# Tractable Term Structure Models: A New Approach

Bruno Feunou Jean-Sébastien Fontaine
Anh Le Christian Lundblad

October 2015

**Abstract**

We greatly expand the space of tractable term structure models (TTSM). We consider one example of TTSM with positive yields together with rich volatility and correlations dynamics. Bond prices can be expressed in closed-form and estimation is straightforward. We find that the early stages of a recession have distinct effects on yield volatility. Upon entering a recession when yields are far from the lower bound, (1) the volatility term structure becomes flatter, (2) the level and slope of yields are nearly uncorrelated, and (3) the second principle component of yields plays a larger role. However, these facts are significantly different when yields are close to the lower bound. Entering a recession in such a setting, (1) the volatility term structure instead steepens, (2) the level and slope factors are strongly correlated, and (3) the second principle component of yields plays a smaller role. Existing DTSMs do not capture the changes in the cyclical responses of the volatility term structure near the lower bound.

**Keywords**: Term Structure, Zero Lower Bound, Stochastic Dominance, No-arbitrage
**JEL codes**: G12

---

Bank of Canada, Bank of Canada, Smeal College of Business, and Kenan–Flagler Business School University of North Carolina Chapel Hill. We thank for comments and suggestions Peter Christoffersen, Antonio Diez de los Rios, Peter Feldhutter, Richard Luger, seminar and conference participants at the 2014 Tremblant Risk Conference, McGill University, the Federal Reserve Bank of New York, Banque de France, 2015 CIRANO finance workshop. Corresponding author: `jsfontaine@bankofcanada.ca`.

# I Introduction

The Absence of Arbitrage (AOA) is an essential feature of bond pricing models. Given a model for bond prices, the managers of a bond portfolio should discard estimates of the portfolio's future risk and returns if the underlying forecasts of bond yields offer arbitrage opportunities. Similarly, its risk manager should attribute zero probability to these configurations of yield forecasts.

Producing yield forecasts that are free of arbitrage is not a trivial task. However, the fundamental theorem of finance can help us transform the problem into a more manageable form. AOA is equivalent to the existence of a strictly positive change of measure from the historical measure $\mathbb{P}$ to a measure $\mathbb{Q}$, where asset prices are martingales when discounted at the risk-free rate; the "risk-neutral" measure (Harrison and Kreps, 1979; Delbaen and Schachermayer, 1994, 1998). Since the terminal price of the risk-free bond is known, the construction of arbitrage-free bond prices based on this equivalence requires that we specify: (i) the one-period risk-free rate $r_t$ (ii) the dynamics of $r_t$ under $\mathbb{P}$ and (iii) a strictly positive change of measure. The price of a bond with any maturity $h$ solves the expectation of future short rates $r_{t+1}, \dots, r_{t+h}$ under $\mathbb{Q}$.

We take a different route, introducing the family of Tractable Term Structure Models (TTSMs). We specify bond prices directly and then check whether these prices are consistent with the AOA. We proceed in a few steps. First, we show that our construction of bond prices is free of dominant trading strategies (Rothschild and Stiglitz, 1970; Levy, 1992). Of course, many arbitrage opportunities are not dominant. We then show that any remaining arbitrage opportunity within our framework must be self-financing and offer no initial payoff (if an opportunity exists). Building on these results, we show that our construction of bond prices precludes arbitrage opportunities as long as transaction costs are non-zero. In other words, our construction is arbitrarily close to the AOA. Finally, we show (in an Appendix) how to construct a sequence of no-arbitrage frictionless models that in the limit converge to our pricing model proposed in Assumption 1.

Our approach comes at a cost. Absent a pricing kernel, the prices of risk remain unspecified. We can still easily compute bond risk premia and Sharpe ratios, but we cannot offer a decomposition of the premia between quantities and prices of risk (inside the model). Yet, our approach nests the widely-used Nelson and Siegel (1987) (NS) term structure representation of yields, providing an instructive example. Notwithstanding the popularity and empirical success of the NS representation, Bjork and Christensen (1999); Filipovic

(1999) show that it is not consistent with the AOA in a context without trading frictions. Our approach shows that the NS representation is arbitrarily close to the AOA, explaining the common observations that models based on NS loadings are "almost" free of arbitrage. Indeed, Coroneo, Nyholm, and Vidova-Koleva (2011) cannot reject the null hypothesis that a dynamic version of NS (Diebold and Li, 2006) is consistent with the AOA . Conversely, the empirical success of the NS decomposition supports the usefulness of TTSMs.

What do we gain from moving toward TTSMs? We emphasize that TTSMs are parsimonious, flexible, and highly tractable. Directly specifying bond prices guarantees tractability, circumventing the need to compute the oftentimes intractable integral under $\mathbb{Q}$. As a prime example, we can easily impose that all yields are positive. Beyond tractability, our deeper motivation follows from the enduring tension between fitting the dynamics of yields and of yield volatility (Dai and Singleton, 2000). Gaussian dynamic term structure models (DTSMs) with constant variance can fit the evidence on bond predictability (the dynamics of yields) but not the volatility of yields. In contrast, a Standard $A_n(m)$ models allows for stochastic volatility but cannot fit the bond risk premium. Recent research address this tension, but in situations when yields are far from the lower bound (Cieslak and Povala 2011,CP 2015). However, approaching (and reaching) the lower bound creates strong implications for the fit of yields and yield volatility. Failing to impose the lower bound produces biased forecasts and misses the dramatic volatility compression of short-term yields exhibited in recent years. Table 1 summarizes how this tension continues to haunt existing models.

| | | Volatility Compression at the Lower Bound | |
|---|---|---|---|
| Cyclical Short-Rate Volatility | | no | yes |
| | no | $A_0(3)$ | Black (1995) |
| | yes | CP (2015) | this paper |

Table 1: Matching yield volatility stylized facts.

Our approach to specifying bond prices relies only on weak technical conditions for the risk factor dynamics. Within our general framework, we introduce a class of TTSMs which are as close as possible to the standard models except for two key departures: (i) yields must be positive and (ii) facilitate a flexible specification of the volatility dynamics.[1] In particular, we specify risk factors with conditionally Gaussian VAR(1) dynamics. To ensure

[1]The lower bound can be any constant, and not necessarily zero.

positivity, we write the short-rate function as a logistic transformation of the usual linear rule. Intuitively, our specification is close in spirit to that in Black (1995). In fact, we adapt the logistic form to nest the linear and the max case if one key parameter tends to either 0 or $\infty$ in the limit, respectively. Finally, we consider a flexible multivariate specification for the conditional volatilities and correlations of the risk factor innovations.

Our results highlight three key contributions. First, we capture existing stylized facts within a traditional model with positive interest rates. In the sample away from the lower bound, our model matches standard risk premium estimates from a standard Gaussian 3-factor model (by construction), but also fits the yield volatility dynamics. We confirm that the volatility term structure slopes upward but with a hump-shape around the 2-year maturity (Piazzesi, 2005), that the volatility of all yields rise in recession; and that the volatility term structure becomes flat or inverted in the early stages of a recession (Cieslak and Povala, 2011) (the volatility of short-term yield increases most).

Second, we find that changes in the volatility term structure alter the nature and the explanatory power of the principal components extracted from yields. Consider the following exercise. Fix a date, simulate the term structure of yields one-period ahead, and compute the principal components across simulations. This corresponds to a Principal Component Analysis applied to the *conditional* correlation matrix of yields.[2] Away from the lower bound, the results show that the explanatory power of the first (conditional) component varies from close to 95 percent in good times to below 85 percent during the recessions. In these periods, the second component plays a correspondingly greater role, reflecting the actions of the Federal Reserve. This effect is entirely captured by changes in the variance of the risk factor, since the factor loadings are constant.

Finally, we find that response of the volatility term structure to economic conditions changes sign near the lower bound. The short-rate volatility falls—instead of rises—and the volatility term structure steepens—instead of flattens—as the Federal Reserve approaches and reaches the lower bound. Existing models do not capture this shift in the volatility dynamics. In addition, the volatility compression also changes the correlation structure between yields. Near the lower bound, the explanatory power of the first component quickly reaches and stays at 95 percent, while the other components become less important. The first components play a greater role because yields are more correlated. This happens either away from the lower bound, where the correlation between the level and slope is large and negative, or near the lower bound, where the correlation between the level and slope is large and positive (as

[2]We use a simulation-based PCA, since yields are non-linear function of the risk factors in our model.

high as 0.8).

Our specification of bond prices is closely related to several existing DTSMs, and we discus these connections in detail. We find that results from an affine Gaussian TTSM and a DTSM with constant variance are indistinguishable. Further, we confirm that risk premia estimates and short-rate forecasts from linear models suffer from severe bias after 2008, especially for short maturities and for short horizons. We also show how to construct close counterparts to the affine-quadratic and Black (1995) max models. Within the standard framework, both approaches pose severe estimation or tractability problems.[3]

Section II discusses our key assumptions and introduces the family of TTSMs. Section III establishes theoretically that our bond prices are arbitrarily close to the absence of arbitrage. Section IV details the specification, identification, and estimation of one class of TTSMs with positive yields. Sections V-VI present the empirical results for linear and non-linear models, respectively, and Section VII concludes.

# II Tractable Term Structure Models

## A Bond prices

This section introduces the family of Tractable Term Structure Models (TTSMs) where bond prices are specified directly. This approach to constructing bond prices is flexible and tractable, guaranteeing that bond yields are available in closed form with only minimal assumptions about the risk factor dynamics. The next section verifies that the resulting bond prices are free of arbitrage.

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

[3] DTSMs based on Black (1995) based on truncated distributions does not deliver closed-form bond prices and yields. Closed-form approximation schemes have been proposed recently, but only for the cases where risk factors have Gaussian VAR(1) dynamics (Krippner, 2011; Christensen and Rudebusch, 2013; Bauer and Rudebusch, 2013; Wu and Xia, 2013).

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

$$P_0(X_t) \equiv 1, \tag{1}$$
$$P_n(X_t) = P_{n-1}(g(X_t)) \times \exp(-m(X_t)), \tag{2}$$

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

The recursive structure of bond prices is the distinctive feature of Assumption 1. The initial value in Equation (1) corresponds to the fact that maturing bonds are redeemed at face value, which we normalize to one dollar. Equation (2) states that the pricing function for the $n$-period bond is given recursively from the pricing function for the ($n$-1)-period bond. Since, zero-coupon bond prices are available in closed form for all maturities, it follows that all forward rates and zero-coupon yields are available in closed form. This calculation is spelled out formally in Theorem 1.

**Theorem 1.** *Assumption 1 implies that the n-period zero-coupon yield is given by:*

$$y_{n,t} = (1/n) \sum_{i=0}^{n-1} m(g^{\circ i}(X_t)), \tag{3}$$

*and that the one-period forward rate n periods in the future is given by:*

$$f_{n,t} = m(g^{\circ n}(X_t)), \tag{4}$$

*where the operator $\circ$ applies the function $g(\cdot)$ iteratively: $g^{\circ i}(X_t) = g(g(\ldots g(X_t)))$ and $g^{\circ 0}(X_t) = X_t$ by convention.*

*Proof.* Starting with $P_0(\cdot) \equiv 1$ and expanding the recursion (2) we get:

$$P_n(X_t) = \exp(-\sum_{i=0}^{n-1} m(g^{\circ i}(X_t))). \tag{5}$$

Equations (3) and (4) follows from the definition of the $n$-period yield and forward rate, $y_{n,t} \equiv -\log(P_n(X_t))/n$ and $f_{n,t} \equiv (n+1)y_{n+1,t} - ny_{n,t}$, respectively. □

The functions $m(\cdot)$ and $g(\cdot)$ are the primitive blocks in the construction of bond prices. Their interpretation is straightforward. The function $m(X_t)$ gives the one-period interest

rate. Setting $n = 1$ we get $y_{1,t} = f_{0,t} = -\log(P_1(X_t)) = m(X_t)$. In turn, the function $g(\cdot)$ embodies how the price of the bond tomorrow is discounted back to its present value today. For instance, using Equation 2 with $n = 2$, the two-period bond price is given by $P_2(X_t) = P_1(g(X_t)) \times \exp(-m(X_t)) = \exp\left(-m(g(X_t)) - m(X_t)\right)$.

# B Properties

Assumption 2 build on economic principles and summarizes key desirable properties for the choice of $(m(\cdot), g(\cdot))$. These additional properties play a central role in our empirical implementation but we emphasize that our theoretical results only rely on Assumption 1.

**Assumption 2.** *Bond prices satisfy the following properties:*

**P1** — **Positivity** $P_n(X_t) \leq 1 \; \forall X \in \underline{X}$ *or equivalently* $y_{n,t} \geq 0$*;*

**P2** — **Discounting distant cash flows** $\lim_{n\to\infty} P_n(X_t) \to 0$*;*

**P3** — **Invertibility** $\exists u(\cdot) : \mathbb{R} \to \mathbb{R}$ *such that* $u^{-1}(f_{n,t}) = a_n + b_n X_t \; \forall n$*.*

**P1** requires that yields to maturity are positive. Nominal interest rates are necessarily bounded below by the costs of alternative stores of value. For instance, negative yields can be avoided costlessly for small amounts by holding cash. **P2** reflects the fact that sure payoffs that are more distant should have a lower price. Finally, **P3** implies that forward rates can be inverted to a linear transformation of $X_t$. **P3** leads to term structure models where estimation is fast and robust despite any inherent nonlinearity and regardless of the dimension of the risk factors $X_t$. In that regard, **P3** ensures that we can "undo" the nonlinearity in forward rates.

**Proposition 1.** *The following choices guarantee that bond prices constructed as in Assumption 1 satisfy Properties* **P1-P3***:*

1. $m(\cdot)$ *is continuous and monotonic with* $m(X) \geq 0 \; \forall X \in \underline{X}$*,*
2. $g(X)$ *is a contraction with unique fixed-point* $g(X^*) = X^*$*,*
3. $g(X) = GX$*.*

*Proof.* See appendix. □

From Equations 3-4, it is clear that $m(X) \geq 0$ guarantees **P1**: all forward rates and all yields to maturity are positive.[4] The second condition is necessary for **P2**, guaranteeing that long-run forward rates do not diverge. Finally, the last condition takes $g(\cdot)$ as a linear function of $X$ and, together with the other two conditions, guarantees **P3** so that we can invert forward rates and proceed with fast and robust estimation. In particular, the requirement that $g(\cdot)$ is a contraction implies that the matrix $G$ has no unit root.

## C Examples

### *Nelson-Siegel*

The direct approach to constructing bond prices in Assumption 1 is reminiscent and overlaps with a longstanding tradition in fitting observed bond prices. For instance, TTSMs includes several 3-factor models where the loadings on yield,

$$y_{t,n} = X_{1,t}b_1(n) + X_{2,t}b_2(n) + X_{3,t}b_3(n),$$

are identical to those given in Nelson and Siegel (1987) (see also Figure 1):

$$\begin{aligned} b_1(n) &= 1 \\ b_2(n) &= \left(\frac{1-\exp(-n\lambda)}{n\lambda}\right) \\ b_3(n) &= \left(\frac{1-\exp(-n\lambda)}{n\lambda} - \exp(-n\lambda)\right). \end{aligned} \tag{6}$$

The mapping with Nelson and Siegel (1987) is formally stated in Proposition 2.

