SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2010-11

Affine Bond Pricing with a Mixture Distribution for Interest Rate Time-Series Dynamics

Torben B. Rasmussen

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Affine Bond Pricing with a Mixture Distribution for Interest Rate Time-Series Dynamics*

Torben B. Rasmussen†
Aarhus University and CREATES

January 13, 2010

Abstract

Starting from the discrete-time affine term structure model by Dai, Le & Singleton (2006), this paper proposes a Radon-Nikodym derivative which implies that factors follow a mixture distribution under the physical measure. The model thus maintains attractive features of an affine relation between yields and factors, while allowing for nonlinear and non-normal time-series dynamics. Empirically the fit of the discrete-time 3-factor affine model is found to be substantially improved by the inclusion of two components to describe the time-series dynamics. Relative to the risk-neutral model, the mixture model is able to let the variance of the one-period rate be higher and faster increasing in the variance factor, and to introduce negative skewness and positive excess kurtosis. When weights on the components depend on factors, the model produces a speed of mean reversion and variance of the one-period rate that both increase fast with higher levels of the yield curve. The added second component is found to capture infrequent relatively large simultaneous shifts in direction of a yield curve that is at a lower level, is steeper, and is more positively curved.

Keywords: Term Structure, Discrete-time models, No-arbitrage pricing, Mixture models.

JEL Classifications: C51, E43, G12

*The author is grateful to Aarhus University Research Foundation for support and to the Bendheim Center for Finance, Princeton University for the hospitality provided during a stay there as a visiting PhD student. I thank Bent Jesper Christensen for comments and suggestions. CREATES (Center for Research in Econometric Analysis of Time Series) is funded by the Danish National Research Foundation.

†Email: trasmussen@econ.au.dk

1 Introduction

The term structure of interest rates is commonly described by a factor model in which zero-coupon bond yields depend linearly on factors. This goes back at least to Litterman & Scheinkman (1991), who showed that the cross-sectional relation of yields with different maturities is well captured by 3 linear factors with loading functions shaped such that the factors impact the level, slope and curvature of the yield curve, respectively. In addition, a complete model of the term structure of interest rates describes the behavior of the yield curve over time by modelling the dynamics of the factors in a way that does not admit arbitrage. Duffie & Kan (1996) establish that an affine relation between yields and factors is implied by models in which the factor drift and squared volatility terms for a risk-neutral investor, as well as the short rate, are affine in the factors. Adding a description of investor preferences by a properly defined market price of risk function leads to the continuous-time affine term structure models, which are detailed in Dai & Singleton (2000).

The time series of interest rates have been documented by several authors to exhibit nonlinearities in both first and second moments. Thus the short rate is found to mean revert much stronger further away from the mean, and the volatility to increase faster than linearly with the level of rates, Chan, Karolyi, Longstaff & Sanders (1992), Aït-Sahalia (1996*a*), Aït-Sahalia (1996*b*), and Boudoukh, Richardson, Stanton & Whitelaw (1999). Yield changes also does not appear to be normal distributed showing negative skewness and positive excess kurtosis. This is for instance captured in models of the short rate that include jumps, Johannes (2004) and Piazzesi (2001).

An affine pricing model is only a consequence of dynamics in the risk-neutral model, and several papers have shown that generalizations of the market price of risk function improves the ability of continuous-time affine models to match time-series properties of yields, Duffee (2002), Duarte (2004), and Cheridito, Filipovic & Kimmel (2007). In continuous-time Wiener driven models differences between the dynamics under the risk-neutral and historical measures are only in the drift term as a consequence of the Girsanov theorem. This naturally limits the ability to reconcile affine pricing models resulting from affine continuous time risk-neutral factor dynamics with some of the observed empirical time-series properties of yields.

This paper attempts to improve on the ability of term structure models with an affine pricing relation to capture the nonlinear and non-normal time-series properties of interest rates. The proposed model is set in discrete time and builds on the framework introduced by Dai et al. (2006), a discrete analog to the continuous-time affine models. By modelling via the conditional moment generating function, their approach conveniently obtains closed-form expressions for the conditional likelihood function even for general market price of risk functions. Still, the formulation of the Radon-Nikodym derivative that they propose implies that only the mean of factors that are Gaussian under $\mathbb{Q}$ is affected, and more generally, that changes in higher moments are fixed when the difference in mean between measures is determined.

