics feature constant volatility. In our approach, setting $m(X) = \max(0, X)$ does not introduce any difficulty for computing bond prices while guaranteeing that interest rates and forward rates are positive.

By swapping the no-arbitrage requirement for the weaker no-dominance requirement, our proposed approach greatly expands the space of tractable dynamic term structure models that we can explore. In essence, we broaden the scope of tractable but accurate term structure models à la Nelson-Siegel. We show in Appendix A.2 how to formulate general linear and quadratic variants of our model. In Section 3, we specialize our framework to a specification where the choice of $m(\cdot)$ imposes a lower bound on interest rates at all maturities. All these
cases produce closed-form bond prices by construction. In addition, Section 3 evaluates the time-series dynamics with either constant or flexible time-varying volatility, which we can afford due to the minimal restrictions in Assumption 2 on the state dynamics.

3 A No-Dominance Model with a Zero Lower Bound

In the spirit of Black (1995), this section specializes the no-dominance model to a specification that imposes a zero lower bound on interest rates. For comparison purposes, this section also describes a no-arbitrage analogue of our model, which is a discrete-time version of the Black (1995) model. For both models, the two most important building blocks are (i) the pricing equation; and (ii) the state dynamics. We summarize each of these blocks below.

3.1 Pricing Equation

We derive the pricing equation for forward rates given the $N \times 1$ state vector $X_t$ in a special case with positive rates. Recall that Assumption 1 constructs the $n$-period zero coupon bond price recursively, from $P_{n-1}(X_t)$ to $P_n(X_t)$, using two primitive functions: (i) the short rate function $m(X_t)$ and (ii) the function $g(X_t)$. For simplicity, we adopt a linear choice for $g(\cdot)$:

$$g(X_t) = K X_t,$$

where $K$ is a $N \times N$ matrix. For the short rate function $m(\cdot)$, we let:

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

where $s_t = \delta_0 + \delta'_1 X_t$ and $\theta$ is a positive scalar. The function $w(x)$ takes positive values, it is strictly increasing and it converges to the 45 degree line ($y = x$) for large positive values of $s$. Applying Proposition 1 using the $m(\cdot)$ and $g(\cdot)$ functions given in equations (6-7) immediately reveals that the pricing equation for the $n$-period one-month forward rate is given by:

$$f_{n,t} = \theta w\left(\frac{\delta_0 + \delta'_1 K^n X_t}{\theta}\right),$$

which implies that the pricing equation for $n$-month zero yields is given by $\frac{1}{n} \sum_{0}^{n-1} f_{k,t}$.

The choice of the function $w(x)$ embeds a reduced-form interpretation that $s_t$ is the
shadow short rate as in a prototypical Black style model. For this, \( w(x) \) captures two key properties of the link between the short rate \( r_t \) and the shadow rate \( s_t \). First, the observed and shadow rates converge to each other for high values of the shadow rate. Indeed, it is apparent that the short rate converges to \( s_t \): \( r_t \approx \theta \frac{\sigma}{\bar{\sigma}} = s_t \) for large positive values of \( s_t \). Second, the short rate gets smaller but stays positive as \( s_t \) gets closer to zero or becomes negative, because \( w(x) \) is positive-valued.\(^8\)

Equation (8) suggests a similar interpretation linking the forward rate \( f_{n,t} \) and a shadow forward rate \( s_{n,t} = \delta_0 + \bar{\delta}_n K^n X_t \). Like the short rate, the observed and shadow forward rates converge to each other for large value of the shadow rate \( s_{n,t} \), but the observed forward rate remains positive for low or negative values of the shadow forward rate.

For implementation, we adopt the following choice for \( w(x) \):

\[
w(x) = x\Phi(x) + \phi(x),
\]

where \( \Phi(x) \) and \( \phi(x) \) are the cumulative probability function and the density function of a standard normal distribution, respectively. The transformation in equation (9) is our preferred choice because it resembles the short-rate approximation proposed by Wu and Xia (2016) for a no-arbitrage Black style model. Figure 2 short rate \( r_t \) for values of \( s_t \) between -2 percent and 5 percent and fixing \( \theta = 0.0070 \), drawing the familiar hockey stick pattern. It is visually apparent (and it can be shown analytically) that this choice for \( w(x) \) satisfies the requirement listed above; the short rate is (i) bounded below at zero (ii) increasing in the shadow rate \( s_t \), and (iii) that it converges to the shadow rate for large values of \( s_t \). This is by no means the only modeling choice for our short rate function that guarantees the positiveness of the short rate and a reduced-form interpretation of a shadow rate. We also considered a logistic transformation that delivers a similar set of properties and our subsequent empirical results seem robust to this choice.\(^9\)

\(^8\)We also impose additional requirements to ensure that near-zero values of the shadow rate \( s_t \), between five and fifty basis points, correspond to values of the short rate near the ZLB. In particular, we require: (i) \( \theta w \left( \frac{\bar{\sigma} \sigma}{\bar{\sigma} \sigma} \right) \geq 0.0005 \); and (ii) \( \theta w \left( \frac{\bar{\sigma} \sigma}{\bar{\sigma} \sigma} \right) \leq 0.0050 \) where \( \sigma_s \) is the volatility of the shadow short rate \( s_t \).

\(^9\)In this case, we use the function \( w(x) = \log(1 + \exp(x)) \), which is simple and guarantees that forward rates are positive. This choice is equivalent to choosing the logistic transformation for zero-coupon bond prices \( P(x) = 1/(1 + e^x) \in [0, 1] \).
Figure 2: Hockey Stick
The hockey stick transformation from the shadow rate $s_t$ to the short rate $r_t$ 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, respectively, the cumulative probability function and the density function of a standard normal distribution.

3.2 The State Dynamics

For simplicity, we consider a discrete-time VAR(1) dynamics for the $N \times 1$ state vector $X_t$:

$$X_{t+1} = K_0 + K_1 X_t + \sqrt{\Sigma_t} \varepsilon_{t+1},$$

(10)

where $\varepsilon_{t+1}$ is i.i.d. standard normal. For comparison with the existing literature on Black style models, we consider a model labeled $B_{nd}$ that has a constant covariance matrix $\Sigma_t = \Sigma_0$. The subscript $nd$ indicates the no-dominance property of the model. We also consider a model labeled $SV-B_{nd}$ where $\Sigma_t$ varies over time:

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

(11)

where $a$ and $b$ are both scalar and $\Delta X_t = X_t - X_{t-1}$. This construction is similar in spirit to a scalar BEKKK construction. The choice of the scalar coefficients $a$ and $b$ is for simplicity. Scalar coefficients have been shown to work successfully in practice. For instance, Noureldin, Shephard, and Sheppard (2011) finds that the scalar case performs well in the challenging case of S&P500 securities volatility. Without loss of generality, we use the first differences
$\Delta X_t \Delta X'_t$ to update the volatility. This is in the spirit of using realized volatility to update volatility (as in Noureldin, Shephard, and Sheppard (2011)). One key benefit is that the VAR parameters in equation (10) do not enter the volatility dynamics in equation (11). In the empirical results below, this specification can be estimated quickly and robustly and it will result in a good fit of yield volatility.

The flexibility in specifying the volatility dynamics is one of the main advantages of our proposed no-dominance framework. Many traditional affine DTSMs, in ensuring the positiveness of rates, adopt a constant variance structure or rely on standard square-root processes. The latter cases are burdened with a significant tension between the fit of the conditional means and of the conditional variances of yields (Dai and Singleton, 2002; Joslin and Le, 2018). Gaussian affine DTSMs relax this tension but allows for negative yields. Creal and Wu (2015) introduce tractable estimation of term structure models with spanned or unspanned stochastic volatility, but this approach allows for negative yields, since it relies on Gaussian yield factors.

3.3 The No-Arbitrage Benchmark

To facilitate comparison, this section describes a no-arbitrage analogue of our model, which is a discrete-time version of the Black (1995) model. We label this model as the $B_{na}$ model where the subscript $na$ indicates the no-arbitrage property of the model. As is typical in no-arbitrage term structure models, the state dynamics follow a first order Gaussian VAR under both the time series and risk-neutral measures with constant variance $\Sigma$. In the original model, the short rate equation is $r_t = (\delta_0 + \delta'_0 X_t)^+$, but this leads to a pricing equation that is not tractable. However, with the simple risk-neutral dynamics above, Wu and Xia (2016) obtain an accurate approximation for the $n$-period forward rate:

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

where the coefficients $A_n$, $B_n$ and $S_n$ are functions of the model’s risk-neutral parameters and where the function $w(x) = x\Phi(x) + \phi(x)$ is constructed using the cumulative probability function $\Phi(\cdot)$ and the density function $\phi(\cdot)$ of a standard normal distribution. Appendix A.3 develops the model and the approximation scheme proposed by Wu and Xia (2016) in detail.

It is instructive to compare the no-dominance pricing Equation 8 and the no-arbitrage pricing equation (12). To begin, the nonlinear transformation $w(\cdot)$ is identical in (8) and (12). Beyond the inherent nonlinearity, there are two key differences between these pricing
equations. First, the intercepts $A_n$ in the no-arbitrage model contain Jensen terms that vary with maturity and depend on the variance parameter $\Sigma$. By contrast, the no-dominance intercept $\delta_0$ are completely invariant to volatility specifications. This difference parallels the commonly observed difference between the constant terms in the Nelson-Siegel model and the Jensen terms in a standard no-arbitrage Gaussian model. Second, the scaling parameters $S_n$ in the no-arbitrage pricing equation (12) vary with maturity and depend strongly on the volatility of yields. For example, when $n = 1$, the term $S_n$ represents the one-month variance of the shadow rate. By contrast, the scaling parameter $\theta$ in the no-dominance equation (8) is invariant to the volatility specification.