**Proposition 2.** *Suppose $X_t \in \mathbb{R}^3$, with $m(\cdot)$ and $g(\cdot)$ given by:*

$$m(X_t) = \begin{bmatrix} 1 & \frac{1-e^{-\lambda}}{\lambda} & \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{bmatrix} X_t, \tag{7}$$

$$g(X_t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda} & \lambda e^{-\lambda} \\ 0 & 0 & e^{-\lambda} \end{bmatrix} X_t, \tag{8}$$

*then, bond prices generated from the construction in Assumption 1 have yields-to-maturity with Nelson-Siegel loadings given by Equation 6.*

[4] Alternatively, one could impose restriction on the support of $X$: i.e., $\underline{\mathrm{X}} = \mathbb{R}^K_+$. We choose instead to maintain broad flexibility in the time-series dynamics for $X$ at little costs.

*Proof.* Direct computation of $(1/n)\sum_{i=0}^{n-1} m(g^{\circ i})$ yields the result. □

Note that we differ from Nelson and Siegel (1987) for several reasons. First, our construction is more general, allowing for choices of $m(\cdot)$ and $g(\cdot)$ beyond Equations (7)-(8). Second, Section III assesses explicitly any arbitrage opportunity within our framework and provides a precise *economic* sense in which these models (including Nelson and Siegel (1987)) are very nearly arbitrage-free. Third, the choice of $m(\cdot)$ and $g(\cdot)$ in Equations (7)-(8) does not satisfy **P1** and **P2**.

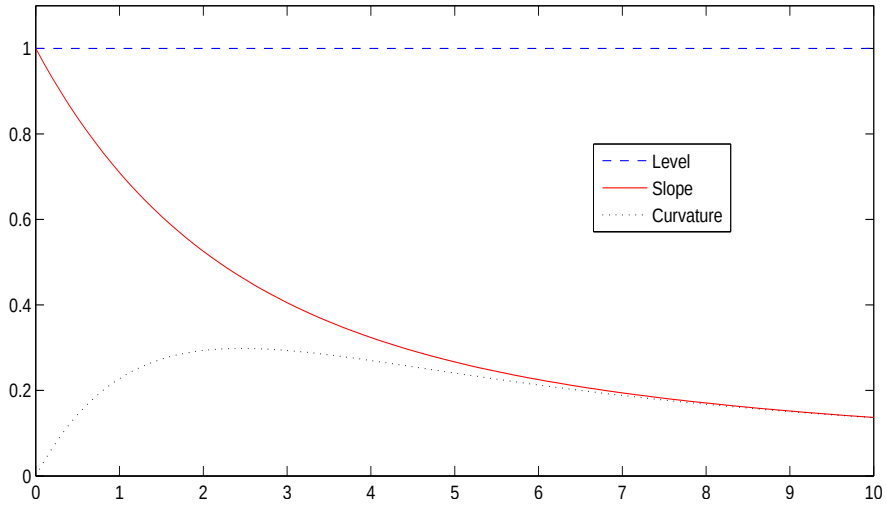



**Figure 1:** Nelson-Siegel Loadings

## *Linear models*

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

$$m(X_t) = \delta_0 + \delta_1' X_t \tag{9}$$

$$g(X_t) = KX_t \tag{10}$$

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

$$y_{n,t} = \delta_0 + (B_n/n)X_t, \tag{11}$$

with $B_n$ given by the recursion:

$$B_n = B_{n-1}K + \delta_1'. \tag{12}$$

For comparison, the $A_0(N)$ Gaussian DTSMs (e.g., Dai and Singleton (2000) and Duffee (2002)) has a linear short-rate equation and risk-neutral dynamics given by:

$$\begin{aligned} r_t =& \delta_0 + \delta_1' X_t \\ X_{t+1} =& K_0 + K_1^{\mathbb{Q}} X_t + \epsilon_{t+1}, \end{aligned} \tag{13}$$

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

$$y_{n,t} = A_n/n + (B_n/n)X_t, \tag{14}$$

with coefficients given by:

$$B_n = B_{n-1}K_1^{\mathbb{Q}} + \delta_1', \tag{15}$$

$$A_n = A_{n-1} + \delta_0 - \frac{1}{2}B_{n-1}\Sigma B_{n-1}'. \tag{16}$$

Clearly, the short rate $r_t$ and the loadings $B_n$'s are identical between these models. The intercept terms $A_n$ for $n > 1$ are different only due to the convexity correction $B_{n-1}\Sigma B_{n-1}'$. This Jensen term is negligible in a typical application (see Figure 2).

### *Black's models*

Black (1995) emphasizes that bondholders have an *option* to hold cash instead of bonds, which implies a lower bound on nominal yields. Black uses the max function to obtain a positive overnight rate, i.e., $r_t = \max(s_t, 0)$ where $s_t$ is the so-called "shadow" rate that would prevail absent the lower bound. In our framework, this yields:

$$s_t = \delta_0 + \delta_1' X_t \tag{17}$$

$$m(X_t) = \max(0, s_t) \tag{18}$$

$$g(X_t) = KX_t, \tag{19}$$

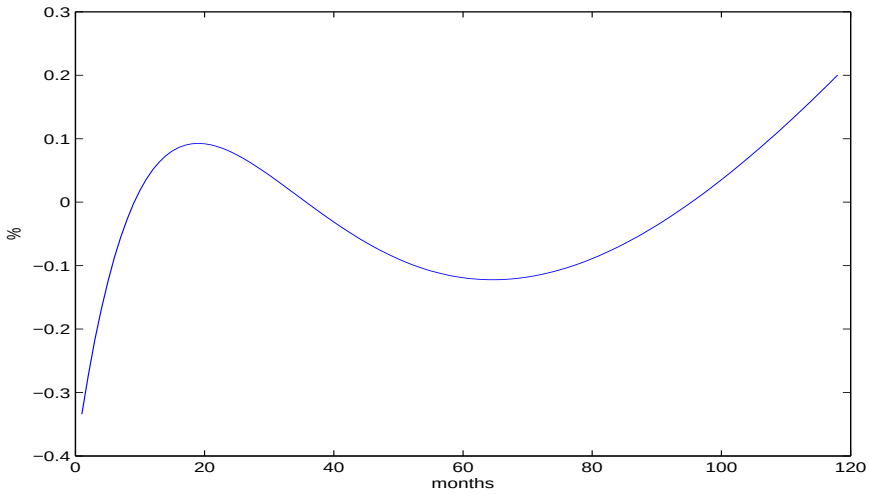



**Figure 2:** The Jensen terms is negligible in Gaussian models. Difference to the Jensen Term $\frac{1}{2}B_{n-1}\Sigma B'_{n-1}$ between the loadings in linear 3-factor models where yield PCA are used as risk factors and parameter. Estimates based on the canonical representation in Joslin, Singleton, and Zhu (2011).

and from Theorem 1, forward rates are simply given by;

$$f_{n,t} = \max(0, K^n X_t), \tag{20}$$

and yields are given by Equation 3. In sharp contrast, DTSM implementations of Black's formulation suffer from severe drawbacks. First, closed-form bond prices are not available and estimation is computationally demanding.[5] Second, Black DTSMs must impose a constant volatility to remain tractable.[6] Finally, this approach does not meet **P3**, since the max function cannot be uniquely inverted.

### *Quadratic models*

The following choice generates linear-quadratic TTSMs:

$$\begin{aligned} m(X_t) =& \delta_0 + \delta_1' X_t + X_t' \delta_2 X_t \\ g(X_t) =& K X_t \end{aligned} \tag{21}$$

[5] Bauer and Rudebusch (2013) use Monte Carlo simulations under the risk-neutral measure. Krippner (2011) modifies no-arbitrage prices to explicitly account for the zero lower bound in a tractable way (see also Christensen and Rudebusch (2013)) but does not check check whether prices are consistent with the AOA.

[6] Kim and Priebsch (2013), in continuous time, and Wu and Xia (2013), in discrete time, introduce tractable second-order approximations based on series expansion of yields.

where $\delta_0$ is a scalar, $\delta_1$ is $N \times 1$ and $\delta_2$ is an $N \times N$ matrix. From Theorem 1, yields are given by:

$$y_{n,t} = \delta_0 + (B_n/n)X_t + X_t'(C_n/n)X_t \tag{22}$$

where the linear and quadratic coefficients $B_n$ and $C_n$ are given by:

$$\begin{aligned} C_n &= K'C_{n-1}K + \delta_2 \\ B_n &= B_{n-1}K + \delta_1'. \end{aligned} \tag{23}$$

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

$$r_t = \delta_0 + \delta_1'X_t + X_t'\delta_2 X_t,$$

and with the risk-neutral dynamics as in (13).[7] The solution for yields in this case is given by:

$$y_{n,t} = A_n/n + (B_n/n)X_t + X_t'(C_n/n)X_t, \tag{24}$$

where the loadings $A_n$, $B_n$ and $C_n$ are given by the following recursions:

$$\begin{aligned} C_n &= K_1^{\mathbb{Q}\prime}C_{n-1}\Omega_{n-1}K_1^{\mathbb{Q}} + \delta_2, \\ B_n &= B_{n-1}\Omega_{n-1}K_1^{\mathbb{Q}} + \delta_1', \\ A_n &= A_{n-1} + \delta_0 - \frac{1}{2}log|\Omega_{n-1}| - \frac{1}{2}B_{n-1}\Omega_{n-1}\Sigma B_{n-1}, \end{aligned} \tag{25}$$

with $\Omega_{n-1} \equiv (I_N - 2\Sigma C_{n-1})^{-1}$. Comparing loadings in (23) and (25) reveals two differences. First, the term $B_{n-1}\Omega_{n-1}\Sigma B_{n-1}$ reflects a convexity adjustment. Second, the matrix $\Omega_{n-1}$ may introduce a wedge between loadings if the quadratic coefficient $\delta_2$ is "large".

[7] See also Longstaff (1989), Beaglehole and Tenney (1991), and Constantinides (1992) (the SAINTS model) for earlier iterations of the affine quadratic models. Realdon (2006) provides a discrete time treatment.

# III Near-Absence of Arbitrage

Directly specifying bond prices stands in contrast to the standard (indirect) construction of term structure models based on computing $\int (M_{t+n}/M_t) d\mathbb{P}$.[8] The standard approach *guarantees* that bond prices rule out arbitrage opportunities if the proposed pricing kernel $M_{t+n}/M_t$ is strictly positive almost surely $(\forall n)$. However, computing the integral in this conditional expectation is not analytically tractable for many interesting choices of $(M_t, \mathbb{P})$.

Our direct approach to bond prices guarantees tractability. On the other hand, we have yet to assess whether it precludes arbitrage opportunities. For instance, Bjork and Christensen (1999) and Filipovic (1999) show that the Nelson and Siegel (1987) model does not ensure the absence of arbitrage. However, it also generally agreed that these loadings are "almost arbitrage-free", as put forth in Diebold and Rudebusch (2012).[9] In the following, we provide economic foundations for the pervasive perception that these models are "almost arbitrage-free".

## *Absence of Dominant Trading Strategies*

Checking for the absence of arbitrage is a difficult problem. It may seem that we have traded one intractable problem for another. However, we can build on the simply structure of bonds. To begin, we check these models for the absence of dominant (AOD) trading strategies – a requirement which is related to, but weaker than, the absence of arbitrage (AOA) condition. Recall that the AOA holds if and only if any portfolio with strictly non-negative payoffs admits a strictly positive price.[10] The AOD also imposes a strictly positive price but only for the set of portfolios with payoffs that are *strictly* positive in *all states*. Theorem 2 establishes that using Assumption 1 to construct bond prices is sufficient to guarantee AOD.

**Theorem 2.** *Assumption 1 guarantees the absence of dominant trading strategies (AOD)*

[8] The integral can be computed given the relevant information set $\mathbb{I}_t$ and dynamics for $M_{t+n}$ under the historical measure $\mathbb{P}$. Equivalently, one can obtain bond prices via $\int \exp(-\sum_{j=0}^{n-1} r_{t+j}) d\mathbb{Q}$ given the dynamics for the risk-free rate $r_t$ under the risk-neutral measure $\mathbb{Q}$.

[9] For an excellent account of this literature, see Diebold and Rudebusch (2012) and references therein. Krippner (2013) tightens this point, showing that the Nelson and Siegel (1987) model can be seen as low-order Taylor approximations of the generic no-arbitrage Gaussian affine term structure models. The difference is small in practice, Coroneo, Nyholm, and Vidova-Koleva (2011) find that the no-arbitrage parameters are not statistically different from those obtained based on Nelson and Siegel (1987). Our line of argument is distinct from the *technical* approaches, such as one based on Taylor approximations proposed by Krippner (2013). It is also much more general, since our results apply to a much richer set of models beyond Nelson and Siegel (1987).

[10] To be precise, any portfolio with positive cash flows for a strictly positive measure and zero cash flows otherwise.

*between bond prices.*

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

$$\sum_n w_n P_n(X_t) = \exp(-m(X_t)) \times \sum_n w_n P_{n-1}(g(X_t)). \tag{26}$$

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

### *Near-Absence of Arbitrage*

Given the result in Theorem 2 that TTSMs ensure AOD, can we ask how "close" we are to the AOA? One way of addressing this question is to ask how negative can the price of portfolios with strictly non-negative payoffs be? In other words, how close are we to guaranteeing that their price is positive (as required by the AOA)? A negative price is a serious violation of the no-arbitrage principle, since one is paid money to own a portfolio that never requires contributions in the future. In contrast, any model that assign a price of zero is close to guaranteeing the AOA. Theorem 3 shows that Assumption 1 is also sufficient to guarantee that portfolios with strictly non-negative payoffs cannot admit negative prices.

**Theorem 3.** *Assumption 1 ensures that portfolios with strictly non-negative payoffs cannot admit negative prices.*

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

$$\sum_n w_n P_n(X_t) = exp(-m(X_t)) \times \sum_n w_n P_{n-1}(g(X_t)). \tag{27}$$

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

Figure 3 illustrates Theorem 3. For portfolios with strictly non-negative payoffs, TTSMs allow for prices on the positive half of the real line, including the origin. The absence of

arbitrage allows for prices on the positive half of the real line, *excluding* the origin. The one difference reduces to one point on the real line (the origin).