The paper suggests a Radon-Nikodym derivative that is a weighted average of Esscher transforms. This has the consequence that the $\mathbb{P}$ distribution of yields will be a mixture distribution with components defined by the individual Esscher transforms and mixture proportions that are the weights from the Radon-Nikodym derivative. Component distributions are known in closed form, by the method used in Dai et al. (2006), and thus the full conditional likelihood function of the mixture model is known as well. Besides the individual market price of risk functions that define each component relative to the risk-neutral model, the mixture proportions may also depend on factors in the model.

By combining individual components the mixture model allows for nonlinearities as well as variation in higher moments of the time-series distribution of yields. This is also possible in models where each component under the physical measure is affine with constant parameters like in the risk-neutral model. Factors come from the standard affine pricing framework, and the variation in moments conditional on these can therefore be interpreted in relation to level, slope, and curvature of the yield curve. Models with regime switching also describe interest rate time series as a combination of distributions. These models have been applied for instance by Hamilton (1988), Ang & Bekaert (2002), Bansal & Zhou (2002). A mixture model is similar to a regime switching model prior to realization of the regime but without dependence in regime over time. That is, the weighing of components does not depend on which component the previous period's realization comes from. The mixture model with factor-dependent weights do allow for periods of higher or lower probability of each "regime", but this dependence runs entirely through the position of factors.

In an empirical implementation the 3-factor affine pricing model with a mixture element in the time-series dynamics is applied to weekly US zero-coupon bond yields. The estimation method uses the standard simplifying assumption that some yields are observed without error to implicitly obtain the factors and allow for maximum likelihood estimation without filtering. Results show that allowing for a two-component distribution in the time series substantially improves the fit to observed yields relative to a single-component model when fit is compared by information criterion statistics to correct for the higher number of parameters in the mixture model. Further, there is evidence in favour of letting weights depend on factors, though this improvement is less than that of adding the second component.

The mixture models estimate a main component fairly similar to the single-component models, while the second component with low weight is quite different. The conditional mean of the additional component is located such that it allows for relatively large changes of the yield curve in a specific direction. This can be described as a simultaneous shift towards a yield curve that has lower level, more steepness and more positive curvature. Due to the high and low weights on each of the two components, such that these represent standard movement and infrequent large moves, the model captures some of the same effects that Poisson-type jumps do in continuous-time models.

The suggested Radon Nikodym derivative allows for substantial differences in other moments than the mean between the risk-neutral and the physical measures. Compared to single-component models, this results in the variance of the one-period rate being lower and more slowly increasing with the variance factor in the risk-neutral part of the mixture models. Thus the pricing implications are changed, although the specification of this part of the model is unchanged. The higher one-period rate variance, which also increases sharper with the variance factor, shows up only in the time-series dynamics of the mixture model. Changes in the one-period rate under the physical measure are very close to being normal distributed for weekly intervals in the single-component models, whereas the mixture models estimate negative skewness and positive excess kurtosis in the one-period rate time series.

With factor-dependent weights, the weight on the second component is higher when the yield curve has a higher level factor, when it has a lower positive curvature factor, and slightly higher when the steepness factor is higher. These results imply clear nonlinear patterns in expected changes of the one-period rate. Thus it is expected to revert faster towards lower rates when the general level of rates is high and when medium term rates are low. Variance increases sharply as well for a higher level factor and for higher negative curvature. Negative skewness and positive excess kurtosis are highest when these two factors are about one standard deviation above their means.

A by-product of the empirical implementation is the observation that the canonical identification scheme for affine models is problematic for some data. The constant terms in the variance of the Gaussian factors are fixed in the canonical rotation to identify the scale of these factors, but they are in general only required to be non-negative. If data requires a model with some of these terms close to zero, then the scale of the related factors becomes weakly identified resulting in problems of estimating standard errors for all parameters affected by scale rotations of the given factor. This result is also applicable for continuous-time models and is easily solvable when discovered by instead identifying the scale of the given factor via restrictions on another parameter.

