Review of Derivatives Research
https://doi.org/10.1007/s11147-024-09202-3

An affine model for short rates when monetary policy is path dependent

Haitham A. Al-Zoubi[1]

Accepted: 25 March 2024

Abstract

I propose an affine model of short rates that incorporates a random walk with stochastic drift. This framework enables my model to capture the behavior of monetary authorities in the short rate market, allowing for minor deviations while reacting strongly to deviations large enough to threaten production. Importantly, my model facilitates the derivation of closed-form bond prices, thereby providing an analytical solution for bond-option prices. I compare my model with nine standard short rate models found in the literature. Among these, five are single-factor models and four are multifactor models. Remarkably, my model outperforms all competing short rate models, including the constant elasticity of volatility, stochastic mean, and stochastic volatility models. Moreover, it yields interest rate forecasts consistent with common term structure priors and surpasses the performance of the naive random walk model. Additionally, my stochastic mean model can explain the unspanned risks documented in the literature.

Keywords Short rates · Stochastic volatility · Continuous-time estimation · Option options

JEL Classification C15 · C32 · G12 · E43 · E47

1 Introduction

Monetary and financial stability (see Orphanides & Wilcox, 2002; Aksoy et al., 2006) have indicated that monetary policy is path-dependent. Path dependence for short rate models implies that if there is mean reversion, the short rates must revert to a nonstationary mean. Central banks research (see Orphanides & Wieland, 2000) also suggests that optimal monetary policy should deviate from standard linear

✉ Haitham A. Al-Zoubi
halzoubi@alfaisal.edu

[1] College of Business, Alfaisal University, P. O. Box 50927, 11533 Riyadh, Saudi Arabia

Published online: 13 May 2024

policies. For short rate models, this implies that short rates may follow a nonlinear process. Both these results could result from an "opportunistic" monetary policy that adjusts expectations for past forecast errors yet actively counters adverse short rate shocks that are large enough to disrupt production. Empirical evidence supports this view of short rate behavior (see Akso, 2006).

Bauer and Rudebusch (2020) demonstrate that market data incorporates or reflects central bank policy towards inflation by modifying conditioning information, the connection between the stochastic means of macroeconomic series and short rates. Balduzzi et al. (1998) take a conditional information approach to clarifying the news content of shocks to bond yield by modifying conditioning information. Thus, I should expect to see nonlinearity and nonstationary mean reversion in short rate data. These short rates follow a nonlinear process that reverts to a nonstationary trend is consistent with recent empirical work by Fama (2006). Fama finds "…that mean reversion is toward a nonstationary long-term mean." (p. 1). This result indicates that the shocks to short rates are permanent, implying that the mean is nonstationary. Fama also documents that the short rate behavior is mean reverting to this nonstationary trend. This line of financial literature is independent of and consistent with that published in the monetary economics literature discussed above.

In this paper, Ipropose a short rate model that incorporates both the above features. First, my proposed model allows for the very high persistence of interest rates resulting from inflation trends and equilibrium real interest rates. Scholars have documented that nominal interest rates encompass a slow-moving common trend component (Campbell and Shiller, 1987). Al-Zoubi (2019) and Bauer and Rudebusch (2020) conclude that accounting for the trend in the short rate reduces forecast error. Duffee (2011) demonstrates that a random walk component is essential to explain the term structure and that incorporating a random walk component leads to better predictability than existing term structure models. Duffee (2018) also extends the results and finds that variations in news reports concerning expected inflation explain about 10% to 20% of variances of yield shocks. Second, my model preserves the mean-reverting characteristic empirically documented and suggested by opportunistic monetary policy. Yield curves do not explain Time-varying bond risk premiums; this is also true for excess returns (see Cochrane and Piazzesi, 2005). Predictors of excess bond returns include; expected inflation (see Cieslak & Povala, 2015), measures of Treasury bond supply (see Greenwood & Vayanos, 2014), the output gap (see Cooper & Priestley, 2009), and macroeconomic fundamentals (see Joslin et al., 2014). On the other hand, Duffee (2018) reports that expected inflation shocks explain a substantial portion (up to 20%) of bond yield variance.[1]

[1] Liu (2018) provides Out-of-sample evidence of predictability of excess returns for economic growth and inflation. However, Bauer and Hamilton (2017) conclude that this predictability is possibly spurious due to size distortions resulting from persistent regressors and independent variables that are not strictly exogenous. Ang et al. (2008) develop a term structure model with a time-varying risk premium. They find that the inflation risk premium explains the upward sloping nominal term structure. Bekaert et al. (2010) also document persistent interest rates where the inflation target is time varying. Duffee (2002) finds that forecasts from a random walk model outperform the Dai and Singleton (2000) affine model.

My model's contribution is to capture the permanent component of short rates, i.e., I model the nonstationary stochastic mean of the short rate as a random walk (with and without drift).[2] However, current stochastic mean models of the term structure of interest rates generally assume that the stochastic mean has reverted to this point (See Balduzzi et al., 1998). My model is in line with the motivation in Fama (2006) and Bauer and Rudebusch (2020) by assuming time-varying long-run stochastic mean. There are two important features of my model that inspire its relative success in simultaneously pricing bonds and bond options. First, I emphasize members of the affine family of short rate models that are encompassed by Duffie and Kan (1996). The second feature of my model is the dependence of the market price of risk on the state of the macroeconomy. My proposed time-varying stochastic mean can capture the premium associated with expected inflation (see Cieslak & Povala, 2015) and the output gap (see Cooper & Priestley, 2009). This extended specification allows for time-variation in the premium associated with unspanned risks, an element that found to be critical for matching bond and options returns (see Baksh et al., 2023a, b). I refer to my model parameterization as the nonstationary mean model (NSM).

Besides model development, this paper makes several other contributions to the literature. First, my model captures both nonlinearity (because the random walk creates nonlinear drift) and non-stationarity because my mean is stochastic and non-mean reverting to permanent shocks. Second, my model has better in-sample and out-of-sample forecasts than the benchmark models (random walk, single-and multi-factor models). Third, my closed-form solutions for bond and bond option prices exhibit less bias and attenuates the downward bias documented in bond option prices (see Phillips & Yu, 2005; Tang & Chen, 2009). I conclude that the downward bias in current bond option pricing models is due to the short rate mean non-stationarity.

My main finding is that my NSM model, better captures short rate dynamics by incorporating a random walk mean. My findings differ from Durham (2003) and Chan et al. (1992), who find that constant elasticity of volatility (CEV) models are superior in modeling the behavior of short rates. I find that CEV models outperform by capturing the nonstationary mean during unstable periods. This advantage does not exist, when compared against my proposed nonstationary mean model, and my model outperforms the CEV-based models. I also find that my proposed model outperforms stochastic mean and volatility models.

Interestingly, in-sample and out-of-sample results show that my model outperforms the Chen (1996) three-factor model. Also, as the literature predicts, it is surprisingly hard to consistently beat the naive random walk prediction, which predicts that the short rate has a martingale behavior. Nevertheless, I note that the forecasts implied by my model are significantly more accurate than the naive random walk model.

Additionally, using my model, I derive closed-form solutions for both bond and bond-option securities. Closed-form solutions provide several benefits, including: (i) allowing for consistent estimation and (ii) allows a full description for bond

[2] Balduzzi et al. (1998) assume a stochastic mean. However, their stochastic mean is mean-reverting. My contribution is that I introduced a model with a random walk stochastic mean.

and option prices over each maturity level without approximation error. I utilize my model to investigate estimation bias in bond and option prices. As a preview of my results, I outline three main findings. First, my model reduces the bias in terms of mean, and reduces standard deviation resulting in a substantial overall gain in RMSE. Second, I relate my findings to the results of Phillips and Yu (2005) and Tang and Chen (2009). If the random walk component is the primary source of misspecification bias, one would expect higher option prices for the NSM model. I find that the NSM model increases the implied option prices by 0.036–9.4% for in-the-money options and by 68% for at-the-money options. Third, the transmission of these biases onto money markets exhibits considerable time variation. Results show a greater bias in estimating bond and option prices with higher interest rates. Lower liquidity properties such as time deposits, commercial paper, and treasuries. Furthermore, bias in the stochastic mean is more considerable when interest rates are volatile.

I further contribute to the literature on the relation between the transitory and permanent components of the short rate. I do this by considering two different trend filtering methods to produce the random walk component. The first method assumes that the permanent component follows a random walk with drift correlated with the stationary component. Intuitively, the stochastic mean depends on the volatility of the short rate. The second type of method assumes that the permanent component is a driftless random walk which is independent from the resulting cyclical component. In this regard, I used three methods. The first is based on the filtering technique introduced in Hodrick and Prescott (1997) and developed in Ravn and Uhlig (2002). The second method is that suggested by Hamilton (2018). The third, is the boosted HP (bHP) method suggested by Phillips and Shi (2021).

The remainder of the paper is structured as follows. Section 2 develops the model and derives closed-form solutions for bond prices. Section 3 develops the estimation techniques. Sections 4 and 5 review the benchmark models and describes the data, respectively. Empirical results are presented in Sect. 6, while Sect. 7 estimates resulting Bond and option prices. Section 8 offers conclusions.

2 Model setup

In this section I derive analytic representations for the short rate and bond price. In the model, the short rate has a Gaussian random walk component. I refer to my model as the NSM model.

2.1 Short rate with a random walk

Assumption 2.1 *As in Duffie and Kan (1996), risk is produced by n independent Brownian motions. The instantaneous short rate, $\dot{r}_t$, is affine in the state:*

$$\dot{r}_t = \partial_0 + \partial_1 r_t$$

The risk-neutral continuous-time analogue of the short rate, r_t, and the Gaussian random walk component, μ_t, can be written as

$$dr_t = -\alpha_1(\mu_t - r_t) + \sigma dZ_t \quad \text{and} \quad d\mu_t = \sigma_\mu dW_t = \eta_t. \tag{1}$$

where *Zt* and *Wt* are two independent Brownian motions and η_t is a white noise uncorrelated with the stationary component of short rate $c_t = r_t - \mu_t$. The corresponding risk-neutral measure determines bond price.

The specification in (1) implies the following limiting distributions of the short rate.

Proposition 2.1 *Let Assumption 2.1holds. Then the expected value of the short rate is given by.*

$$E\left[-\int_t^T r_u du\right] = \frac{r_t - \mu_t}{\alpha_1}\left(1 - e^{\alpha_1(T-t)}\right) - \mu_t(T-t).$$

Proof See Appendix 1. □

Proposition 2.2 *Let Assumption 2.1 holds. Then the variance of the short rate is given by.*

$$\begin{aligned} Var\left[-\int_t^T r_u du\right] &= Var\left[\int_t^T (c_u + \mu_u) du\right] \\ &= \frac{\sigma_c^2}{2\alpha_1^3}\left(2\alpha_1(T-t) + 3 - 4e^{\alpha_1(T-t)} + e^{2\alpha_1(T-t)}\right) + \sigma_\mu^2 \frac{(T-t)^3}{6}. \end{aligned}$$

where σ_c^2 and σ_μ^2 are the variances of the stationary and permanent components, respectively.

Proof See Appendix 1. □

Note that the variance of the short rate of my model can be written as:

$$= \sigma_{VAS}^2 + \sigma_\mu^2 \frac{(T-t)^3}{6}$$

where σ_{VAS}^2 is the variance of the short rate under the Vasicek (1977) model.

2.2 Bond price with a random walk

The derivation details of the bond pricing formula are contained in Appendix 1. In this section, I only show the main results.

Theorem 2.1 *Let the expected value and variance of the short rate as in Propositions 2.1 and 2.2. Define the price of a zero-coupon bond with maturity T at time t,P (t, T), as.*

$$P(t,T) = E\left[exp\left(-\int_t^T r_u du\right)|F_t\right] = E\left[exp\left(-\int_t^T r_u du\right)|r_t\right]$$

where $\{F_t\}$ *is standard filtration.*

$$P\left(t,T,r_t\right) = Exp\left(\left[-\int_t^T r_u\left(r_t\right)du\right] + \frac{1}{2}Var\left[-\int_t^T r_u\left(r_t\right)du\right]\right)$$

by the results from Propositions 2.1. and 2.2. I get

$$= Exp\left(A(t,T)r_t - \mu_t(A(t,T) + (T-t)) + B(t,T) + D(t,T)\right)$$

where

$$A(t,T) = \left(\frac{1-e^{\alpha_1(T-t)}}{-\alpha_1}\right)$$

$$B(t,T) = \frac{(A(t,T) + (T-t))}{2\alpha_1^2} + \frac{A(t,T)^2}{4\alpha_1}\left[\frac{\sigma^2}{(1+q)}\right]$$

and.

$$D(t,T) = \left(\frac{q\sigma^2(T-t)^3}{12(1+q)}\right).$$

where q is the signal-to-noise ratio defined as

$$q = \frac{\sigma_{d\mu}^2}{\sigma_c^2} = \frac{\sigma_\eta^2}{\sigma_c^2} = \frac{\sigma_\mu^2}{\sigma_c^2}.$$

Proof See Appendices 1 and 2. □

Remark 2.1 *In the case that the stochastic mean,* μ_t*, is a target that the Federal reserve wants the rate to converge to, then* μ_t *depends on the current interest rate environment, i.e. whether it is volatile or not. Hence, I also build a correlation between the driving Brownian motions shown in Assumption 2.2 below.*

Assumption 2.2 *The continuous-time analogue of the short rate,* r_t*, and the Gaussian random walk component,* μ_t*, can be written as.*

$$dr_t = -\alpha_1(\mu_t - r_t) + \sigma dZ_t \quad \text{and} \quad d\mu_t = \sigma_\mu dW_{t=}\eta_t.$$

.

I now assume that Zt and Wt are correlated, such that $dZtdWt = \rho dt$, where the correlation, ρ, is some constant. Thus η_t is a white noise correlated with the stationary component of short rate $c_t = r_t - \mu_t$.

The corresponding bond price is given by

$$P\left(t, T, r_t\right) = Exp\left(\left[-\int_t^T r_u\left(r_t\right)du\right] + \frac{1}{2}Var\left[-\int_t^T r_u\left(r_t\right)du\right]\right)$$
$$= Exp\left(A(t, T)r_t - \mu_t(A(t, T) + (T - t)) + B(t, T) + D(t, T)\right)$$

Where

$$A(t, T) = \left(\frac{1 - e^{\alpha_1(T-t)}}{-\alpha_1}\right)$$

$$B(t, T) = \left[\frac{A(t, T) + (T - t)}{2\alpha_1^2} + \frac{A(t, T)^2}{4\alpha_1}\right]\sigma_c^2$$

and

$$D(t, T) = \frac{\sigma_\mu^2(T - t)^3}{12}.$$

Proof See Appendix 1. □

2.3 Term structure with a random walk

From proposition (2.2), the zero rate at time t for period $T - t$ is

$$R(t, T) = -\frac{1}{T - t}A(t, T)r_t + \frac{1}{T - t}[A(t, T) + (T - t)]\mu_t - \frac{1}{T - t}B(t, T) - \frac{1}{T - t}D(t, T))$$

In my model, one needs to specify only the parameters of the model and the permanent component of the short rate to determine the entire term structure at time t. The zero rate $R(t, T)$ is linearly dependent on both r_t and μ_t. This implies that r_t and μ_t determine the level of the term structure at time t. The shape of the term structure depends solely on T.

2.4 Nonstationary stochastic mean as a proxy for unspanned risks

The consensus in literature is that the spanning hypothesis, the yield curve contains all information relevant for predicting future bond returns, can be rejected by the observed data (see, for example, Bakshi et al. (2023a, b).[3] The evidence comes from predictive regressions for bond yields on several macroeconomic factors, controlling for the level, slope, and curvature of the yield curve. The most important factors that have been found to help predict bond returns in such regressions are trends in inflation (Cieslak & Povala, 2015), the output gap (Cooper & Priestley, 2009), and economic growth and inflation (Joslin et al., 2014). In this section I provide the link between these macroeconomic factors and my proposed stochastic mean.

Remark 2.2 *In the case that the Federal reserve follows Taylor's rule, then the stochastic mean, μ_t, is the natural rate of interest that contains information about the output gap and the inflation gap. Accordingly, the permanent components of inflation and output can be explained by the stochastic mean.*

Following Taylor (1993) the deviations of the short rate from the desired nominal policy rate, μ_t, are proportional to deviations of a target variables Z_t, from its target Z_t^*.

$$r_r - \mu_t = \ + \acute{\emptyset}(Z_t - Z_t^*)$$

Among the alternative target variables suggested, Taylor (1993) examined two as more likely to result in better economic conditions. These two factors are the inflation gap, $\left(\pi_t - \pi_t^*\right)$, studied by Cieslak & Povala (2015) and the output gap, $\left(y_t - y_t^*\right)$ studied by Cooper and Priestley (2009). The variation of the short rate from its baseline path can be written as:

$$r_r - \mu_t = \ + \acute{\emptyset}_\pi\left(\pi_t - \pi_t^*\right) + \acute{\emptyset}_y\left(y_t - y_t^*\right)$$

In my model, the transitory component of short rate, $c_t = r_t - \mu_t$, captures these two unspanned factors as the permanent component captures trends in inflation, π_t^*, and potential output, y_t^*.