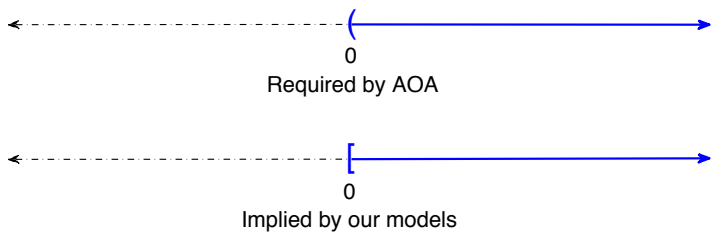



**Figure 3:** Prices of portfolios with strictly non-negative payoffs.

The following example provides another way to see intuitively how the AOD is close to the AOA. Consider a portfolio $w$ that pays one dollar in some states (with a strictly positive measure) and zero otherwise. If markets are complete, we can construct a portfolio $w_\epsilon$ that pays at least $\epsilon$ dollars in all states (i.e., either 1 or $1+\epsilon$ dollar). The AOD implies that the price $p_\epsilon$ of $w_\epsilon$ must be positive ($w_\epsilon$ earns strictly positive cash flows in every state). Varying $\epsilon$ closer to zero, we obtain a sequence of portfolios $w_\epsilon$ approaching portfolio $w$. Each of these portfolios has a positive price. With continuous prices, this example suggests that the price of portfolio $w$ – the limit of the sequence of positive prices $p_\epsilon$ as $\epsilon$ tends to zero – cannot be negative. For convenience, this heuristic argument invoked assumptions regarding market completeness and price continuity. Fortunately, the result in Theorem 3 only relies on Assumption 1.

### *Transaction Costs*

At this point, we conclude that our framework may allow for the possibility of self-financing portfolios (with price zero) paying non-negative cash flows. Does this represent a serious violation of the no-arbitrage principle? Theorem 4 addresses this question formally, showing that, as long as there is some (however small) transaction cost in short-selling bonds, all arbitrage opportunities are ruled out by our models.

**Theorem 4.** *Bond prices constructed in Assumption 1 combined with any non-zero level of transaction costs rule out all arbitrage opportunities.*

*Proof.* See appendix. □

The presence of small transaction costs preventing self-financing strategies is *economically* plausible. Again, consider a portfolio $w$ that pays one dollar in some states (with a strictly

positive measure) and zero otherwise. If this portfolio is self-financing, we must sell short some bonds to cover for costs of the long positions.[11] But this is costly. A substantial literature has documente the costs to establish and maintain short Treasury bond positions, even for the most liquid issues. (See e.g., Duffie (1996), Krishnamurthy (2002), Vayanos and Weill (2008), and Banerjee and Graveline (2013).) In other words, this portfolio may not give rise to an arbitrage opportunity after shorting (and other transactions) costs are taken into account. Consistent with this observation, Liu and Longstaff (2004) show that risk-averse arbitrageurs do not fully exploit arbitrage opportunities in the Treasury market.

We emphasize that the magnitude of transaction costs is inconsequential for the above argument to go though. That is, we do not require transaction costs to be of certain sizes (after all, with sufficiently large transaction costs, no trading strategy can be profitable.) The key of our argument lies in the fact that to finance even very small setup costs, the price of a portfolio must be in the negative region. And it is the sign, not the magnitude, of the price that triggers Theorem 3.

# IV Estimation

## A Data

We use data sampled at a monthly frequency between January 1990 and December 2014, but we only use data up to December 2008—when the overnight rate reached its lower bound—when estimating affine models. Our sample includes forward rates and macro variables. We include forward rates with quarterly maturities between three months and ten years. We use data from GSW for maturities longer than 6 months but we extract the 3-month forward from CRSP data. Figure 4 shows selected 3-month forward rates with maturities of 6 months, 1 year, 5 years and 10 years.

We also use unspanned macro-economic variables that are relevant for the dynamics of the risk factors (Joslin, Priebsch, and Singleton, 2013). Specifically, our sample also includes one-year ahead forecasts of inflation and of GDP growth from Blue Chips surveys of professional forecasters. This provides forward-looking information useful to forecast future interest rates (Chun, 2010).

[11] To be more precise, a long position on any given zero coupon bond commands a positive price, by construction our model (see Equation 5). By the same token, any portfolio combining long-positions only will have a positive price models. It follows that any portfolio to which our model assign a price of zero or less must include short positions.

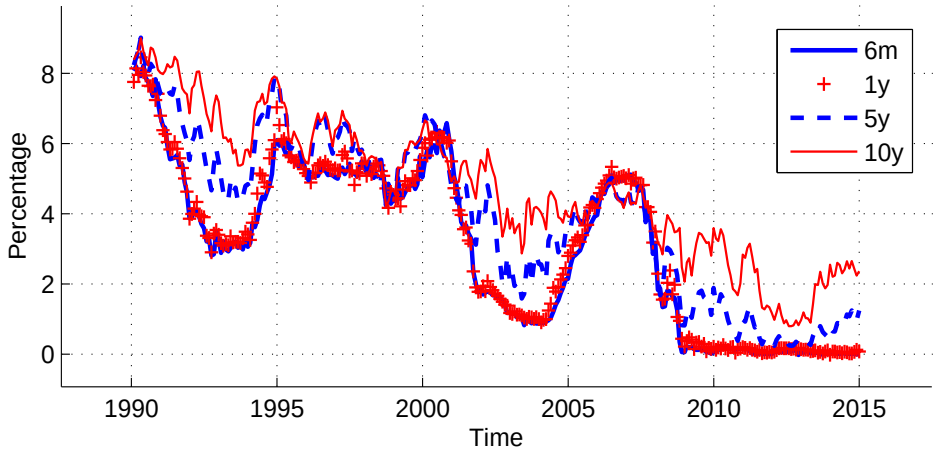



**Figure 4:** Forward rates reached their lowest range at different dates. Monthly data from 1990 to 2014.

## B Identification

In line with Assumption 1 and Assumption 2, we introduce the following parametrization for the $m(\cdot)$ and $g(\cdot)$ function:

$$m(X_t) = u(\theta, \delta_0 + \delta_1' X_t) \tag{28}$$

$$g(X_t) = KX_t, \tag{29}$$

where $X_t \in \mathbb{R}^N$, $\delta_0 + \delta_1' X_t$ is a scalar and $\theta$ is a parameter (possibly a vector) summarizing non-linear features in $m(\cdot)$. Taking $u(\cdot)$ as the identity function then collapses to the linear case (in which case $\theta$ is empty). From Equation 28, the $n$-period forward rate $f_{n,t} = m(g^{\circ n}(X_t))$ is given by:

$$f_{n,t} = u(\theta, \delta_0 + \delta_1' K^n X_t), \tag{30}$$

**P3** requires that $u(\theta, \cdot)$ is invertible with respect to the second argument, implying that we can transform to the familiar linear form,

$$\tilde{f}_{n,t} = u^{-1}(\theta, f_{n,t}) = \delta_0 + \delta_1' K^n X_t. \tag{31}$$

We use the standard normalization of Joslin, Singleton, and Zhu (2011) to identify $\delta_0$, $\delta_1$ and $K$, Specifically, $\delta_1 = \iota$ is a vector of ones, $K$ is an ordered Jordan form and $g(X)$ has no constant. To see why this normalization applies to our case, consider an arbitrary $K$ matrix

with decomposition $K = U\lambda U^{-1}$, where $\lambda$ is the diagonal matrix of real eigenvalues. Thus, Equation (31) is observationally equivalent to:

$$\begin{aligned}\tilde{f}_n =& \delta_0 + \delta_1' U \lambda^n U^{-1} X_t \\ =& \delta_0 + \iota_K' \lambda^n \text{diag}(\delta_1' U) U^{-1} X_t \\ =& \delta_0 + \iota_K' \lambda^n Z_t \end{aligned} \tag{32}$$

with the restrictions that $\delta_1 = \iota_K$, $K_1$ is diagonal and $Z_t \equiv \text{diag}(\delta_1' U) U^{-1} X_t$. The second equality uses the fact that $\lambda$ is diagonal. Similar arguments can be made for cases where $K_1$ has complex eigenvalues (along the line of Joslin, Singleton, and Zhu (2011)).

## C Yield Portfolios as Risk Factors

Stack $J$ forwards in the vector $Y_t = (f_{n_1,t}, ..., f_{n_J,t})'$ and construct the corresponding vector of transformed forwards by $\tilde{Y}_t = (\tilde{f}_{n_1,t}, ..., \tilde{f}_{n_J,t})'$:

$$\tilde{Y}_t = A_X + B_X X_t \tag{33}$$

where the coefficients are obtained by stacking coefficients in Equation (31). Note that the well-known "shadow rate", $s_t$, is a special case with the inversion applied to the one-period forward rate:

$$s_t \equiv \tilde{f}_0 = u^{-1}(\theta, f_{0,t}) = \delta_0 + \delta_1' X_t.$$

Next, consider $N \leq J$ portfolios of $\tilde{Y}_t$ with an $N \times J$ loadings matrix, $W$,

$$\mathcal{P}_t = W \tilde{Y}_t, \tag{34}$$

which are measured without errors, $\mathcal{P}_t = \mathcal{P}_t^o$ as in Joslin, Singleton, and Zhu (2011), and which can be used to uncover the risk factors:

$$X_t = (W B_X)^{-1} (\mathcal{P}_t - W A_X). \tag{35}$$

Substituting in Equation (31), we can express the transformed forward rates, $\tilde{f}_{n,t}$, in terms of the portfolios $\mathcal{P}_t$,

$$\tilde{f}_{n,t} = \delta_0 + \iota_K' K_1^n (W B_X)^{-1} (\mathcal{P}_t - W A_X) = A_{n,\mathcal{P}} + B_{n,\mathcal{P}} \mathcal{P}_t, \tag{36}$$

with coefficients given by:

$$A_{n,\mathcal{P}} = \delta_0 - B_{n,\mathcal{P}} W A_X$$
$$B_{n,\mathcal{P}} = \iota_K' K_1^{\mathbb{Q}^n} (W B_X)^{-1}. \tag{37}$$

Stacking (36) for the relevant maturities $(n_1, \ldots, n_J)$, we obtain a version of the pricing function but with $\mathcal{P}_t$ as risk factors on the right hand side:

$$\tilde{Y}_t = A_{\mathcal{P}} + B_{\mathcal{P}} \mathcal{P}_t. \tag{38}$$

# D Historical Dynamics

Assumption 1 is sufficient to establish the absence of exploitable arbitrage opportunities. In the following, we also rely on Assumption 2 to guide our specification of the physical dynamics for the states $X_t$. In particular, it is essential that the support of $X_t$ under the time series measure coincide with the space $\underline{X}$ over which our bond prices are defined.[12] Furthermore, for valid statistical inferences, we need $X_t$ to be stationary and ergodic under the time series measure. Assumption 3 summarizes these requirements formally.

**Assumption 3.** *The time series dynamics of $X_t$ admits $\underline{X}$ as support and is such that yields for all maturities $y_{n,t} \equiv -log(P_n(X_t))/n$ have a joint distribution that is stationary and ergodic.*

Assumption 3 imposes mild restrictions on the evolution of $X_t$ and accommodates virtually all common stationary dynamics. Importantly, the dynamics of $X_t$ may involve factors beyond $X_t$. For instance, the conditional mean at time $t$, $E_t[X_{t+1}]$, may not be completely spanned by $X_t$. This allows for notions of unspanned risks introduced by Joslin, Priebsch, and Singleton (2013) Duffee (2011) and Feunou and Fontaine (2014). Likewise, the conditional variances, $V_t[X_{t+1}]$, can be constant, as in standard Gaussian DTSMs, can depend on $X_t$ itself, as in the $A_M(N)$ models of Dai and Singleton (2000), can depend on the history $\{X_t, X_{t-1}, \ldots\}$, in the spirit of the ARCH literature pioneered by Engle (1982), or can depend on the history of other risk factors, capturing the notion of unspanned stochastic volatility in Collin-Dufresne and Goldstein (2002), Li and Zhao (2006) and Joslin (2011).

Hence, our framework allows for general time-series specifications for $E_t[X_{t+1}]$ and $V_t[X_{t+1}]$

[12] This requirement is analogous to the requirement that the time series and risk-neutral measures must admit the same support (they are equivalent measure) in no-arbitrage models.

without introducing overly restrictive connection between the mean and variance. This is a significant departure from affine no-arbitrage term structure models that are burdened by a significant tension in simultaneously fitting the conditional means and the conditional variances of yields (Dai and Singleton, 2002; Joslin and Le, 2013). Instead, we choose to let the data speak concerning the inter-relation between mean and variance in the term structure. Our approach also contrasts with most implementations of Blacks' non-affine no-arbitrage models where the variance is assumed constant.

To operationalize this, consider a vector of macro variables $U_t$ that are not spanned by yields. For simplicity, we consider joint VAR dynamics for $\mathcal{P}_t$ and $U_t$:

$$\begin{pmatrix} \mathcal{P}_{t+1} \\ U_{t+1} \end{pmatrix} = K_0^{\mathbb{P}} + K_1^{\mathbb{P}} \begin{pmatrix} \mathcal{P}_t \\ U_t \end{pmatrix} + \sqrt{\Sigma_t} \begin{pmatrix} \varepsilon_{\mathcal{P},t+1} \\ \varepsilon_{U,t+1} \end{pmatrix}, \tag{39}$$

where $\varepsilon_t \equiv (\varepsilon_{\mathcal{P},t+1},\ \varepsilon_{U,t+1})'$ is a standard (conditionally) gaussian innovation with time-varying variance-covariance matrix, $\Sigma_t$. We choose a robust and parsimonious specification for the time-varying covariance matrix, $\Sigma_t$. First, the $N \times 1$ vector of variances $\sigma_t^2 = \text{diag}(\Sigma_t)$ has dynamics given by:

$$\ln \sigma_t^2 = (I - B) \ln \bar{\sigma}^2 + B \ln \sigma_{t-1}^2 + A\varepsilon_t + \gamma \left( |A\varepsilon_t| - E[|A\varepsilon_t|] \right), \tag{40}$$

where $A$ is a $N \times N$ matrix, and $B$ and $\gamma$ are $N \times N$ diagonal matrices. This vector-EGARCH allows innovations to any of the risk factors to affect the variance of every other risk factor, since the matrix $A$ is not restricted to be diagonal. Second, we use a Dynamic Conditional Correlation (DCC) model for time-varying correlations (Engle, 2002). Specifically, the correlation matrix $C_{\mathcal{P},t}$ and $C_{U,t}$ are defined via the following normalization of the matrix $Q_{\mathcal{P},t}$ and $Q_{U,t}$:

$$C_{i,t} = Q_{i,t} ./ \left( diag\left[Q_{i,t}\right]^{-1/2} \otimes diag\left[Q_{i,t}\right]^{-1/2} \right) \tag{41}$$

to guarantee a well-defined correlation matrix.[13] Each of $Q_{\mathcal{P},t}$ and $Q_{U,t}$ follows

$$\begin{aligned} Q_{\mathcal{P},t} &= (1 - a_{\mathcal{P}} - b_{\mathcal{P}})\, \bar{Q}_{\mathcal{P}} + a_{\mathcal{P}} \varepsilon_{\mathcal{P},t} \varepsilon_{\mathcal{P},t}' + b_{\mathcal{P}} Q_{\mathcal{P},t-1} \\ Q_{U,t} &= (1 - a_U - b_U)\, \bar{Q}_U + a_U \varepsilon_{U,t} \varepsilon_{U,t}' + b_U Q_{U,t-1}, \end{aligned} \tag{42}$$

[13] The operator ./ is the element-by-element division of matrix, $\otimes$ is the Kronecker product and $diag\left[Q_t\right]$ is the vector whose elements are the main diagonal of $Q_t$

where $a_i > 0$, $b_i > 0$, $a_i + b_i < 1$ and where $\bar{Q}_i$ is a symmetric positive definite.[14] Finally, the matrix of cross-correlations between elements of $\varepsilon_{\mathcal{P},t}$ and elements of $\varepsilon_{U,t}$ is set to zero. Preliminary estimates of the cross-correlations matrix $C_{\mathcal{P}U,t}$ were consistently small and exhibited little time-series variation.