Section 2 of the paper introduces the discrete-time affine pricing model by Dai et al. (2006). The conditions that must be satisfied by the Radon-Nikodym derivative in this framework to properly define the physical dynamics are discussed in Section 3. In Section 4 the suggested form of this function is described and the consequences for the $\mathbb{P}$ distribution of factors is presented. Section 5 discusses how to estimate the proposed type of model under the assumption that some yields are observed without error. The empirical study of different risk price specifications in the 3-factor model with one factor affecting variance is in Section 6 and Section 7 concludes.

2 Discrete-Time Affine Term Structure Model

Dai et al. (2006) suggest a class of discrete-time affine term structure models analogous to the well known affine models set in continuous time. In this section, I briefly review the part of their model that determines the cross-sectional relation between zero-coupon bond prices and thus the yield curve. As is standard, this is determined by a risk-neutral model of the short rate, or in discrete time the one-period rate. The model describes the one-period rate as affine in N factors and then specifies the dynamics of these factors under the risk-neutral measure, $\mathbb{Q}$. The factors follow a first order affine Markov process in which M factors are allowed to affect conditional variances. Thus the discrete affine risk-neutral model is termed $DA_M^{\mathbb{Q}}(N)$.

The N-dimensional Markov process for factors, X, is affine since the conditional moment generating function of X_{t+1} given X_t is exponentially affine

$$\phi_t^{\mathbb{Q}}(u) = E_t^{\mathbb{Q}}\left[e^{u'X_{t+1}}\right] = e^{a(u)+b(u)'X_t}, \tag{1}$$

where u is an $N \times 1$ vector, the function $a(\cdot)$ takes scalar values, and $b(\cdot)$ is an $N \times 1$ vector of functions. In the $DA_M^{\mathbb{Q}}(N)$ model the conditional factor distributions are in particular chosen such that the model has as its continuous-time limit the continuous affine $A_M^{\mathbb{Q}}(N)$ model specified by Dai & Singleton (2000). Therefore arrange the N factors in two groups, $X_t' = [Z_t', Y_t']$, where Z_t is a discrete time counterpart to M correlated Cox-Ingersoll-Ross processes, and Y_t corresponds to $N - M$ Vasicek processes. To simplify the presentation, I restrict Z_t to be scalar, i.e., the model is only described here for $M \leq 1$. This includes the empirically applied $DA_1^{\mathbb{Q}}(3)$ model, but the extension suggested in the paper can equivalently be applied to models with multiple variance factors.

The variance factor Z is an autonomous process, independent of the Y factors at all times, and it follows the exact discrete time counterpart of a CIR process. Thus the conditional distribution of the variance factor at $t+1$ given Z_t can be expressed by

$$\frac{2Z_{t+1}}{c}|Z_t \sim \chi^2\left(2\nu, \frac{2\rho Z_t}{c}\right), \tag{2}$$

where $\chi^2(k, \lambda)$ is the non-central chi-square distribution with k degrees of freedom and non-centrality parameter λ. The parameters must satisfy $0 < \rho < 1$, $\nu > 0$, and $c > 0$.

Conditional on variables at t, the $N - 1$ non-variance factors Y_{t+1} are independent of Z_{t+1} and normally distributed,

$$Y_{t+1}|X_t \sim N(\omega_{Yt}, \Omega_{Yt}). \tag{3}$$

The mean is affine in all factors, $\omega_{Yt} = \mu_0 + \mu_Z Z_t + \mu_Y Y_t$, for $(N-1) \times 1$ vectors μ_0 and

μ_Z, and an $(N-1)\times(N-1)$ matrix μ_Y. The variance is affine in Z_t alone, and this is parameterized by $\Omega_{Yt} = \Sigma_Y S_{Yt} \Sigma_Y'$, where Σ_Y is a non-singular $(N-1)\times(N-1)$ matrix and $S_{Yt} = diag(\alpha + \beta Z_t)$ for $(N-1)\times 1$ parameter vectors α and β which both are non-negative. Occasionally, it is convenient to express the variance as $\Omega_{Yt} = h_0 + h_Z Z_t$, for the symmetric, positive semi-definite matrices $h_0 = \Sigma_Y diag(\alpha)\Sigma_Y'$ and $h_Z = \Sigma_Y diag(\beta)\Sigma_Y'$.