3 Model estimation

As in Fama (2006) I assume the level of short rates is mean reverting to its long-term conditional mean. I also assume that this long-term conditional mean is subject to permanent shocks. Thus, I decompose the short rate process, r_t, into its two constituent parts: (i) a random walk stochastic mean (μ_t) and (ii) a transitory component with zero mean (c_t):

$$r_t = c_t + \mu_t$$

[3] Bakshi, Gao, and Xue (2023) use of options on the 10- and 30-year Treasury bond to estimate the expected return of bond futures. These measures exhibit forecasting ability for future returns, surpassing the predictive power of the level, slope, and curvature variables typically derived from the yield curve.

I consider two different trend filtering methods to produce the random walk component that is consistent with Assumptions 2.1 and 2.2. The first method is the Beveridge and Nelson (1981) decomposition which assumes that the permanent component follows a random walk with drift, and it is correlated with the stationary component. Intuitively, the stochastic mean is depending on the volatility of the short rate (Assumption 2.2). The second group of methods assume that the permanent component is driftless random walk which is independent from the resulting cyclical component (Assumption 2.1). In this regard, I used three techniques, the first is based on the filtering technique introduced in Hodrick and Prescott (1997) and developed in Ravn and Uhlig (2002). The second method of trend filtering method used is suggested in Hamilton (2018). He argues that the HP filter generates spurious dynamics, is two-sided and its application does not fit with random walk. The third method is the boosted HP (bHP) method suggested in Phillips and Shi (2021). The proposed method delivers consistent estimation of stochastic trends that involve unit root processes and shows better characteristics than the autoregression method of Hamilton.

3.1 Case 1: correlated permanent and transitory components

3.1.1 Beveridge and Nelson (BN) decomposition

I define the random walk with drift component of BN decomposition as

$$\mu_t = \lim_{M\to\infty} E\left[r_{t+M-}M_{Tx}/\Omega_t\right]$$

where $Tx = E\left(\Delta r_t\right)$ is the deterministic drift and Ω_t is the information set.

Empirically, the permanent component is often calculated using an Autoregressive integrated moving average (ARIMA) model that is designed to capture the autocovariance structure of the short rate. The state space representation for the short rate is given by:

$$r_t = c_t + \mu_t$$

$$\mu_{t+1} = \Lambda + \mu_t + \eta_{t,}\eta_t \sim NID\left(0, \sigma_\eta^2\right)$$

$$c_{t+1} = \rho c_t + \sum_{j=1}^{p} \rho_j c_{t-j} + v_{t+1}, v_t \sim NID\left(0, \sigma_v^2\right)$$

$$Corr\left(\eta_t, v_t\right) = \rho_{\eta v}$$

I follow the approach in Anderson et al. (2006) and write the serially uncorrelated innovation of the transitory component as

$$v_t = b\eta_t,$$

where b is a parameter that lets permanent and transitory innovations to have different signs and variances even though being driven by identical underlying shock. Note that this representation implies that the permanent and transitory innovations are observable and can be measured using the forecast error from the reduced-form ARIMA representation for r_t. This framework uses perfect correlation between permanent and transitory components which implies that shocks to short rate will affect both components as suggested in Assumption 2.2. Stock and Watson (1988) point out that policy makers need to recognize the substantial trend component, even if they are principally interested in short term fluctuations. Empirically, the short-term fluctuations in the transitory component of the short rate are almost insignificant, and its serial dependence properties are usually very weak.

3.2 Case 2: uncorrelated permanent and transitory components

3.2.1 HP decomposition

As in Fama (2006) I assume the level of short rates is mean reverting to its long-term conditional mean. I also assume that this long-term conditional mean is subject to permanent shocks. Thus, I decompose the short rate process, r_t, into its two constituent parts: (i) a random walk stochastic mean (μ_t) and (ii) a transitory component with zero mean (c_t):

$$r_t = c_t + \mu_t$$

I specify the random walk component, μ_t, by minimizing the loss function as in Hodrick and Prescott (1997) and Ravn and Uhlig (2002):

$$\sum\nolimits_{t=1}^{T} \left[r_t - \mu_t\right]^2 + q\sum\nolimits_{t=2}^{T} \left[\nabla^2 \mu_t\right]^2, \tag{2}$$

Let L be the lag operator, which implies that $\nabla = 1 - L$, the second difference lag operator. T is the sample size and q is the signal-to-noise ratio. I follow Ravn and Uhlig (2002) and chose $q = 43,\ 200$ for monthly data. The loss function in Eq. (2) imposes two penalties. First, mean reversion of the expected return on bonds is penalized via the first term, while the variability in the mean, which is in part due to the Federal Reserve actions to control money supply, is penalized by the second term.

The minimization problem in Eq. (2) possesses a unique solution. The first order condition for μ_t is:

$$-2(r_t - \mu_t) - 4\frac{(\mu_{t+1} - 2\mu_t + \mu_{t-1})}{q} + 2\frac{(\mu_t - 2\mu_{t-1} + \mu_{t-2})}{q} + 2\frac{(\mu_{t+2} - 2\mu_{t+1} + \mu_t)}{q} = 0$$

which simplifies to:

$$r_t = \left(\frac{1}{q} \left(1 - L^{-1}\right)^2 (1 - L)^2 + 1 \right) \mu_t. \tag{3}$$

The above equation allows us to estimate μ_t (and, in turn, c_t) using the short rate series r_t, subject to stipulating q, I describe the choice of q at the end of the section.

The discrete-time analogue of the short rate in my model can be written as

$$r_{t+1} - r_t = \alpha_1 \left(r_t - \mu_t\right) + \varepsilon_{t+1} \tag{4a}$$

$$\Delta \mu_{t+1} = \eta_{t+1} \sim NID\left(0, \sigma_\eta^2\right)$$

$$\varepsilon_{t+1} \sim NID(0, \sigma^2)$$

where $NID(0,\ \sigma^2)$ denotes normally and independently distributed with mean zero and variance σ^2.

My analysis with a nonstationary mean can be motivated by the work of de Jong and Sakarya (2016). They analyze the decomposition in my framework and show that the transitory component follows weak dependence properties when the procedure is applied to a stationary mixing process with a random walk. Accordingly, suppose that the short rate is a stationary mixing stochastic process with a random walk. The resulting transitory component from Eq. (3), c_t, possesses weak dependence properties with mean zero. Thus, c_t is an AR(p) process:

$$c_{t+1} = \rho c_t + \sum_{j=1}^{p} \rho_j c_{t-j} + \nu_{t+1}, \nu_t \sim NID\left(0, \sigma_\nu^2\right) \tag{4b}$$

with

$$\Delta \mu_{t+1} = \eta_{t+1} \sim NID\left(0, \sigma_\eta^2\right) \tag{4c}$$

The error terms ν_{t+1} and η_{t+1} are mutually independent, hence:

$$E\left(\nu_t \eta_s\right) = 0, \quad \forall t, s,$$

Subtracting r_t from both sides of Eq. (2) after taking one period ahead versions and substituting for the Eqs. (4b) and (4c), I get:

$$r_{t+1} - r_t = \mu_{t+1} - \mu_t + (\rho - 1)c_t + \sum_{j=1}^{p} \rho_j c_{t-j} + \nu_{t+1},$$

Defining $\alpha_1 = \rho - 1$, I have:

$$r_{t+1} - r_t = \alpha_1 c_t + \sum_{j=1}^{p} \rho_j c_{t-j} + \varepsilon_{t+1} \tag{5a}$$

$$r_{t+1} - r_t = \alpha_1 (r_t - \mu_t) + \pi_t \tag{5b}$$

where $\varepsilon_{t+1} = \eta_{t+1} + \nu_{t+1}$ such that $\varepsilon_{t+1} \sim NID(0,\sigma^2)$. Equation (5b) follows from Eq. (5a).

$$\pi_{t+1} = \sum_{j=1}^{p} \rho_j c_{t-j} + \varepsilon_{t+1}$$

is weakly dependent under general conditions of the lag polynomials given by Eq. (5a) and Eq. (5b). Note that Eq. (5b) can be written as:

$$r_{t+1} = \alpha r_t + u_{t+1} \tag{6}$$

Since Eq. (6) is identical to Eq. 1 in Phillips (1987) and$u$ satisfies Assumption 2.1 in Phillips (1987), my estimation of the model parameter is consistent even with error terms that exhibit serial correlation and heteroskedasticity (see Phillips, 1987, Theorem 3.1). Since the parameter estimate of Eq. (5b) is consistent, Phillips and Perron (1988) suggest the Parzen estimator of Gallant (1987) as a consistent estimate for parameter variance.

Define q, a smoothing constant to estimate Eq. (3), to be the variance ratio of η_{t+1} and c_{t+1}. I derive the expression for q in Appendix 2. Thus, q can be written as:

$$q = \sigma_\eta^2 / \sigma_v^2 = \sigma_\eta^2 / \left(1 - \rho^2 - \sum_{j=1}^{p} \rho_j^2\right) \sigma_c^2.$$

3.2.2 Hamilton decomposition (Hamilton, 2018)

I replicate each estimation but using the filtering approach suggested in Hamilton (2018). Specifically, the random walk component is approximated by estimating Eq. (4a) as

$$r_{t+1} - r_t = \alpha_1 (r_t - \mu_t) + \varepsilon_{t+1}$$

where

$$\mu_t = b_0 + b_1 r_{t-7} + b_2 r_{t-8} + b_3 r_{t-9} + b_4 r_{t-10}$$

and

$$c_t = r_t - \mu_t.$$

Hamilton (2018, proposition 4) demonstrates that the series $c_t = r_t - \mu_t$ is stationary given that fourth differences of the series r_{t+1} are stationary. Hamilton method has several advantages over HP filter. First, c_t is a one-sided filter. It is often criticized that the stationary component resulting from filter is HP is two sided-filter. Second, c_t resulting from Hamilton method is uncorrelated with the lagged values Hamilton (2018, footnote 17). Third, the estimated value of c_t is model-free as the long as the fourth differences of the process r_{t+1} are covariance stationary.

3.2.3 Boosted HP (bHP) decomposition (Phillips & Shi, 2021)

Phillips and Shi (2021) propose a solution to establish consistent estimation of stochastic trends that involve unit root processes. If the transitory component c_t still displays trending behavior after HP filtering, they suggest applying the HP filter to c_t to remove the residue trend residual. After m fitting, the transitory component can be written as

$$c^{(m)} = \left(I_n - S\right)c^{(m-1)} = \left(I_n - S\right)x$$

$$f^{(m)} = x - c^{(m)}$$

where m is the number of iterations performed. I follow Phillips and Shi (2021) and used the Bayesian Information Criterion (BIC) as a terminating criterion to stop the iteration.

Phillips and Shi (2021) show that the bHP filter consistently estimates the stochastic trend. This result delivers two advantages: First, it indicates that the Cogley and Nason (1995) critique of the possible presence of spurious cycles no longer holds (see Phillips and Sainan, 2021) for further discussion). Second, the boosting procedure is valid and appropriate when applied to a random walk, as it can accelerate the convergence to the actual trend. Phillips and Shi (2021) also show that the bHP filter numerically and empirically outperforms Hamilton's (2018) autoregressive filter.

3.3 Measurement error

Cochrane and Piazzesi (2009, pp.14 find that the better performance of the moving of the forward rate is attributable to IID measurement errors. Cochrane and Piazzesi (2005) argue that:

> Since bond prices are time-t expectations of future nominal discount factors, it is very difficult for any economic model of correctly measured bond prices to produce dynamics in which lagged yields help to forecast anything. If, however, the risk premium moves slowly over time but there is measurement error, moving averages will improve the signal to noise ratio on the right-hand side.

Note that the error term $\pi_{t+1} = \sum_{j=1}^{p} \rho_j c_{t-j} + v_{t+1} + \eta_{t+1}$ resulting from the HP, bHP, and BN filters is serially correlated. Estimating (5b) directly results in measurement error. For example, the transitory component of the HP filter is weakly dependent stochastic process (de Jong & Sakarya, 2016) and it is an ARMA process in the case of BN filter. Note that, However, that the transitory component of Hamilton's filter is unpredictable (See Hamilton, 2018, footnote 17).

I propose a version of Eq. (5b) with White noise errors. The idea stems from the following system derived from Eq. (4b) and Eq. (5a):

$$\sum_{j=1}^{p} \rho_j c_{t-j} + v_{t+1} = c_{t+1} - \rho c_t$$

$$\pi_{t+1} = \sum_{j=1}^{p} \rho_j c_{t-j} + v_{t+1} + \eta_{t+1}.$$

It follows that

$$\pi_{t+1} = c_{t+1} - \rho c_t + \eta_{t+1}$$

and the variance follows immediately

$$\sigma_\pi^2 = \left(1 - \rho^2\right)\sigma_c^2 + \sigma_\eta^2.$$

Because $\sigma_c^2 = \frac{\sigma_v^2}{1-\rho^2-\sum_{j=1}^{p}\rho_j^2}$ I obtain,

$$\sigma_\pi^2 = \frac{\left(1 - \rho^2\right)}{\left(1 - \rho^2 - \sum_{j=1}^{p}\rho_j^2\right)}\sigma_v^2 + \sigma_\eta^2.$$

Defining $\Omega = 1 - \rho^2$ and $\phi = -\sum_{j=1}^{p}\rho_j^2$ we can write,

$$\sigma_\pi^2 = \frac{\Omega}{\Omega + \phi}\sigma_v^2 + \sigma_\eta^2,$$

which can be written as,

$$\sigma_\pi^2 = -\frac{\phi}{\Omega + \phi}\sigma_v^2 + \sigma^2, \tag{7}$$

Because $q = \frac{\sigma_\eta^2}{\sigma_v^2}$ we obtain,

$$\sigma_v^2 = \left(\frac{1}{1+q}\right)\sigma^2.$$

Plug this in (7) we get,

$$\sigma_\pi^2 = \left(1 - \frac{\phi}{(\Omega + \phi)(1 + q)}\right)\sigma^2.$$

Because $\Omega = 1 - \rho^2$ and $\beta = \rho - 1$, we can write: $\Omega = {\alpha_1}^2 - 2\alpha_1$, then we have,

$$\sigma^2 = \left(\frac{\left({\alpha_1}^2 - 2\alpha_1\right) + \frac{q}{(1+q)}\phi}{\left(\phi + {\alpha_1}^2 - 2\alpha_1\right)}\right)\sigma_\pi^2. \tag{8}$$

I derive the expression for q in Appendix 3.
I estimate my model as

$$r_{t+1} - r_t = \alpha_1\left(r_t - \mu_t\right) + \pi_{t+1} \tag{9}$$

$$E\left[\pi_{t+1}\right] = 0 \text{ and } \sigma_\pi^2 = \left(\frac{\left(\alpha_1^2 - 2\alpha_1\right) + \frac{q}{(1+q)}\phi}{\left(\phi + \alpha_1^2 - 2\alpha_1\right)} \right)\sigma^2$$

If the HP and bHP filters are used.

Also, I estimate the model as

$$E\left[\pi_{t+1}\right] = 0 \text{ and } \sigma_\pi^2 = \left(1 - \frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)}x\right)\sigma^2, \text{ where } x = \frac{\sigma_v^2}{\sigma^2}$$

if the BN filter is used.

Note that the parameter φ captures the AR terms in the transitory component equation regardless the number of lags. By Eq. (8) the coefficients relating the stochastic mean of the model depends on the parameter $\varphi = \sum_{j=1}^{p} \rho_j^2$ which, in turn, describes the sensitivity of the estimated parameters to the serial correlations in the transitory component. I use Eq. (8) for reasons of measurement errors.

I compute the nonstationary mean of HP and bHP by choosing the smoothing constant q such that $1/q = 43,200$. This is equivalent to the Ravn and Uhlig (2002) adjustments of the third power of the frequency of observations.[4]

I estimate my model using the GMM developed in Chapman and Pearson (2000) and Chan et al. (1992). Details for the GMM estimations used in this paper appear in Appendix 4. I show that my main results are robust to using the four methods. However, using the BN and bHP methods lead to additional impressive gains in the in- and out-of-sample forecasting power.

4 Benchmark models

4.1 Single-factor models

To calibrate and measure the success of my model, I compare it with five alternative single-factor models and three multifactor models of the short rate family. The equation for the single-factor short rate

$$dr_t = \left(\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1}\right)dt + \sigma r_t^\gamma dZ_t \tag{10}$$

where r_t is the short rate; Z_t is a standard Brownian motion; α_i for $i = 0$ to 3 are parameters that capture the drift of the process; and σ and γ are parameters that capture the volatility of the process, which may depend on the short rate (captured by γ). Each single-factor model studied in this paper can be represented by Eq. (10) by incorporating constraints on the parameters α, σ, and γ. The single factor models I benchmark against include: the VAS model (Vasicek, 1977), the CIR model (Cox et al., 1985), the CKLS model (Chan et al., 1992), the AG model (Ahn & Gao,

[4] Ravn and Uhlig (2002) point out, the standard constant for monthly data is 1/129,600 or 1/43,200 when the fourth and third power of the number of months in a quarter are used, respectively.