4 Estimation and Implementation Strategy

Bond pricing models constructed following Black (1995) are non-linear. This adds further complications to the problem of estimating term structure model—a problem already rife with difficulties. An important insight in JSZ is that numerical implementation of linear models can be made easy by constructing the model with observable economic objects, as opposed to using pure latent variables. Following this insight, estimation of affine term structure model becomes much more convenient and robust if the model is rotated to an observationally equivalent representation in which the state vector are the first $N$ principal components (PCs) of bond yields/forwards. Black style models are not linear and do not belong to the affine class of models studied by JSZ. In this section, 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. Within this linear space, we can take advantage of the same insight in JSZ to achieve convenient and robust estimation of our model. First, following Joslin, Singleton, and Zhu (2011), we invert the pricing equation to recover observable pricing factors. In the case of the Black models, the inversion recovers a shadow short rate and shadow forward rates that we use to construct observable portfolios as risk factors with linear dynamics. Second, given these portfolios of shadow forwards as state variables, we develop an equivalent representation of Black model and provide the closed-form conditional likelihood of the data based on this representation. Finally, we also want to preserve from JSZ the appealing feature that VAR(1) conditional mean parameters can be compartmentalized in the likelihood and estimated using least squares in a separate step. We show how to analytically concentrate out these parameters from the likelihood, which is feasible in our case because the mean and volatility parameters
are not mingled in the specifications that we consider.

4.1 Portfolios of Shadow Forwards as Risk Factors

The no-dominance pricing equation (8):

\[ f_{n,t} = \theta w \left( \frac{\delta_0 + \delta'_t K^n X_t}{\theta} \right) \]

is strictly increasing and has a well-defined inverse \( f^{-1}(\cdot) \). Applying the inverse \( f^{-1}(\cdot) \) to both sides, we can write:

\[ s_{n,t} = \theta w^{-1} \left( \frac{f_{n,t}}{\theta} \right) = \delta_0 + \delta'_t K^n X_t. \]  

Equation (13)

We refer to the left hand side \( s_{n,t} \) of Equation (13) as the shadow forward implied in the Black style model. Shadow forwards are linear in the states, and thus can turn negative, just as the shadow short rate. In the case \( n = 0 \), the shadow forward is the same as the shadow rate \( s_{0,t} = s_t \).

Given a value of the scaling parameter \( \theta \), the portfolios of shadow forwards are directly observable within our model from Equation 13. Now consider \( J \) shadow forwards \( s_{n,t} \) with maturities \( n = n_1, n_2, \ldots, n_J \) stacked into a \( J \times 1 \) vector \( \bar{s}_t \):

\[ \bar{s}_t = A_X + B_X X_t \]  

Equation (14)

where \( A_X \) and \( B_X \) stack the intercepts and loadings of Equation 13, respectively. Equation (14) is linear. Therefore, we can apply the insight of JSZ and rotate the latent state vector \( X_t \) into linear combinations of shadow forwards. Specifically, we construct \( N \) portfolios of shadow forwards \( \mathcal{P}_t = W \bar{s}_t \) for a given \( N \times J \) matrix \( W \). We can recover the latent state \( X_t \) as a linear function of the portfolio \( \mathcal{P}_t \),

\[ X_t = (W B_X)^{-1}(\mathcal{P}_t - W A_X), \]

which can be substituted into (14) to obtain a pricing equation for the shadow forwards in terms of \( \mathcal{P}_t \):

\[ \bar{s}_t = A_P + B_P \mathcal{P}_t, \]  

Equation (15)
where $B_P = B_X(WB_X)^{-1}$ and $A_P = A_X - B_P(WA_X)$. Equipped with the shadow forwards, we can evaluate the actual forwards simply by $f_{n,t} = \theta w \left( \frac{s_{n,t}}{\theta} \right)$, which only involves the parameters $\theta, \delta_0, \delta_1$ and $K$ governing bond pricing.

4.2 Identification

The no-dominance model that we propose in Section 3 has the parameters $\theta, \delta_0, \delta_1$ and $K$ governing bond pricing in addition to the parameters $K_0, K_1, \Sigma_0, a,$ and $b$ that are responsible for the time series dynamics. Not all of the parameters are econometrically identified, because the state variables are latent. We adopt the JSZ canonical form whereby all identification assumptions are implemented on the pricing side. Specifically, we assume that $\delta_1 = \iota$ is a vector of ones and $K$ has an ordered Jordan form. The JSZ canonical form leaves the time series dynamics completely unconstrained. Therefore, this identification strategy is applicable to the rich set of time-series dynamics satisfying Assumption 2. This identification approach will also play an important role to construct the likelihood and to analytically concentrate some parameters in the estimation. Based on the insights of JSZ, we anticipate that estimation of the $\mathcal{P}$-representation will be robust.

4.3 Conditional Likelihood Construction

It’s easy to see that $\mathcal{P}_t$ will inherit the VAR(1) structure as well as the scalar BEKK volatility specification from the $X$-dynamics, since the correspondence between $X_t$ and $\mathcal{P}_t$ is linear. We have:

$$\mathcal{P}_{t+1} = K_{0,\mathcal{P}} + K_{1,\mathcal{P}} \mathcal{P}_t + \sqrt{\Sigma_{t,\mathcal{P}}} \varepsilon_{t+1}, \quad (16)$$

$$\Sigma_{t,\mathcal{P}} = \Sigma_{0,\mathcal{P}} + a \Sigma_{t,\mathcal{P}} + b \Delta \mathcal{P}_t \Delta \mathcal{P}'_t. \quad (17)$$

Therefore, the primitive parameters of our model governing the dynamics of $\mathcal{P}$ are $K_{0,\mathcal{P}}, K_{1,\mathcal{P}}, \Sigma_{0,\mathcal{P}}, a,$ and $b$. Importantly, due to the JSZ normalization, parameters of the $\mathcal{P}$-dynamics are unconstrained, since parameter of the $X$-dynamics are unconstrained. The parameters $\theta, \delta_0$ and $K$ governing the pricing equation remain unchanged.

We estimate our models by maximizing the log-likelihood of the observed forwards.10

---

10Recent advances provide alternative estimation methods. See, for example, Joslin, Singleton, and Zhu (2011), Joslin, Le, and Singleton (2013), Hamilton and Wu (2011), Adrian, Crump, and Moench (2013), and Diez de los Rios (2015).
That is, we need to compute for each $t$:

$$
P(f_t^o | I_{t-1})
$$

where $f_t^o$ denotes the $J \times 1$ vector $(f_{n_1,t}^o, f_{n_2,t}^o, \ldots, f_{n_J,t}^o)'$ of observed forwards and $I_t$ denotes the information set generated by $f_t^o$ up to time $t$. The superscript $^o$ differentiates observed quantities from their theoretical constructs. The overall likelihood is simply obtained by $\prod_t P(f_t^o | I_{t-1})$. We construct this likelihood as follows.

We make the assumption that the $N$ portfolios of shadow forwards $P_t^o$ are priced without error and identical to their model counterparts at each point in time: $P_t^o = W \bar{s}_t \equiv W \bar{s}_t = P_t$. Next, we assume that $J - N$ combinations of forwards $W_e f_t$ are priced with i.i.d. errors:

$$
W_e (f_t^o - f_t) \sim N(0, \sigma_e^2 I_{J-N}).
$$

(18)

given some $(J - N) \times J$ loading matrix $W_e$. These two assumptions are similar to case C in JSZ where the likelihood of coupon bonds is used to estimate the model using $N$ combinations of zero-coupon yields measured without errors, which are a non-linear transformation of the coupon bond yields. The analogy is that here the shadow forwards are non-linear transformation of the forwards.

Combining these two measurement assumptions, we can write the one-step-ahead conditional likelihood of forwards $f_t^o$:

$$
P(f_t^o | I_{t-1}) = P(W_e f_t^o | P_t, I_{t-1}) \times P(P_t | I_{t-1}) \times \left| \frac{\partial h(f_t^o)}{\partial f_t^o} \right|,
$$

(19)

which have three components on the right hand side.\textsuperscript{11} The first component $P(W_e f_t^o | P_t, I_{t-1})$ captures the cross-sectional fit of the model: its ability to explain the observed forwards $W_e f_t^o$ based on $N$ portfolios of shadow forwards $P_t$ observed contemporaneously. This component can be computed easily using the distribution of pricing errors assumed in (18). The second component $P(P_t | I_{t-1})$ captures the time series fit of the model: the predictive density of observing $P_t$ given the information set one period earlier. This component can also be derived in a straightforward manner, since $P_t$ is measured without errors and governed by the conditional Gaussian VAR(1) dynamics in (16-17). The final term is a Jacobian adjustment to

\textsuperscript{11}Except for the SV-B_n model, it is straightforward to derive the model-implied likelihood of the data for the more general case in which all forwards are observed with errors, using the Kalman filter.
account for the nonlinear dependence between forward rates and their shadow counter-parts:

\[ h(f_t^o) = \begin{pmatrix} W_e f_t^o \\ W \theta w^{-1}(f_t^o / \theta) \end{pmatrix}. \]

To avoid a singular likelihood, the choice of the matrices \( W \) and \( W_e \) must be such that the Jacobian adjustment term is non-zero. For simplicity, we choose \( W \) as the loadings on the first \( N \) PCs of \( f_t^o \) and \( W_e \) the remaining \( J - N \) PCs of \( f_t^o \).

**Analytical Concentration of Parameters**

The maximum likelihood estimates of \( K_{0P} \) and \( K_{1P} \) can be derived analytically. This is possible because these conditional mean parameters do not mix with volatility parameters in the expression for the log likelihood. Specifically, \( K_{0P} \) and \( K_{1P} \) enters the log likelihood via a quadratic form because time series innovations are conditionally Gaussian. Therefore, the first-order derivative of the log likelihood must be linear and the maximum likelihood estimates \( K_{0P} \) and \( K_{1P} \) are given:

\[
vec([\hat{K}_{0P}, \hat{K}_{1P}]) = E_T[\mathcal{P}_t^a \mathcal{P}_t^{a'} \otimes \Sigma_t^{-1}]^{-1} vec(E_T[\Sigma_t^{-1} \mathcal{P}_{t+1} \mathcal{P}_t^{a'}])
\]

where \( E_T[.] \) denotes sample average, \( \otimes \) is the Kronecker product, \( \mathcal{P}_t^a = (1, \mathcal{P}_t')' \) and where the \( vec(\cdot) \) operate converts a matrix into a column vector by stacking its columns (from left to right).