For the factor distributions of the $DA_1^{\mathbb{Q}}(N)$ model, the conditional moment generating function is on the exponential-affine form (1) with the a and b functions given by

$$\begin{aligned} a(u) &= -\nu \ln(1-u_Z c) + u_Y' \mu_0 + \frac{1}{2} u_Y' h_0 u_Y \\ b(u) &= \left[\frac{u_Z}{1-u_Z c}\rho + \frac{1}{2} u_Y' h_Z u_Y + u_Y' \mu_Z \ , \ u_Y' \mu_Y \right]'. \end{aligned} \tag{4}$$

Here, u_Z and u_Y are the scalar and $(N-1)$-vector in $u' = [u_Z, u_Y']$, and it must hold that $u_Z < 1/c$.

When parameters are chosen appropriately the model converges to the continuous affine $A_1^{\mathbb{Q}}(N)$ model as the period length shrinks to zero. Suppose that parameters in a Δ-period discrete model are set equal to

$$\begin{aligned} &\rho = 1-\kappa_{ZZ}\Delta, && \nu = 2\kappa_{ZZ}\theta_Z/\sigma_Z^2, && c = \sigma_Z^2\Delta/2, \\ &\mu_0 = [\kappa_{YZ}, \kappa_{YY}]\,\theta\Delta, && \mu_Z = -\kappa_{YZ}\Delta, && \mu_Y = I - \kappa_{YY}\Delta, \\ &\alpha = \alpha^c\Delta, && \beta = \beta^c\Delta, && \end{aligned} \tag{5}$$

where dimensions of right hand side parameters follow directly from the discrete model parameters. Then it can be shown that for $\Delta \to 0$ the discrete model converges to

$$dX_t = \kappa(\theta - X_t)\,dt + \Sigma\sqrt{S_t^c}\,dW_t, \tag{6}$$

where

$$\kappa = \begin{pmatrix} \kappa_{ZZ} & 0 \\ \kappa_{YZ} & \kappa_{YY} \end{pmatrix}, \qquad \theta = \begin{pmatrix} \theta_Z \\ \theta_Y \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_Z^2 & 0 \\ 0 & \Sigma_Y \end{pmatrix}. \tag{7}$$

and $S_t^c = diag([Z_t, \alpha^c + \beta^c Z_t])$.[1]

The affine relation between the one-period interest rate and the factors is written as

$$r_t = \delta_0 + \delta_X' X_t, \tag{8}$$

for a scalar δ_0 and an $N\times 1$ vector δ_X. The entries in δ_X are denoted $(\delta_Z, \delta_1, ..., \delta_{N-1})'$, i.e., the numbers indicate which Y factor the parameters affect.

[1] The continuous model parameters α^c, β^c, and matrix S_t^c have been given superscript c to distinguish them from the corresponding discrete model parameters α, β, and matrix S_{Yt}, which are all scaled relative to the period length Δ.

Analogous to the result in Duffie & Kan (1996) for continuous time models, the affine specifications in (1) and (8) imply that the price of a zero-coupon bond with maturity n is exponentially affine in the factors,

$$P_t^n = e^{-A_n - B_n' X_t}, \tag{9}$$

where the scalars A_n and the $N \times 1$ vectors B_n are determined recursively according to[2]

$$\begin{aligned} A_n &= A_{n-1} + \Delta\delta_0 - a\left(-B_{n-1}\right) \qquad & A_0 = 0 \\ B_n &= \Delta\delta_X - b\left(-B_{n-1}\right) \qquad & B_0 = 0. \end{aligned} \tag{10}$$

This result follows since under the martingale measure $\mathbb{Q}$ discounted prices must satisfy the no-arbitrage condition

$$P_t^n = E_t^{\mathbb{Q}}\left[e^{-\Delta r_t} P_{t+1}^{n-1}\right]. \tag{11}$$

For a risk-neutral model determined by the short rate (8) and factor dynamics (1) this is satisfied when prices are on the form (9) - (10). The n-period zero-coupon yield at time t (continuously compounded and measured in per annum terms) is given by $y_t^n = -\log\left(P_t^n\right) / \left(n\Delta\right)$. The yield curve at time t when factors are equal to X_t is therefore

$$y_t^n = \delta_0^n + \delta_X^{n\prime} X_t, \tag{12}$$

where the scalars $\delta_0^n = A_n / \left(n\Delta\right)$ and the $N \times 1$ vectors $\delta_X^n = B_n / \left(n\Delta\right)$ have been named for their analogy with parameters in the one-period rate, $r_t = y_t^1 = \delta_0^1 + \delta_X^{1\prime} X_t = \delta_0 + \delta_X' X_t$.