1999), and the CP model (Chapman & Pearson, 2000). Table1 summarizes the various benchmark single-factor models used in this paper.

The above models have the following properties. Single-factor models like VAS, CIR, CP, AG, and CKLS assume a stationary short rate. The CKLS (heteroscedastic) model outperforms the VAS, CIR, CP, and AG models, however, CKLS does not use a stochastic process that allows for the derivation of closed-form bond prices. The conclusion of CKLS is that volatility is more important than the drift in capturing short rates.

4.2 Multifactor models

In addition to single factor models, I also compare my model to three multifactor models, including.

1) The two factor stochastic mean model (BDF) of Balduzzi et al. (1998) is given by:

$$dr_t = \alpha_1\left(r_t - \theta_t\right)dt + \sqrt{\sigma_0 + \sigma_1}r_t dZ_t$$

where r_t is the short rate, θ_t is the stochastic mean, and Z is a standard Brownian motion. In turn, θ_t evolves over time according to the stochastic differential equation

$$d\theta_t = \left(\alpha_{\theta 0} + \alpha_{\theta 1}\theta_t\right)dt + \sqrt{s_0 + s_1\theta_t}dW_t$$

where Z_t, W_t are two independent Brownian motions. The stochastic mean θ_t can be proxied by:

$$\theta_t = \alpha_{\theta 0} + \alpha_{\theta 1}\left[B\left(T_2\right)T_1 y\left(T_2\right) - B\left(T_1\right)T_2 y\left(T_1\right)\right]$$

where

$$B(T) = \frac{2\left(e^{\delta T} - 1\right)}{(\delta + k)\left(e^{\delta T} - 1\right) + 2\delta}$$

and

$$\delta = \sqrt{\left(\alpha_1^2 + \sigma_1^2\right)}$$

where $y\left(T_1\right)$ and $y\left(T_2\right)$ are two bond yields of maturities T_1 and T_1. I follow Balduzzi et al. (1998) and consider two cases. Both cases result from specific parameter choices; first by setting $\sigma_1 = 0$ I get the stochastic mean Vasicek model and second by setting $\sigma_0 = 0$ I get the stochastic mean CIR model.

2) Heston (1993) defines a two-factor stochastic volatility model as:

$$dr_t = \left(\alpha_0 + \alpha_1 r_t\right)dt + \sqrt{V_t}dZ_t$$

Table 1 Short-rate model parameters

	Model	α_0	α_1	α_2	α_3	γ	σ	σ_0	σ_1	α_{v0}	α_{v1}	σ_v	$\alpha_{\theta 0}$	$\alpha_{\theta 1}$	σ_θ
Single factor models	VAS	+	–	0	0	0	+	0	0	0	0	0	0	0	0
	CIR	+	–	0	0	0.5	+	0	0	0	0	0	0	0	0
	CLKS	+	–	0	0		+	0	0	0	0	0	0	0	0
	AG	0	+	–	0	1.5	+	0	0	0	0	0	0	0	0
	CP						+	0	0	0	0	0	0	0	0
Multi factor models	BDF–VAS	0	–	0	0	0	0	+	0	0	0	0	0	0	0
	BDF–CIR	0	–	0	0	0.5	0	0	+	0	0	0	0	0	0
	HEST	+	–	0	0	0	1	0	0				0	0	0
	CHEN	0	–	0	0	0	1	0	0						

This table provides a list of short rate models and their respective parameterization that are com- pared to my proposed stochastic mean model. Single factor models are nested within the following general model:

$dr_t = \left(\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1}\right)dt + \sigma r_t^{\gamma} dZ_t$

where r_t is the short rate; Z_t is a standard Brownian motion; α_i for $i=0$ to 3 are parameters that capture the drift of the process; and σ and are parameters that capture the volatility of the process. The single factor models I benchmark against include: the VAS model (Vasicek, 1977), the CIR model (Cox et al., 1985), the CLKS model (Chan et al., 1992), the AG model (Ahn & Gao, 1999), and the CP model (Chapman & Pearson, 2000).

Multifactor models of Chen (1996) and Heston (1993) are nested within the following general model:

$dr_t = \alpha_1\left(r_t - \theta_t\right)dtdt + \sqrt{V_t}dZ_t$

$dV_t = \left(\alpha_{v0} + \alpha_{v1} v_t\right)dt + \sigma_v \sqrt{V_t}dW_t$

$d\theta_t = \left(\alpha_{\theta 0} + \alpha_{\theta 1}\theta_t\right)dt + \sigma_\theta \sqrt{\theta_t}dX_t$

Multifactor BDF-VAS and BDF-CIR models (Balduzzi et al., 1998) are nested within the following general model:$dr_t = \alpha_1\left(r_t - \theta_t\right)dt + \sqrt{\sigma_0 + \sigma_1 r}dZ_t$

where θ_t is calculated from long term yields of two bonds. Each model studied in this paper can be represented by this general short rate model by incorporating con- straints on the parameters. For unconstrained parameters + and—give the expected sign of a coefficient if the model is to exhibit mean reversion. A blank entry implies that the parameter is unconstrained in that model

$$dV_t = \left(\alpha_{v0} + \alpha_{v1} v_t\right) dt + \sigma_v \sqrt{V_t} dW_t$$

where the interest rate has mean reversion speed of mean α_1 and the mean reversion speed of the variance α_{v1}. To avoid nonstationarity, both these parameters are required to be positive. The mean of the variance is given by α_{v0} and the volatility of the variance is given by α_{v1}. As is standard, both Z_t and W_t represent scalar Brownian motions over a probability measure. Z_t and W_t are allowed to be correlated. The SV model is explored further by Chen (1996), Andersen, and Lund (1997) and Ait-Sahalia and Kimmel (2007).

3) Chen (1996) three-factor Model (CHEN) is given by:

$$dr_t = \alpha_1 \left(r_t - \theta_t\right) dt + \sqrt{V_t} dZ_t$$

$$dV_t = \left(\alpha_{v0} + \alpha_{v1} v_t\right) dt + \sigma_v \sqrt{V_t} dW_t$$

$$d\theta_t = \left(\alpha_{\theta 0} + \alpha_{\theta 1} \theta_t\right) dt + \sigma_\theta \sqrt{\theta_t} dX_t$$

where the stochastic volatility of the short rate is represented by v_t and the stochastic mean of the short rate is represented by θ_t. Z_t, W_t, and X_t are independent Brownian motions. The Chen (1996) model I implement is further explored in Dai and Singleton (2000).

4) Al-Zoubi (2019) two factor integrated random walk I(2) model is given by

$$dr_t = \alpha_1 \left(r_t - \mu_t\right) dt + \sigma dZ_t$$

$$d\mu_t = \vartheta_t dt, d\vartheta_t = \sigma_\mu dW_t$$

where the stochastic mean is the result of an integrated random walk I(2) process. The Al Zoubi (2019) model maintains resilience as it assumes that older shocks induce a stronger effect than newer shocks.

5 Data

I use three-month Treasury bill rates and bond yields for 1- and 2- year maturities collected from the Federal Reserve Economic Database (FRED). Chicago Fed National Activity Index, industrial production index, and of inflation (1-year CPI inflation expectations) from the FRED available from the Federal Reserve Bank of St. Louis website. I also use discount-bond prices to compute continuously compounded yields for 1- and 2-year maturities. The discount-bond prices used to compute the continuously compounded yields for the 1-, 2- and 3-year maturities are from the Center for Research in Security Prices (CRSP) FAMABLIS file. I collect data from July 1952 through December 2018 for the interest rates and from January 1982 to December 2023 for the unspanned risk factors.

Table 2, Panel A provides the following summary statistics: number of observations (N), mean, standard deviations (Std Dev), and the first five lagged autocorrelations ($\rho_i, i=1$ to 5) for both the yield and the changes in the yield. The mean of the short-term riskless rate over the period is 3.49% with a standard deviation of 3.17%. For the change in yield, the corresponding numbers are a mean of 0.00% with a standard deviation of 0.36%. Correlation factors for the yield rate are close to 1, while those for changes in yield are small, with the exception of the first lag which is 33.61%.

The Augmented Dickey-Fuller (ADF) test in Dickey and Fuller (1979) and the Augmented Weighted Symmetric (WS) test in Pantula et al. (1994) are provided in Table 2 Panel B.[5] Both the ADF and WS tests do not reject the null of a unit root as p-values are above the 5% significance level.[6] Both the ADF and WS test suggest the short rate is persistent.[7]

Figures 1, 2, and 3 plot the short rate, its nonstationary mean, and its stationary component using the BN, HP, bHP and Hamilton procedures, respectively.[8]

6 Empirical results

6.1 In-sample forecast comparison

This section provides parameter estimates for the benchmark models and for my nonstationary stochastic mean (NSM) model. I use methods developed in Hodrick and Prescott (1997), Beveridge and Nelson (1981), Phillips and Shi (2021), and Hamilton (2018) to locate the nonstationary means. These methodologies are described in Sect. 3. I estimate models using GMM technique of Hansen (1982). The orthogonality conditions for the CP and AG models are defined according to Chapman and Pearson (2000), and I follow Chan et al. (1992) to define the orthogonality conditions for the Vasicek, CIR, and CKLS models. I also follow Cai and Swanson (2011) to define the orthogonality conditions of BDF, Heston, and CHEN models. Appendix 4 provides details concerning the relevant moment conditions.

[5] I compute the asymptotic p-values for ADF and WS via the MacKinnon (1994) approximation. These p-values are robust to size distortion. Results are provided for the null of driftless unit root; qualitatively similar results are obtained under the null of unit root with drift.

[6] Several papers find the nominal short rate follows a unit root process (See Perron (1989), Aït-Sahalia(1996a), and Bandi (2002)). However, Bierens (1997) and Al-Zoubi (2009) propose this conclusion is possibly flawed for two reasons: First, negative values should be realizable in a random walk with no drift; Second, given a positive drift a random walk would converge to infinity. Observed short rates do not exhibit these characteristics.

[7] See Perron (1989).

[8] I implement the one-sided Hodrick–Prescott filter introduced by Mehra (2004) to calculate the nonstationary mean. I employ a smoothing constant of 1/q = 43,200 for both the HP and bHP filters, this smoothing constant corresponds to the Ravn and Uhlig (2002) adjustments of the third power for the frequency of observations.

Table 2 Treasury bill and bond summary statistics and nonstationary tests

Variables	N	Mean	St Dev	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
Panel A: descriptive statistics								
r_t	799	0.0417	0.0293	0.9919	0.9781	0.9652	0.9535	0.9419
$rt+1-rt$	798	0.0000	0.0037	0.3446	−0.0594	−0.0667	−0.0099	0.0348
$r1,t$	799	0.0457	0.0301	0.9908	0.9791	0.9683	0.9584	0.9499
$r1,t+1-r1,t$	798	0.0000	0.0041	0.1342	−0.0520	−0.0440	−0.0872	0.0797
$r2,t$	799	0.0476	0.0296	0.9925	0.9825	0.9733	0.9651	0.9579
$r2,t+1-r2,t$	798	0.0000	0.0036	0.1606	−0.0551	−0.0623	−0.0701	0.0388
				WS (p-value)			ADF (p-value)	
Panel B: nonstationary tests-constant in the fitted regression								
r_t				−2.4653 (0.3147)			−2.7989 (0. 1973)	
$r1,t$				−2.2363 (0.4764)			−2.4560 (0. 3503)	
$r2,t$				−2.0035 (0.6461)			−2.0035 (0. 4585)	

Panel (A) provides summary statistics for the 3-month T-bill rate and 1- and 2- year bond yields, using the entire sample period from July 1952 through December 2018. There are 797 annualized monthly observations. I report the number of observations (N), the mean, and the standard deviation (Std Dev). In addition, I report monthly autocorrelation coefficients ρ_j for $j=1$ to 5 that denote the autocorrelation at lag j of the short rate in levels and first differences.

Panel (B) provides the results of random walk tests applied to the 3-month T-bill rate, 1- and 2- year bonds yields inspected in this study. The two tests: (1) the Augmented Dickey-Fuller test (ADF) in Dickey and Fuller (1979) and (2) the Augmented Weighted Symmetric test (WS) in Pantula et al. (1994). For the ADF test, I consider the ratio $\hat{\alpha}/S_{\hat{\alpha}}$ from the estimated model $\Delta r_{t+1} = a + \alpha r_t + \sum_{i=1}^{n} \phi_i \Delta r_t + u_{t+1}$, where $S_{\hat{\alpha}}$ is the standard error of the parameter estimate $\hat{\alpha}$. I use the Akaike information criterion (AIC) is used to determine the optimal lag length n. The test statistic (WS) is given by $\hat{\rho}_{WS} = \frac{\sum_{t=2}^{n} r_t r_{t+1}}{\sum_{t=2}^{n-1} r_{t+1}^2 + n^{-1} \sum_{t=1}^{n} r_{t+1}^2}$. I reject the null hypothesis of a unit root if the p-value is below a relevant significance level. p-values for the relevant statistics are computed using the approximation of MacKinnon (1994) on the basis of a regression surface. I report simulated p-values based on 1,000 replications drawn from a normal distribution with zero mean and OLS squared residual variances, i.e., I use the Wild bootstrap

The general GMM approach employed in this paper is explained in detail by Chan et al. (1992) and Cai and Swanson (2011).

I compare the in-sample and out-of-sample performance of the nine models with each other and with the stochastic mean NSM model. I also compare the performance of the models with the naive random walk model. To evaluate the in-sample forecast performance of each model I define the following goodness-of-fit measures based on the Bergstrom (1986, 1989) generalized discrete form:

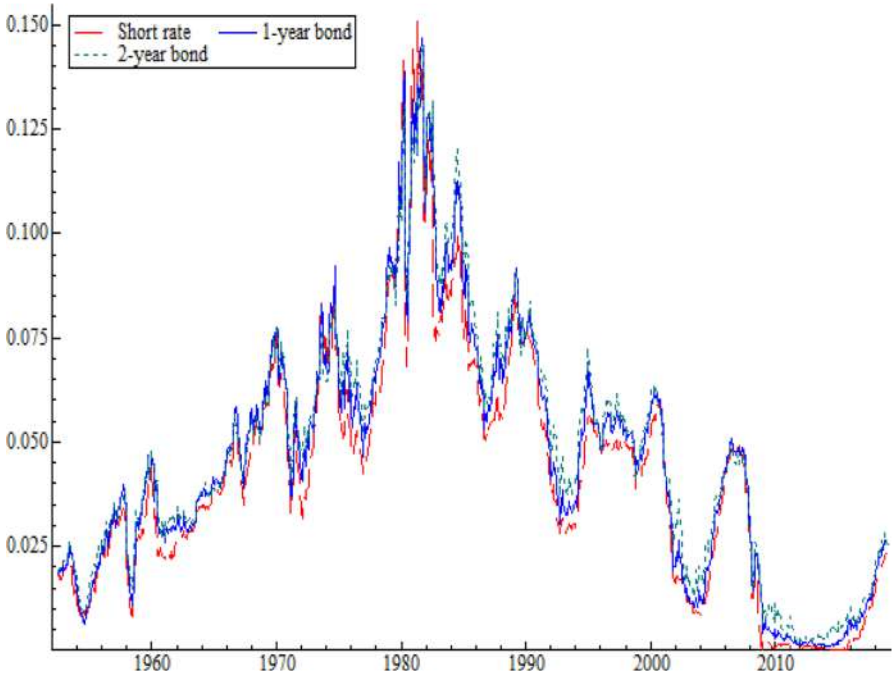


Fig. 1 Interest rates from July 1952 through December 2018. This figure compares the short rate time series to the one-year and two-year bond rates. The Federal Reserve Bank of St. Louis three-month secondary market T-bill rate is used for the short rate. The FAMABLIS file from the Center for Research in Security Prices (CRSP) tape is used for one- and two-year bond rates. The sample is from July 1952 through December 2018 (i.e., 797 annualized monthly observations)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(r-\hat{r}\right)^2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|r-\hat{r}\right|$$

$$MPE = \frac{1}{n}\sum_{t=1}^{n}\frac{r-\hat{r}}{r}$$

where n is the forecast period, and the forecast value of r is given by $\hat{r}_t$. The root-mean-square error, mean absolute error, and mean percentage error are provided by RMSE, MAE, and MPE, respectively.