## E Maximum likelihood

The parameters for the pricing function (36) are $\Theta_p = \{\theta, \delta_0, K\}$, subject to the restriction that $K$ has the Jordan form; the parameters of the conditional mean in (39) are $\Theta_m = \{K_0^{\mathbb{P}}, K_1^{\mathbb{P}}\}$; and the parameters of the conditional variance are $\Theta_v = \{A, B, \gamma, a, b\}$. The full parametrization is given by $\Theta = \{\Theta_p, \Theta_m, \Theta_v\}$. For each choice of maturities $n_1, \ldots, n_J$ and loading matrix $W$, these parameters are identified.[15]

Estimation of $\Theta$ can be implemented by maximizing the log-likelihood of the observed data $D_t = (Y_t^{o\prime}, U_t^{o\prime})'$,

$$\begin{aligned}\mathbb{P}(D_t|I_{t-1}) =& \mathbb{P}(h(D_t)|I_{t-1}) \left|\frac{\partial h(D_t)}{\partial D_t}\right| \\ =& \mathbb{P}\left(W_e Y_t^o | \mathcal{P}_t, U_t^o, I_{t-1}\right) \times \mathbb{P}\left(\mathcal{P}_t, U_t^o | I_{t-1}\right) \left|\frac{\partial h(D_t)}{\partial D_t}\right| \end{aligned} \tag{43}$$

where $I_{t-1} = \{D_1, D_2, ...D_{t-1}\}$ denotes the information generated by the history of the data, $h(D_t)$ is a transformation of the data $D_t$, and $W_e$ is a $J - N \times J$ matrix to be defined. The first term in the likelihood corresponds to the cross-section of yields measured with error

$$W_e(Y_t^o - \hat{Y}_t) \sim N(0, \sigma_e^2 I_{J-N}), \tag{44}$$

where $\hat{Y}_t$ is a function of $\mathcal{P}_t$ and measurement errors are $i.i.d.$. We assume that $\mathcal{P}_t$ and $U_t$ are observed without measurement errors ($\mathcal{P}_t^o = \mathcal{P}_t$ and $U_t^o = U_t$). The second term $\mathbb{P}\left(\mathcal{P}_t, U_t | I_{t-1}\right)$ is also a Gaussian density which can be easily derived from Equations (39)-(42). The final

[14] In practice, we target $\bar{\sigma}^2$ and $\bar{Q}$ using the sample covariance $\bar{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$.

[15] We can then applies recent advances in the estimation of affine models. See, for example, Joslin, Singleton, and Zhu (2011), Joslin, Le, and Singleton (2013), Hamilton and Wu (2011), Adrian, Crump, and Moench (2013), and de los Rios (2015).

term in the likelihood is the Jacobian of the transformation

$$h(D_t) = \begin{pmatrix} W_e f_t^o \\ W u^{-1}(\theta, f_t^o) \\ U_t \end{pmatrix}, \tag{45}$$

where we note that the second entry $Wg(f_t^o, \theta)$ gives the observed portfolios $\mathcal{P}_t$. The Jacobian is given by evaluating the first-order derivative of $h(D_t)$ with respect to $D_t$:

$$\frac{\partial h(D_t)}{\partial D_t} = \begin{pmatrix} \begin{pmatrix} W_e \\ W \frac{\partial u^{-1}}{\partial f}(f_t^o, \theta) \end{pmatrix} & 0_{J \times M} \\ 0_{M \times J} & I_M \end{pmatrix}, \tag{46}$$

leading to:

$$\left| \frac{\partial h(D_t)}{\partial D_t} \right| = \left| \begin{pmatrix} W_e \\ W \frac{\partial u^{-1}}{\partial f}(f_t^o, \theta) \end{pmatrix} \right|. \tag{47}$$

In practice, the choice of the matrix $W$ and $W_e$ must satisfy the restriction that Equation 47 is different from zero, otherwise the likelihood in Equation 43 would be singular. In other words, the matrix

$$\begin{pmatrix} W_e \\ W \frac{\partial u^{-1}}{\partial f}(f_t^o, \theta) \end{pmatrix}$$

must be invertible almost everywhere. In practice, we choose $W$ as the loadings on the first $N$ yield principal components, and we choose $W_e Y_t$ selecting $J - N$ elements of $Y_t$.

## F Term Structure Models with Positive Yields

Finally, we choose a specification of $m(\cdot) = u(\theta, \cdot))$, nesting both the linear and the max cases, that is consistent with **P1-P3**. Specifically,

$$\begin{aligned} m(X_t) &= lb(\theta) + \theta_1 \log(1 + \theta_2 \exp((\delta_0 + \delta_1' X_t)/\theta_1)) \\ lb(\theta) &= -\theta_1 \log(1 + \theta_2). \end{aligned} \tag{48}$$

where the parameter $\theta = (\theta_1, \theta_2)'$ controls the lower bound and the curvature of the mapping between the shadow rate $s_t \equiv \delta_0 + \delta_1' X_t$ and the observed rate $u(\theta, s)$. We calibrate $\theta_2$ so that $m(\cdot) \geq lb$, where $lb$ is the lowest sample value for the one-period yield. Equation 48 is a

generalization of the logistic function.[16]

Figure 5 shows the one-period bond yield $u(\theta, s)$ as a function of $s$ for different values of the lower bound and curvature parameters. We have that $u(\theta, s_t)$ always crosses the origin when $s_t = 0$ and that its curvature quickly disappears for large values of $s_t$. One can easily show that $u(\theta, s)$ converges to the linear function as $\theta_1 \to \infty$ (keeping $\theta_2$ fixed). We also have that $u(\theta, s)$ converges to the max function as $\theta_1 \to 0$ (again, keeping $\theta_2$ fixed). Therefore, our specification is closely related both to standard affine models and to the formulation in Black (1995) based on the max function for the short rate.

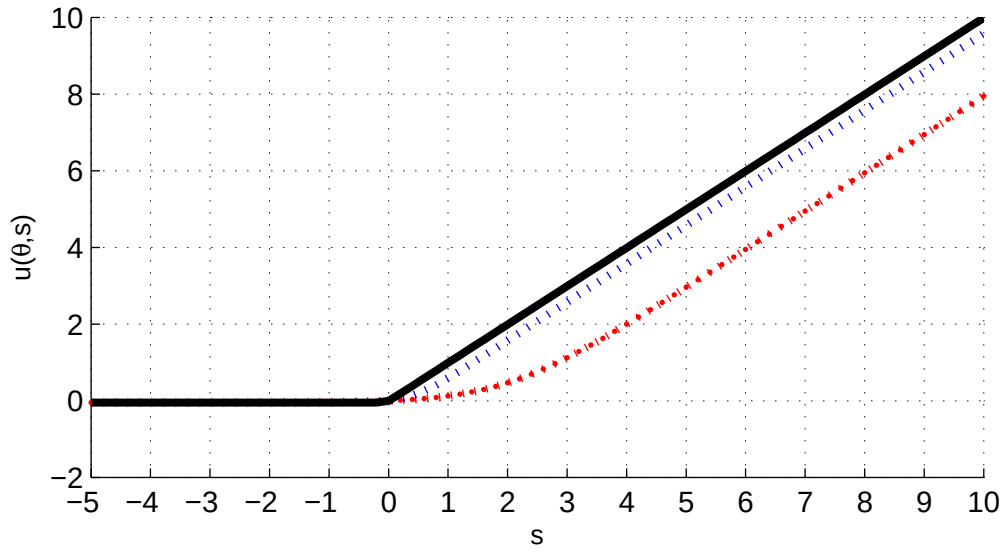



**Figure 5:** Different shapes of the short-rate function $u(\theta, s)$ with corresponding lower bound -0.5%, -0.25% and -0.05%, respectively. The values for $\theta$ are $(0.80, 0.87)$, $(0.54, 0.59)$ and $(0.032, 0.032)$

.

Black (1995) uses the max function to obtain a positive overnight rate, i.e., $r_t = \max(s_t, 0)$. This choice is economically appealing but it is not consistent with **P3** (invertibility) because of the kink in the max function. Our approach captures the essence of Black's formulation. The difference between the one-period yield in Equation 48 and the max function is analogous to the difference between the option value and its payoff. The *overnight* rate exhibits a kink $\max(s_t, 0)$ at the strike $s_t = 0$ but short-term bond yields (contracts linked to the overnight rate) may be positive even if $s_t = 0$. Hence, yields combine the intrinsic value and time value of an option. This observation is a common feature of term structure models with a short-rate equation governed by a max function (e.g., see Kim and Singleton, 2012, Figure

[16] The logistic transformation $P_1(s_t) = 1/(1 + e^{s_t})$ leads to one natural choice $m(s_t) = \log(1 + \exp(s_t))$ where all forward rates remain positive.

5a).

# V Results–Affine Models Before 2009

This section shows that affine Gaussian DTSM and TTSM with constant variance are indistinguishable, but that an extended affine TTSM with flexible second moments captures time-varying volatility and correlations of yields. Cieslak and Povala (2011) show that even a 4-factor $A_1(4)$ does not capture these stylized facts.

## A Model Nomenclature

We consider term structure models with $K = 3$ risk factors, where the matrix of portfolio weights in Equation 35 is obtained from Principal Component Analysis applied to our sample of yields. Figure 6 shows the time series of risk factors.[17] We estimate three classes of affine models. First, we estimate the standard 3-factor Gaussian DTSM using the canonical form in Joslin, Singleton, and Zhu (2011), which we label $A$ (this corresponds to the $A_0(3)$ model in Dai and Singleton (2000)). This model imposes two important restrictions: (i) the computation of bond prices under the risk-neutral measure connects the mean and variance of the risk factors via the pricing equation and (ii) the variances and correlations of the risk factors are constant $\Sigma_t = \Sigma$. Second, we estimate an affine version our 3-factor TTSM with constant variance, which we label $AT$ (*Affine TTSM*) and an extended version with unrestricted EGARCH and DCC dynamics (Equations 40-42), which we label $ATV$ (*Affine TTSM* with time-varying Variance). In all cases, the short-rate equation is affine (i.e., $\theta \to \infty$). We also exclude macro variables in this section for comparability with existing results. We estimate non-affine versions of these models, with and without macro variables, in the following section.

## B Fitting Yields and Bond Risk Premium

Table 2 provides summary statistics on the pricing errors from models $A$, $AT$ and $ATV$, respectively. These models are indistinguishable based on pricing errors. Since the pricing factors are the same across these models, the small differences between pricing errors imply that factor loadings are similar across models. Figure 17 of the appendix compares the factor

[17] Inspection of the weights (not reported) reveal that these portfolios have the standard interpretation in terms of level, slope and curvature factors.

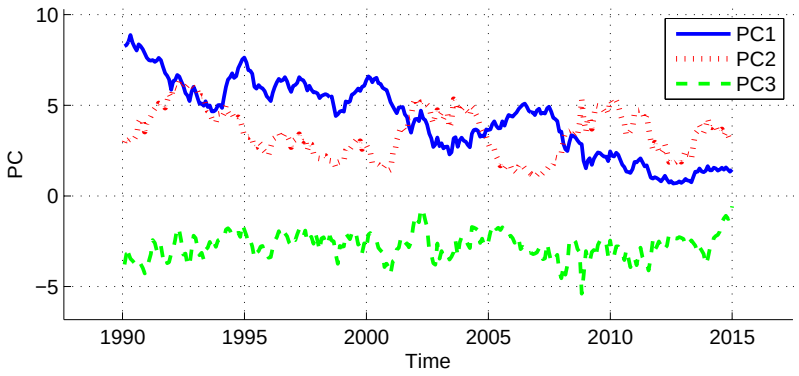



**Figure 6:** Pricing Portfolios. Monthly data from January 1990 to December 2014.

loadings, $B_{n,\mathcal{P}}$. The differences are so small that we cannot distinguish them visually.

| | $A$ | | $AT$ | | $ATV$ | |
|---|---|---|---|---|---|---|
| | RMSE | ME | RMSE | ME | RMSE | ME |
| 1yr | 6.1 | -1.5 | 6.1 | -1.5 | 6.1 | -1.5 |
| 3yr | 2.5 | -0.7 | 2.5 | -0.7 | 2.5 | -0.7 |
| 5yr | 2.3 | -0.4 | 2.3 | -0.4 | 2.3 | -0.4 |
| 10yr | 3.0 | -1.0 | 3.0 | -1.0 | 3.0 | -1.1 |

Table 2: Yield pricing errors from models $A$, $AT$ and $ATV$ are indistinguishable. Root Mean Squared Error (RMSE) and Mean Error (ME) in annualized basis points.

Following Dai and Singleton (2002), we use Cambpell-Shiller (CS) regressions (Campbell and Shiller 1991) to next check wether each model captures variations in bond risk premium. Figure 7 shows OLS estimates of the coefficients from CS predictability regressions in our sample alongside estimates derived from three term structure models. We repeat the exercise for monthly and quarterly returns. In both cases, the model-implied coefficients are remarkably close to each other.[18]

Summing up, the linear model, with constant variance, constructed based on an explicit no-arbitrage argument is indistinguishable from a linear model with constant variance constructed within our new framework. This was expected. Moving between models $A$ and $AT$, the *only* difference is the introduction of the Jensen's term in the equation for $A_{n,X}$. Figure 2

[18]The $A_0(3)$ model does not match the OLS coefficients as closely as reported by Dai and Singleton (2002) in a sample from 1970 to 1995. Noticeably, the expectation hypothesis is not rejected for the shortest maturity in our sample, which could be due to the FOMC's increased transparency since the early 1990's.

above shows that this term is very small. Then, moving between models $AT$ and $ATV$, the pricing equations for yields are *identical* except for differences in parameter estimates. But the estimates for $\delta_0$ and $K$ are nearly identical across models (the factor structure in the cross-section of yields is measured very precisely).

## C Yields' Volatilities and Correlations

Existing affine DTSMs face a well-known tension in fitting the cross-section and the (time-series) variance of yields. This section assesses whether the $ATV$ model can *also* capture stylized fact of yields' variances and correlations. Figure 8 compares model-implied conditional volatilities with EGARCH volatility estimates for the 1-year and 5-year yields in Panels (a)-(b), respectively.[19] Our framework produces a close fit to the conditional volatility of yields. The fit from the $ATV$ model is remarkably close throughout the sample and across yield maturities.

The results show that the volatilities of yield peak at times when the Federal Reserve is loosening its target rate in the midst of a recession. We also find significant variation across the term structure of yield volatility. To see this, Figure 8 compares the model-implied volatility for the 1-month, 1-year, and 10-year yields (Panels c and d). The volatility of the 1-year yield is generally higher than the volatility at both shorter and longer maturities (i.e., the red line generally sits on top of the others) but this hump in the term structure of volatility varies substantially over the sample.