3 Restrictions in Choice of Physical Dynamics

A complete term structure model must also describe the behavior of interest rates over time. As seen from (12) the risk-neutral model determines the cross-sectional relation between yields in terms of the factors at each point in time. Therefore, the model is completed by also modelling the time-series behavior of the factors, i.e., determining the dynamics of factors under the actual, or physical, probability measure, $\mathbb{P}$. An asset pricing model is free of arbitrage if there exists an equivalent probability measure under which discounted prices are martingales. The approach used here to modelling the term structure is somewhat backwards, but standard, since dynamics under the martingale measure $\mathbb{Q}$ is determined first and then the relation to the actual behavior of rates is determined in the second step. To preclude arbitrage it is therefore only left to ensure that the physical probability measure $\mathbb{P}$ is equivalent to $\mathbb{Q}$, which then is an equivalent martingale measure per definition. This

[2]The period length Δ enters in front of δ_0 and δ_X since the one-period rate is measured in per annum terms, see the appendix.

section clarifies the necessary restrictions to ensure a well defined arbitrage-free model when specifying the physical dynamics on top of a given discrete-time risk-neutral model, such as the affine pricing model of Section 2.

Before turning to the multi-period model, consider first a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$. Then by the Radon-Nikodym theorem a measure absolutely continuous to $\mathbb{Q}$ can be defined via a non-negative, $\mathcal{F}$-measurable random variable ξ by

$$\mathbb{P}(E) = \int_E \xi d\mathbb{Q} \qquad \forall \quad E \in \mathcal{F},$$

and $\xi = \frac{d\mathbb{P}}{d\mathbb{Q}}$ is called the Radon-Nikodym (RN) derivative of $\mathbb{P}$ with respect to $\mathbb{Q}$. The new measure $\mathbb{P}$ is a probability measure if it integrates to 1 on Ω, which is satisfied if the RN derivative has $\mathbb{Q}$ expected value 1, i.e., $E^{\mathbb{Q}}(\xi) = 1$. To ensure that $\mathbb{Q}$ is also absolutely continuous to $\mathbb{P}$, such that the measures are equivalent, it is further required that $\xi > 0$, $\mathbb{Q}$-a.s. With these additional conditions an equivalent probability measure can be defined by an RN derivative.

In the discrete-time model let the RN derivative of $\mathbb{P}$ with respect to $\mathbb{Q}$ at time t for variables to be realized at $t+1$ be $\xi_{t,t+1}$. For tractability reasons I restrict $\xi_{t,t+1}$ to be a function of X_t and X_{t+1} only, such that the first order Markov property is preserved. Applying the restrictions found above, the function must satisfy

$$\begin{aligned} E_t^{\mathbb{Q}}\left[\xi_{t,t+1}\right] &= 1 \\ \xi_{t,t+1} > 0, &\quad \mathbb{Q}_{t+1} - a.s. \end{aligned} \tag{13}$$

to ensure that the new measure it defines is an equivalent probability measure. The distribution of variables in the risk-neutral model was specified through the conditional moment generating function, (1). By definition of the RN derivative it is now straightforward to obtain the corresponding function under the physical measure

$$E_t^{\mathbb{P}}\left(e^{u'X_{t+1}}\right) = E_t^{\mathbb{Q}}\left(\xi_{t,t+1}e^{u'X_{t+1}}\right). \tag{14}$$

Defining all one-period RN derivatives in the discrete model to satisfy (13) will ensure that the change of measure is also well defined over longer periods. By application of iterated expectations and (14) for each period it follows that the RN derivative at time t for variables at $t+k$ must be the product of those for each period,

$$\xi_{t,t+k} = \prod_{i=1}^{k} \xi_{t+i-1,t+i}. \tag{15}$$

Again, by iterated expectations it is straightforward to check that $\xi_{t,t+k}$ has $\mathbb{Q}_t$-expected value 1, and that it is strictly positive on sets with positive mass under $\mathbb{Q}_{t+k}$, since this holds for all the single period RN derivatives. Thus the discrete multi-period model can consistently be defined under the physical measure by specifying all one-period conditional RN derivatives.