This result arises because we use the first difference \( \Delta \mathcal{P}_t \) to update volatility in equation (17), instead of the innovations \( \mathcal{P}_{t+1} - (K_{0P} + K_{1P} \mathcal{P}_t) \), which would be the case in the standard BEKK volatility. This alternative construction ensures that the parameters \( K_{0P} \) and \( K_{1P} \) do not enter the dynamics of \( \Sigma_t \). Otherwise, the first-order conditions of the log likelihood with respect to \( K_{0P} \) and \( K_{1P} \) would be highly nonlinear and closed-form estimates of \( K_{0P} \) and \( K_{1P} \) would likely be much harder to obtain.

**5 Model Evaluation in a Simulated Environment**

In this section, we evaluate bond pricing models when yields are near the ZLB and away from the ZLB. Our strategy is to compare these models across a large number of simulated yields data sets featuring ZLB episodes. The main advantage of model comparisons in a
simulation environment is the fact that all true predictive moments of the simulated data are known with certainty at each point in time, unlike those of actual historical data, which must be estimated (e.g., forecasts of yields, volatilities or Sharpe ratios). Hence, we can evaluate performance knowing the true optimal forecasts in each simulated sample, which is not feasible in the observed data. Importantly, the same assessment can be carried out for any sub-samples, such as the ZLB episodes, without concerns about statistical uncertainties that might arise due to the size of the sub-samples.\footnote{Note there is no standard re-sampling bootstrap strategy to generate ZLB episode}

There is another benefit from using a realistic data generating process to evaluate the performance of other largely unrelated models. This simulation-based approach captures the inescapable reality that the true data generating process will always remain unknown in an actual empirical implementation. To that extent, it seems sensible to base our assessments partly on the ability of the models to cope with the risks of misspecification, by introducing an element of uncertainty into our controlled environments.

## 5.1 The Simulated Environment

We choose the linear-rational term structure model of \cite{Filipovic_2017} as a realistic environment to simulate yield data and benchmark competing models. Of course, this is not the only candidate data generating process.\footnote{We also experimented with a gaussian quadratic term structure model featuring bond yields consistent with the ZLB and with time-varying volatility and we found very similar results.} We make this choice partly because the linear-rational model is known to capture salient features of bond data. These features include the ability to generate an upward sloping yield curve, a downward term structure of yields volatility, the predictability of bonds excess returns, the time variation in yields volatility, and most importantly, the ability to produce ZLB-consistent bond yields. Tractability is what differentiates our approach from linear-rational models, which can be challenging to estimate, requiring swap and swaption data, a non-linear Kalman filtering and a quasi-maximum likelihood estimator.

In the following, we assess the model performance across one hundred simulated sample of yield data. Every simulation sample has thirty years in length and features some ZLB episodes toward the end of the thirty year period. The parameter values that we use in the simulation are those published in \cite{Filipovic_2017} for the flexible LRSQ(3,3) specification, in which three factors explain the cross-section of bond yields and three factors drive unspanned yields volatility. Appendix A.4 provides a more detailed
description of the model, the model estimates used in our analysis, and how yields are simulated from the model.

The panel on the left in Figure 3 shows simulated yields in one of the simulated samples. We see that yield volatilities are strongly time-varying. It is also clear that all yields are bounded below at zero. By design of the simulation, the zero lower bound becomes relatively more binding about thirty years into the sample. In this graph, we plot yields over a sixty year period for a “big picture” view of the data. However, only the first thirty years of the data will be used in estimation. The panel on the right of Figure 3 shows that the average yield curve is upward sloping and that the average yield volatilities are downward sloping—consistent stylized facts in the U.S. data.

![Simulated yields](image1)

![Average means and volatilities of simulated yields](image2)

Figure 3: Simulated Yields

The left-panel reports the time-series of yields with one month, five years and ten years to maturity in the first simulated sample. The right-panel reports the term structure of average yields and the term structure of average yield volatilities in the first simulated sample.

5.2 Sharpe Ratios

We use the ability to forecast Sharpe ratios as a criteria to evaluate and rank models. It is the ratio of its conditional expected excess return divided by the conditional volatility for any given portfolio and horizon. Intuitively, the Sharpe ratio captures the market price of risk commanded by the average investor for each unit of risk exhibited by its portfolio—where risk is measured by the standard deviation of returns. Needless to say, the usefulness of
a term structure model depends to a large extent on its ability to accurately forecast the Sharpe ratios of a range of fixed income investments.

Our comparison includes the $B_{nd}$ and SV-$B_{nd}$ variants of the no-dominance models developed in Section 3, the no-arbitrage counter part $B_{na}$ to the $B_{nd}$ model with constant variance, and the standard affine model, which we label $G_{na}$.\footnote{All of these models have three factors with the same form of VAR(1) dynamics. The base case results reported in the paper assumes that three portfolios $P_t$ are priced without errors and can be used as pricing factors. For robustness, we also check that our key results are unchanged if we assume that the pricing portfolios are measured with errors and use a Kalman filter to derive the likelihood.} We ask two simple questions. First, we ask what are the improvements in term structure forecasts in the SV-$B_{nd}$ model, which is the most flexible model, consistent with the ZLB and time-varying volatility. Second, we ask whether moving from the no-arbitrage $B_{na}$ model to the no-dominance $B_{nd}$ framework induces differences in term structure forecasts. This mirrors existing work comparing the DNS model with the no-arbitrage $G_{na}$ model. We extend this comparison to Black style model and also check the relative performances in forecasting volatilities and Sharpe ratios.

For every date and every simulation, we compute the model-implied conditional Sharpe ratios and the difference with true Sharpe ratios. We assess Sharpe ratio forecast errors over 3-month, 6-month and 1-year investment horizons and across the 6-month, 2-year, 5-year and 9-year maturities. Here, we fix the maturity of the bond standing at the end of the investment horizon. For example, the 9-year maturity together with the 1-year horizon corresponds to a one-year position in a 10-year zero coupon bond. When this position is closed out after one year, the bond has 9 years remaining until maturity. Three horizons and four bond maturities give rise to twelve distinct fixed income investments. Given the potentially important role of the covariances across these cases, we also consider bond portfolios in Section 5.5.

Table 1 reports the performance of each model when forecasting conditional Sharpe ratios. We report the root median squared errors (RMedSE) between true and model-implied Sharpe ratios.\footnote{The median is taken both along the time dimension (within each simulated sample). We note that our use of the median statistics mitigates the risk that one of the models performs poorly overall because of very poor performances in a handful of simulations.} The first column of Table 1 reports the values implied by the $G_{na}$ model. Smaller values in this column indicate that Sharpe ratios are closer to the true Sharpe ratios, on average.

The next three columns of Table 1 report the same statistics implied by the $B_{na}$, $B_{nd}$ and SV-$B_{nd}$ models, respectively, but scaled with the values reported for the $G_{na}$ model in the first column. A value lower than one in these three columns indicates an improved accuracy...
relative to the $G_{na}$. The final four columns of Table 1 report the same set of statistics as the first four columns, but for the ZLB sub-samples. The ZLB sub-samples are defined as the collection of dates for which the one-month rate is twenty five basis points or less.

The first key finding is that the no-dominance 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 no-arbitrage Gaussian model, ranging from 10 to 30 percent. The $SV-B_{nd}$ gains are smaller relative to the $B_{nd}$ model, but larger relative to the $B_{na}$. We will return to this result below.

When specializing to the ZLB samples, the $SV-B_{nd}$ model still delivers the best performance for all horizon-maturity combinations with the exception of the 6-month bonds. The RMedSE gains are large relative to either of the NA models, ranging between 20 to 50 percent relative to the $G_{na}$ model or more relative to $B_{na}$ model, except in one case.

<table>
<thead>
<tr>
<th>h</th>
<th>m</th>
<th>Full-Sample</th>
<th></th>
<th></th>
<th>ZLB Sample</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$G_{na}$</td>
<td>$B_{na}$</td>
<td>$B_{nd}$</td>
<td>$SV-B_{nd}$</td>
<td>$G_{na}$</td>
<td>$B_{na}$</td>
</tr>
<tr>
<td>3-mth</td>
<td>6-mth</td>
<td>0.2</td>
<td>1.62</td>
<td>1.27</td>
<td>0.84*</td>
<td>0.3</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>2-yr</td>
<td>0.5</td>
<td>1.59</td>
<td>0.98</td>
<td>0.75*</td>
<td>0.6</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>5-yr</td>
<td>0.8</td>
<td>1.42</td>
<td>0.93</td>
<td>0.72*</td>
<td>1.0</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>9-yr</td>
<td>1.2</td>
<td>1.24</td>
<td>0.93</td>
<td>0.72*</td>
<td>1.3</td>
<td>1.01</td>
</tr>
<tr>
<td>6-mth</td>
<td>6-mth</td>
<td>0.2</td>
<td>1.71</td>
<td>1.35</td>
<td>0.90*</td>
<td>0.2</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>2-yr</td>
<td>0.4</td>
<td>1.59</td>
<td>0.97</td>
<td>0.77*</td>
<td>0.5</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>5-yr</td>
<td>0.7</td>
<td>1.46</td>
<td>0.95</td>
<td>0.72*</td>
<td>0.7</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>9-yr</td>
<td>1.1</td>
<td>1.22</td>
<td>0.92</td>
<td>0.69*</td>
<td>1.1</td>
<td>1.09</td>
</tr>
<tr>
<td>1-yr</td>
<td>6-mth</td>
<td>0.2</td>
<td>1.59</td>
<td>1.19</td>
<td>0.92*</td>
<td>0.2</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>2-yr</td>
<td>0.3</td>
<td>1.57</td>
<td>1.01</td>
<td>0.79*</td>
<td>0.4</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>5-yr</td>
<td>0.6</td>
<td>1.42</td>
<td>0.95</td>
<td>0.68*</td>
<td>0.6</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>9-yr</td>
<td>0.9</td>
<td>1.24</td>
<td>0.98</td>
<td>0.67*</td>
<td>0.9</td>
<td>1.12</td>
</tr>
</tbody>
</table>

Table 1: **Bond’s Sharpe Ratio Forecast Errors**

$G_{na}$ and $B_{na}$ refer to the no-arbitrage Gaussian and Black models, respectively. $B_{nd}$ and $SV-B_{nd}$ refer to the no-dominance Black models with and without BEKK volatility, respectively. For the $G_{na}$ model, we report the root median squared errors (RMedSE) between true and model-implied forecasts. For other models we report this statistics but scaled by the values reported for the $G_{na}$ model in the first column. Forecasts horizons are 3, 6, and 12 months and bond maturities are 6 months, 2, 5 and 9 years. The symbol * indicates the best performance for a given combination of forecast horizon, bond maturity, and sample.
The second key finding is obtained from a direct comparison between the no-dominance \(B_{nd}\) model with constant variance and its no-arbitrage \(B_{na}\) counter-part. It is striking that the \(B_{nd}\) model delivers strictly more accurate Sharpe ratios than the \(B_{na}\) model, on average, for every investment horizon and bond maturity that we consider, without exception. In full samples results, the RMeadSE deteriorations can reach as high as 60 percent. In fact, the \(B_{na}\) model even underperforms relative to the the Gaussian \(G_{na}\) model. This contrast is rather stark over the ZLB samples. Whereas the \(B_{nd}\) model handily outperforms the Gaussian model \(G_{na}\), the \(B_{na}\) model performs consistently worse than the Gaussian \(G_{na}\) model, barring one exception.