Figure 9 reports the difference between the volatility of 2-year and 1-month yields, as implied by the $ATV$ 3-factor model. This provides a direct measure of the term structure hump. The hump declines substantially in 1991, in 2001 and in 2007, which correspond to the peaks in the level of volatility in Figure 8.

Necessarily, variations in the humps are entirely driven by the time-varying volatility and correlation of the risk factors, since the factor loadings are constant. What is less evident, and which is still undocumented (to the best of our knowledge), is that the changes in the volatility term structure implies that the nature and explanatory power of the principal components of yields change over time. Figure 10 shows the explanatory power of the first two (conditional) components at each date. We stress that these components do not correspond to the unconditional (full-sample) principal components used to construct the risk factors

[19] The EGARCH is used as a benchmark by Kim and Singleton (2012) and Joslin (2014). For comparability between models, the yield innovations in the unrestricted EGARCH(1,1) are computed relative to the projection of current yields on the lagged principal components, as in the VAR(1).

$\mathcal{P}_t$.[20] The importance of the level factor varies between 85 and 95 percent over the sample. By construction, the explanatory power of the other components must vary in the opposite direction. We find that the importance of the slope factor rises by as much as 10 percent in absolute terms (to 15 percent) in these episodes where the short-rate is relatively more volatile in Figure 9 (the hump-shape in volatility is less pronounced).

Our model captures the following stylized facts about the yield volatility term structure. First, the early stages of a recession are characterized by lower yields (higher bond prices) but higher volatility.[21] Second, we confirm that the volatility of yields exhibit a downward-sloping term structure but with a hump-shape around the maturity of one or two years (Piazzesi, 2005). Third, the volatility of the short rate increases relative to other rates during episodes where the Federal Reserve loosens its policy rate (consistent with Cieslak and Povala (2015)). Finally, the explanatory power of the first conditional principal component falls in these episodes, but the explanatory power of the second component rises. Note that the conditional volatility implied by our model is not spanned by contemporaneous yields, which is consistent with Collin-Dufresne, Goldstein, and Jones (2009), but derived from the history of shocks to the yield curve using the EGARCH-DCC filter.

# VI Results– Non-affine Models

This section uses the full-sample period, including six years of data when the target rate is tied to a lower bound. Consistent with Joslin, Priebsch, and Singleton (2013), we also add macro variables to the dynamics of the risk factors. We estimate three term structure models. First, we re-estimate the $ATV$ model. Second, we estimate one non-linear model where the short-rate is given by Equation 48, but assume constant variances for the risk factors. This model is closely comparable to existing implementations of Black's term structure model. Finally, we estimate a non-linear model allowing for time-varying covariances as in Equations (39)-(42). We label these two non-linear versions $PT$ and $PTV$, respectively.

[20] At each date, we simulate several paths of the risk factors and compute the corresponding yields. We then apply PCA to the cross-section of yields across simulated paths. This produces principal component with different loading at each date. This procedure measures the strength of the factor structure in yields.

[21] This contrasts with results based on data covering the late 1970s and early 1980s when the level and volatility of yields were positively correlated.

# A Model Fit and Bond Risk Premium

Table 3 compares the Root Mean Squared Errors (RMSEs) and Mean Errors (MEs), all in annualized basis points, from the *ATV*, *PT* and *PTV* models. Overall, pricing errors are very similar across models. If anything, non-linear models provide small reduction in bias.

| | | *ATV* | | *PT* | | *PTV* | |
|---|---|---|---|---|---|---|---|
| | | RMSE | ME | RMSE | ME | RMSE | ME |
| Yields | 1yr | 5.5 | -1.2 | 5.6 | -1.2 | 5.6 | -1.1 |
| | 3yr | 2.5 | -1.1 | 2.5 | -0.9 | 2.5 | -0.9 |
| | 5yr | 2.9 | -0.6 | 2.9 | -0.5 | 2.9 | -0.5 |
| | 10yr | 3.8 | -1.7 | 3.6 | -1.5 | 3.6 | -1.5 |

Table 3: Full-sample summary statistics of pricing are similar across models. RMSE = root mean squared error; ME = mean error; (all in basis points).

We also compare differences between term premia across models.[22] We define the term premium as the difference between the $\mathbb{Q}$-expectation and the $\mathbb{P}$-expectation. We can compute both terms form the models:

$$tp_t^{(n)} = \frac{12}{n}\left((E_t^{\mathbb{Q}}\left[\sum_{i=0}^{n-1} r_{t+i}\right] - E_t^{\mathbb{P}}\left[\sum_{i=0}^{n-1} r_{t+i}\right]\right) = \hat{y}_t^{(n)} - \frac{12}{n}E_t^{\mathbb{P}}\left[\sum_{i=0}^{n-1} r_{t+i}\right]. \tag{49}$$

Figures 11 shows that the 1-year and 2-year term premium estimates are very different in the affine model (Panels a-b, respectively). Term premium estimates appear biased upward by as much as 30 bps between 2009 and 2011 and, after switching sign, biased downward by as much as 40 bps between 2011 and 2013. Term premium from the non-affine models are more stable and do not reach as deeply into negative territory.

The differences between term premium estimates are driven by differences between long-horizon forward forecasts (the second term in Equation 49). Figure 11 also compares 1-year and 2-year ahead forecasts of the short rate from different models (Panels c-d, respectively). One-year ahead forecasts of the short rate turn negative in the linear model, reaching as low as -0.5%. In addition, short-rate forecasts from the affine model can diverge substantially and extend far in the future. Figure 12 shows the term structure of short-rate forecasts on two given dates—December 2010 and August 2012—across horizons from one month up to

[22] Section V emphasized comparison with the linear risk CS premium model. Yields become non-linear near the lower bound and this benchmark becomes inappropriate. Section 2 reports significant changes in estimated CS coefficients due to the lower bounds.

ten years ahead. Figure 13 reports the Sharpe ratio for an investment horizon of one year and for bonds with two and 10 years to maturity, computed from the different models.

As expected, the bias in term structure estimates from the linear model is due to severe downward bias in short-rate forecasts. The poor forecasts suggest that the model estimates can fit the cross-section of yields near the lower bound, avoiding breaches of the lower bound, but only at the added costs of poor time-series properties.[23] Their yield forecasts are based on a distribution that have significant mass below zero and the precision of their forecasts is poor (Kim and Priebsch, 2013; Christensen and Rudebusch, 2013; Bauer and Rudebusch, 2013).

Figure 14 displays the expected liftoff for the $PT$ and $PTV$ models, starting in 2009. Of course, the expected liftoff time remains zero until the end of 2008 when the shadow rate first passes below zero. The results show that the expected liftoff time is very similar across models, slowly increasing until the summer 2011, and reaching 1.5 years at the time. This is consistent with survey forecasts by primary dealers. The median liftoff forecast was for the fourth quarter of 2012. At the end of our sample, liftoff estimates were pointing towards the middle of 2015, again consistent with survey forecasts made at the same time.

## B Volatility

The presence of a lower bound changes the dynamics of yields volatility fundamentally. The lower bound compresses the volatility of yields with short maturities but, on the other hand, the volatility of yields with long maturities remains elevated and continue to reveal the changing variance of the risk factors. For instance, Swanson and Williams (2014) show that the response of long-term yields to macroeconomic news remain high even when the response of short-term yields is pinned down to zero.

Figure 15 shows the volatilities of the 1-year and 10-year yields throughout the whole sample (Panels a-b) and in the period where the short-rate is near its lower bond (Panels c-d). The $ATV$ and $PTV$ provide a very similar picture of volatilities *before* the 2008. This should be expected, since the short-rate equation is nearly linear in the $PTV$ model when yields are high (see Figure 5) and since these models share the same variance dynamics for the risk factors.

Of course, these two models draw a very different picture after 2008, where the linear $AT$

[23] Note that forecasts from 2-factor affine models, which has less parameters, do not diverge as much, providing additional evidence that the 3-factor affine model uses additional parameters to offer a better fit in the cross-section of yields but at a costs in the time-series (unreported).

model cannot capture the volatility compression for short-maturity yields. Panels (a) and (c) illustrates this contrast in the case of the 1-year yield. The volatility from the $AND3-DCC$ model is *twice* the volatility in the $PTV$ model. Repeating this exercise for shorter maturities reveal an even starker contrast.

The $PT$ model implies a constant variance for the risk factor. Nonetheless, Panel (c) shows that accounting for the lower bound is enough to capture the volatility of short-term yields since 2008. The volatility compression implies that the term structure of volatility is strictly upward sloping (the hump disappears). But Panel (d) shows that the assumption of constant risk factor volatility (embedded in the $PT$ model) fails to capture the volatility dynamics at longer maturities. The 10-year yield volatility implied by the model is almost constant.

Summing up, our new TTSM resolves two tensions that existing models face when fitting the volatility of yields. To see the first tension, compare the periods before and after 2008. Tractable implementations of the Black term structure models can match the volatility dynamics of (short-term) yields after but not before 2008.[24] In contrast, affine models with time-varying volatility can match the volatility terms structure before but not after 2008. To see the second tension, compare the volatility of short-term and long-term yields after 2008. A linear model with changing volatility captures the volatility dynamics of long-term yields but does not match the volatility compression of short term-term yields. Black's term structure models achieve exactly the opposite. The $PTV$ can match the yield volatility at the opposite ends of the yield curve, both before and after 2008 (see Panels c- d).

## C Correlations

The volatility compression at short maturities also changes the correlations between yields. As in Section VC, we use simulations to compute the explanatory power of the (conditional) principal components of yields at each date. Figure 16 reports the results for the first two components (Panels a and b, respectively). The explanatory power of the first component is always high, typically close to 90 percent, but with three significant declines to less than 85 percent around 1991, 2000 and 2007. Our full-sample estimates produce pre-2008 results that are consistent with those presented in Section V (see Figure 10). By construction, this reduced explanatory content is captured by other components. Panel (b) shows that

[24] Priebsch (2013), in continuous time, and Wu and Xia (2013), in discrete time, introduce tractable second-order approximations based on series expansion of yields. These approximations rely on the assumption of constant variance for the risk factors.

the second component explanatory explains more than 15 percent in these episodes (the explanatory power of the third PC is below 1 percent—unreported). Obviously, linear or Black model with constant volatility fails to capture the changing correlation structure of yields in that period.

The explanatory power of the first component quickly rises to 95 percent after 2008 in both *PT* and *PTV* models, but the linear *ATV* model fails to capture this sharp increase. The explanatory power of the second component is close to 5 percent and that of the third component is close to 1 percent. Intuitively, the volatility compression in the front of the term structure implies that the level and slope factors become highly correlated. Panel (c) confirms this intuition, showing the correlation between the first and second components of yields in our simulation. The conditional correlation between the first and second PCs varies over every business cycle, going from positive to negative during most loosening cycles and increasing toward an average near 0.2 at the end of tightening cycle. The recent period is simply an extreme example of this changing correlation.[25]

# VII Conclusion

We introduce a family of Tractable Term Structure Models where bond prices are analytical by construction and very nearly arbitrage-free. Our results show how this new class of model captures the dynamics of yields and yield volatility before and after 2008, when yields reach the lower bound. Variations in the volatility term structure remain a challenge for existing models. The family of TTSMs is large and permits flexible specifications of the dynamic interactions between yield and macro variables. This should lead future research to revisit several results involving the trade-off between the risk premium and yield volatility faced by investors, the influence of conventional and unconventional policy actions on this trade-off (including QE and forward guidance), and the correlations among international term structures (when far from or near to their respective lower bounds).

[25]There is no necessary connection between the average conditional correlation and the unconditional correlation. The law of iterated covariance $\text{cov}(X,Y) = E[\text{cov}(X,Y \mid Z)] + \text{cov}(E[X \mid Z], E[Y \mid Z])$ also involves terms from the conditional mean equation.

# References


Adrian, T., Crump, R. K., Moench, E., 2013. Pricing the term structure with linear regressions. Journal of Financial Economics, forthcoming.

Ahn, D., Dittmar, R., Gallant, A., 2002. Quadratic term structure models: Theory and evidence. Review of Financial Studies 15, 243–288.

Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61(4), 821–856.

Banerjee, S., Graveline, J. J., 2013. The cost of short-selling liquid securities. Journal of Finance.

Bauer, M. D., Rudebusch, G. D., 2013. Monetary policy expectations at the zero lower bound. Working paper Federal Reserve Bank of San Francisco.

Beaglehole, D. R., Tenney, M. S., 1991. General solutions of some interest rate-contingent claim pricing equations. Journal of Fixed Income 9, 69–83.

Bjork, T., Christensen, B. J., 1999. Interest rate dynamics and consistent forward rates curves. Mathematical Finance 9(4), 323–348.

Black, F., 1995. Interest rates as options. Journal of Finance 50, 1371–1376.

Campbell, J., Shiller, R., 1991. Yield spreads and interest rate movements: A bird's eye view. Review of Economic Studies 58, 495–514.

Christensen, J. H. E., Rudebusch, G. D., 2013. Estimating shadow-rate term structure models with near-zero yields. Working paper Federal Reserve Bank of San Francisco.

Chun, A., 2010. Expectations, bond yields and monetary policy. Unpublished working paper. forthcoming, *Review of Financial Studies*.

Cieslak, A., Povala, P., 2011. Understanding bond risk premia. Unpublished working paper. University of Lugano.

——— , 2015. Information in the term stucture of yield curve volatility. Journal of Finance forthcoming.

Collin-Dufresne, P., Goldstein, R., Jones, C., 2009. Can interest rate volatility be extracted from the cross section of bond yields ?. Journal of Financial Economics 94, 47–66.

Collin-Dufresne, P., Goldstein, R. S., 2002. Do bonds span the fixed income markets ? Theory and evidence for 'unspanned' stochastic volatility. Journal of Finance 57, 1685–1730.

Constantinides, G., 1992. A theory of the nominal term structure of interest rates. Review of Financial Studies 5, 531–552.

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

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

——— , 2002. Expectations puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics 63, 415–441.

de los Rios, A. D., 2015. A new linear estimator for gaussian dynamic term structure models. Unpublished working paper. 2.

Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. Mathematische Annalen 300(1), 463–520.

Delbaen, F., Schachermayer, W., 1998. The fundamental theorem of asset pricing for unbounded stochastic processes. Mathematische Annalen 312(2), 215–250.

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

Diebold, F. X., Rudebusch, G. D., 2012. Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. Princeton University Press, .

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

——— , 2011. Information in (and not in) the term structure. Review of Financial Studies 24, 2895–2934.

Duffie, D., 1996. Special repo rates. Journal of Finance 51, 493–526.

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

——— , 2002. Dynamic conditional correlation - a simple class of multivariate GARCH models. Journal of Business and Economic Statistics 17, 339–350.

Feunou, F., Fontaine, J., 2014. Non-markov gaussian term structure models: The case of inflation. Unpublished working paper. 5.

Filipovic, D., 1999. A note on the nelson-siegel family. Mathematical Finance 9, 349–359.

Hamilton, J., Wu, J., 2011. Identification and estimation of affine term structure models. Working paper University of California, San Diego.

Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory 20, 381–408.

Joslin, S., 2011. Can unspanned stochastic volatility models explain the cross section of bond volatilities?. Unpublished working paper. USC.