To determine the physical dynamics of interest rates by a change of measure in the initial risk-neutral model in the way just described, the RN derivative must further be chosen such that the resulting $\mathbb{P}$ distributions are econometrically tractable. In the discrete-time model tractability is only necessary on a period by period basis, i.e., concerning distributions arising from (14). The distributions that result for variables over longer periods from $E_t^{\mathbb{P}}\left(e^{u'X_{t+k}}\right) = E_t^{\mathbb{Q}}\left(\xi_{t,t+k}e^{u'X_{t+k}}\right)$ may be complicated and not known in closed form due to the way $\xi_{t,t+k}$ is obtained by (15). This is different from continuous-time models for which tractability at the frequency of observations automatically implies this at longer frequencies, as well.

The physical dynamics of factors in the discrete-time term structure model can be determined by a measure change from the affine risk-neutral model by defining in each period the conditional RN derivatives, $\xi_{t,t+1}$. These must satisfy conditions (13) to ensure that $\mathbb{P}$ is an equivalent probability measure to the martingale measure $\mathbb{Q}$, such that the model is free of arbitrage. Further, $\xi_{t,t+1}$ should define a conditional moment generating function under $\mathbb{P}$, (14), such that the distribution of variables $X_{t+1}|X_t$ is econometrically tractable. Besides these considerations, $\xi_{t,t+1}$ can be chosen to get a $\mathbb{P}$ model with a good fit to empirically observed interest rate behavior.

4 Mixture Model

This section suggests a flexible approach to defining the physical dynamics of interest rates on top of the discrete-time risk-neutral model. Thus it shows that if the one-period Radon-Nikodym derivatives are weighted averages of variables that all satisfy the conditions discussed in the previous section, then the resulting model under $\mathbb{P}$ is a mixture model. Start with k variables $\xi_{t,t+1}^1, .., \xi_{t,t+1}^k$ that satisfy (13) and therefore define equivalent probability measures to $\mathbb{Q}$, say $\mathbb{P}_1, ..., \mathbb{P}_k$.[3] Then let the RN derivative that defines $\mathbb{P}$ be the convex combination of these k variables,

$$\xi_{t,t+1} = \sum_{j=1}^{k} w_t^j \xi_{t,t+1}^j, \qquad \sum_{j=1}^{k} w_t^j = 1, \tag{16}$$

where the non-negative weights are allowed to be functions of the conditioning set of variables at time t, $w_t^j = w^j(X_t) \geq 0$. The convex combination preserves the properties (13), satisfied

[3] Here, k is generally small and in the later empirical application k is 2, but the notation for arbitrary k is convenient.

by each $\xi^j_{t,t+1}$,

$$
\begin{aligned}
E^{\mathbb{Q}}_t\left[\xi_{t,t+1}\right] &= \sum_{j=1}^{k} w^j_t E^{\mathbb{Q}}_t\left[\xi^j_{t,t+1}\right] = \sum_{j=1}^{k} w^j_t = 1 \\
\xi_{t,t+1} &= \sum_{j=1}^{k} w^j_t \xi^j_{t,t+1} > 0 \quad \mathbb{Q}_{t+1} - a.s.,
\end{aligned}
$$

so the combined RN derivative properly defines a probability measure $\mathbb{P}$ equivalent to $\mathbb{Q}$. The implied conditional moment generating function under $\mathbb{P}$ is

$$
\phi^{\mathbb{P}}_t(u) = E^{\mathbb{Q}}_t\left[\xi_{t,t+1} e^{u' X_{t+1}}\right] = \sum_{j=1}^{k} w^j_t E^{\mathbb{Q}}_t\left[\xi^j_{t,t+1} e^{u' X_{t+1}}\right] = \sum_{j=1}^{k} w^j_t \phi^j_t(u),
$$

where $\phi^j_t(u)$ is the conditional moment generating function under $\mathbb{P}_j$, the probability measure defined by the j'th individual RN derivative, $\xi^j_{t,t+1}$.