What can explain the poor performance of the no-arbitrage \(B_{na}\) model relative to the no-dominance models? It is easy to understand why the SV-\(B_{nd}\) model does a better job, given its flexible volatility dynamics. But what is responsible for the difference between the \(B_{na}\) and \(B_{nd}\) models? The answer can be found in the pricing equations, which is the only difference between these models. Recall the pricing equation (12) of the \(B_{na}\) model:

\[
f_{n,t} \approx S_n w \left( \frac{A_n + B'_n X_t}{S_n} \right)
\]

where \(S_n\), \(A_n\), and \(B_n\) are dependent on the parameters governing the risk-neutral dynamics of states (see Section A.3). As we discuss in Section 3.3, \(S_n\) is tightly linked to the volatility of the forwards due to the no-arbitrage restrictions. To this extent, some of the parameters that constitute \(S_n\) play a dual role in the \(B_{na}\) models. They govern the pricing of forwards, but they also play an important role in fitting the volatility of forwards and yields. It is possible that this dual role creates a tension reducing the ability of the model to fit the multiple dimensions of the data. Notably, this type of tension is largely absent from the pricing equation (8) of the \(B_{nd}\) model (as well as the SV-\(B_{nd}\) model). The scaling parameter \(\theta\) of the \(B_{nd}\) model does not play a direct role in shaping the volatility dynamics of states.\(^{16}\)

5.3 Volatility Forecasting

To examine this tension further, this section directly investigates the models’ ability to forecast the volatility of bond returns. (The next section looks at the returns forecasts directly). Parallel to the Sharpe ratio results, we find, unsurprisingly, that (i) the SV-\(B_{nd}\)

\(^{16}\)Strictly speaking, we requires that \(\theta w \left( \frac{0+\sigma}{\sigma} \right) \geq 0.0005\) and \(\theta w \left( \frac{\sigma+\sigma}{\sigma} \right) \leq 0.0050\). However, these inequality constraints are much milder, only constraining the variations of the shadow rate locally around zero, relative to the no-arbitrage restrictions tying \(S_n\) at every maturity with the underlying volatility parameters.
model delivers the best volatility forecasts but, what is more surprising, that (ii) there is no clear winner between the no-dominance model $B_{nd}$ and its no-arbitrage counterpart $B_{na}$ in volatility forecasts.

It is important to recognize that the superior performance of a model in delivering the most accurate Sharpe ratios (such as the SV-$B_{nd}$ model in the previous section) does not automatically translate to a superior ability to forecast bond returns and bond returns' volatility. For a numerical illustration of this possibility, consider an investment with an expected excess return and expected volatility of 0.3 and 0.1, respectively, so that its true Sharpe ratio is 3. A model that overestimates both the expected excess return and return volatility by 0.1 implies a conditional Sharpe ratio of $0.4/0.2 = 2$, a difference of 1 relative to the true ratio. By contrast, a model that underestimates both the expected excess return and return volatility by 0.05 implies a conditional Sharpe ratio of $0.35/0.05 = 7$, a difference of 2 relative to the true ratio. Although the first model misses returns and returns volatility forecasts with errors larger in magnitude, its Sharpe ratio prediction is relatively more accurate.

Table 2 reports the performance of each model in forecasting yields volatility. The structure of Table 2 is identical to that of Table 1. The first column reports the median RMedSE from the Gaussian model $G_{na}$. The next three columns report the same statistics implied by the $B_{na}$, $B_{nd}$, SV-$B_{nd}$ models but scaled by the values reported for the $G_{na}$ model in the first column. The last four columns repeat these statistics but for the ZLB sub-sample.

First, the volatility comparison between the SV-$B_{nd}$ model and other models is consistent with Sharpe ratio results. The SV-$B_{nd}$ model consistently delivers the best volatility forecasts over the full sample, for all combinations of horizons and bond maturities. The results are not surprising, since the SV-$B_{nd}$ model has a built-in stochastic volatility construction. This model is able to bring down the RMedSE substantially relative to the $G_{na}$ benchmark, by as much 50-70 percent in many cases. The magnitude of these gains underlines the importance of time-varying yield volatilities.

Second, in contrast with the Sharpe ratio results, there is no clear winner between the no-dominance $B_{nd}$ and its $B_{na}$ counterpart in full-sample volatility forecast results. If anything, the $B_{nd}$ model seems to do a relatively better job at the one-year forecasting horizon. There is no clear winner in the ZLB results either. In fact, it is remarkable that the $B_{nd}$ and $B_{na}$ deliver similar volatility forecasting performance relative to the SV-$B_{nd}$ model in ZLB samples, even or better performances for a good number of horizon and bond maturity combinations. What is remarkable is that both models assume constant variances for the state dynamics.
Table 2: Bond Yield Volatility Forecast Errors

$G_{na}$ and $B_{na}$ refer to the no-arbitrage Gaussian and Black models, respectively. $B_{nd}$ and SV-$B_{nd}$ refer to the no-dominance Black models with and without BEKK volatility, respectively. For the $G_{na}$ model, we report the root median squared errors (RMedSE), in basis points, between true and model-implied forecasts. For other models we report that statistics but scaled by the values reported for the $G_{na}$ model in the first column. Forecasts horizons are 3, 6 and 12 months and bond maturities are 1 month, 2 years, 5 years, and 10 years. The symbol * indicates the best performance for a given combination of forecast horizon, bond maturity, and sample.

The results in volatility forecasting clearly show the effect of the volatility compression around the ZLB. The only source of time-varying volatility for the $B_{nd}$ and $B_{na}$ models comes from the convexity of the nonlinear transformation between forwards and shadow forwards. Intuitively, as the shadow forwards venture deep into the negative region, the corresponding forwards have little room to “wiggle”, thus their conditional volatilities are “compressed” toward zero. On the other hand, when far away from the ZLB, the forward rates inherit the constant variance property of the shadow shadow forwards within the $B_{nd}$ and $B_{na}$ models. The results reported in Table 2 suggest that the “compression” channel present in Black models is be a good way to capture time-varying volatility during ZLB periods.

5.4 Yield Forecasting

Last, but not least important, this section investigates the models’ ability to forecast bond excess returns. This is a standard exercise featured in Christensen, Diebold, and Rudebusch...
among others, that helped support the popularity of the DNS model. This is also
the forecast for the numerator of the Sharpe ratio, which shed further lights on the Sharpe
ratio results, and the tensions we uncover there. Table 3 reports each model’s performance.\footnote{Note that the excess return on an \( n \)-period zero coupon bond is given by: \( ny_{n,t} - (n - 1)y_{n-1.t+1} - y_{t,t} \). Since \( y_{n,t} \) and \( y_{1,t} \) belong to the information set at time \( t \), a model’s ability to forecast this excess return is largely dependent on its ability to forecast the future yield \( y_{n-1,t+1} \). As a result, ranking models based on their ability to forecast future yields is equivalent to ranking based on the ability to forecast bonds’ excess returns.}
The structure of Table 3 is identical to that of Table 1. The first column reports the median
RMedSE from the Gaussian model \( G_{na} \). The next three columns report the same statistics
implied by the \( B_{na}, B_{nd}, SV-B_{nd} \) models but scaled with the values reported for the \( G_{na} \)
model in the first column. The last four columns repeat these statistics but for the ZLB
sub-sample.

<table>
<thead>
<tr>
<th>h</th>
<th>m</th>
<th>Full sample</th>
<th></th>
<th>ZLB sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( G_{na} )</td>
<td>( B_{na} )</td>
</tr>
<tr>
<td>3-mth</td>
<td>1-mth</td>
<td>10.8 0.91 0.96 0.75*</td>
<td>7.8 0.69 0.58* 0.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-yr</td>
<td>8.0 0.96 0.93 0.81*</td>
<td>7.5 0.76 0.49* 0.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-yr</td>
<td>7.9 1.00 0.90 0.82*</td>
<td>7.2 1.01 0.74 0.62*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10-yr</td>
<td>7.8 1.00 0.97 0.75*</td>
<td>6.8 0.91 0.91 0.65*</td>
<td></td>
</tr>
<tr>
<td>6-mth</td>
<td>1-mth</td>
<td>17.9 0.91 0.90 0.79*</td>
<td>14.2 0.68 0.53* 0.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-yr</td>
<td>14.4 0.98 0.94 0.79*</td>
<td>13.1 0.87 0.58* 0.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-yr</td>
<td>15.1 1.01 0.90 0.77*</td>
<td>13.3 1.02 0.79 0.58*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10-yr</td>
<td>14.6 1.02 0.97 0.79*</td>
<td>12.9 0.94 0.94 0.68*</td>
<td></td>
</tr>
<tr>
<td>1-yr</td>
<td>1-mth</td>
<td>27.5 0.94 0.92 0.82*</td>
<td>24.0 0.67 0.47* 0.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-yr</td>
<td>26.1 0.97 0.91 0.77*</td>
<td>23.4 0.87 0.66 0.54*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-yr</td>
<td>27.8 1.00 0.91 0.72*</td>
<td>24.1 0.97 0.81 0.61*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10-yr</td>
<td>26.9 1.01 0.96 0.75*</td>
<td>21.3 1.00 1.04 0.70*</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Bond Yield Forecast Errors

\( G_{na} \) and \( B_{na} \) refer to the no-arbitrage Gaussian and Black models, respectively. \( B_{nd} \) and SV-\( B_{nd} \) refer to the no-dominance Black models with and without BEKK volatility, respectively. For the \( G_{na} \) model, we report the root median squared errors (RMedSE), in basis points, between true and model-implied forecasts. For other models we report that statistics but scaled by the values reported for the \( G_{na} \) model in the first column. Forecasts horizons are 3, 6, and 12 months and bond maturities 1 month, 2 years, 5 years, and 10 years. The symbol * indicates the best performance for a given combination of forecast horizon, bond maturity, and sample.