Because of the persistence of the short rate (Duffee, 2013), the naive random walk model found to outperform most models when forecasting short rates. To

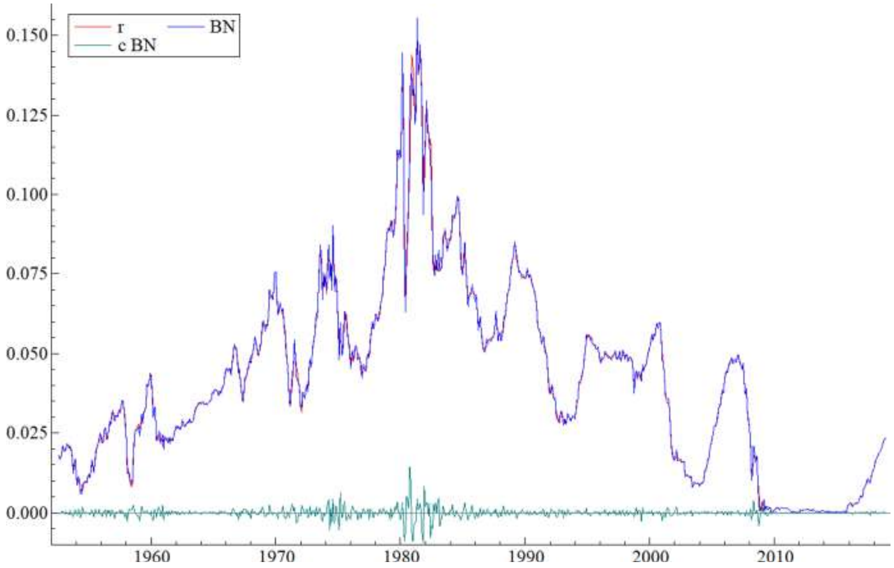


Fig. 2 Short rate, stationary and non-stationary components using Beveridge and Nelson (BN) filter. This figure compares the short rate time series to its nonstationary mean and stationary components produced using the BN filter. The BN permanent component follows a random walk with drift process. The Federal Reserve Bank of St. Louis three-month secondary market T-bill rate is used to proxy short rates. The sample is from July 1952 through December 2018 (i.e., 797 annualized monthly observations)

evaluate the forecast performance of each model against the naive model I use the following statistic developed by Theil (1958)

$$Theil's\ U = \frac{\sum_{t=1}^{n} \frac{\hat{r}_{t+1} - r_{t+1}}{r}}{\sum_{t=1}^{n} \frac{r_{t+1} - r_t}{r_t}}.$$

I use three criteria to measure in-sample fit of each model: the GMM overidentifying test used in Chan et al. (1992), the RMSE, MAE, and MPE statistics of Bergstrom (1986, 1989), and the Theil's U statistic. Table 3 reports the parameter estimates, the asymptotic p-values of the individual parameters, the GMM overidentifying (χ^2) test and the goodness-of-fit tests (RMSE, MAE, and ME) for the models over the period spanning July 1952 through December 2018.

The estimated parameters are given in Table 3. It should first be noted that the VAS, CIR, CKLS, and AG models have no drift coefficients significant at the 5% level. The CKLS and AG models even have volatility parameter estimates that are not significantly different from zero. Thus, the CKLS and AG models have no significant parameters. In contrast, the VAS and CIR models have statistically significant volatility parameters. These two models suggest that short rate dynamics is solely due to volatility. These results indicate that the VAS, CIR, CKLS, and AG models are not able to capture short rate dynamics. On the other hand, the CP model has all drift parameters statistically different from zero. However, the CP model

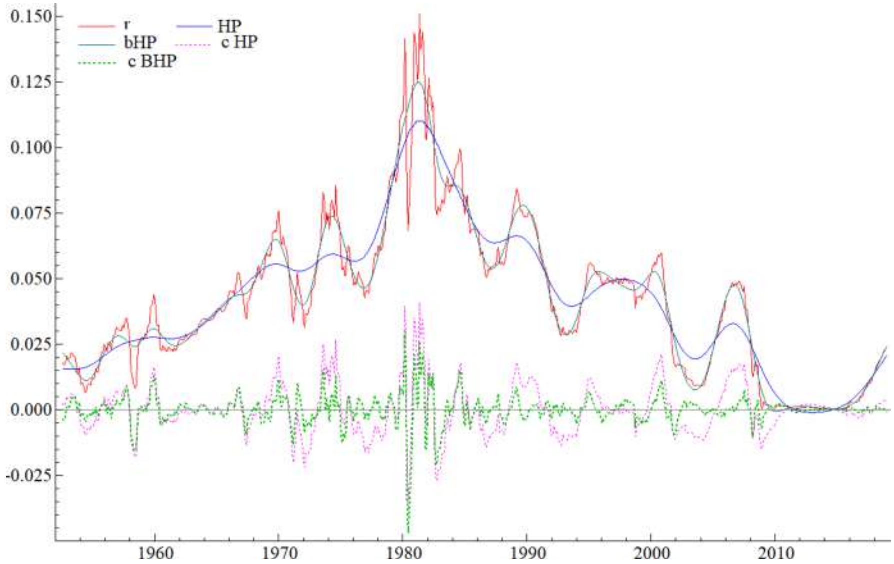


Fig. 3 Short rate, stationary and non-stationary components using HP and boosted HP (bHP) filters. This figure compares the short rate time series to its nonstationary mean and stationary components produced using the HP and bHP filters. The Federal Reserve Bank of St. Louis three-month secondary market T-bill rate is used to proxy short rates. A smoothing constant (1/q) of 43,200 is utilized, which is equivalent to the Ravn and Uhlig (2002) adjustment of the third power of the frequency of observations. The sample is from July 1952 through December 2018 (i.e., 797 annualized monthly observations)

volatility term is not statistically different from zero. Thus, the CP model indicates that short rate dynamics are solely due to drift. Only the NSM model has both drift and volatility significant, indicating that both drift and volatility contribute to short rate dynamics.

By Eq. (8) the coefficients relating the stochastic mean of the NSM model depends on the parameter $\varphi = \sum_{j=1}^{p} \rho_j^2$ which, in turn, describes the sensitivity of the estimated parameters to the serial correlations in the transitory component. In the empirical application, though, I find that the estimated parameter to be largely insensitive to φ. In fact, I find that φ is equal to zero. Which implies that that the transitory component appears close to AR (1). I use Eq. (8) for reasons of measurement errors, although the gain seems to be minor.

For an overidentified system the *p*-value of the overidentifying test of Hansen (1982) (Over ID test) must be above 5% to be correctly identified. Thus, the VAS and CIR models are rejected by the GMM test as their *p*-values on the Over ID test are 0.61% and 0.71%, respectively. The sign of $\alpha1 > 0$ in the VAS model implies, counter-intuitively, that short rates are exploding rather than mean reverting. My results are consistent with Aït-Sahalia (1996b) and Bandi (2002), who argue that neither of these models are correctly specified. The AG model has the best GMM *p*-value at 23%, however, this model has no coefficient that is statistically significant at the 5% level. Thus, the AG model seems to imply, again counter-intuitively, that

Table 3 Short rate single-factor model estimation

MOD	α_0 (*p*-val)	α_1 (*p*-val)	α_2 (*p*-val)	α_3 (*p*-val)	σ2 (*p*-val)	γ (*p*-val)	φ (*p*-val)	OID (*p*-val)	RMSE	MAE	MPE
CP	−0.0021	0.0843	−0.7076	0	0.1293	1.6355		EXACT	0.00387	0.00229	−6.3118
	(0.0442)	(0.0209)	(0.0355)	(0.0284)	(0.4225)	(0.0000)					
AG		0.0026	−0.0313		0.0522	1.5		4.3085	0.00373	0.00201	−1.9887
		(0.7727)	(0.8437)		0.0000			(0.2300)			
CKLS	0.0012	−0.0222			0.162	1.6794		EXACT	0.00375	0.00216	8.1393
	(0.2335)	(0.3002)			(0.4218)	0.0000					
CIR	0.0001	−0.0004			0.0017	0.5		6.0327	0.00374	0.00201	1.3623
	(0.8859)	(0.9824)			0.0000			(0.0071)			
VAS	0.0000	0.0019			0.0000	0.0000		7.2362	0.00374	0.00201	−1.3468
	(0.9861)	(0.9213)			0.0000			(0.0061)			
NSM-BN		0.62574			0.00005			3.1177	0.00354	0.00196	−1.5769
		(0.0000)			(0.0000)			(0.2104)			
NSM-HP		−0.0576			0.0000		0.0000	4.9560	0.00363	0.00200	−3.1588
		(0.0050)			(0.0099)		(0.9999)	(0.0839)			
NSM-bHP		0.18611			0.00003		0.0000	1.92375	0.00355	0.00193	−3.5739
		(0.0000)			(0.0159)		(0.9952)	(0.3822)			
NSM-HAM		−0.0005			0.56451			15.6387	0.00368	0.00201	−1.3413
		(0.9855)			(0.9857)		(0.0001)				

This table presents GMM estimates for the five single-factor short rate models (MOD), namely: the AG model of Ahn and Gao (1999), the CKLS model of Chan et al. (1992), the CP model of Chapman and Pearson (2000), the CIR model of Cox et al.(1985), and the VAS model of Vasicek (1977) for the period starting in July 1952 through December 2018. The models I consider can all be nested within the following model:

$dr_t = \left(\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1}\right)dt + \sigma r_t^{\gamma} dZ_t$ where Z_t is the standard Brownian motion. I examine the following the following discrete-time econometric specification:

$$r_{t+1} - r_t = \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1} + \varepsilon_{t+1}$$

with $E(\varepsilon_{t+1}) = 0$ and $E(\varepsilon_{t+1}^2) = \sigma^2 r_t^{2\gamma}$

I follow Chapman and Pearson (2000) to estimate the models via GMM. The relevant orthognality conditions are explained in Appendix 4.

GMM estimates are also provided for my proposed stochastic mean model, i.e., the NSM model. I follow Ravn and Uhlig (2002) and de Jong and Sakarya (2016) to approximate the nonstationary mean, μt, with smoothing constants of q = 1/43,200. Also, I report results for my proposed stochastic mean model using Phillips and Shi (2021) boosted HP. I denote this model as NSM-bHP. Also, I report results for my proposed stochastic mean model using Hamilton proposed trend. I denote this model as NSM-H. I impose various restrictions on the parameters (see Table 1) to obtain standard parametric models for the short rate dynamics. GMM estimates of the coefficients, overidentifying tests (OID) of Hansen (1982), their significance levels, and goodness-of-fit tests (RMSE, MAE, and MPE). The corresponding *p*-values are in parentheses. I follow Inoue and Shintani (2006) by using the Parzen kernel of Gallant (1987) with two lags to compute the moment-weighting matrix. The covariance matrix is robust to heteroskedasticity and autocorrelation

short rates have no mean reversion, and that volatility is not related to short rate dynamics. The NSM model is the only model that satisfies the Over ID test and has coefficients on both drift and volatility that are significant at the 5% level.

I compare in-sample forecasting performance of the nonstationary stochastic-mean (NSM) model using the methods suggested in Hodrick and Prescott (1997), Beveridge and Nelson (1981), Phillips and Shi (2021), and Hamilton (2018). I denote the models by NSM-BN, NSM-HP NSM-bHP, and NSM-HAM, respectively. I test in-sample fit using the RMSE, MAE, and MPE statistics. Comparing the benchmark affine models to the NSM model, I find consistently larger RMSE errors. For example, the CP model has an 5.4% increase in RMSE compared to NSM-bHP. This is true for all the benchmark models with RMSE error increases between 1.6 and 8.4%. Similar results are realized under the MAE measure. The NSM model (NSM-bHP in particular) consistently outperforms. Increases in MAE are between 0.5% for the AG model to as high 45% for the CP model. Results for the MPE measure are even more remarkable with MPE increases between −6.3 for the CP model to as large as 8.14 for the CKLS model.

Table 4 reports parameter estimates for the three multifactor models. As shown, none of these models outperform the NSM model. All three goodness-of-fit measures (i.e. RMSE, MAE, and MPE) suggest that my model with a random walk mean outperforms the three multifactor models. No model is consistently best across all goodness-of-fit tests except the NSM (NSM-bHB followed by NSM-BN). Although the CHEN, Al-Zoubi, and BDF-CIR models perform well with RMSE and MAE. The BDF-CIR failed the MPE test. I conclude that CHEN model is the second best in terms of RMSE and Al-Zoubi model is the second best in terms of MAE.

Table 5 reports the results of the Theil's U-statistic for each model against the naive random walk model. As shown, the NSM-bHP model reports the lowest Theil's U-statistic among all models. This is followed by NSM-BN model, Al-Zoubi model and NSM-HP model. I find that the forecasts from my stochastic mean model (except the NSM-HAM) are substantially and statistically more accurate than the naive random walk model. I have not found results for other models to be qualitatively similar, with the Al-Zoubi model being the exception. Only two single factor models (Vasicek and CIR) are found to slightly outperform the random walk model by the U test as their statistic are 0.9804 and 0.9431, respectively. For multifactor models, two models forecasting technique is about as good as guessing. Those are the Chen and BDF-VAS models. Other models underperform the random walk model.

6.2 Out-of-sample forecast comparison

In this section I analyze out-of-sample (OOS) forecasts of my model and competing models. Each benchmark model is compared over the sample period. I use data for the period starting in July 1952 through December 2008 for estimation and reserve data for the period starting in January 2009 through December 2018 for out-of-sample forecasting.

First, I test out-of-sample fit using the RMSE, ME, MAE, and Theil's U statistics. As shown in Table 6, comparing the benchmark affine models to my model, I find consistently larger RMSE and MAE errors. Most notably, the NSM-bHP model shows impressive gains in forecasting power. For example, NSM-bHP

Table 4 Shor rate multifactor model estimation

MOD	α_0 (p-val)	α_1 (p-val)	$\alpha\theta$,0 (p-val)	$\alpha\theta$,1 (p-val)	$\alpha\nu$,0 (p-val)	$\alpha\nu$,1 (p-val)	2 (p-val)	σ2θ (p-val)	σ2ν (p-val)	OID	RMSE	MAE	MPE
BDF-VAS		−0.0000	0.0030	0.0000			0.0000			15.00	0.00375	0.00201	−1.7426
		(0.9320)	(0.7200)	(0.6560)			(0.0000)			(0.1820)			
BDF-CIR		−0.0010	0.0014	0.0007			5.8842			285.5	0.00371	0.00203	−8.7475
		(0.0000)	(0.0000)	(0.9820)	(0.0000)		(0.9150)			(0.0000)			
AL-Zoubi		−0.0572					0.00001			5.7821	0.00372	0.00197	−3.721
		(0.0049)					(0.0099)			(0.0732)			
HEST	0.0007	−0.0122			0.0000	0.1771		0.2598		19.20	0.00373	0.00206	5.9327
	(0.3397)	(0.4070)			(0.0685)	(0.0000)		(0.0000)		(0.0000)			
CHEN		−0.0009	0.0000	−0.4348	0.0000	−0.1148	1.0	0.0000	0.0000	121.7	0.00370	0.00201	−2.1832
		(0.6820)	(0.0000)	(0.0000)	(0.0000)	(0.0000)		(0.1894)	(0.0000)	(0.0000)			

This table presents GMM estimates for the four multifactor short rate models (MOD), namely: the BDF-VAS and BDF-CIR models of Balduzzi et al., 1998), Al-Zoubi Model of Al-Zoubi (2019), the HEST model of Heston (1993), the CHEN model of Chen 1996 for the period starting in July 1952 through December 2018. I impose various restrictions on the parameters (see Table 1) to obtain standard parametric models for the short rate dynamics. GMM estimates of the coefficients, overidentifying tests (Over ID) of Hansen (1982), their significance levels, and goodness-of-fit tests (RMSE, MAE, and MPE). The corresponding p-values are in parentheses. I follow Inoue and Shintani (2006) by using the Parzen kernel of Gallant (1987) with two lags to compute the moment-weighting matrix. The covariance matrix is robust to heteroskedasticity and autocorrelation. The relevant orthogonality conditions are shown in Appendix 4

Table 5 Theil's U statistics in-sample forecast accuracy: naive random walk models vs NSM, single and multifactor models

	Model	Theil′sU
	NSM-BN	0.9574*
	NSM-HP	0.9097*
	NSM-bHP	0.8972*
	NSM-HAM	6.7936
Single	CP	27.9627
Factor	AG	1.0703
Models	CKLS	4.1208
	CIR	0.9431*
	VAS	0.9804*
Multi	BDF-VAS	0.9967
Factor	BDF-CIR	1.3595
Modes	Al-Zoubi	0.9107*
	HEST	2.2701
	CHEN	0.9993

The table presents the results of the Theil's U statistics applied to forecast errors of the NSM models and benchmark models against the naive random walk model. I use the entire sample period from July 1952 through December 2018. There are 797 annualized monthly observations.

$Theil's\ U = \frac{\sum_{t=1}^{n} \frac{\hat{r}_{t+1} - r_{t+1}}{r}}{\sum_{t=1}^{n} \frac{r_{t+1} - r_t}{r_t}}.$

*The model outperforms the naive random walk model

lowers RMSE by 40 percent relative to the CHEN, CIR and Vasicek models. The improvement relative to RW is also notable where Theil's U statistic is 65 percent.