——— , 2014. Can unspanned stochastic volatility models explain the cross section of bond volatilities?. Management Science.

Joslin, S., Le, A., 2013. Interest rate volatility and no-arbitrage affine term structure models. working paper.

Joslin, S., Le, A., Singleton, K., 2013. Why gaussian macro-finance term structure models are (nearly) unconstrained factor-vars. Journal of Financial Economics 109(3), 604–622.

Joslin, S., Priebsch, M., Singleton, K., 2013. Risk premiums in dynamic term structure models with unspanned macro risks. Journal of Finance, forthcoming.

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

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

Kim, D. H., Singleton, K. J., 2012. Term structure models and the zero bound: An empirical investigation of japanese yields. Journal of Econometrics 170, 32–49.

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

——— , 2013. A theoretical foundation for the Nelson-Siegel class of yield curve models. Journal of Applied Econometrics.

Krishnamurthy, A., 2002. The bond/old-bond spread. Journal of Financial Economics 66, 463–506.

Leippold, M., Wu, L., 2003. Design and estimation of quadratic term structure models. .

Levy, H., 1992. Stochastic dominance and expected utility: Survey and analysis. Management Science 38(4), 555–593.

Li, H., Zhao, F., 2006. Unspanned stochastic volatility: Evidence from hedging interest rate derivatives. Journal of Finance 61, 341–378.

Liu, J., Longstaff, F. A., 2004. Losing money on arbitrage: Optimal dynamic portfolio choice in markets with arbitrage opportunities. 17(3), 611–641.

Longstaff, F. A., 1989. A nonlinear general equilibrium model of the term structure of interest rates. Journal of Financial Economics 2, 195–224.

Nelson, C., Siegel, A., 1987. Parsimonious modelling of yield curves. Journal of Business 60, 473–489.

Piazzesi, M., 2005. Bond yields and the federal reserve. Journal of Political Economy 113, 311–344.

Realdon, M., 2006. Quadratic term structure models in discrete time. Finance Research Letters 3, 277–289.

Rothschild, M., Stiglitz, J., 1970. Increasing risk: I. A definition. Journal of Economic Theory 2(3), 225–243.

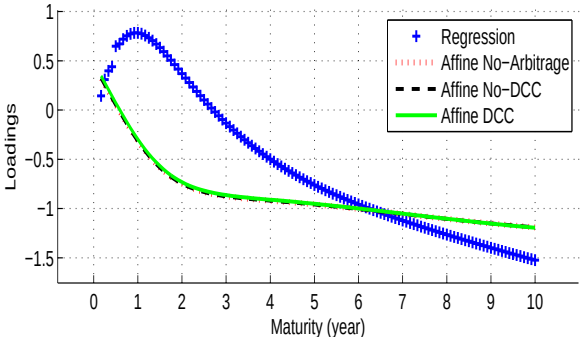
Swanson, E. T., Williams, J. C., 2014. Measuring the effect of the zero lower bound on medium- and longer-term interest rates. The American Economic Review 104(10), 3154–3185.

Vayanos, D., Weill, P.-O., 2008. A search-based theory of the on-the-run phenomenon. Journal of Finance 63, 1361–1398.

Wu, C., Xia, F. D., 2013. Measuring the macroeconomic impact of monetary policy at the zero lower bound. Working paper University of Chicago Booth School of Business.

# Figures

**(a)** 1-month




**(b)** 3-month

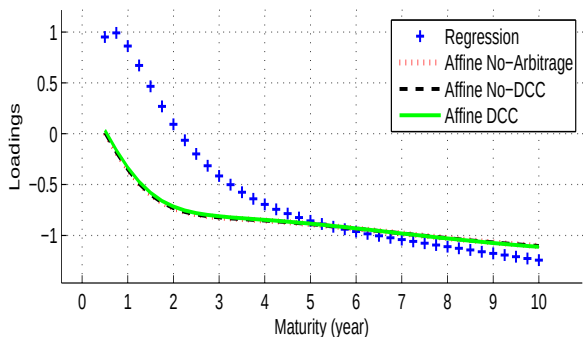



**Figure 7:** Model-implied Campbell-Shiller coefficients are close to each other. Results from monthly and quarterly CS predictability regressions estimated with OLS and coefficient implied from different 3-factor models: $A$, $AT$ and $ATV$. Monthly data, December 1990 to June 2008.

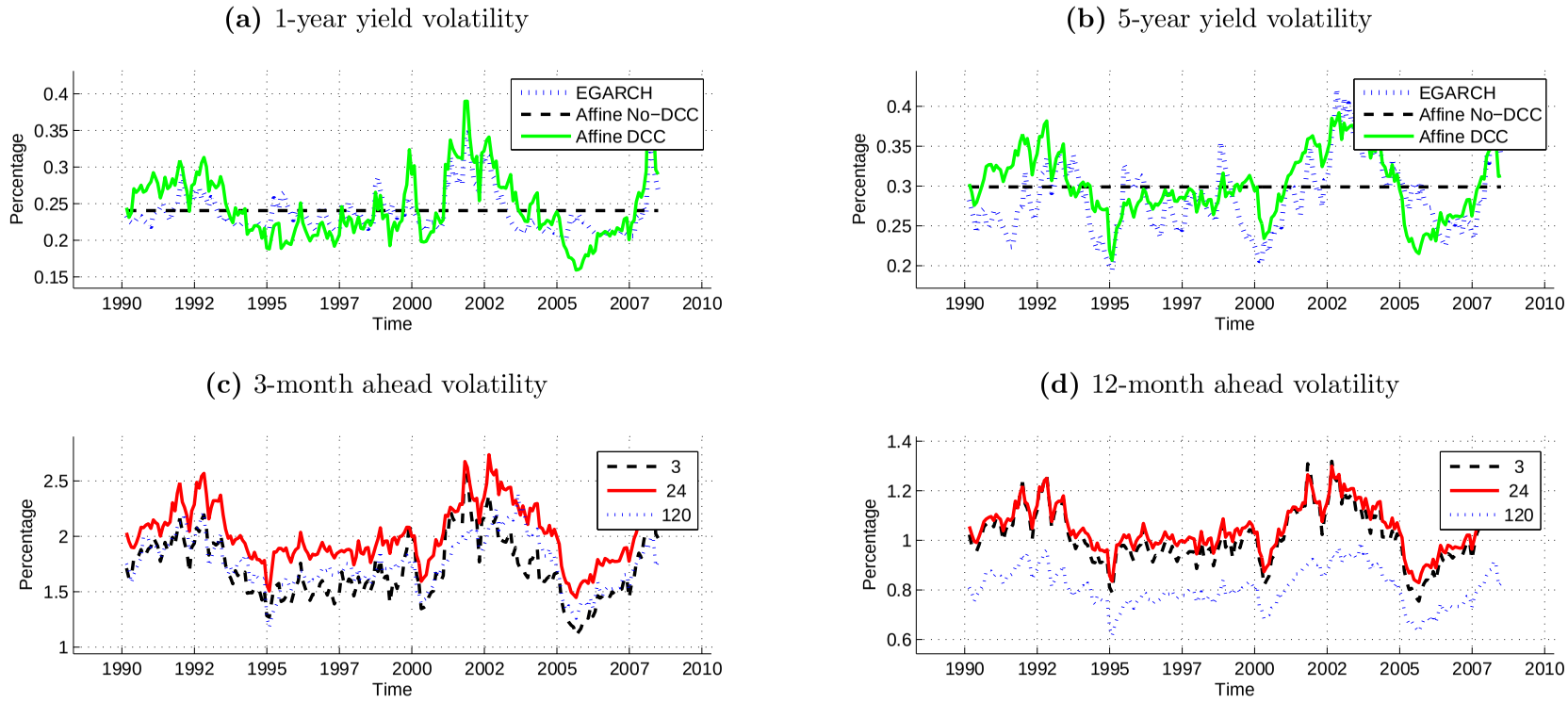



**Figure 8:** Model-implied yield volatilities. Panels 8a-8b: 1-month volatility estimates are close to EGARCH estimates. Conditional volatility for 1-year and 5-year yields estimated from univariate EGARCH and from the *AT* and *ATV* 3-factor models. Panels 8c-8d: the term structure of yields conditional volatility changes over time, displaying a counter-cyclical hump-shape. Volatility estimates 3-month and 12-month ahead for the 1-month, 1-year and 10-year yields, implied by the *ATV* 3-factor model. Monthly data, December 1990 to December 2008.

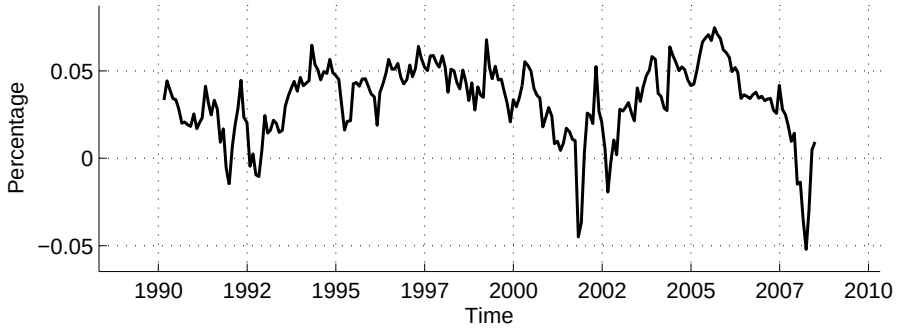
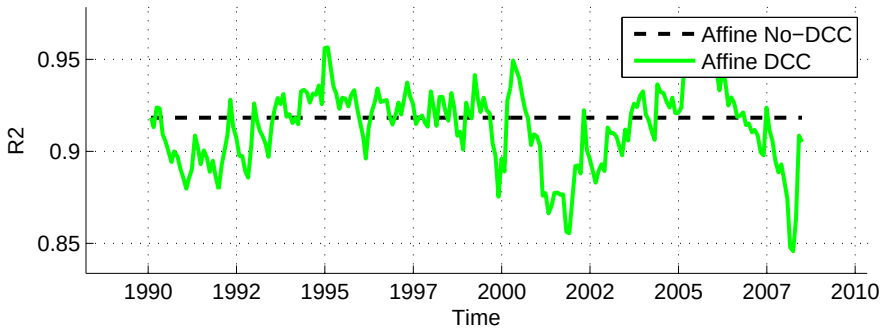



**Figure 9:** The yield volatility term structure hump changes over time. Difference between the 12-month conditional volatility for the 1-year and 1-month yield, respectively, implied by the *ATV* 3-factor model.

**(a)** PC1




**(b)** PC2

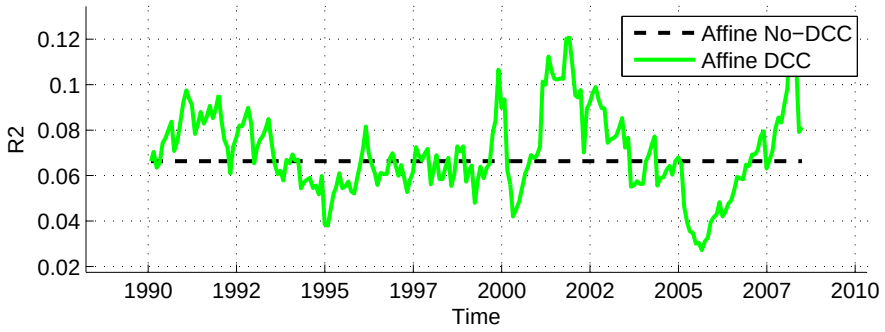



**Figure 10:** The explanatory power of the level and slope factors changes over time. Results from *conditional* principal components analysis of yields based from the model. At each date, we simulate yields one-step ahead and apply PCA across the simulated paths. We report the percentage of variance explained by the first two principal components at each date. Monthly data, December 1990 to December 2008.

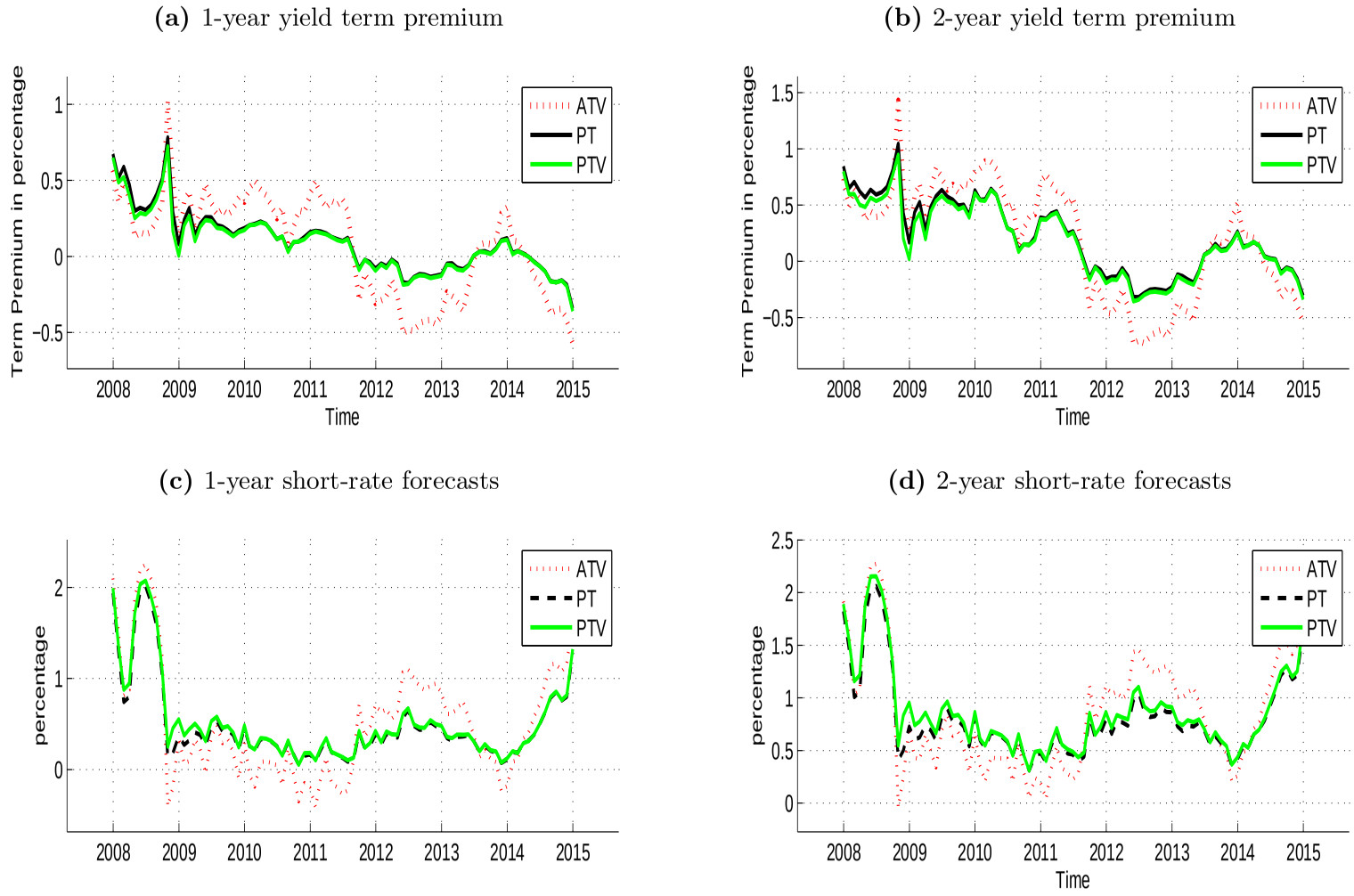
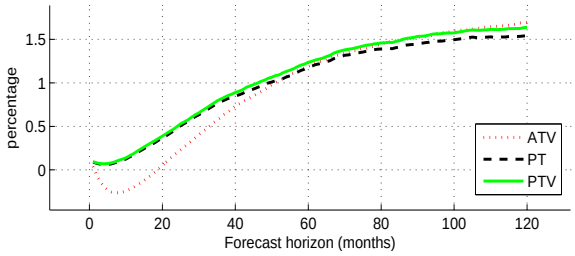



**Figure 11:** Term Premium.

Affine models produce large negative forecasts that biases term premium estimates upward. Panels 11a-**??**: term premium in 1-year and 2-year yields, respectively, implied by different models. Panels 11c-11d: one year ahead and two years ahead forecasts of the short rate.