First, it is remarkable that the SV-\( B_{nd} \), delivers strictly superior forecasts of bond yields over full samples relative any other model that we consider. The improvements range between
20 and 30 percent relative to the no-arbitrage Gaussian model $G_{na}$ and around 20-25 percent relative to the no-arbitrage Black model $B_{na}$. What is remarkable is that the improvements in forecasting bond yields must be attributed to better forecasts of bond yield volatilities. Recall from Table 2 that, over the full sample, the SV-$B_{nd}$ model captures the volatility dynamics particularly well. In a typical maximum likelihood estimation, it is well-known that (i) the variance of forecast errors depends on the variance of the parameter estimator and (ii) that controlling for conditional variance reduces the variance of the parameter estimator (e.g., GLS). Intuitively, *ex ante* noisier signals have less influence on the estimation. Superior knowledge of the volatility dynamics can assign smaller (larger) weights to forecast errors in the likelihood function around heightened (lower) volatility. In contrast, constant-variance models assign the same weights to forecast errors observed during period of high and low volatility, thereby treating noisy signals and high quality signals equally.

In this light, much of the difference in the yield forecasting performances between the $B_{nd}$ and SV-$B_{nd}$ models is likely owing to the econometric efficiency gain obtained by the latter through its superior ability to match yield volatility over the full sample. This means that these two models should offer forecasts with similar accuracy when they offer similar volatility forecast performances. Indeed, the two no-dominance models display quite comparable yield forecasting performances in the ZLB samples, precisely when Table 2 shows that the $B_{nd}$ model is able to quite reasonable time variation in volatility. As a result, the $B_{nd}$ model is likely able to generate as much econometric efficiency gain as the SV-$B_{nd}$ model through the “compression” channel. It fact, during the ZLB episodes, the $B_{nd}$ model seems to do a better job at yields forecasting for the shorter bond maturities (1-month and 2 year). Interestingly, these are essentially the categories over which the $B_{nd}$ model provides relatively superior volatility forecasts on average.

It would seem that the same logic should apply to the $B_{na}$ model. Turning to a direct comparison between the no-arbitrage and no-dominance Black style models, it is interesting to note that the $B_{na}$ model is the winner for almost all horizon and maturity combinations. We note that the improvements of the $B_{nd}$ model, relative to the $B_{na}$ model, are more sizable and concentrated in the ZLB-sample results. This difference in performance likely reflects the type of tension inherent in the $B_{na}$ model that we discussed earlier (between the scaling parameter $S_n$ and the volatility parameters of the model).
Table 4: Bond Portfolio Sharpe Ratio Forecast Errors

$G_{na}$ and $B_{na}$ refer to the no-arbitrage Gaussian and Black models, respectively. $B_{nd}$ and $SV-B_{nd}$ refer to the no-dominance Black models with and without BEKK volatility, respectively. For the $G_{na}$ model, we report the root median squared errors (RMedSE) between true and model-implied forecasts, errors (RMedSE), in basis points, between true and model-implied forecasts. For other models we report that statistics but scaled by the values reported for the $G_{na}$ model in the first column. Forecasts horizons are 3, 6, and 12 months. The symbol $\ast$ indicates the best performance for a given combination of forecast horizon, bond maturity, and sample.

<table>
<thead>
<tr>
<th>horizon</th>
<th>maturity</th>
<th>Full sample</th>
<th>ZLB sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$G_{na}$</td>
<td>$B_{na}$ $B_{nd}$</td>
<td>$SV-B_{nd}$</td>
</tr>
<tr>
<td>3-m</td>
<td>6m-9y</td>
<td>0.5 1.63 1.01 0.77*</td>
<td>0.7 1.04 0.49 0.45*</td>
</tr>
<tr>
<td></td>
<td>6m-4yr</td>
<td>0.8 1.46 0.98 0.74*</td>
<td>1.0 1.10 0.73 0.49*</td>
</tr>
<tr>
<td></td>
<td>5yr-9yr</td>
<td>1.0 1.33 0.93 0.71*</td>
<td>1.2 1.04 0.76 0.52*</td>
</tr>
<tr>
<td>6-m</td>
<td>6m-9y</td>
<td>0.5 1.65 1.01 0.79*</td>
<td>0.6 1.20 0.58 0.46*</td>
</tr>
<tr>
<td></td>
<td>6m-4yr</td>
<td>0.7 1.46 0.96 0.70*</td>
<td>0.8 1.22 0.81 0.53*</td>
</tr>
<tr>
<td></td>
<td>5yr-9yr</td>
<td>0.9 1.35 0.94 0.69*</td>
<td>0.9 1.15 0.86 0.58*</td>
</tr>
<tr>
<td>1-yr</td>
<td>6m-9y</td>
<td>0.4 1.61 1.01 0.76*</td>
<td>0.4 1.39 0.72 0.52*</td>
</tr>
<tr>
<td></td>
<td>6m-4yr</td>
<td>0.7 1.42 0.99 0.67*</td>
<td>0.7 1.24 0.79 0.53*</td>
</tr>
<tr>
<td></td>
<td>5yr-9yr</td>
<td>0.8 1.31 0.97 0.64*</td>
<td>0.8 1.27 0.95 0.58*</td>
</tr>
</tbody>
</table>

5.5 Bond Portfolios

In this sub-section, we evaluate the ability of the considered models to forecast Sharpe ratios of bond portfolios. Embedded in these portfolio statistics is the ability of a model to forecast the correlations among future yields over various horizons. We form three simple equal-weighted portfolios consisting of zero coupon bonds with maturities of 6-month, 1-, 2-, ..., 9-year.\textsuperscript{18} The first portfolio combines all ten maturities, the second portfolio combines the first five short-term maturities between 6-month and 4-year and the third portfolio combines the last five maturities between 5-year and 9-year. Table 4, sharing the same column structure as Table 3, 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

\textsuperscript{18} We stop at the 9-year maturity so one-year forecasts of returns, volatility and correlations are not extrapolative relative to the data in our sample; the longest maturity in our sample is 10-year.
bonds reported in Table 1. Part of this improved performance can be related to better forecasts of the correlations between bond yields (unreported). In addition, as we argue before, the econometric efficiency gain induced by better volatility fitting likely plays a role in explaining the superior performance by the SV-\(B_{nd}\) model. Considering the Black models with constant variance, the \(B_{na}\) model’s performance is strictly inferior, even when compared to the Gaussian model \(G_{na}\). In particular, we find that the \(B_{na}\) model fails to accurately forecast correlations (unreported). The no-dominance model \(B_{nd}\), on the other hand, performs rather well, particularly during the ZLB sub-samples.

6 Empirical Illustrations

In this section, we illustrate the performances of the \(B_{na}\), \(B_{nd}\) and SV-\(B_{nd}\) models in U.S. bond data sampled monthly between January 1970 and December 2015. We report pricing errors, shadow rate estimates, yield forecasts, volatility forecasts as well as term premium estimates, relative to the \(G_{na}\) model. For estimation, we use rates on one-month forward loans with 13 different maturities: 1 month, 3 month, 6 month, 1 year, 2 year, and annually until 10 year. Following Wu and Xia (2016), we use the Kalman filter estimate of the \(B_{na}\), \(B_{nd}\) and the \(G_{na}\) models in our analysis. For the SV-\(B_{nd}\) model, we use the estimate obtained by assuming the portfolios of shadow forwards are priced perfectly. Estimate using this approach is easy to implement and converges quickly, which is often not the case when using the Kalman filter. Results are unchanged if we estimate every model with shadow forward portfolios priced without error.

6.1 Pricing Errors

Table 5 reports pricing error RMSEs for selected maturities in the full sample and in the sub-sample where the short rate is at the ZLB. First contrast the results between the ND Black models with and without time-varying volatility, SV-\(B_{nd}\) and \(B_{nd}\). The SV-\(B_{nd}\) model with time-varying volatility shows lower pricing errors for the short rate in the whole sample, 39.3 bps compared with 49.6 bps. This relatively large gain appears to trade-off small pricing

---

19We use the data provided by Gurkaynak, Sack, and Wright (2006) for maturities longer than 6 months. For maturities less than or equal to 6-month, we bootstrap for the forward rates from Treasury bond prices provided by the Center for Research in Security Prices (CRSP).

20Joslin, Le, and Singleton (2013) show that the Kalman filter estimate of affine Gaussian models are highly similar to the estimate obtained by assuming the first three PCs of bond yields/forwards are assumed priced perfectly.
error deteriorations, between 1.5 and 2.5 basis points, for a few intermediate maturities. It is interesting that allowing for time-varying volatility yields some improvement in form of smaller pricing errors.

Second, contrast the results between the no-arbitrage and no-dominance Black models with constant volatility, $B_{na}$ and $B_{nd}$. We find that the pricing errors are very close, within a fraction of 1 bps 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 5 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 ZLB period.