I next utilize the Pesaran and Timmermann (1992) directional forecast accuracy test to predict the change in direction of the short rate. I denote the predicted short rate by $\hat{r}_t$. The Pesaran-Timmermann (PT) test statistic can be written as:

$$S_n \frac{\hat{P} - \widehat{P_*}}{\sqrt{\hat{V}\left(\hat{P}\right) - \hat{V}(\widehat{P_*})}}$$

where:

$$\hat{P} = n^{-1} \sum_{t=1}^{n} I\left(r_t, \hat{r}_t\right)$$

$$\widehat{P_*} = \hat{P}_{\mathrm{r}} \hat{P}_{\hat{r}} + (1 - \hat{P}_{\mathrm{r}})(1 - \hat{P}_{\hat{r}})$$

n is the number of out-of-sample period and the estimate of the sample variance is given by:

Table 6 Out-of-sample forecast accuracy test: NSM model vs random walk, single and multifactor models

	Model	RMSE	MAE	MPE	Theil′sU
	NSM-BN	0.00056	0.00035	−9.570142	0.7912*
	NSM−HP	0.00059	0.00039	−13.07865	0.8903*
	NSM-bHP	0.00021	0.00013	−9.2397	0.6517*
	NSM-HAM	0.00057	0.00037	10.11236	1.2461
	RW with drift	0.00052	0.00031	−59.9451	1.34175
	RW without drift	0.00054	0.00035	−3.7551	
Single	CP	0.04348	0.02668	11,393.5	348.0352
Factor	AG	0.00053	0.00034	−13.9737	1.5657
Models	CKLS	0.00113	0.00103	193.6967	4.2208
	CIR	0.00051	0.00036	6.6227	1.4234
	VAS	0.00053	0.00034	−16.6293	1.5931
Multi	BDF-VAS	0.00052	0.00033	−13.5009	1.5613
Factor	BDF-CIR	0.00068	0.00061	−103.662	2.2555
Models	Al-Zoubi	0.00062	0.00042	−13.9603	0.9213*
	HEST	0.00068	0.00061	94.5644	2.2452
	CHEN	0.00054	0.00035	−14.0696	1.5668

*The model outperforms the naive random walk model

The table presents goodness-of-fit tests (RMSE, MAE, and MPE) and the Theil's U statistics applied to forecast errors of the NSM model and benchmark models. I use data for the period starting in July 1952 through December 2008 for estimation and reserve data for the period starting in January 2009 through December 2018 for out-of-sample forecasting

$$\hat{V}\left(\hat{P}\right) = n^{-1}\widehat{P_*}(1 - \widehat{P_*})$$

$$\hat{V}\left(\widehat{P_*}\right) = n^{-1}\left(2\widehat{P_{\hat{r}}} - 1\right)^2\widehat{P_r}\left(1 - \widehat{P_r}\right) + n^{-1}\left(2\widehat{P_r} - 1\right)^2\widehat{P_{\hat{r}}}\left(1 - \widehat{P_{\hat{r}}}\right) + 4n^{-2}\widehat{P_{\hat{r}}}\widehat{P_r}\left(1 - \widehat{P_{\hat{r}}}\right)\left(1 - \widehat{P_r}\right)$$

$$\widehat{P_{\hat{r}}} = n^{-1}\sum_{t=1}^{n} I\left(\hat{r}_t\right), \widehat{P_r} = n^{-1}\sum_{t=1}^{n} I\left(r_t\right)$$

and $I(.)$ is the indicator function or the directional accuracy (DA) of the forecast that can be defined as.

$$I(.) = \begin{cases} 1 & if. > 0 \\ 0 & otherwise \end{cases},$$

The null hypothesis maintains that the model does not forecast the direction of the change in the short rate, while the alternative hypothesis maintains that the probability the model forecast will correctly predict the direction of change in the short rate is greater than 50 percent. I set the level of significance at 5%.

As a robustness test, I also utilize the following directional forecast accuracy measures suggested in Blaskowitz and Herwartz (2011) and evaluated in Bergmeir et al. (2014):

Table 7 Out-of-sample directional forecast accuracy test: NSM model vs random walk, single and multi-factor models

	Model	MDA	MDV	MDPV	PT
	NSM-BN	0.550	0.000010	12.2421	5.4996* (0.0000)
	NSM-HP	0.450	0.000063	4.4785	5.4996* (0.0000)
	NSM-bHP	0.547	0.000012	13.3168	5.49962* (0.0000)
	NSM-HAM	0.408	−0.000001	4.9334	2.6453* (0.0041)
	RW with drift	0.483	0.000199	2.57618	
	RW without drift	0.542	−0.00002	−8.30454	3.8399* (0.0000)
Single	CP	0.317	−0.000062	−5.49166	−0.5695 (0.7154)
Factor	AG	0.558	0.000175	16.0515	5.4996* (0.0000)
Models	CKLS	0.458	0.000184	−3.84899	−0.0663 (0.4672)
	CIR	0.508	0.000184	10.4784	2.6453* (0.0041)
	VAS	0.575	0.000185	16.9117	5.4996* (0.0000)
Multi	BDF-VAS	0.558	0.000176	−16.0515	5.4996* (0.0000)
Factor	BDF-CIR	0.517	0.000194	11.7162	5.4996* (0.0000)
	Al-Zoubi	0.450	0.000067	4.4873	5.4996* (0.0000)
Modes	HEST	0.467	0.000181	−3.8303	0.33049 (0.6557)
	CHEN	0.533	0.000131	15.61828	5.4996* (0.0000)

The table presents out-of-sample directional forecast accuracy tests (MDA, MDV, and MDPV) of Blaskowitz and Herwartz (2011) and the PT test of Pesaran and Timmermann (1992) applied to the NSM model and benchmark models. I use data for the period starting in July 1952 through December 2008 for estimation and reserve data for the period starting in January 2009 through December 2018 for out-of-sample forecasting

*Significant at 5% level

$$MDA = n^{-1} \sum_{t=1}^{n} I(.)$$

$$MDV = n^{-1} \sum_{t=1}^{n} \left| r_{t+1} - r_t \right| I(.)$$

$$MDPV = n^{-1} \sum_{t=1}^{n} \left| \frac{r_t - r_t}{r_{t+1}} \right| I(.)$$

where MDA is the Mean Directional Accuracy, MDV is the Mean Directional Value, and the MDPV is the Mean Directional Percentage Value.

Tables 7 provide the results comparing the competing models against my model for the short rate. I find that the directional forecasts from my nonstationary stochastic mean model are statistically more accurate than the competing models, (again, Al-Zoubi model is an exception). For example, Models NSM-bHP and NSM-BN demonstrate superior OOS forecasting ability compared to standard models used in the literature across all measures. The improvements

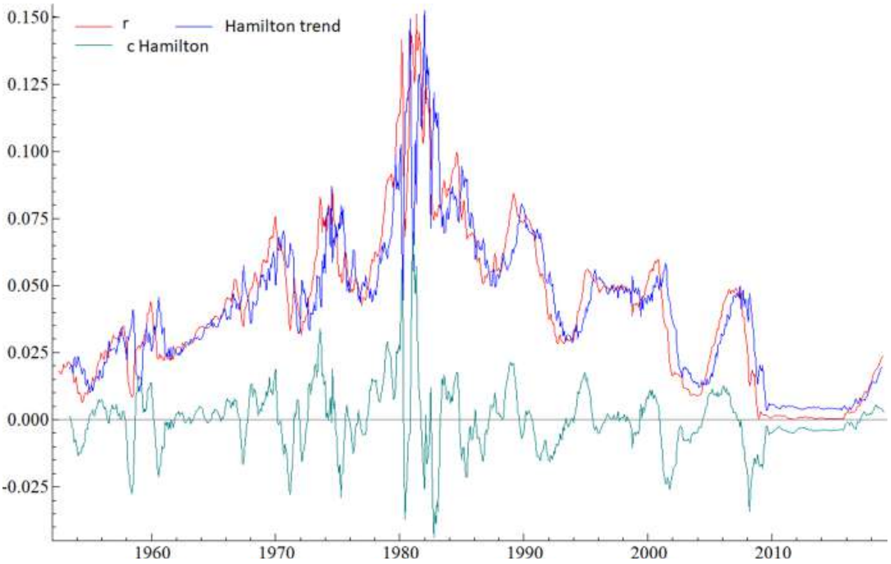


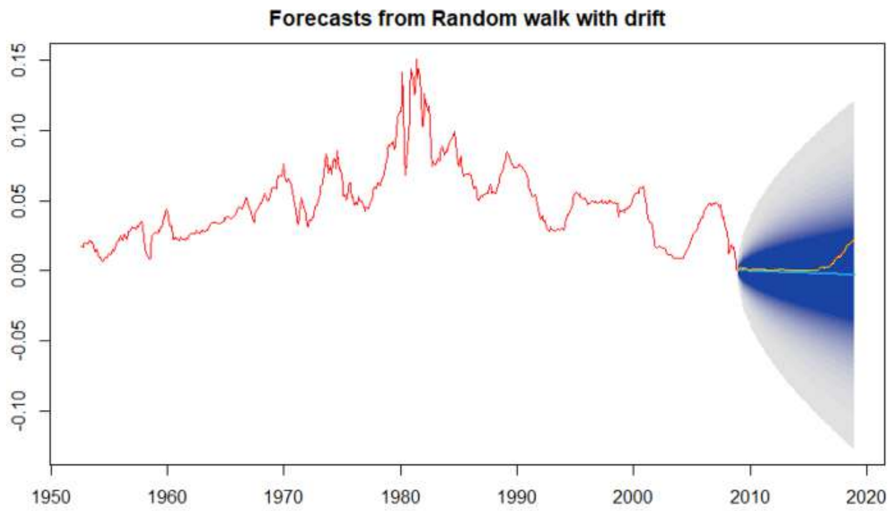
Fig. 4 Short rate, stationary and non-stationary components using Hamilton filter. This figure compares the short rate time series to its nonstationary mean and stationary components produced using the Hamilton (2018) filter. The Federal Reserve Bank of St. Louis three-month secondary market T-bill rate is used to proxy short rates. The OLS autoregression of order four that reads: $r_{t+1} = a_0 + a_1 r_{t-7} + a_2 r_{t-8} + a_3 r_{t-9} + a_4 r_{t-10} + \varepsilon_{t+1}$. The short rate sample is from July 1952 through December 2018 (i.e., 797 annualized monthly observations)

relative to the CHEN and RW models are smaller but still of interest. At the same time, CHAP, CKLS, and Heston models perform poorly in predicting the direction of the short rate where the null hypothesis cannot be rejected at 5% significance level. My results suggest that, relative to stationary short rate models, more accurate short rates forecasts can be obtained by allowing for a random walk component mean in the short rate.

In Figs. 4, 5 I present fan plots for out-of-sample forecasts from the NSM-bHP and random walk models. The figure compares the short rate (red line) with the forecasts from the NSM-bHP model (yellow line) and random walk model (blue line). Shaded blue areas show intervals. The figures show that the forecast from the NSM-bHP model is very close to the true short rate in comparison to the random walk model.

6.3 Unspanned risks are the drivers of the nonstationary stochastic mean.

The consensus in literature is that the spanning hypothesis, the yield curve contains all information relevant for predicting future bond returns, can be rejected by the observed data (see Cooper & Priestley, 2009) and Joslin et al., 2014). Evidence of unspanned risks comes from regressions of the form



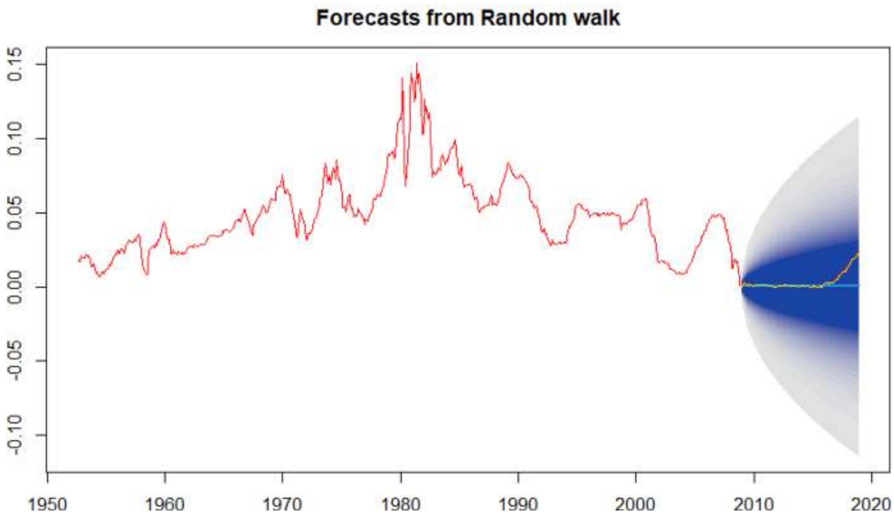


Fig. 5 Fan plots for out-of-sample forecasts from the NSM-bHB model vs random walk model. This figure compares the short rate (red line) with the forecasts from the NSM model (yellow line) and random walk model (blue line). I use data for the period starting in July 1952 through December 2008 for estimation and reserve data for the period starting in January 2009 through December 2018 for out-of-sample forecasting

$$rx_{t+1} = a + b_1 PC1_t + b_2 PC2_t + b_3 PC3_t + e_{t+1}$$

$$rx_{t+1} = a + b_1 PC1_t + b_2 PC2_t + b_3 PC3_t + b_4 F1_t + b_5 F2_t \cdots + e_{t+1}$$

Table 8 Spanned and unspanned risk factors inference on interest rates and their components

	Inflation	Growth	Output gap
r1,t	2.5572	−0.00361	0.00229
	(0.0000)	(0.0000)	(0.0647)
μ1,t	2.5888	−0.00389	0.00164
	(0.0000)	(0.0298)	(0.2017)
c1,t	(−0.0316)	(0.0003)	(0.0007)
	(0.5919)	(0.4988)	(0.0028)

This table reports regressions for the of one-year bond returns. The dependent variables are one-year bond returns and their transitory and permanent components. The independent variables the Chicago Fed National Activity Index (JPS1) and expected inflation (JPS2) in Joslin et al (2014) or the output gap in Cooper and Priestley (2009). I use data for the period starting in January 1982 through December 2023. The corresponding p-values are in parentheses. I follow Bakshi et al., (2023a, 2023b) by using the Newy-West (1994) heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimator. The permanent component is calculated using bHP method in Phillips and Shi (2021)

where rx_{t+1} is the bond excess returns, PC1, PC2, and PC3 are the first three principal components of yields (level, slope, and curvature), and F1, F2,… are the unspanned factors. The null hypothesis of spanning hypothesis is

$$\mathrm{H0} : b4, b5 \ldots . = 0$$

My model is founded on the perception that the excess bond returns are determined by the nonstationary stochastic mean of the short rates. My theory is that the unspanned risks should not be able to predict the stationary component of the interest rate. Specifically, I consider the following test specifications, in line with Joslin et al. (2014) and Cooper and Priestley (2009):

$$\begin{gathered}
cx_{t+1} = a + b_1 cPC1_t + b_2 cPC2_t + b_3 cPC3_t + e_{t+1} \\
cx_{t+1} = a + b_1 cPC1_t + b_2 cPC2_t + b_3 cPC3_t + b_4 JPS1_t + b_5 JPS2_t + e_{t+1} \\
\ldots \\
cx_{t+1} = a + b_1 cPC1_t + b_2 cPC2_t + b_3 cPC3_t + b_4 GAP + e_{t+1} \\
\mu x_{t+1} = a + b_1 pPC1_t + b_2 pPC2_t + b_3 pPC3_t + e_{t+1} \\
a + b_1 pPC1_t + b_2 pPC2_t + b_3 pPC3_t + b_4 JPS1_t + b_5 JPS2_t + e_{t+1} \\
\mu x_{t+1} = a + b_1 pPC1_t + b_2 pPC2_t + b_3 pPC3_t + b_4 GAP + e_{t+1}
\end{gathered}$$

where cx_{t+1} and μx_{t+1} are the stationary and permanent components of the excess bond returns, respectively. $cPCi_t$ and $pPCi_t$ are the stationary and permanent components of the three principle components of the yield curve, $JPS1$ is the one year expected inflation rate, $JPS2$ is a measure of economic growth (the 3-month moving average of the Chicago Fed National Activity Index), GAP is the output gap.

The analysis of the model based on the permanent component is further supported by regressions of one-year interest rate bond yield in Table 8. It is observed that Joslin et al. (2014) factors are unable to predict the stationary component of the interest rate; rather, they predict its permanent component. Additionally, the factor proposed by Cooper and Priestley (2009) (the stationary component of output, or output gap) predicts the stationary component of the interest rate but not its permanent component..