In a finite mixture distribution the moment generating function is the weighted average of the component moment generating functions with weights that are the mixture proportions. It follows that factors in a $\mathbb{P}$ model defined by (16) have mixture distributions in which components are the distributions obtained for factors under each $\mathbb{P}_j$. The mixture proportions are the weights, w^j_t, used to combine the individual RN derivatives. Thus in terms of conditional densities

$$
f^{\mathbb{P}}\left(X_{t+1}|X_t\right) = \sum_{j=1}^{k} w^j_t f^j\left(X_{t+1}|X_t\right), \tag{17}
$$

for densities f^j obtained under $\mathbb{P}_j$. Then if each $\xi^j_{t,t+1}$ is defined such that conditional densities of variables under $\mathbb{P}_j$ are known in closed form, the $\mathbb{P}$ model is econometrically tractable using standard methods for finite mixture distributions.

A way to obtain component densities in closed form is to let the individual RN derivatives be Esscher transforms as suggested in the single-component case by Dai et al. (2006). Thus for N-vector functions $\Lambda^j_t = \Lambda^j(X_t)$ for which the expectation $E^{\mathbb{Q}}_t\left(e^{\Lambda^{j\prime}_t X_{t+1}}\right)$ exists, define the component RN derivatives as

$$
\xi^j_{t,t+1} = \frac{e^{\Lambda^{j\prime}_t X_{t+1}}}{E^{\mathbb{Q}}_t\left(e^{\Lambda^{j\prime}_t X_{t+1}}\right)}, \qquad j = 1, ..., k. \tag{18}
$$

This has $E^{\mathbb{Q}}_t\left[\xi^j_{t,t+1}\right] = 1$ by definition, and positivity follows if the exponent in the numerator is finite with probability one. The restrictions that each Λ^j_t must satisfy for this method to be well-defined are therefore

$$-\infty < \Lambda^j_{Zt} < 1/c \quad and \quad \left|\Lambda^j_{Yt}\right| < \infty, \qquad \mathbb{Q}-a.s., \tag{19}$$

where Λ^j_t has been separated into the parts that affect variance and non-variance factors, $\Lambda^{j\prime}_t = \left(\Lambda^j_{Zt}, \Lambda^{j}_{Yt}{}'\right)$.

The transformation (18) is convenient because the resulting component distributions are on the same form as those under $\mathbb{Q}$, but with some parameters varying dependent on the value of Λ^j_t. The moment generating function for the j'th component is

$$\begin{aligned}\phi^j_t(u) &= E^j_t\left[e^{u'X_{t+1}}\right] = E^{\mathbb{Q}}_t\left[\xi^j_{t,t+1}e^{u'X_{t+1}}\right] \\ &= e^{a\left(u+\Lambda^j_t\right)-a\left(\Lambda^j_t\right)+\left[b\left(u+\Lambda^j_t\right)-b\left(\Lambda^j_t\right)\right]'X_t},\end{aligned} \tag{20}$$

and this will only in special cases be affine, i.e., on the form (1). For the factor distributions in the $DA^{\mathbb{Q}}_M(N)$ model, though, it is possible even for general Λ^j_t functions to write (20) as that under $\mathbb{Q}$ with time-varying parameters. Let parameters in the risk-neutral model be $\Theta = \{\rho, \nu, c, \mu_0, \mu_Z, \mu_Y, \Sigma_Y, \alpha, \beta\}$, then manipulations of (20) show that

$$\phi^j_t(u) = e^{a\left(u;\Theta\left(\Lambda^j_t\right)\right)+b\left(u;\Theta\left(\Lambda^j_t\right)\right)'X_t}, \tag{21}$$

where a and b are the function in (4). The parameters in $\Theta\left(\Lambda^j_t\right)$ that depend on Λ^j_t are

$$\begin{aligned} c\left(\Lambda^j_t\right) &= c\cdot\left(1-\Lambda^j_{Zt}c\right)^{-1}, & \rho\left(\Lambda^j_t\right) &= \rho\cdot\left(1-\Lambda^j_{Zt}c\right)^{-2}, \\ \mu_0\left(\Lambda^j_t\right) &= \mu_0 + h_0\Lambda^j_{Yt}, & \mu_Z\left(\Lambda^j_t\right) &= \mu_Z + h_Z\Lambda^j_{Yt},\end{aligned} \tag{22}$$

while the remaining ones are unaffected. Writing the conditional factor density under $\mathbb{Q}$ as $f^{\mathbb{Q}}(X_{t+1}|X_t;\Theta)$, the mixture density under $\mathbb{P}$ can be written as

$$f^{\mathbb{P}}(X_{t+1}|X_t) = \sum_{j=1}^{k} w^j_t f^{\mathbb{Q}}\left(X_{t+1}|X_t;\Theta\left(\Lambda^j_t\right)\right), \tag{23}$$

with the parameters in each component that differ from their values under $\mathbb{Q}$ given by (22).