### 6.2 Shadow Short Rates

Figure 4 plots 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.\(^{21}\)

For much of the ZLB period, it is rather interesting that the shadow rates implied by the two no-dominance models, despite their different volatility specifications, remain close to one another. Overall, the no-arbitrage model $B_{na}$ shadow rates remain within within a range of 50 bps. But with one clear exception. For much of the year 2014, the shadow rate implied

---

\(^{21}\)Before 2008, the shadow short rates implied by all models are all very close to the short rate, unsurprisingly, since yields are far from the ZLB.
Figure 4: **Shadow rates**

One-month rate and shadow rate in the sample 2008:Jan-2015:Dec. $G_{na}$ and $B_{na}$ refer to the no-arbitrage Gaussian and Black models, respectively. $B_{nd}$ and $SV-B_{nd}$ refer to the no-dominance Black models with and without BEKK volatility, respectively.

by the no-arbitrage model $B_{na}$ falls much deeper into the negative region, often by more than 100 bps, relative to the two ND shadow rates. The pattern in Figure 4 suggests that while there has been much interest in studying the economic content of shadow rates, this must be done with caution because shadow rates can be highly model-dependent. This is an important observation already documented in Bauer and Rudebusch (2013).

6.3 Yield Forecasts

Figure 5 shows models’ forecasts of the 1-month rate, focusing on ZLB period after 2008. We report results for the 3-months, 6-months and 1-year forecast horizons. Two remarks are in order. First, the forecasts implied by the Gaussian model (the green lines) are visibly

---

22 The estimated shadow rate for the $B_{na}$ model is broadly similar to results reported on the web site of the Federal Reserve Bank of Atlanta based for the WX model. In particular, they also report a shadow rate at -3 percent in 2014.
distinct from the forecasts implied by the Black style models. In particular, the Gaussian forecasts can significantly violate the zero lower bound (at all horizons). Second, regardless of whether it is a no-arbitrage or a no-dominance model, whether it is a constant variance or stochastic volatility model, the forecasts by the Black style models seem visibly close to one another. The closeness of these forecasts suggests that it can be challenging to accurately compare the forecast performance of the different models in one short ZLB subsample. This observation strengthens the case for the exercise of Section 5 where the models are evaluated in a simulated environment. Results for other maturities lead to similar observations (not reported).
Figure 5: Short Rate Forecasts
Forecasts of the one-month rate over sample period 2008:Jan - 2015:Dec. $G_{na}$ and $B_{na}$ refer to the no-arbitrage Gaussian and Black models, respectively. $B_{nd}$ and $SV-B_{nd}$ refer to the no-dominance Black models with and without BEKK volatility, respectively.
6.4 Volatility Forecasts

Figure 6 plots the volatility forecasts of the 24-month yield implied by the models over three forecast horizons: 3-month, 6-month, and 1-year. It is known that volatility dynamics can be estimated precisely, even in short samples. We compute the realized variances of the 24-month yield, using three months of daily data leading up to each point in time \( t \), as a (noisy) proxy for the “true” volatility dynamics. We then overlay each subplot of Figure 6 with this RV series, adjusting for the different forecasting horizons.

While we have argued that it might be challenging to accurately compare a model’s ability in matching the first conditional moments of bond yields in a short sample, the same argument does not apply fully to evaluating a model’s ability to forecast bond volatility. Outside the ZLB period, the SV-\( B_{nd} \) model does a good job matching the time variation of the RV series, which exhibits wild fluctuations (ranging from as low as 40 bps to as high as 300 bps). This justifies the simple construction for the volatility dynamics in Equation 11.

By contrast, and as expected, it is clear that the yield volatilities implied by the constant-variance Black models are essentially a flat line for the non-ZLB periods. Loosely speaking, the flat lines corresponds to the the average volatility levels away from the ZLB 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 no-arbitrage model \( B_{na} \) is higher than the level implied by the no-dominance 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 (12) 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 RMSEs reported in Table 6. Averaging across maturities, the SV-\( B_{nd} \) model is able to reduce volatility forecast errors by 30-40 percent during the full sample, and by 20-35 percent during the ZLB sample. Comparing the \( B_{na} \) and \( B_{nd} \) models, the volatility forecast errors exhibited by the \( B_{nd} \) model, is on average, 20-25 percent lower than that of the \( B_{na} \) model. Nevertheless, the ordering is the opposite during the ZLB sample: the \( B_{na} \) model becomes better than the \( B_{nd} \) model at matching the RV dynamics.
Figure 6: Volatility Forecasts
Forecasts of the 2-year bond yield volatility over sample period 1970:Jan - 2015:Dec. $G_{na}$ and $B_{na}$ refer to the no-arbitrage Gaussian and Black models, respectively. $B_{nd}$ and SV-$B_{nd}$ refer to the no-dominance Black models with and without BEKK volatility, respectively.
### Table 6: Volatility Forecasts

Root median squared difference between volatility forecasts and our realized volatility forecasts. $B_{na}$ refers to the no-arbitrage Black model. $B_{nd}$ and $SV-B_{nd}$ refer to the no-dominance Black models with and without BEKK volatility, respectively. For the $B_{na}$ model, we report the root median squared errors (RMedSE), 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 $B_{na}$ model. Forecasts horizons are 3, 6, and 12 months.

<table>
<thead>
<tr>
<th>h</th>
<th>m</th>
<th>$B_{na}$</th>
<th>$B_{nd}$</th>
<th>$SV-B_{nd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-month</td>
<td>40.68</td>
<td>0.76</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>25.16</td>
<td>0.65</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>3-m</td>
<td>5-year</td>
<td>17.90</td>
<td>0.77</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>10-year</td>
<td>16.33</td>
<td>0.81</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>25.02</td>
<td>0.75</td>
<td>0.58</td>
</tr>
<tr>
<td>6-m</td>
<td>1-month</td>
<td>55.52</td>
<td>0.79</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>2-year</td>
<td>34.28</td>
<td>0.68</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>5-year</td>
<td>24.76</td>
<td>0.78</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>10-year</td>
<td>23.42</td>
<td>0.79</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>34.50</td>
<td>0.76</td>
<td>0.59</td>
</tr>
<tr>
<td>12-m</td>
<td>1-month</td>
<td>71.16</td>
<td>0.84</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>2-year</td>
<td>46.62</td>
<td>0.73</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>5-year</td>
<td>33.84</td>
<td>0.81</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>10-year</td>
<td>31.61</td>
<td>0.85</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>45.81</td>
<td>0.81</td>
<td>0.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>h</th>
<th>m</th>
<th>$B_{na}$</th>
<th>$B_{nd}$</th>
<th>$SV-B_{nd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-month</td>
<td>15.85</td>
<td>0.97</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>7.75</td>
<td>1.93</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>3-m</td>
<td>5-year</td>
<td>10.65</td>
<td>1.05</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>10-year</td>
<td>12.21</td>
<td>0.82</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>11.61</td>
<td>1.19</td>
<td>0.64</td>
</tr>
<tr>
<td>6-m</td>
<td>1-month</td>
<td>24.24</td>
<td>0.96</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>2-year</td>
<td>13.78</td>
<td>1.68</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>5-year</td>
<td>14.00</td>
<td>1.14</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>10-year</td>
<td>17.84</td>
<td>0.80</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>34.50</td>
<td>0.76</td>
<td>0.59</td>
</tr>
<tr>
<td>12-m</td>
<td>1-month</td>
<td>40.97</td>
<td>1.01</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>2-year</td>
<td>26.22</td>
<td>1.35</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>5-year</td>
<td>21.08</td>
<td>1.04</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>10-year</td>
<td>27.08</td>
<td>0.77</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>28.84</td>
<td>1.04</td>
<td>0.78</td>
</tr>
</tbody>
</table>

#### 6.5 Yield Decomposition

This section compares the term premium from each model computed as:

$$y_{n,t} = -\frac{1}{n} E_t \left[ \sum_{k=0}^{n-1} r_{t+k} \right],$$

in the case of the 10-year bond (i.e., $n = 120$). As the above formula suggests, whether a model generates a reasonable term premium dynamic is dependent on its ability to produce reasonable forecasts of the short rate. Overall, we find that every model implies a countercyclical term premium. To emphasize the differences between models, Figure 7 reports the
term premium components for the 10-year zero coupon bond, derived from the $B_{na}$, $B_{nd}$, and $SV-B_{nd}$ models minus the estimate derived from the $G_{na}$ model. The constant-variance models $B_{na}$ and $B_{nd}$ exhibit small differences, mirroring the small difference in a linear setting between a DNS model and the corresponding no-arbitrage Gaussian model.

![Figure 7: Term Premium](image)

The term premium component in the 10-year yield implied by each model minus the term premium component implied by the Gaussian model $G_{na}$. $B_{na}$ refers to the no-arbitrage Black model. $B_{nd}$ and $SV-B_{nd}$ refer to the no-dominance Black models with and without BEKK volatility, respectively. Sample period is 1970:Jan - 2015:Dec.

Interestingly, the term premium component derived from the $SV-B_{nd}$ model exhibits notable cyclical differences. In particular, the $SV-B_{nd}$ term premium becomes lower following the end of every NBER-dated recession (illustrated by the gray bars in Figure 7). In other words, it falls faster at the onset of the expansionary phase. This pattern is most pronounced when comparing the $SV-B_{nd}$ model with the $G_{na}$ model. However, similar patterns can also be observed comparing the $SV-B_{nd}$ model against the other two constant-variance Black models, $B_{na}$ and $B_{nd}$. Based on the $SV-B_{nd}$ model estimate, the term premium embedded in the 10-year yield is almost 60 basis points lower after the 1991 recession, almost 70 basis points lower after the 2001 recessions and around 80 basis points lower after the 2008-2009 great recession.
What can explain the differences? For one, the presence of stochastic volatility, as argued earlier, can bring about econometric efficiency gains, thereby improving the precision in the estimate of the time series parameters.

Moreover, the non-linear feature of the Black models opens an additional channel for volatility to affect the term premium components. Intuitively, when the ZLB is near binding, the conditional distribution of the short rate becomes heavily skewed because the short rate cannot go below zero. As a result, short rate forecasts will be positively dependent on volatility. If volatility is high, it is more likely that the short rate will have a chance to break away from the ZLB. On the other hand, if volatility is low, the short rate is expected to stay at the ZLB for relatively longer. It should be noted that this channel can be at work even outside of the ZLB as long as a ZLB period can be reached within the forecast horizons, for instance soon after the 2001 recession when the target rate reaches one percent.