**(a)** December 2010




**(b)** August 2012

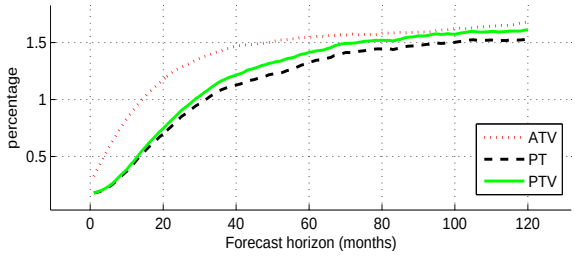



**Figure 12:** Forecasts of the short rate from affine models can differ even at long horizons. Forecasts of the short rate from different models as of December 2010 and as of August 2012 across forecast horizons between one month and then years.

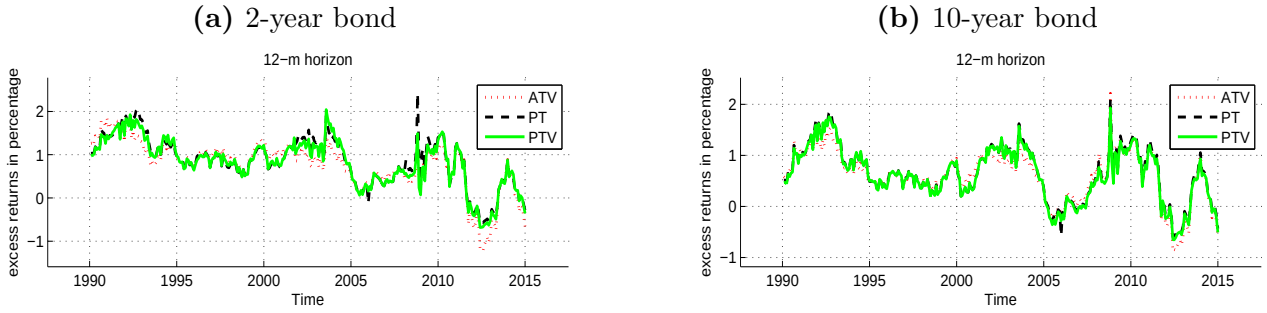



**Figure 13:** Conditional Sharpe Ratios. Time-series of 1-year conditional Sharpe ratios for a 2-year and 10-year bond. Forecasts for the period between July 2007 and December 2013.

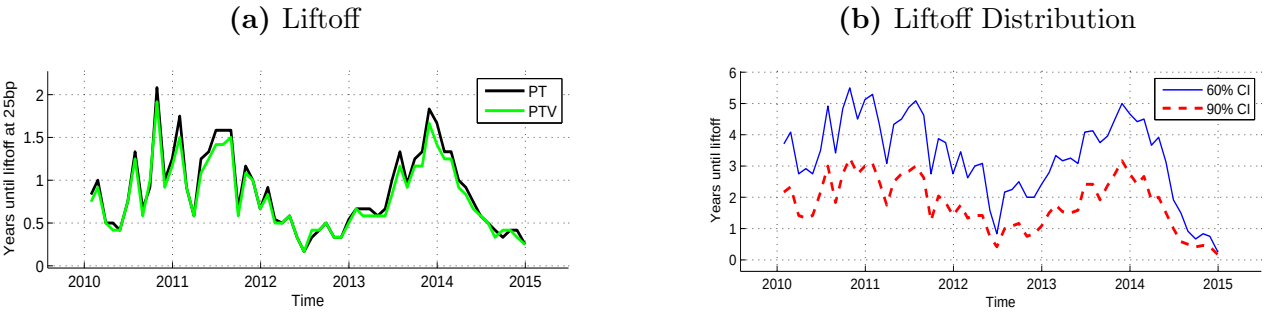



**Figure 14:** Expected time until the short-rate liftoff and distribution of time until the short-rate liftoff. Monthly data August 2007 to December 2014.

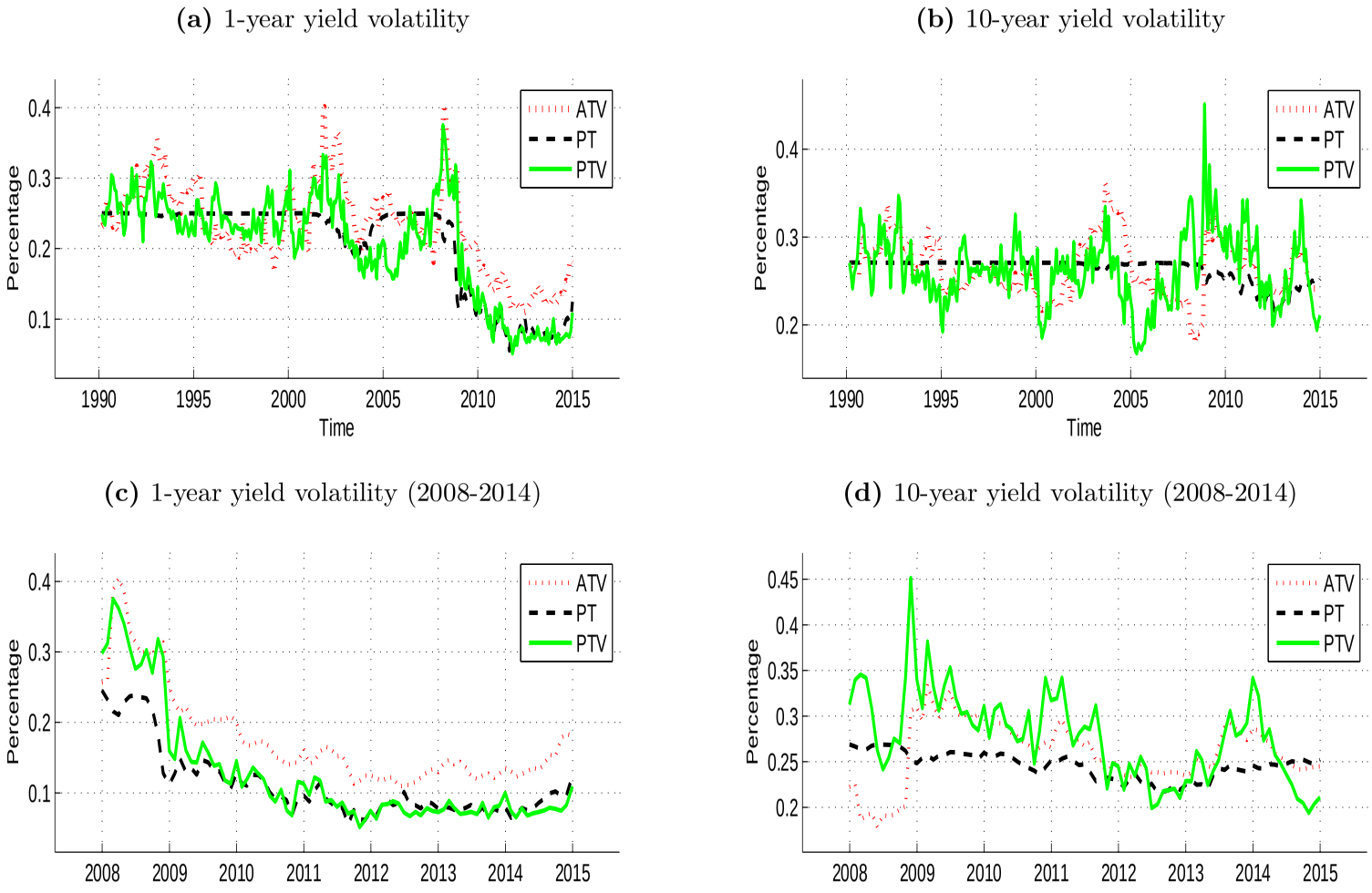
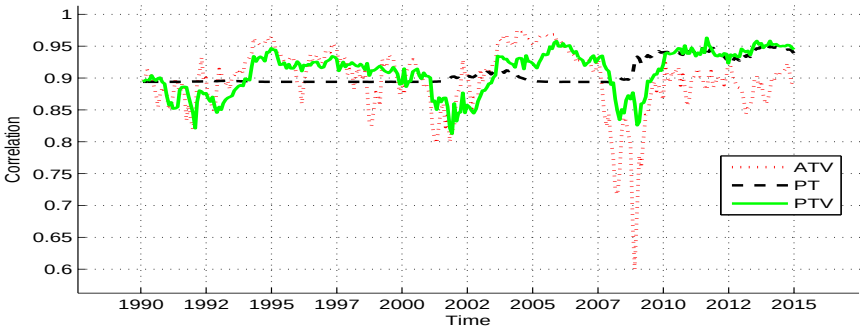


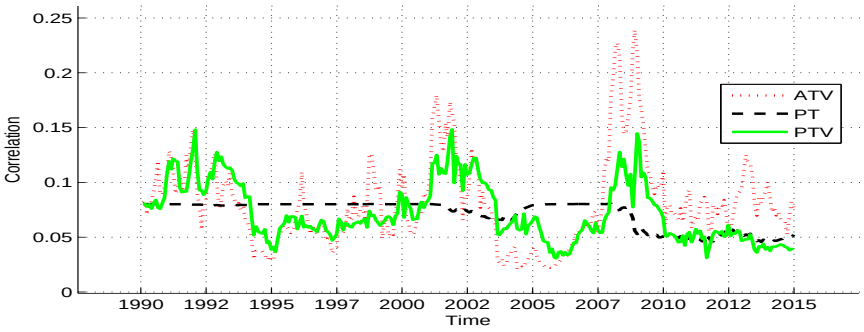

**Figure 15:** Yield volatility dynamics change at the lower bound

Conditional volatility of the 1-year and 10-year yields in the full sample (Panels (a)-(b)) and in the shorter sample where the short-rate was near its lower bond (Panels (c)-(d)). One-month conditional volatility implied by models; monthly data 1990-2014.

**(a)** PC1




**(b)** PC2




**(c)** corr(PC1,PC2)

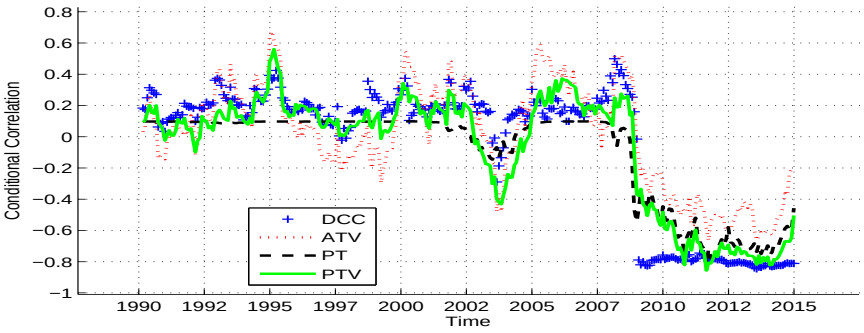



**Figure 16:** The first component of yields plays a greater role near the lower bound. Panels (a)-(b): explanatory power of the first and second *conditional* principal components of yields. Results from *conditional* principal components analysis of yields based from the model. At each date, we simulate yields one-step ahead and apply PCA across the simulated paths. We report the percentage of variance explained by the first two principal components at each date. Panel (c): conditional correlations between the first and second components of yields. Monthly data 1990-2014.

# A Additional Results

## 1 Loadings

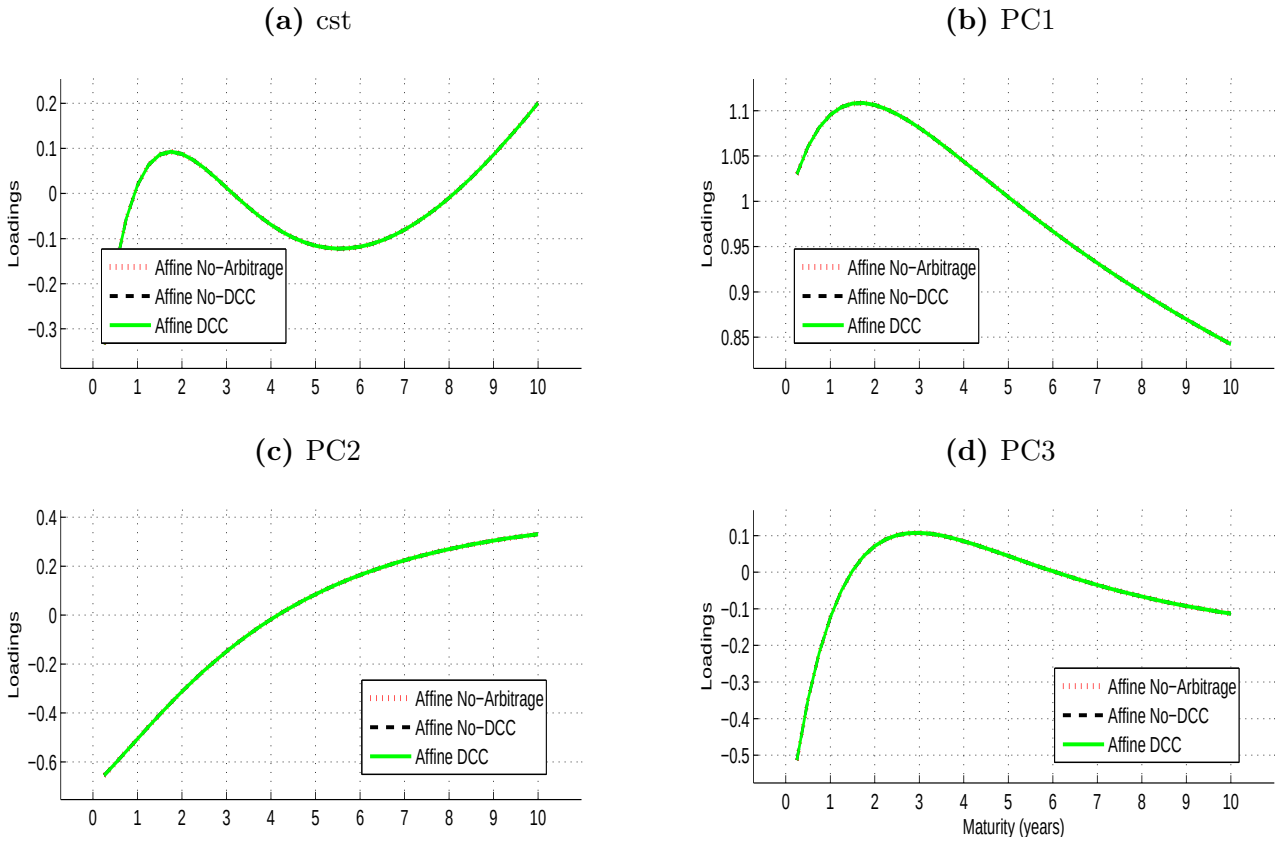



**Figure 17:** Portfolios loadings are almost identical across linear Gaussian models. Constant and risk factor loadings $A_{n,\mathcal{P}}$ and $B_{n,\mathcal{P}}$ across models $A$, $AT$ and $ATV$. Monthly data from CRSP and GSW, January 1990-June 2008.