An intriguing question to explore is the following: How effectively do the unspanned risk factors predict the component of excess bond returns, given the estimation results of the interest rate model? As shown in Table 9, The permanent components of the three principal components of the yield do not play any role in predicting excess bond returns. The expected inflation (JPS1) helps predict the stationary component of excess bond returns and the growth (JPS2) helps predict the permanent component of the excess returns. Consistent with Bauer and Hamilton (2017) I do not find evidence that the output gap predicts excess bond returns.

7 Bond and option prices

An advantage of my model is that closed-form solutions for bond and option prices can be derived. Closed-form solutions provide several benefits, including: (i) allowing for consistent estimation and (ii) allowing a full description for bond and option prices over all maturities without approximation error. I utilized these closed-form solutions to investigate estimation bias.

Estimation bias in pricing options is significant. For example, while estimation bias only leads to a 1% downward bias in the bond price, it results in a 24.4% downward bias in option prices (see Phillips and Yu (2005)). Using the Vasicek and the CIR models, Tang and Chen (2009) document large underestimation in option values. This estimation bias remains significant even after reducing it via a jackknife or bootstrap algorithm. The reason why estimation error remains is that misspecification error accounts for most of the observed bias. Misspecification becomes even more severe if either fewer observations are available or longer maturity options are priced.

I implement analytical formulas for option prices for the VAS and NSM models. This allows us to address the issue of biased estimation that may result from nonstationary mean models for spot rates. Let $C_{t,T,S,K}(\theta)$ represents the theoretical price of a European call option at time t, with strike price K, and maturity S where the underlying bond has price $P_{t,T}(\theta)$ that matures at $T > S$. Sect. 6.1derives the analytic formula for the bond price for my NSM model. I utilize the analytic formula in Vasicek (1977) for the VAS model. First, I compute the implied bond prices using parameter estimates for α1 and σ2 from the VAS and the NSM model with t = 0 and T = 3. Comparing the implied bond prices with observed bond prices, I use the goodness-of-fit test (RMSE) to evaluate the relative performance of the models. Finally, I apply implied option prices from Jamshidian (1989) for the models using the analytic formula a European call option with a face value of a bond = \$100, K = (\$90, \$95, \$99, \$100, \$101), and $S = 1$.

I relate the implied option prices of the VAS and NSM models to results of Phillips and Yu (2005) and Tang and Chen (2009). If the random walk component

Table 9 Spanned and unspanned risk factors inference on excess returns and their components

	PC1	PC2	PC3	Inflation	Growth	GAP
Panel A: three-factor						
Excess returns (rx)						
	0.0183	−0.4383	0.09299			
	(0.0000)	(0.0000)	(0.0003)			
HAC	(0.0000)	(0.0000)	(0.0155)			
Transitory component of excess returns (cx)						
	−0.06904	−0.36543	0.34892			
	(0.0151)	(0.0000)	(0.0004)			
HAC	(0.1712)	(0.0826)	(0.0555)			
Permanent component of excess returns (px)						
	0.011329	0.030593	−0.12868			
	(0.0000)	(0.0000)	(0.0000)			
HAC	(0.1898)	(0.4012)	(0.4622)			
Panel B: Joslin et al. (2014)						
Excess returns (rx)						
	0.01064	−0.43433	0.09074	0.07018	−0.00062	
	(0.02397)	(0.00000)	(0.00061)	(0.06114)	0.00680	
HAC	(0.10927)	(0.0000)	(0.03391)	(0.14082)	(0.04316)	
Transitory component of excess returns (cx)						
	−0.12225	−0.34454	0.34581	0.20788	0.00020	
	(0.00000)	(0.0000)	(0.0001)	(0.0000)	(0.5896)	
HAC	(0.0022)	(0.0683)	(0.0400)	(0.0000)	(0.6559)	
Permanent component of excess returns (px)						
	0.01155	0.03172	−0.12440	−0.00103	−0.00044	
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
HAC	(0.3333)	(0.5759)	(0.5256)	(0.9626)	(0.0474)	
Panel C: Cooper and Priestley (2009)						
Excess returns (rx)						
	0.01858	−0.43151	0.09468			−0.00013
	(0.0000)	(0.0000)	(0.0003)			(0.0511)
HAC	(0.0000)	(0.0000)	(0.0143)			(0.0979)
Transitory component of excess returns (cx)						
	−0.04906	−0.34195	0.34376			−0.00044
	(0.0000)	(0.0000)	(0.0003)			(0.0511)
HAC	(0.3039)	(0.0749)	(0.0421)			(0.1182)
Permanent component of excess returns (px)						
	0.01138	0.03212	−0.12840			−0.00003
	(0.0000)	(0.0000)	(0.0000)			(0.0821)
HAC	(0.0997)	(0.1489)	(0.2433)			(0.5472)

Table 9 (continued)

This table reports regressions for one-year excess bond returns. The dependent variables are one-year excess returns and their transitory and permanent components. In Panel A the independent variables are the spanned risk factors (level, slope, and curvature of yield curve). In Panel B the independent variables are the spanned risk factors and Chicago Fed National Activity Index (JPS1) and expected inflation (JPS2) in Joslin et al (2014). In Panel C the independent variables are the spanned risk factors and output gap in Cooper and Priestley (2009). I use data for the period starting in January 1982 through December 2023. The corresponding p-values are in parentheses. I follow Bakshi et al., (2023a, 2023b) by using the Newy-West (1994) heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimator

is the main source of misspecification bias, one would expect higher option prices for the NSM model.

Table 7 reports the estimation results for bond and option prices for both the VAS and NSM models. As shown, the nonstationary mean plays an important role in bond pricing. In fact, my model not only reduces the bias in terms of mean by 1.7% (hence bias-reduced), but also reduces standard deviation resulting in a substantial overall gain in RMSE. As shown, the RMSE of NSM-HP and NSM-bHP bond prices are 2.6% and 2% smaller than that of the VAS model.

It is worth noting that assuming a correlation between the nonstationary stochastic mean and the transitory components reduces the forecasting power in bond pricing. As shown in the Table 10, NSM-BN model reports higher RMSE. I conclude that modeling the stochastic mean as independent to the current level of the short rate is more appropriate in bond pricing. Consistent with the results in Tables 5 and 6, NSM-HAM model reports higher RMSE. Phillips and Shi (2021) show that the bHP filter numerically and empirically outperforms Hamilton's (2018) autoregressive filter and accelerates the convergence to the actual trend when applied to a random walk.

Compared to the VAS model, the NSM-bHP model increases the implied option prices by 0.036%—9.4% for in-the-money options and by 68%% for at-the-money option. The NSM-BN and NSM-HAM models are also able to increase the implied option prices. The bias reduction of my proposed model is comparable to the size of the bias reported by Phillips and Yu (2005, Table 5), who find that the estimation bias of the mean-reversion parameter leads up to 36.2% downward bias in the in-the-money bond option price and up to 1.84% downward bias in the discount bond price. These results confirm that the nonstationary mean is an important source of downward bias in implied bond option prices.

Figure 6 shows the implied option price on the discount bond over the sample period. I can see that the NSM-HAM model delivers the highest prices during periods of both high and low interest rates. I can also see that the NSM-bHP model produces higher option prices than the Vasicek model during periods of high interest rates. However, the Vasicek model delivers higher option prices than NSM-bHP model during periods of low interest rates. Consequently, I expect that bias in the stochastic mean is larger when interest rates are volatile. Furthermore, the impact depends on the level of interest rates. The higher the interest rate, the higher the sensitivity. Finally, the nonlinear magnitude indicates that the bias in

Table 10 Zero-coupon bond and european call option prices

			$\hat{C}$				
	Fama-Bliss P	$\hat{P}$	$K=90$	$K=95$	$K=99$	$K=100$	$K=101$
Panel A: VAS model estimation							
Avg	86.1146	88.5134	3.3059	0.9316	0.1006	0.0177	0.0005
St Dev	7.7023	7.5621	3.4892	1.6625	0.2709	0.05104	0.0017
Min	62.6787	63.4882	0	0	0	0	0
Max	99.1348	99.9690	9.9784	4.9788	0.9888	0.2161	0.0083
RMSE	3.0406						
Bias	2.4116						
Panel B: NSM-BN model estimation							
Avg	86.1146	88.5722	3.3296	0.9688	0.0995	0.0026	0
St Dev	7.7023	7.5747	3.5107	1.7262	0.2819	0.0420	0
Min	62.6787	62.3395	0	0	0	0	0
Max	99.1348	100.8038	10.8497	5.9597	1.8543	0.8547	0
RMSE	3.1323						
Bias	2.4566						
Panel C: NSM-HP model estimation							
Avg	86.1146	88.4743	3.2982	0.9179	0.0994	0.0201	0.0012
St Dev	7.7023	7.4883	3.4602	1.6408	0.2704	0.0588	0.0032
Min	62.6787	64.2167	0	0	0	0	0
Max	99.1348	99.9911	9.9982	4.9986	1.0144	0.2532	0.0161
RMSE	2.9879						
Bias	2.3501						
Panel D: NSM-bHP model estimation							
Avg	86.1146	88.5603	3.3189	0.93239	0.1101	0.0299	0.0035
St Dev	7.7023	7.4844	3.4639	1.6537	0.2892	0.0828	0.0103
Min	62.6787	0.6475	0	0	0	0	0
Max	99.1348	0.9996	9.9763	4.9772	1.0244	0.3174	0.0431
RMSE	3.0219						
Bias	2.3614						
Panel E: NSM-HAM model estimation							
Avg	86.1146	88.6803	3.4306	1.0060	0.1372	0.0424	0.0061
St Dev	7.7023	7.5772	3.5509	1.7527	0.3491	0.1147	0.0174
Min	62.6787	63.7238	0	0	0	0	0
Max	99.1348	100.195	10.197	5.1969	1.2233	0.4364	0.0730
RMSE	3.2341						
Bias	2.6678						

The table provides bond and option prices for the NSM and Vasicek (1977) models. $\widehat{P}$ and $\widehat{C}$ are the estimated zero-coupon bond and European call option prices respectively, and P is the observed-zero coupon bond price. Avg, Std Dev, and RMSE are the estimated average, standard deviation, and root-mean-square error, respectively. I use parameter estimates for the entire sample to estimate $\widehat{P}$ and $\widehat{C}$. To test model's performance, I compare $\widehat{P}$ with the Fama-Bliss data. The data are interpolated three years zero-coupon bond prices (see Fama & Bliss, 1987). I compute $\widehat{P}$ and $\widehat{C}$ monthly using the observed

Table 10 (continued)

three-month Treasury-bill rate for the NSM and Vasicek (1977) models.

For the NSM model, the bond price is given by $\hat{P}(t,T,r_t) = Exp(A(t,T)r_t - [A(t,T) + (T-t)]\mu_t + B(t,T) + D(t,T))$

where

$A(t,T) = \left(\frac{1-e^{\alpha_1(T-t)}}{\alpha_1}\right)$, $B(t,T) = \left[\frac{A(t,T)+(T-t)}{2\alpha_1^2} + \frac{A(t,T)^2}{4\alpha_1}\right]\sigma_c^2$, and $D(t,T) = \frac{\sigma_\mu^2(T-t)^3}{12}$.

For the case of HP and bHP filters

$B(t,T) = \left[\frac{(A(t,T)+(T-t))}{2\alpha_1^2} + \frac{A(t,T)^2}{4\alpha_1}\right]\left[\frac{\sigma^2}{(1+Q)}\right]$, and $D(t,T) = \left(\frac{q\sigma^2(T-t)^3}{12(1+q)}\right)$.

*The number of observations is 787 due to loss of 11 observations.

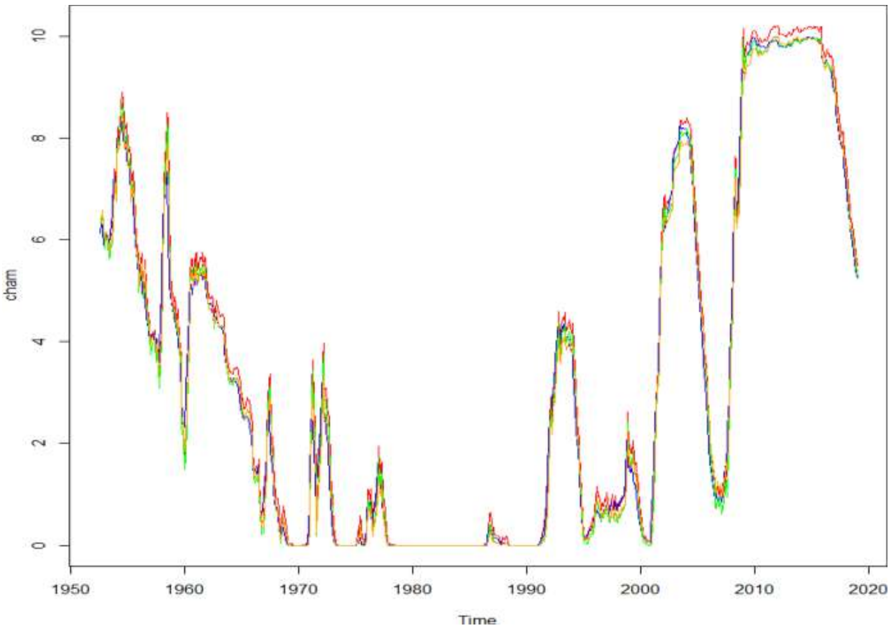


Fig. 6 Option prices implied from Vasicek, NSM-HP, NSM-bHP, and NSM-HAM models. This figure compares options prices implied from the Vasicek model (green line), the NSM-HP model (orange line), the NSM-bHP model (blue line) and NSM-HAM (red line). I use data for the period starting in July 1952 through December 2018 to calculate option prices

estimating bond option prices can also arise from the errors in estimating the stochastic trend in the short rate.

8 Conclusion

In this paper, I develop a short rate model that embeds a nonstationary mean in the short rate process. Relative to several well-known models, my model provides the best fit to the data. The improved performance of my model relative to other models in the literature is due to the inclusion of a random walk component in the short rate process. A nonstationary mean allows for possible nonlinearity in the drift function.

My model accords with the assertion of Fama (2006) and Bauer and Rudebusch (2020) that the short rate exhibits strong evidence of a nonstationary mean. Empirical results show the procedure of imposing a random walk component is highly effective and offers substantial in-sample and out-of-sample pricing improvements. In fact, I find that the random walk mean governs a substantial majority of the dynamics of bond and bond option prices. In particular, the stochastic mean of the interest rates can capture the unspanned risks.

Appendix 1: Bond and option prices

This appendix provides details of the derivations related to bond and option prices under random walk mean for the short rate process.

***Proof of Proposition* 2.1** Consider the NSM model of the short rate introduced in Sect. 2. Under the risk-neutral measure, the short rate dynamics are given by

$$dr_t = -\alpha_1\left(\mu_t - r_t\right)dt + \sigma dZ_t, \tag{11}$$

$$d\mu_t = \sigma_\mu dW_t.$$

It can be verified using Ito's formula that

$$r_t = e^{\alpha_1}\left[r_0 - \int_0^t \alpha\mu_t e^{\alpha_{1u}}du + \sigma\int_0^t e^{\alpha_{1u}}dZ_u\right]$$

is a solution to the stochastic differential equation (SDE) in (11). It can be shown using integration by parts for the second term that

$$\begin{aligned}
r_t &= e^{\alpha_1 t}\left[r_0 - \mu_t\int_0^t e^{-\alpha_1 u}du + \alpha_1\int_0^t d\mu_t\left(\int_0^t e^{-\alpha_1 t}dt\right)du + \sigma\int_0^t e^{-\alpha_1 u}dZ_u\right]\\
&= e^{\alpha_1 t}\left[r_0 + \mu_t\left(e^{-\alpha_1 t} - 1\right) - \int_0^t \sigma_m dW\left(e^{-\alpha_1 t} - 1\right)du + \sigma\int_0^t e^{-\alpha_1 u}dZ_u\right]\\
&= e^{\alpha_1 t}\left[r_0 + \mu_t\left(e^{-\alpha_1 t} - 1\right) + \sigma\int_0^t e^{-\alpha_1 u}dZ_u\right].
\end{aligned}$$

The expectation follows immediately from the equation given above,

$$E\left[r_t\right] = e^{\alpha_1 t}\left[r_0 + E\left[\mu_t\right]\left(e^{-\alpha_1 t} - 1\right) + E\left[\sigma\int_0^t e^{-\alpha_1 u}dZ_u\right]\right],$$

$$E\left[r_t\right] = e^{\alpha_1 t}\left[r_0 + \mu_t\left(e^{-\alpha_1 t} - 1\right)\right].$$

Clearly, $\lim_{t\to\infty} E\left[r_t\right] = E\left[\mu_t\right]$.