Following the above logic, if volatility of the shadow rates tends to be higher following the end of recessions (for example, due to uncertainties regarding future direction of monetary policies), then a model that correctly captures the cyclical nature of the volatility of the shadow rates will tend to produce a higher short-rate forecasts. These higher forecasts in turn translate into a higher expectation component, $\frac{1}{n}E_t [\sum_{k=0}^{n-1} r_{t+k}]$, and thus a lower term premium component as observed from Figure 7.

7 Conclusion

We introduce a family of tractable no-dominance term structure models where bond prices are analytically available by construction and very nearly arbitrage-free. Our results show how this new class of models can do a reasonable job, relative to existing models, in capturing the dynamics of yields and yield volatility 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 macro variables. 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).
References


A Appendix

A.1 Proofs

This section contains proofs not included in the main text.

A.1.1 Proposition 1 — Bond Prices

Starting with $P_0(\cdot) \equiv 1$ and expanding the recursion (2), we get

$$P_n(X_t) = \exp(-\sum_{i=0}^{n-1} m(g^i(X_t))).$$

(21)

Equation 3 follow 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.

A.1.2 Proposition 2 — Nesting Nelson-Siegel

Let $\Delta$ denote the discrete time interval and consider a three factor model $N = 3$ with the following one-period interest rate:

$$m(X_t) = \Delta(1, 1, 0)'X_t$$

and the $g(\cdot)$ function

$$g(X_t) = \left( \begin{array}{ccc} 1 & 0 & 0 \\ 0 & a & 1-a \\ 0 & 0 & a \end{array} \right) X_t.$$

Using the pricing equations in Proposition 1, the per annum yield for a bond with maturity $m = n\Delta$ is given by:

$$y_{n,t}/\Delta = X_{1,t} + (X_{2,t} + X_{3,t}) \frac{1-a^n}{1-a}/n - a^{n-1}X_{3,t}.$$  

(22)

Then, using notations of Nelson and Siegel (1987): $a = 1 - \Delta/\tau$ where $1/\tau$ can be interpreted as modulating the frequency of the factors, we can use the L'Hôpital’s rule to show that as $\Delta \to 0$,

$$a^{n-1} \to e^{-m/\tau} \quad \text{and} \quad \frac{1-a^n}{1-a}/n \to \left(1 - e^{-m/\tau}\right)/(m/\tau),$$

and, therefore, that the yield in (22) approaches:

$$X_{1,t} + (X_{2,t} + X_{3,t}) (1 - e^{-m/\tau})/(m/\tau) - e^{-m/\tau}X_{3,t}.$$  

This last expression is identical to the model of Nelson and Siegel (1987) with $X_{1,t}$, $X_{2,t}$, and $X_{3,t}$ corresponding respectively to their original parameters of $\beta_0$, $\beta_1$, and $\beta_2$.

A.1.3 Theorem 1 — No Dominant Strategies

Let $w_n$ denote the amount (in face value) invested in each $n$-period bond. Suppose that this portfolio guarantees positive payoffs: $\sum_n w_n P_{n-1}(X_{t+1}) > 0 \ \forall X_{t+1} \in X$. From the pricing recursions in Equation (2), the price of this portfolio is given by

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

(23)
Since \( g(X_t) \in \mathbb{X} \) and since \( \sum_n w_n P_{n-1}(X_{t+1}) > 0 \) for all \( X_{t+1} \in \mathbb{X} \), it follows that the price of this portfolio is strictly positive. Thus, a dominant trading strategy does not exist.

### A.1.4 Theorem 2—Non-Negative Payoffs

Let \( w_n \) denote the amount (in face value) invested in each \( n \)-period bond. Consider a portfolio with strictly non-negative payoffs: \( \sum_n w_n P_{n-1}(X_{t+1}) \geq 0 \) \( \forall X_{t+1} \in \mathbb{X} \). From the pricing recursion in Equation (2), the price of this portfolio is given by:

\[
\sum_n w_n P_n(X_t) = \exp(-m(X_t)) \times \sum_n w_n P_{n-1}(g(X_t)).
\]

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 \( g(X_t) \in \mathbb{X} \) and \( \sum_n w_n P_{n-1}(X_{t+1}) \geq 0 \) for all \( X_{t+1} \in \mathbb{X} \). Figure 1 illustrates Theorem 2. For portfolios with strictly non-negative payoffs, TTSMs allow for prices on the positive half of the real line, including the origin. The absence of arbitrage allows for prices on the positive half of the real line, excluding the origin. The difference reduces to one point on the real line (the origin).

The following example provides another way to see intuitively how ND is close to NA. Consider a portfolio \( w \) that pays one dollar in some states (with a strictly positive measure) and zero otherwise. If markets are complete, we can construct a portfolio \( w_\epsilon \) that pays at least \( \epsilon \) dollars in all states (i.e., either 1 or 1 + \( \epsilon \) dollars). ND implies that the price \( p_\epsilon \) of \( w_\epsilon \) must be positive (\( w_\epsilon \) earns strictly positive cash flows in every state). Varying \( \epsilon \) closer to zero, we obtain a sequence of portfolios \( w_\epsilon \) approaching portfolio \( w \). Each of these portfolios has a positive price. With continuous prices, this example suggests that the price of portfolio \( w \) – the limit of the sequence of positive prices \( p_\epsilon \) as \( \epsilon \) tends to zero – cannot be negative. For convenience, this heuristic argument invoked assumptions regarding market completeness and price continuity. Fortunately, the result in Theorem 2 only relies on Assumption 1.

### A.1.5 Theorem 3 — Convergence of NA models to an ND model

**Assumption 3.** The \( n \)-period bond price \( P_n \) is determined by the following recursion:

\[
P_0(X) \equiv 1, \\
P_n(X) = \frac{1}{J} \sum_{i=1}^{J} P_{n-1}(g_i(X)) \times \exp(-m_i(X)),
\]

for functions \( g_i(\cdot) \) and \( m_i(\cdot) \) with \( i = 1 \ldots J \) such that \( g_i(X) \in \mathbb{X} \) for every \( X \in \mathbb{X} \). Let \( v_i(X) \) denote the price vector: \( v_i(X) = (P_0(g_i(X)), P_1(g_i(X)), \ldots, P_{J-1}(g_i(X)))' \) and assume that the matrix obtained by stacking the \( v_i \) column by column \((v_1, v_2, \ldots, v_J)\) is full rank for all \( X \in \mathbb{X} \).

Consider a portfolio with non-negative payoffs: \( \sum_n w_n P_{n-1}(X) \geq 0 \) for all \( X \in \mathbb{X} \). According to (26), the price of this portfolio for each state is given by:

\[
\sum_n w_n P_n(X) = \frac{1}{J} \sum_i \left( \exp(-m_i(X)) \times \sum_n w_n P_{n-1}(g_i(X)) \right).
\]

Because \( g_i(X) \in \mathbb{X} \), it follows that each term \( \sum_n w_n P_{n-1}(g_i(X)) \) is a possible payoff from the portfolio and must be non-negative. Therefore, the price of the this portfolio must be non-negative. For the price of the portfolio to be zero, each of the summations \( \sum_n w_n P_{n-1}(g_i(X)) = w \cdot v_i(X) \) must be zero, where \( w = (w_1, w_2, \ldots, w_J)' \). That is, \( w \cdot (v_1, v_2, \ldots, v_J) = 0 \). Because \((v_1, v_2, \ldots, v_J)\) is full rank, it follows that the price of the portfolio can only be zero when \( w \equiv 0 \). This means that the portfolio’s payoff must be uniformly zero across all states.
Finally, consider \( g_t(X_t) \equiv g(X_t) \) and \( m_1(X_t) \equiv m(X_t) \) and for \( i > 1, \exp(-m_i(X_t)) \equiv a \) for some constant \( a > 0 \), then we have

\[
P_n(X_t) = P_{n-1}(g(X_t)) \times \exp(-m(X_t)) + a \left( \sum_{i=2}^{J} P_{n-1}(g_i(X_t))/J \right).
\]

(28)

Letting \( a \to 0 \), we obtain the model in Assumption 1.

A.2 Examples

Linear models

Suppose \( X_t \in \mathbb{R}^N \). The following natural specification leads to affine Gaussian TTSMs:

\[
m(X_t) = \delta_0 + \delta_1'X_t
\]

(29)

\[g(X_t) = KX_t,
\]

(30)

where \( \delta_0 \) is a scalar, \( \delta_1 \) is an \( N \times 1 \) vector and \( K \) is an \( N \times N \) matrix. From Proposition 1, yields are linear:

\[
y_{n,t} = \delta_0 + (B_n/n)X_t,
\]

(31)

with \( B_n \) given by the recursion \( B_n = B_{n-1}K + \delta_1' \). For comparison, the \( A_0(N) \) Gaussian DTSM (e.g., Dai and Singleton, 2000 and Duffee, 2002) has a linear short-rate equation and risk-neutral dynamics given by

\[
\begin{align*}
r_t &= \delta_0 + \delta_1'X_t \\
X_{t+1} &= K_0 + K_1QX_t + \epsilon_{t+1},
\end{align*}
\]

(32)

where \( \epsilon_{t+1} \sim N(0, \Sigma) \). The solution for yields in that standard case is given by

\[
y_{n,t} = A_n/n + (B_n/n)X_t,
\]

(33)

with coefficients given by

\[
\begin{align*}
B_n &= B_{n-1}K + \delta_1' \quad (34) \\
A_n &= A_{n-1} + \delta_0 - \frac{1}{2}B_{n-1}\Sigma B'_{n-1}.
\end{align*}
\]

(35)

Clearly, the short rate \( r_t \) and the loadings \( B_n \) are identical between these models. The intercept terms \( A_n \) for \( n > 1 \) are different only because of the convexity correction \( B_{n-1}\Sigma B'_{n-1} \). This Jensen term is negligible in a typical application (see Figure 8).