## 2 CS Regressions before and after 2008

The CS risk premium regressions are linear and does not provide a reliable benchmark in a sample including episodes with interest rate at the lower bound. If the short-rate is fixed, then a steeper slope predicts a faster decline of longer-term yields as they mature and CS regressions should produce large negative coefficients after 2008. To check this, Figure 18 compares regression estimates of CS coefficients in subsamples before and after December 2008, respectively. Coefficients in the first sample correspond to the benchmark case in Figure 7 and displays the well-known pattern. Estimates are close to one for very short maturities–a higher slope predicts a higher short rate–but the estimates decline gradually to values around -2 for longer maturities–a higher slope predicts higher bond returns. The pattern is very different in the second subsample. Estimates start close to zero for the shortest maturity. This was expected: the slope has little predictive power when the short-rate is pinned down to its lower bound. Estimates decline rapidly for longer maturities, ranging between -2 and -7 at maturities between one and ten years.

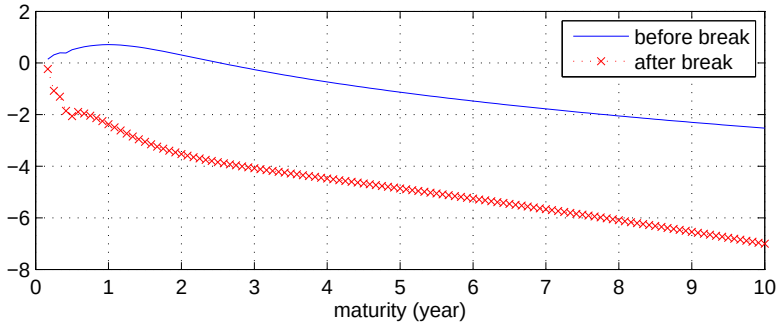



**Figure 18:** Coefficients from Campbell-Shiller regressions are much more negative after 2008, reflecting the effect of non-linearities the short rate reaches its lower bound. Results from regressions with a one-month horizon for bonds with maturities between 3 months and 10 years, in two samples before and after December 2008, respectively. Monthly data from 1990 to 2014.

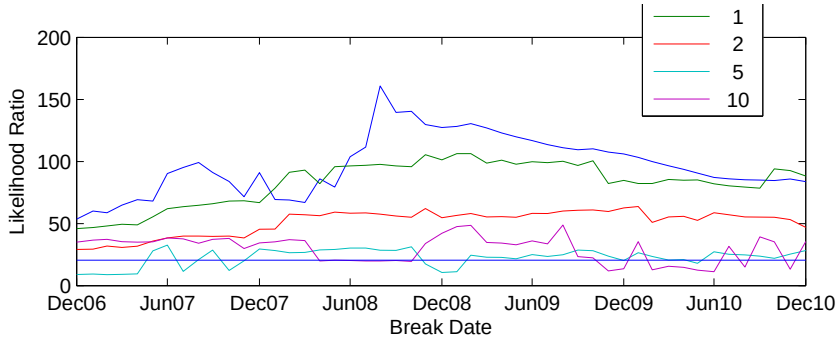



**Figure 19:** The volatility of yields exhibit a clear break in 2008. Likelihood Ratio test statistics for break in all parameters of univariate EGARCH estimated for yields with maturity of 3 month, 1, 2, 5 and 10 years, where yields innovations are obtained from projections on lagged principal components from all yields. Monthly data from 1990 to 2014.

## 3 Breaks in Volatility

The behavior of yields' volatilities also change abruptly after 2008. To see this, we estimated univariate EGARCH across different maturities, but allowing for one break in every parameter. Since the optimal break date is unknown, we re-estimate the EGARCH but varying the break date between December 2006 and December 2010. Figure 19 reports the likelihood ratio (times 2) between the EGARCH estimated either without a break or with one break on December 2008, across a range of maturities. For every maturity, the *sup* test from Andrews (1993) reject the null (absence of a break) at standard significance level for every maturity, yielding $p$-values that are essentially zero for maturities of two years or less.[26]
The test places the break date between June 2008 and June 2009 in almost every case.

[26]We estimate univariate EGARCH models based on the forecasting errors from a projection of individual yields on lagged yield PCs. We allow for a break in every parameter. With 8 parameters, the critical values are 20.5 and 25.2 a the 1% and 5% significance levels, respectively.

# B Proofs

## 1 Proof of Proposition 1

### 1.1 *Discussions on* $\lim_{n\to\infty} P_n(X_t)$

Given that

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

$$\lim_{n\to\infty} P_n(X_t) \to 0$$

if and only if

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} m(g^{\circ i}(X_t)) \to +\infty.$$

Given that $m(X) \geq 0$, $\sum_{i=0}^{n-1} m(g^{\circ i}(X_t))$ is a positive term series, and hence either converge or diverge to $+\infty$. If we further assume that $m(X) \geq lb$, with $lb > 0$, then we can check easily that $\sum_{i=0}^{n-1} m(g^{\circ i}(X_t))$ diverge to $+\infty$, and $\lim_{n\to\infty} P_n(X_t) \to 0$. Indeed, if $m(X) \geq lb$, with $lb > 0$, we have $n \times lb \leq \sum_{i=0}^{n-1} m(g^{\circ i}(X_t))$, which implies that $\sum_{i=0}^{\infty} m(g^{\circ i}(X_t)) = \infty$.

### 1.2 *Discussions on* $\lim_{n\to\infty} f_{n,t}$

Next, we are going to discuss convergence of the forward rate and yield. To do that, we need to make further assumptions on function $g(\cdot)$.

The forward rate is

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

which converges if the sequence $g^{\circ n}(X_t)_{n=0}^{\infty}$ converge. $g^{\circ n}(X_t)_{n=0}^{\infty}$ is also known as the *Picard sequence*, and its convergence have been studied in the mathematic literature. If $\underline{\mathrm{X}}$ is a complete subset (i.e. that every Cauchy sequence converges in $\underline{\mathrm{X}}$) of $\mathbf{R}^K$, and $g(\cdot)$ a contraction (i.e. there exists $L \in [0, 1)$ such that $\|g(X) - g(Y)\| \leq L\|X - Y\|$ for all $X, Y \in \underline{\mathrm{X}}$), then the "Contraction mapping principle", also known as the Banach fixed-point theorem states that:

1. $g(\cdot)$ has a unique fixed point, say $X^*$ in $\underline{\mathrm{X}}$,
2. the sequence $g^{\circ n}(X_t)_{n=0}^{\infty}$ in $\underline{\mathrm{X}}$ converges to $X^*$.

Hence if $g(\cdot)$ is a contraction, and $m(\mathrm{cot})$ is a continuous function, then the forward rate $f_{n,t}$ converges to $m(X^*)$.

### 1.3 *Discussions on* $\lim_{n\to\infty} y_{n,t}$

Turning next to yield, the $n$-period zero-coupon yield is given by:

$$y_{n,t} = (1/n) \sum_{i=0}^{n-1} f_{i,t}.$$

To proceed forward, let recall the following definition and result:

- Definition: Two sequences $x_{nn\geq 0}$ and $y_{nn\geq 0}$ of positive real numbers will be called asymptotic if $\lim_{n\to\infty} \frac{x_n}{y_n} = 1$. It is then denoted by $x_n \sim y_n$.
- Result: If $x_n \sim y_n$, and either sequence of partial sums, $\sum_{k=1}^{n} x_k$ or $\sum_{k=1}^{n} y_k$, ends to $\infty$ as $n \to \infty$, then both partial sum sequences tend to $\infty$ and the partial sum sequences are asymptotic.

Applying this result, we have that $f_{n,t} \sim m(X^*)$ (i.e $\lim_{n\to\infty} \frac{f_{n,t}}{m(X^*)} = 1.$), and $\sum_{i=0}^{\infty} f_{i,t} = \infty$ we have that $\sum_{i=0}^{n-1} f_{i,t} \sim \sum_{i=0}^{n-1} m(X^*)$. Hence

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} f_{i,t}}{n \times m(X^*)} = 1,$$

thus the $n$-period zero-coupon yield $y_{n,t}$ converges to $m(X^*)$.

### *1.4 What happens if the lower bound on $m(\cdot)$ is negative?*

In the discussions on $\lim_{n\to\infty} P_n(X_t)$ and $\lim_{n\to\infty} y_{n,t}$, we have assumed that the lower bound on function $m(\cdot)$, denoted by $lb$, is strictly positive. What happens if $lb \leq 0$?

If we assume that $g(\cdot)$ is a contraction, and $m(\cdot)$ is continuous, then the forward rate $(m(g^{\circ n}(X_t)))$ converges to $m(X^*)$ where $X^*$ is the fixed point of $g(\cdot)$ (i.e. the unique point in $\underline{X}$ satisfying $g(X^*) = X^*$).

If we further assume that $m(X^*) > 0$, it implies that there exists a strictly positive number (say $\epsilon > 0$) with $m(X^*) > \epsilon$.

Applying the definition of the convergence of $m(g^{\circ n}(X_t))$ to $m(X^*)$ implies that, there exists an $N(\epsilon)$ such that for all $n \geq N(\epsilon)$, $|m(g^{\circ n}(X_t)) - m(X^*)| < \epsilon$.

$|m(g^{\circ n}(X_t)) - m(X^*)| < \epsilon$ is equivalent to

$$m(X^*) - \epsilon < m(g^{\circ n}(X_t)) < m(X^*) + \epsilon.$$

We can then split $\sum_{n=0}^{\infty} m(g^{\circ i}(X_t))$ in two terms:

$$\sum_{n=0}^{\infty} m(g^{\circ i}(X_t)) = \sum_{n=0}^{N(\epsilon)-1} m(g^{\circ i}(X_t)) + \sum_{n=N(\epsilon)}^{\infty} m(g^{\circ i}(X_t)).$$

Because $m(g^{\circ n}(X_t)) > m(X^*) - \epsilon > 0$ when $n \geq N(\epsilon)$, we have $\sum_{n=N(\epsilon)}^{\infty} m(g^{\circ i}(X_t)) = +\infty$.

Thus

$$\lim_{n\to\infty} P_n(X_t) = \exp\left(-\sum_{n=0}^{N(\epsilon)-1} m(g^{\circ i}(X_t))\right) \exp(-\infty) = 0,$$

and $y_{n,t}$ converges to $m(X^*)$.

## 2 Proof of Theorem 4

Consider again a portfolio in which the amount (in face value) invested in each n-period bond is given by $w_n$. To account for transaction costs, let's assume that the setup costs at time $t$ of the portfolio is $C_0$ and to realize the cash flows at time $t+1$, the transaction cost is given by $C_1$. The existence of transaction costs means that $C_0$ and $C_1$ cannot be jointly zeros. The one-period ahead cash-flows net of transaction

costs are given by:

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

Recall that the absence of arbitrage is *equivalent* to requiring that any portfolio with non-negative payoffs must command a positive price. Thus, the key question here is: if $CF(X_{t+1}) \geq 0$ for every $X_{t+1} \in \underline{X}$, can we show that the price of the portfolio, net of transaction costs:

$$Price(X_t) = \sum_n w_n P_n(X_t) + C_0$$

must be strictly positive? From Equation (2), we have:

$$\begin{aligned} Price(X_t) =& exp(-m(X_t)) \sum_n w_n P_{n-1}(g(X_t)) + C_0, \\ =& exp(-m(X_t)) \times (CF(g(X_t)) + C_1) + C_0, \\ =& exp(-m(X_t)) \times CF(g(X_t)) + exp(-m(X))C_1 + C_0. \end{aligned} \tag{50}$$

The first term on the right hand side of (50) is non-negative since $g(X_t) \in \underline{X} \Rightarrow CF(X_t) \geq 0$. Additionally, the last two terms must add up to a strictly positive number since $C_0$ and $C_1$ cannot be jointly zeros. Thus, $Price(X_t) > 0$ as needed.

## 3 Ruling Out Arbitrage Opportunities

This section constructs a model that rules out arbitrage opportunities in frictionless markets. We then construct a sequence of such models that in the limit approach the model proposed in Section II.

**Assumption 4.** *The n-period bond price $P_n$ is determined by the following recursion for each n:*

$$P_0(X) \equiv 1, \tag{51}$$

$$P_n(X) = \frac{1}{J} \sum_{i=1}^{J} P_{n-1}(g_i(X)) \times exp(-m_i(X)), \tag{52}$$

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

**Theorem 5.** *Assumption 4 rules out all arbitrage opportunities.*

*Proof.* Consider a portfolio with non-negative payoffs: $\sum_n w_n P_{n-1}(X) \geq 0$ for all $X \in \underline{X}$. According to (52), the price of this portfolio for each state is given by:

$$\sum_n w_n P_n(X) = \frac{1}{J} \sum_i \left( exp(-m_i(X)) \times \sum_n w_n P_{n-1}(g_i(X)) \right). \tag{53}$$

Because $g_i(X) \in \underline{X}$, it follows each of the $\sum_n w_n P_{n-1}(g_i(X))$ terms is a possible payoff from the portfolio thus these terms must be non-negative. Therefore the price of the considered portfolio must be non-negative. For the price of the portfolio to be zero, each of the summations $\sum_n w_n P_{n-1}(g_i(X)) = w \cdot v_i(X)$ must be zero, where $w = (w_1, w_2, ..., w_J)'$. That is, $w \cdot (v_1, v_2, ..., v_J) = 0$. Because $(v_1, v_2, ..., v_J)$ is full rank, it follows that the price of the portfolio can only be zero when $w \equiv 0$. This means that the portfolio's payoff must be uniformly zero across all states. □

Finally, consider $g_1(X_t) \equiv g(X_t)$ and$m_1(X_t) \equiv m(X_t)$ and for $i > 1$, $exp(-m_i(X_t)) \equiv a$ for some constant $a > 0$, then we have:

$$P_n(X_t) = P_{n-1}(g(X_t)) \times exp(-m(X_t)) + a \left( \sum_{i=2}^{J} P_{n-1}(g_i(X_t))/J \right). \tag{54}$$

Letting $a \to 0$, we obtain the model proposed in Section II in the limit.