***Proof of Proposition* 2.2** Write

$$c_u = r_u - \mu_u \tag{12}$$

where c_u is a solution of the Ornstein–Uhlenbeck equation:

$$dc(t) = \alpha_1 c(t) + \sigma_c dZ_t$$

Applying Ito's lemma, the c_u process is given by

$$c_u = e^{\alpha_1 u}\left(c_0 + \int_0^u \sigma_c e^{-\alpha_1 s} dZ_s\right). \tag{13}$$

Using Eq. (11) I obtain:

$$\begin{aligned} Cov\left[c_t, c_u\right] &= \sigma_c^2 e^{\alpha_1(u+t)} E\left[\int_0^t e^{-\alpha_1 s} dZ_s \left[\int_0^u e^{-\alpha_1 s} dZ_s\right]\right] \\ &= \sigma_c^2 e^{\alpha_1(u+t)} \int_0^{u\Lambda t} e^{-2\alpha_1 s} ds = \frac{\sigma_c^2}{-2\alpha_1} e^{\alpha_1(u+t)}\left(e^{2\alpha_1(u\Lambda t)} - 1\right). \end{aligned}$$

Similarly,

$$\begin{aligned} Cov\left[\mu_t, \mu_u\right] &= \sigma_m^2 E\left[\int_0^t dW_s \int_0^u dW_s\right] \\ &= \sigma_m^2 \left[\int_0^{t\Lambda u} ds\right] = \sigma_m^2 (t\Lambda u). \end{aligned}$$

Consequently,

$$\begin{aligned} Var\left[\int_0^t c_u du\right] &= Cov\left(\int_0^t c_u du, \int_0^t c_u ds\right) \\ &= \int_0^t \int_0^t Cov\left(c_u, c_s\right) duds = \int_0^t \int_0^t \frac{\sigma_c^2}{-2\beta} e^{\alpha_1(u+s)}\left(e^{-2a(u\Lambda s)} - 1\right) duds = \frac{\sigma_c^2}{2\alpha_1^3}\left(2\alpha_1 t + 3 - 4e^{\alpha_1 t} + e^{2\alpha_1 t}\right). \end{aligned}$$

Similarly,

$$Var\left[\int_0^t \mu(u) du\right] = \sigma_m^2 \int_0^t \int_0^t \frac{(s\Lambda u)^2}{2} duds = \sigma_m^2 \frac{t^3}{6}.$$

From Eq. (12), I have

$$E\left[-\int_0^t r_u du\right] = E\left[-\int_0^t \left(c_u + \mu_u\right) du\right]$$

Therefore,

$$E\left[-\int_t^T r_u du\right] = \frac{r_t - \mu_t}{\beta}\left(1 - e^{\alpha_1(T-t)}\right) - \mu_t(T-t) \tag{14}$$

Furthermore,

$$\begin{aligned} Var\left[-\int_t^T r_u du\right] &= Var\left[\int_t^T \left(c_u + \mu_u\right) du\right] \\ &= \frac{\sigma_c^2}{2\alpha_1^3}\left(2\alpha_1(T-t) + 3 - 4e^{\alpha_1(T-t)} + e^{2\alpha_1(T-t)}\right) + \sigma_m^2 \frac{(T-t)^3}{6} \end{aligned} \tag{15}$$

Proof of Theorem 2.1 Consider the expected value and variance of short rate specified in Propositions 2.1 and 2.2. I specify the price of a zero-coupon bond with maturity T at time t, $P(t, T)$using the risk neutral valuation framework:

$$P(t,T) = E\left[exp\left(-\int_t^T r_u du\right)|F_t\right] = E\left[exp\left(-\int_t^T r_u du\right)|r_t\right]$$

where $\left\{F_t\right\}$ is standard filtration.

Combining Eqs. (11) to (15), the bond price is given by

$$\begin{aligned} P\left(t,T,r_t\right) &= \exp\left(\left[-\int_t^T r_u\left(r_t\right) du\right] + \frac{1}{2}Var\left[-\int_t^T r_u\left(r_t\right) du\right]\right) \\ &= Exp\left(\frac{r_t - \mu_t}{\beta}\left(1 - e^{\alpha_1(T-t)}\right) - \mu_t(T-t)\right. \\ &\quad \left. + \frac{\sigma_c^2}{4\alpha_1^3}\left(2\alpha_1(T-t) + 3 - 4e^{\alpha_1(T-t)} + e^{2\alpha_1(T-t)} + \frac{\sigma_\mu^2}{\sigma_c^2}\frac{\alpha_1^3(T-t)^3}{3}\right)\right) \\ &= Exp\left(\left(\frac{1 - e^{\alpha_1(T-t)}}{\alpha_1}\right) r_t - \left(\frac{1 - e^{\alpha_1(T-t)}}{\alpha_1} + T - t\right)\mu_t\right. \\ &\quad \left. + \frac{\sigma_c^2}{4\alpha_1^3}\left(2\alpha_1(T-t) + 3 - 4e^{\alpha_{1(T-t)}} + e^{2\alpha_{1(T-t)}}\right) + \frac{\sigma_\mu^2}{\sigma_c^2}\frac{\alpha_1^3(T-t)^3}{3}\right) \\ &= Exp\left(A(t,T)r_t - \mu_t(A(t,T) + (T-t)) + B(t,T) + D(t,T)\right). \end{aligned} \tag{16}$$

where

$$A(t,T) = \left(\frac{1 - e^{\alpha_1 (T-t)}}{-\alpha_1} \right)$$

$$B(t,T) = \left[\frac{A(t,T) + (T-t)}{2\alpha_1^2} + \frac{A(t,T)^2}{4\alpha_1} \right] \sigma_c^2$$

and

$$D(t,T) = \frac{\sigma_\mu^2 (T-t)^3}{12}.$$

If HP and bHP trends are used, the bond price will be estimated using the signal to noise ratio (q). The bond price is given by

$$= Exp\left(\frac{r_t - \mu_t}{\alpha_1} \left(1 - e^{\alpha_1 (T-t)}\right) - \mu_t (T-t) + \frac{\sigma^2}{4\alpha_1^3 (1+q)} \left(2\alpha_1 (T-t) + 3 - 4e^{\alpha_1 (T-t)} + e^{2\alpha_1 (T-t)} + q \frac{\alpha_1^3 (T-t)^3}{3} \right) \right)$$

$$= Exp\left(\left(\frac{1 - e^{\alpha_1 (T-t)}}{\alpha_1} \right) r_t - \left(\frac{1 - e^{\alpha_1 (T-t)}}{\alpha_1} + (T-t) \right) \mu_t + \frac{\sigma^2}{2\alpha_1^2 (1+q)} \left(\frac{1 - e^{\alpha_1 (T-t)}}{\alpha_1} \right) + \frac{\sigma^2}{2\alpha_1^2 (1+q)} (T-t) \right.$$
$$\left. + \frac{\sigma^2}{2\alpha_1 (1+q)} \left(\frac{1 - 2e^{\alpha_1 (T-t)} + e^{2\beta (T-t)}}{2\alpha_1^2} \right) + \left(\frac{q\sigma^2 (T-t)^3}{12(1+q)} \right) \right)$$

$$= Exp\left(A(t,T) r_t - \mu_t (A(t,T) + (T-t)) + \frac{\sigma^2}{2\alpha_1^2 (1+q)} A(t,T) + \frac{\sigma^2}{2\alpha_1^2 (1+q)} (T-t) \right.$$
$$\left. + \frac{\sigma^2}{2\alpha_1 (1+q)} A(t,T)^2 + \left(\frac{q\sigma^2 (T-t)^3}{12(1+q)} \right) \right)$$

$$= Exp\left(A(t,T) r_t - \mu_t (A(t,T) + (T-t)) + B(t,T) + D(t,T) \right)$$

where

$$A(t,T) \left(\frac{1 - e^{\alpha_1 (T-t)}}{\alpha_1} \right),$$

$$B(t,T) \equiv \frac{\sigma^2}{2\alpha_1^2 (1+q)} (A(t,T) + (T-t)) + \frac{\sigma^2 A(t,T)^2}{2\beta (1+q)},$$

and

$$D(t,T) = \left(\frac{q\sigma^2 (T-t)^3}{12(1+q)} \right).$$

Appendix 2: Signal–noise ratio

This appendix provides details of the derivations related to the signal–noise ratio,

$$q = \sigma_\eta^2/\sigma_v^2 = \sigma_\eta^2/\left(1 - \rho^2 - \sum_{j=1}^{p} \rho_j^2\right)\sigma_c^2.$$

The resulting transitory component from (3), c_t, possesses weak dependence properties with mean zero. Thus, c_t is an AR(p) process:

$$c_{t+1} = \rho c_t + \sum_{j=1}^{p} \rho_j c_{t-j} + v_{t+1}, v_t \sim NID\left(0, \sigma_v^2\right).$$

The variance is

$$= E\left(c_{t+1}^2\right) - \left(Ec_{t+1}\right)^2.$$

It follows immediately that

$$\sigma_c^2 = \frac{\sigma_v^2}{1 - \rho^2 - \sum_{j=1}^{p} \rho_j^2}. \tag{17}$$

Now, consider the case in which μ_t is a random walk process:

$$\Delta\mu_{t+1} = \eta_{t+1} \sim NID\left(0, \sigma_\eta^2\right).$$

I specify μ_t as a driftless random walk process:

$$\mu_{t+1} = \mu_t + \eta_t,$$

which can be written as

$$\mu_{t+1} = \mu_0 + \sum_{i=1}^{t} \eta_t.$$

The variance follows immediately

$$\sigma_\mu^2 = (t+1)\sigma_\eta^2,$$

and

$$\sigma_{\Delta\mu_{t+1}}^2 = (t+1)\sigma_\eta^2 + (t)\sigma_\eta^2 - 2cov\left(\mu_{t+1}, \mu_t\right) = (t+1)\sigma_\eta^2 + (t)\sigma_\eta^2 - 2(t+1-1)\sigma_\eta^2 = \sigma_\eta^2$$

Because the HP filter assume that W_t and Z_t are two independent Brownian motions I have,

$$\sigma^2 = \sigma_\mu^2 + \sigma_c^2, \tag{18}$$

Because μ_t is driftless a random walk process, it follows that

$$\sigma_{\mu}^{2} = (t + 1 - t)\sigma_{\eta}^{2},$$

and

$$\sigma_{\Delta\mu}^{2} = \sigma_{\eta}^{2}$$

.

Therefore,

$$q = \frac{\sigma_{\Delta\mu}^{2}}{\sigma_{c}^{2}} = \frac{\sigma_{\eta}^{2}}{\sigma_{c}^{2}} = \frac{\sigma_{\mu}^{2}}{\sigma_{c}^{2}}. \tag{19}$$

Appendix 3: Autocorrelated error term

This appendix provides details of derivation related to the variance of the autocorrelated error term, π_{t+1},, and the variance of the White noise error term, ε_{t+1}, of the stochastic mean NSM model.

The HP filter case

From Eq. (4a) and Eq. (5a) I have

$$\sum_{j=1}^{p} \rho_j c_{t-j} + v_{t+1} = c_{t+1} - \rho c_t$$

$$\pi_{t+1} = \sum_{j=1}^{p} \rho_j c_{t-j} + v_{t+1} + \eta_{t+1} = \sum_{j=1}^{p} \rho_j c_{t-j} + \varepsilon_{t+1}.$$

It follows that

$$\pi_{t+1} = c_{t+1} - \rho c_t + \eta_{t+1},$$

Following Hodrick and Prescott (1997) and Ravn and Uhlig (2002), I make the assumption of independence between the permanent and transitory shocks, such that $\sigma_{c,\eta}^{2} = 0$. (See Kohn & Ansley, 1987). Hence, the variance follows immediately

$$\sigma_{\pi}^{2} = \sigma_{c}^{2} + \rho^{2}\sigma_{c}^{2} - 2\rho\mathrm{cov}\left(c_{t+1}, c_t\right) + \sigma_{\eta}^{2}.$$

$$\sigma_{\pi}^{2} = \sigma_{c}^{2} + \rho^{2}\sigma_{c}^{2} - 2\rho^{2}\sigma_{c}^{2} + \sigma_{\eta}^{2}.$$

$$\sigma_\pi^2 = \left(1-\rho^2\right)\sigma_c^2 + \sigma_\eta^2.$$

Using (19) I obtain,

$$\sigma_\pi^2 = \frac{\left(1-\rho^2\right)}{\left(1-\rho^2-\sum_{j=1}^{p}\rho_j^2\right)}\sigma_v^2 + \sigma_\eta^2.$$

Defining $\Omega = 1-\rho^2$ and $\phi = -\sum_{j=1}^{p}\rho_j^2$ we can write,

$$\sigma_\pi^2 = \frac{\Omega}{\Omega+\phi}\sigma_v^2 + \sigma_\eta^2,$$

which can be written as,

$$\sigma_\pi^2 = \frac{\Omega}{\Omega+\phi}\sigma_v^2 + \sigma_\eta^2 + \sigma_v^2 - \sigma_v^2$$

$$= \left(\frac{\Omega}{\Omega+\phi} - 1\right)\sigma_v^2 + \sigma_\eta^2 + \sigma_v^2,$$

$$= \left(\frac{\Omega}{\Omega+\phi} - 1\right)\sigma_v^2 + \sigma^2,$$

$$= \left(\frac{\Omega}{\Omega+\phi} - 1\right)\sigma_v^2 + \sigma^2,$$

$$\sigma_\pi^2 = -\frac{\phi}{\Omega+\phi}\sigma_v^2 + \sigma^2,$$

Because $\Omega = 1-\rho^2$ and $\alpha_1 = \rho - 1$, I can write: $\Omega = \alpha_1^2 - 2\alpha_1$, then I have, $\sigma_\pi^2 = \left(-\frac{\varphi}{\left(\varphi+\alpha_1^2-2\alpha_1\right)}\right)\sigma_v^2 + \sigma^2\cdot$

Because $q = \frac{\sigma_\eta^2}{\sigma_v^2}$ we obtain,

$$\sigma_\eta^2 = q\sigma_v^2,$$

$$\sigma^2 = (1+q)\sigma_v^2,$$

$$\sigma_v^2 = \left(\frac{1}{1+q}\right)\sigma^2.$$

Plug this in (19) I get,

$$\sigma_\pi^2 = \left(-\frac{\phi}{\Omega+\phi}\right)\left(\frac{1}{1+q}\right)\sigma^2 + \sigma^2,$$

$$\sigma_\pi^2 = \left(1 - \frac{\phi}{(\Omega+\phi)(1+q)}\right)\sigma^2.$$

Because $\Omega = 1-\rho^2$ and $\alpha_1 = \rho - 1$, we can write: $\Omega = \alpha_1^2 - 2\beta$, then we have,

$$\sigma_\pi^2 = \left(1 - \frac{\phi}{\left(\phi + \alpha_1^2 - 2\beta\right)(1+q)}\right)\sigma^2,$$

$$\sigma_\pi^2 = \left(\frac{\left(\alpha_1^2 - 2\alpha_1\right)(1+q) + q\phi}{\left(\phi + \alpha_1^2 - 2\alpha_1\right)(1+q)}\right)\sigma^2,$$

$$\sigma_\pi^2 = \left(\frac{\left(\alpha_1^2 - 2\alpha_1\right) + \frac{q}{(1+q)}\phi}{\left(\phi + \alpha_1^2 - 2\alpha_1\right)}\right)\sigma^2$$

Let

$$\partial = \left(\frac{\left(\alpha_1^2 - 2\alpha_1\right) + \frac{q}{(1+q)}\phi}{\left(\phi + \alpha_1^2 - 2\alpha_1\right)}\right)$$

Then

$$\sigma_\pi^2 = \partial\sigma^2.$$

The BN filter case

Because the BN filter assumes that $Corr\left(\eta_t, v_t\right) = \rho_{\eta v} = 1$, I have

$$\sigma_\pi^2 = \left(-\frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)}\right)\sigma_v^2 + \sigma_v^2 + \sigma_\mu^2 + 2\sigma_v\sigma_\mu$$

$$\sigma_\pi^2 = \left(1 - \frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)}\right)\sigma_v^2 + \sigma_\mu^2 + 2\sigma_v\sigma_\mu$$

$$\sigma_\pi^2 = \left(1 - \frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)}\right)\sigma_v^2 + \sigma^2 - \sigma_v^2$$

$$\sigma_\pi^2 = \left(-\frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)}\right)\sigma_v^2 + \sigma^2$$

$$\sigma_\pi^2 = \left(1 - \frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)} + 2\frac{\sigma_\mu}{\sigma_v}\right)\sigma_v^2 + \sigma_\mu^2$$

$$\sigma_\pi^2 = \left(-\frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)} + 2\frac{\sigma_\mu}{\sigma_v}\right)\sigma_v^2 + \sigma_v^2 + \sigma_\mu^2$$

Define $\lambda = \frac{\sigma_v^2}{\sigma^2}$ I obtain,

$$\sigma_\pi^2 = \left(-\frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)}\right)\lambda\sigma^2 + \sigma^2$$

$$\sigma_\pi^2 = \left(1 - \frac{\varphi}{\left(\varphi + \alpha_1^2 - 2\alpha_1\right)}\lambda\right)\sigma^2$$