Quadratic models

The following choice generates linear-quadratic TTSMs:

\[
m(X_t) = \delta_0 + \delta_1'X_t + X_1'd_2X_t
\]

(36)

\[g(X_t) = KX_t,
\]

where \( \delta_0 \) is a scalar, \( \delta_1 \) is a \( N \times 1 \) vector and \( d_2 \) is a \( N \times N \) matrix. From Proposition 1, yields are given by

\[
y_{n,t} = \delta_0 + (B_n/n)X_t + X_1'(C_n/n)X_t,
\]

(37)

45
Figure 8: The Jensen term \(\frac{1}{2}B_{n-1}\Sigma B_{n-1}'\) is negligible in Gaussian models. Differences between the loadings in linear 3-factor models where yield PCA are used as risk factors and parameter estimates are based on the canonical representation in Joslin, Singleton, and Zhu (2011).

where the linear and quadratic coefficients \(B_n\) and \(C_n\) are given by

\[
\begin{align*}
C_n &= K'C_{n-1}K + \delta_2 \\
B_n &= B_{n-1}K + \delta_1'.
\end{align*}
\]

(38)

Compare this with the affine-quadratic DTSMs developed by Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2003), where the short-rate equation is quadratic:

\[
r_t = \delta_0 + \delta'_1X_t + X'_t\delta_2X_t,
\]

and with the risk-neutral dynamics as in (32).\(^{23}\) The solution for yields in this case is given by

\[
y_{n,t} = A_n/n + (B_n/n)X_t + X'_t(C_n/n)X_t,
\]

(39)

where the loadings \(A_n\), \(B_n\) and \(C_n\) are given by the following recursions:

\[
\begin{align*}
C_n &= K'^Q'C_{n-1}\Omega_{n-1}K_1^Q + \delta_2, \\
B_n &= B_{n-1}\Omega_{n-1}K_1^Q + \delta_1', \\
A_n &= A_{n-1} + \delta_0 - \frac{1}{2}log|\Omega_{n-1}| - \frac{1}{2}B_{n-1}\Omega_{n-1}\Sigma B_{n-1},
\end{align*}
\]

(40)

with \(\Omega_{n-1} \equiv (I_N - 2\Sigma C_{n-1})^{-1}\). Comparing loadings in (38) and (40) reveals two differences. First, the term \(B_{n-1}\Omega_{n-1}\Sigma B_{n-1}\) reflects a convexity adjustment. Second, the matrix \(\Omega_{n-1}\) may introduce a wedge between loadings if the quadratic coefficient \(\delta_2\) is “large.” posed in Section 2 in the limit.

A.3 Implementation of Wu and Xia (2016)

Wu and Xia (2016) implement a short-rate equation \( r_t = \max(\underline{r}, s_t) \) that relies on the \( \max \) function. For convenient estimation, it is preferable that the short rate function is invertible. To circumvent this issue, we start with their NA model generated at a daily frequency but that we implement with a monthly sampling frequency. In this case, the one-month short rate can be inverted for all values of the shadow rate \( s_t \).

Assume that the daily risk-neutral dynamics are given by:

\[
X_{t+1} = \mu^Q + \rho^Q X_t + \sigma^Q \epsilon_{t+1},
\]

with \( \epsilon_{t+1} \sim N(0, I) \), that suppose linear shadow rate \( s_t = \delta_0 + \delta_1 X_t \) and that the short rate is given by \( r_t = \max(\underline{r}, s_t) \) where \( \underline{r} \) is the lower bound. We choose the standard normalization \( \mu^Q = 0 \) and \( \rho^Q \) with Jordan form. This implies a monthly dynamics

\[
X_{t+1} = K^Q_t X_t + \Sigma^Q \epsilon_{t+1},
\]

and forward rate starting at time \( n - 1 \) and maturing at time \( n \) are given in closed-form:

\[
f_{n-1,n,t} = \underline{r} + S_n g \left( \frac{A_n + B_n X_t - \underline{r}}{S_n} \right),
\]

where \( g(z) = z\Phi(z) - \phi(z) \). For the purpose of estimation, note that the forward rate pricing function can be easily inverted:

\[
\underline{r} + S_n g^{-1} \left( \frac{f_{n,n+1,t} - \underline{r}}{S_n} \right) = A_n + B_n X_t,
\]

since \( g(\cdot) \) is monotone and, therefore, the inverse function \( g^{-1}(\cdot) \) is well-defined. The coefficients \( S_n, A_n \) and \( B_n \) are given by:

\[
S_n = \sum_{i=(n-1)\times30}^{n\times30-1} \sigma_i, \quad A_n = \sum_{i=(n-1)\times30}^{n\times30-1} a_i, \quad B_n = \sum_{i=(n-1)\times30}^{n\times30-1} b_i.
\]  

(41)

with \( a_i, b_i \) and \( \sigma_i \) given by:

\[
a_n = \delta_0 - \frac{1}{2} \delta_1' \left( \sum_{j=0}^{n-1} (\rho^Q)^j \right) \delta_1, \\
b_n = \delta_1' (\rho^Q)^n, \\
\sigma_n^2 = \sum_{j=0}^{n-1} \delta_1'(\rho^Q)^j \sigma(\rho^Q)^j \delta_1.
\]
and

\[ \rho^Q = K_1^{Q1/30}, \]

\[ vec(\sigma) = \left( \sum_{i=0}^{29} \rho^{Q_i} \otimes \rho^{Q} \right)^{-1} vec(\Sigma), \]

\[ \delta'_1 = t' \left( \sum_{n=0}^{29} \rho^{Q_n} \right)^{-1}, \]

\[ 30\delta_0 = \frac{1}{2} \delta'_1 \left[ \sum_{n=0}^{29} \left( \sum_{j=0}^{n-1} (\rho^{Q_j})^j \right) \Sigma_d \left( \sum_{j=0}^{n-1} (\rho^{Q_j})^j \right)^j \right] \delta_1 + r^Q, \]

so that, at the monthly frequency, the model is fully characterized by the parameter set \( \{ K_0, K_1, \Sigma, r^Q, K_1^Q \} \). Additionally, \( r \) can be fixed to a known value or can be free parameter to be estimated.

### A.4 Simulation Details

#### Data generating process

We use the linear rational term structure models of Filipovic, Larsson, and Trolle (2017) to simulate sample of yields data. In particular, we use their LRSQ(3,3) specification with three factors \( Z_t \) explaining the cross-section of bond yields and three factors \( U_t \) driving unspanned yields volatility. The state 6 \( \times \) 1 state vector \( X_t \) follows an extended square root process:

\[ dX_t = (b - \beta X_t)dt + diag(\sigma) \sqrt{X_t} dB_t, \]

the volatility factors \( U_t \) correspond to the last three elements of \( X_t \):

\[ U_t = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} X_t, \]

and three cross-sectional factors \( Z_t \) correspond to the sum of the first three elements of \( X_t \) and the last three element of \( X_t \):

\[ Z_t = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} X_t. \]

#### Pricing functions

The pricing functions involve five parameters: \( \alpha, \phi, \psi, \theta, \kappa \). The price of a zero-coupon bond with \( \tau \) years to maturity is given:

\[ p_t = e^{-\alpha_\tau} \phi + \psi'(\theta + e^{-\tau\kappa}(Z_t - \theta)) \frac{\phi + \psi'Z_t}{\phi + \psi'Z_t}, \]

which is the ratio of two linear functions of \( Z \), hence the name linear rational. The corresponding yield to maturity is given by:

\[ y_t = -\frac{\log(p_t)}{\tau}, \]
and, taking the limit $\tau = 0$, the instantaneous short rate is given by:

$$ r_t = \alpha + \frac{\psi' \kappa (Z_t - \theta)}{\phi + \psi' Z_t}. $$

### Parameter values

We use parameter estimates and normalized values provided in Filipovic, Larsson, and Trolle (2017) $\alpha = 0.0566$, $\phi = 1$, $\psi' = (1,1,1)'$, $\theta' = (0.2985, 0.5738, 0.1601)'$ and

$$ \kappa = \begin{bmatrix} 0.1895 & 0 & 0 \\ -0.2460 & 0.1280 & 0 \\ 0 & -0.1846 & 0.6616 \end{bmatrix}, $$

$$ \beta = \begin{bmatrix} 0.1895 & 0 & 0 & 0 & 0 & 0 \\ -0.2460 & 0.1280 & 0 & 0 & 0 & 0 \\ 0 & -0.1846 & 0.6616 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1895 & 0 & 0 \\ 0 & 0 & 0 & 0.2460 & 0.1280 & 0 \\ 0 & 0 & 0 & 0 & -0.1846 & 0.6616 \end{bmatrix}, $$

$$ \sigma = \begin{bmatrix} 0.3177 \\ 0.2676 \\ 0.0509 \\ 0.7920 \\ 0.9135 \\ 0.3585 \end{bmatrix}, \quad b \times 1000 = \begin{bmatrix} 44.5325 \\ 0.0076 \\ -0.0222 \\ -0.0222 \\ 12.0333 \\ 0.0078 \end{bmatrix}. $$

### Generating ZLB samples

We discretize the dynamics of $X_t$:

$$ X_{t+\Delta} = \max(X_t + (b - \beta X_t)\Delta + \text{diag}(\sigma \sqrt{\Delta}) \sqrt{X_t} \epsilon_{t+\Delta}, 10^{-8}), $$

where the frequency is daily $\Delta = 1/365$ and where $\epsilon_{t+\Delta}$ is an i.i.d. standard normal variable. Using this dynamics, we first run a burn-in sample of 10 years and then generate a daily sample of 1000 years. Out of this sample, we randomly choose a ZLB episode, defined as any single day for which the short rate is less than or equal to one basis point. Starting from this day, we go back by 30 years and collect end-of-month yields. This design implies that ZLB episodes are typically found toward the end of the simulated samples. We repeat this process until we obtain 100 different simulated samples, each of which featuring the ZLB behavior of interest rates. If needed, we generate another daily sample of 1000 years with different simulation seeds until we reach 100 different simulated samples.