Appendix 4: GMM estimation

Define λ as the entire parameter vector. I have the following orthogonality conditions:

1. The Aït-Sahalia Model: $\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_0 - \alpha_1 r_t - \alpha_2 r_t^2 - \alpha_3 r_t^{-1}\right]$. The moment conditions are given by:

$$h\left(r_{t+1}, \lambda\right) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_t, \varepsilon_{t+1} r_t^2, \varepsilon_{t+1} r_t^{-1}, \varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}, \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}\right) r_t\right]$$

2. **The CKLS model:** $\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_0 - \alpha_1 r_t\right]$. The moment conditions are given by:

$$h\left(r_{t+1}, \lambda\right) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_{t,} \varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}, \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}\right) r_t\right]$$

3. **The AG model:** $\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_0 - \alpha_1 r_t\right]$. The moment conditions are given by:

$$h(r_{t+1},\lambda) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_{t,} \varepsilon_{t+1} r_t^2, \varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}, (\varepsilon_{t+1}^2 - \sigma^2 r_t^3) r_t, (\varepsilon_{t+1}^2 - \sigma^2 r_t^3) r_t^3\right]$$

4. **The CIR Model:** $\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_0 - \alpha_1 r_t\right]$. The moment conditions are given by:

$$h(r_{t+1},\lambda) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_{t,} \varepsilon_{t+1}^2 - \sigma^2 r_t, (\varepsilon_{t+1}^2 - \sigma^2 r_t) r_t\right]$$

5. **The Vasicek Model:** $\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_0 - \alpha_1 r_t\right]$. The moment conditions are given by:

$$h(r_{t+1},\lambda) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_{t,} \varepsilon_{t+1}^2 - \sigma^2, (\varepsilon_{t+1}^2 - \sigma^2) r_t\right]$$

6. **The BDF Model:** $\varepsilon_{t+1} = \left[r_{t+1} - r_t + \alpha_1(\theta_t - r_t)\right]$. The moment conditions are given by:

$$h(r_{t+1},\lambda) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_{t,} \varepsilon_{t+1} r_{1,t}, \varepsilon_{t+1} r_{2,t,} \varepsilon_{t+1}^2 - \sigma^2, (\varepsilon_{t+1}^2 - \sigma^2) r_t\right]$$

where

$$\theta_t = \alpha_{\theta,1}\left[B(T_2) T_1 r_{1,t} - B(T_1) T_2 r_{2,T}\right]$$

$$B(T) = \frac{2\left(e^{\delta T} - 1\right)}{(\delta + k)\left(e^{\delta T} - 1\right) + 2\delta}$$

$$\sigma^2 = \sigma_0^2 + \sigma_1^2 r_t$$

and

$$\delta = \sqrt{\left(\delta^2 + \sigma_1^2\right)}$$

where $\sigma_1 = 0$ if BDF-VAS is considered and $\sigma_0 = 0$ if BDF-CIR is considered.

7. **The Heston Model:** $\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_0 - \alpha_1 r_t\right]$ and $\varepsilon_{v,t+1} = \left[v_{t+1} - v_t - \alpha_v - \alpha_v v_t\right]$. The moment conditions are given by:

$$h(r_{t+1},\lambda) = \left[\varepsilon_{t+1}, \varepsilon_{v,t+1}, \varepsilon_{t+1} r_t, \varepsilon_{v,t+1} v_t, \varepsilon_{v,t+1}^2 - \sigma_v^2, \left(\varepsilon_{v,t+1}^2 - \sigma_v^2\right) v_t\right]$$

I follow Andersen, and Lund (1997) and specify v_{t+1} as a GARCH(1,1) model.

8. **The Chen Model:** $\varepsilon_{t+1} = \left[r_{t+1} - r_t + \alpha_1(\theta_t - r_t)\right]$, $\varepsilon_{v,t+1} = \left[v_{t+1} - v_t - \alpha_{vo} - \alpha_{v1} v_t\right]$, and $\varepsilon_{\theta,t+1} = \left[\theta_{t+1} - \theta_t - \alpha_{\theta o} - \alpha_{\theta 1}\theta_t\right]$. The moment conditions are given by: $h(r_{t+1},\lambda) = \left[\varepsilon_{t+1}, \varepsilon_{v,t+1}, \varepsilon_{\theta,t+1}, \varepsilon_{t+1} r_{t,} \varepsilon_{v,t+1} v_{t,}, \varepsilon_{\theta,t+1}\theta_{t,}, \varepsilon_{v,t+1}^2 - \sigma_v^2, \left(\varepsilon_{v,t+1}^2 - \sigma_v^2,\right) v_t, \varepsilon_{\theta,t+1}^2 - \sigma_\theta^2, \left(\varepsilon_{\theta,t+1}^2 - \sigma_\theta^2,\right)\theta_t\right]$. I follow Andersen, and Lund (1997) and specify v_{t+1} as a GARCH(1,1) model.

9. **The Al-Zoubi** (2019) **I(2) model**: $\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_1 c_t\right]$.

$$h(r_{t+1},\lambda) = \left[\pi_{t+1}, \varepsilon_{t+1} c_t, \varepsilon_{t+1}^2 - \sigma^2, (\varepsilon_{t+1}^2 - \sigma^2) c_t\right]$$

10. **The NSM-HP and NSM-bHP Models:** $\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_1 c_t\right]$. Letting:

$$\partial = \left(\frac{\left(\alpha_1^2 - 2\alpha_1\right) + \frac{q}{(1+q)}\phi}{\left(\phi + \alpha_1^2 - 2\alpha_1\right)} \right)$$

then

$$h\left(r_{t+1}, \lambda\right) = \left[\pi_{t+1}, \pi_{t+1}c_t, \pi_{t+1}^2 - \partial\sigma^2, \left(\pi_{t+1}^2 - \partial\sigma^2\right)c_t\right]$$

.

11. **The NSM-HAM Model:**
$\varepsilon_{t+1} = \left[r_{t+1} - r_t - \alpha_1\left(r_t - b_0 - b_1 r_{t-7} - b_2 r_{t-8} - b_3 r_{t-9} - b_4 r_{t-10}\right)\right]$. The moment conditions are given by:
$h\left(r_{t+1}, \lambda\right) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_t, \varepsilon_{t+1}^2 - \sigma^2, \left(\varepsilon_{t+1}^2 - \sigma^2\right) r_t\right]$.

To test the validity of my model, I minimize the GMM criterion of the form,

$$\frac{1}{T}\left[f\left(x_{t+1}, Y_t, \lambda\right)\right]' W_T \frac{1}{T}\left[f\left(x_{t+1}, Y_t, \lambda\right)\right], \tag{18}$$

where W_T is a consistent estimate of $\left(\mathrm{var}\left[(1/T)\left(f\left(x_{t+1}, Y_t, \lambda\right)_t\right)\right]\right)^{-1}$ and Y_t is a K-dimensional vector of instrumental variables.

Under the null hypothesis that GMM restrictions are valid, I have that:

$$\frac{1}{T}\left[f\left(x_{t+1}, Y_t, \lambda\right)\right]' W_T \frac{1}{T}\left[f\left(x_{t+1}, Y_t, \lambda\right)\right] \overset{a}{\sim} \chi_{L-k}^2. \tag{19}$$

For my model to be robust with respect to heteroskedasticity and autocorrelation variance, I follow Inoue and Shintani (2006) and use the Parzen kernel of Gallant (1987) with two lags to calculate the moments weighting matrix.

Acknowledgement I would like to thank anonymous referees, Pietro Veronesi, Robert de Jong and Jun Yu for their helpful comments and suggestions. The project is partially supported by Alfaisal University research grant N. 18102.

Fund No funding was received to assist with the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Ahn, D.-H., & Gao, B. (1999). A parametric nonlinear model of term structure dynamics. *Review of Financial Studies, 12*(4), 721–762.

Aït-Sahalia, Y. (1996a). Nonparametric pricing of interest rate derivatives securities. *Econometrica, 64*(3), 527–560.

Aït-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *Review of Financial Studies, 9*(2), 385–426.

Aït-Sahalia, Y. and Kimmel, R., (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of financial economics*, *83*(2), 413–452.

Aksoy, Y., Orphanides, A., Small, D., Wiel, V., & Wilcox, D. (2006). A quantitative exploration of the opportunistic approach to disinflation. *Journal of Monetary Economics, 53*(8), 1877–1893.

Al-Zoubi, H. A. (2009). Short-term spot rate models with nonparametric deterministic drift. *Quarterly Review of Economics and Finance, 49*(3), 731–747.

Al-Zoubi, H.A., 2019. Bond and option prices with permanent shocks. *Journal of Empirical Finance*, *53*, 272–290.

Andersen, T.G. and Lund, J., (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of econometrics*, *77*(2), 343–377.

Ang, A., Bekaert, G., & Wei, M. (2008). The term structure of real rates and expected inflation. *The Journal of Finance, 63*(2), 797–849.

Bakshi, G., Crosby, J., Gao, X., & Hansen, J. W. (2023a). Treasury option returns and models with unspanned risks. *Journal of Financial Economics, 150*(3), 103736.

Bakshi, G., Gao, X., & Xue, J. (2023). Recovery with applications to forecasting equity disaster probability and testing the spanning hypothesis in the treasury market. *Journal of Financial and Quantitative Analysis, 58*(4), 1808–1842.

Balduzzi, P., Das, S., & Foresi, S. (1998). The central tendency: A second factor in bond yields. *The Review of Economics and Statistics, 80*(1), 62–72.

Bandi, F. M. (2002). Short-term interest rate dynamics: A special approach. *Journal of Financial Economics, 65*(1), 73–110.

Bauer, M. D., & Hamilton, J. D. (2017). Robust bond risk premia. *The Review of Financial Studies, 31*(2), 399–448.

Bauer, M. D., & Rudebusch, G. D. (2020). Interest rates under falling stars. *American Economic Review, 110*(5), 1316–1354.

Bekaert, G., Cho, S., & Moreno, A. (2010). New Keynesian macroeconomics and the term structure. *Journal of Money, Credit and Banking, 42*(1), 33–62.

Bergstrom, A. R. (1986). The estimation of open higher-order continuous time dynamic models with mixed stock and flow data. *Econometric Theory, 2*(3), 350–373.

Bergstrom, A. R. (1989). Optimal forecasting of discrete stock and flow data generated by a higher order continuous time system. *Computers & Mathematics with Applications, 17*(8/9), 1203–1214.

Beveridge, S. and Nelson, C.R., (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics*, *7*(2), 151–174.

Bierens, H. J. (1997). Testing the unit root with drift hypothesis against nonlinear trend stationary, with an application to the us price level and interest rate. *Journal of Econometrics, 81*(1), 29–64.

Blaskowitz, O., & Herwartz, H. (2011). On economic evaluation of directional forecasts. *International Journal of Forecasting, 27*(4), 1058–1065.

Bleich, D., Fendel, R., & Rülke, J.-C. (2012). Inflation targeting makes the difference: Novel evidence on inflation stabilization. *Journal of International Money and Finance, 31*(5), 1092–1105.

Bergmeir, C., Costantini, M., & Benítez, J. M. (2014). On the usefulness of cross-validation for directional forecast evaluation. *Computational Statistics & Data Analysis, 76*, 132–143.

Cai, L., & Swanson, N. R. (2011). In-and out-of-sample specification analysis of spot rate models: Further evidence for the period 1982–2008. *Journal of Empirical Finance, 18*(4), 743–764.

Campbell, J.Y. and Shiller, R.J., (1987). Cointegration and tests of present value models. *Journal of political economy*, *95*(5), 1062–1088.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., & Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance, 47*(3), 1209–1227.

Chapman, D. A., & Pearson, N. D. (2000). Is the short rate drift actually nonlinear? *Journal of Finance, 55*(1), 355–388.

Chen, L., (1996). Stochastic mean and stochastic volatility: a three-factor model of the term structure of interest rates and its applications in derivatives pricing and risk management. Blackwell publishers

Cieslak, A., & Povala, P. (2015). Expected returns in treasury bonds. *The Review of Financial Studies, 28*(10), 2859–2901.

Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *American Economic Review, 95*(1), 138–160.

Cogley, T. and Nason, J.M., (1995). Output dynamics in real-business-cycle models. *The American Economic Review*, 492–511.

Cooper, I., & Priestley, R. (2009). Time-varying risk premiums and the output gap. *The Review of Financial Studies, 22*(7), 2801–2833.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica, 53*(2), 385–407.

Dai, Q., & Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance, 55*(5), 1943–1978.

de Jong, R. M., & Sakarya, N. (2016). The econometrics of the Hodrick–Prescott filter. *Review of Economics and Statistics, 98*(2), 310–317.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association, 74*(366), 427–431.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance, 57*(1), 405–443.

Duffee, G. R. (2011). *Forecasting with the term structure: The role of no-arbitrage*. Johns Hopkins University.

Duffee, G. (2013). Forecasting interest rates. In Handbook of economic forecasting (Vol. 2, pp. 385–426). Elsevier.

Duffee, G. R. (2018). Expected inflation and other determinants of treasury yields. *The Journal of Finance, 73*(5), 2139–2180.

Duffie, D. and Kan, R., (1996). A yield-factor model of interest rates. *Mathematical finance*, *6*(4), 379–406.

Durham, G. B. (2003). Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics, 70*(3), 463–487.

Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 680–692.

Fama, E. (2006). The behavior of interest rates. *Review of Financial Studies, 19*(2), 359–379.

Gallant, A. R. (1987). *Nonlinear statistical models*. New York: Wiley.

Greenwood, R., & Vayanos, D. (2014). Bond supply and excess bond returns. *The Review of Financial Studies, 27*(3), 663–713.

Hamilton, J. D. (2018). Why you should never use the Hodrick–Prescott filter. *Review of Economics and Statistics, 100*(5), 831–843.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica, 50*(4), 1029–1054.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*(2), 281–291.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: An empirical investigation. *Journal of Money, Credit, and Banking, 29*(1), 1–16.

Inoue, A., & Shintani, M. (2006). Bootstrapping GMM estimators for time series. *Econometrics, 133*(2), 531–555.

Jamshidian, F. (1989). An exact bond option formula. *The Journal of Finance, 44*(1), 205–209.

Joslin, S., Priebsch, M., & Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance, 69*(3), 1197–1233.

Kohn, R., & Ansley, C. F. (1987). Signal extraction for finite nonstationary time series. *Biometrika, 74*(2), 411–421.

Liu, R. (2018). Forecasting bond risk premia with unspanned macroeconomic information. *Quarterly Journal of Finance, 9*(1), 1940001.

MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics, 12*(2), 167–176.

Mehra, Y. P. (2004). The output gap expected future inflation and inflation dynamics: Another look. *B.E. Journal of Macroeconomics*. https://doi.org/10.2202/1534-5998.1194

Orphanides, A., & Wieland, V. (2000). Inflation zone targeting. *European Economic Review, 44*(7), 1351–1387.

Orphanides, A., & Wilcox, D. W. (2002). An opportunistic approach to disinflation. *International Finance, 5*(1), 47–71.

Pantula, S. G., Gonzalez-Farias, G., & Fuller, W. A. (1994). A comparison of unit-root test criteria. *Journal of Business and Economic Statistics, 12*(4), 449–459.

Perron, P. (1989). The great crash, the oil-price shock, and the unit-root hypothesis. *Econometrica, 57*(6), 1361–1402.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica, 55*(2), 277–301.

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75*(2), 335–346.

Phillips, P. C. B., & Sainan, J. (2021). Business cycles, trend elimination, and the HP filter. *International Economic Review, 62*(2), 469–520.

Phillips, P. C., & Shi, Z. (2021). Boosting: Why you can use the HP filter. *International Economic Review, 62*(2), 521–570.

Phillips, P. C. B., & Yu, J. (2005). Jackknifing bond option prices. *Review of Financial Studies, 18*(2), 707–742.

Phillips, P. C. B., & Yu, J. (2009). Maximum likelihood and gaussian estimation of continuous time models in finance. In T. Mikosch & J. P. Krei (Eds.), *Handbook of financial time series* (pp. 497–530). Berlin: Springer.

Piazzesi, M., & John, C. (2009) Decomposing the yield curve. In *2009 Meeting Papers, no. 18. Society for Economic Dynamics*.

Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick–Prescott filter for the frequency of observations. *Review of Economics and Statistics, 84*(2), 371–376.

Stock, J. H., & Watson, M. W. (1988). Variable trends in economic time series. *Journal of Economic Perspectives, 2*(3), 147–174.

Tang, C. Y., & Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics, 149*(1), 65–81.

Taylor, J.B., (1993), December. Discretion versus policy rules in practice. In *Carnegie-Rochester conference series on public policy* (Vol. 39, pp. 195–214). North-Holland.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics, 5*(2), 177–188.

Yu, J. (2012). Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics, 169*(1), 114–122.