Consider the additional flexibility obtained in the $\mathbb{P}$ model with multiple components relative to the single-component case. The conditional distributions may depend nonlinearly on the factors as $\Lambda(X_t)$ can be any function that satisfies (19). Still, from the results (21) and (22), when $k=1$ the form of the conditional factor distributions under $\mathbb{P}$ will be the same as under $\mathbb{Q}$, i.e., normal for Y and non-central χ^2 for Z. Further, these distributions can differ from under $\mathbb{Q}$ only by one degree of freedom per factor, since conditional on X_t the shift is determined by the value of Λ^1_t which has dimension equal to the number of factors. Specifically, this implies that if the single-component model is to match a certain difference in the mean for each factor, then the difference between the two measures in higher moments is automatically fixed. This property is conditional on X_t and thus not dependent on how the $\Lambda(X_t)$ function is chosen.

In the multi-component model the conditional moment generating function under $\mathbb{P}$,

$$\phi_t^{\mathbb{P}}(u) = \sum_{j=1}^{k} w_t^j e^{a\left(u;\Theta\left(\Lambda_t^j\right)\right)+b\left(u;\Theta\left(\Lambda_t^j\right)\right)' X_t}, \tag{24}$$

cannot in general be reduced to the single exponential affine form (21), so now also the form and not only the parameters in the conditional distributions may change relative to $\mathbb{Q}$. The uncentered moments of factors under $\mathbb{P}$ are easily obtained from (24) as weighted averages of component uncentered moments,

$$E_t^{\mathbb{P}}\left[X_{t+1}^n\right] = \sum_{j=1}^{k} w_t^j E_t^{\mathbb{Q}}\left[X_{t+1}^n | \Theta\left(\Lambda_t^j\right)\right].$$

This illustrates that in the mixture model there is additional degrees of freedom for the distributions to differ between the two measures, since several moments of the conditional factor distribution can now differ independently. Thus a given first moment shift for each factor can be matched by a multitude of combinations of Λ_t^j's and there is then additional freedom to also match differences in higher moments.

This difference between $k = 1$ and $k > 1$ is especially clear for the conditional normally distributed variables Y_t, since, as seen from (22), only the mean is affected by a single-component Esscher transform. Thus

$$E_t^{\mathbb{Q}}\left[Y_{t+1} | \Theta\left(\Lambda_t^j\right)\right] = E_t^{\mathbb{Q}}\left[Y_{t+1}\right] + V_t^{\mathbb{Q}}\left[Y_{t+1}\right] \Lambda_{Yt}^j, \tag{25}$$

while the variance and all higher central moments remain unchanged. With multiple components other moments than the mean will differ from their values under $\mathbb{Q}$. For instance, the variance in the mixture equals

$$V_t^{\mathbb{P}}\left[X_{t+1}\right] = \sum_{j=1}^{k} w_t^j V_t^j\left[X_{t+1}\right] + \sum_{j=1}^{k} w_t^j \left(E_t^j\left[X_{t+1}\right] - E_t^{\mathbb{P}}\left[X_{t+1}\right]\right)\left(E_t^j\left[X_{t+1}\right] - E_t^{\mathbb{P}}\left[X_{t+1}\right]\right)', \tag{26}$$

i.e., the average component variance plus the average outer product of the distance between component means and the common mean. For the Y factors the component variances are the same but the means may differ in a variety of ways for the same first $\mathbb{P}$ moment, and thus differences between $V_t^{\mathbb{P}}(Y_{t+1})$ and $V_t^{\mathbb{Q}}(Y_{t+1})$ are possible in the multi-component model. That only the first moment of the conditional distribution is affected by the Esscher transform is a special property of the normally distributed variables that does not hold for the non-central chi-square variables, Z_t.[4] It still holds for this type of variable that in a single-component Esscher transform, the shift in all conditional moments is determined by the single variable Λ^j_{Zt}, and that this constraint is relaxed in the mixture formulation.

The ability in the discrete model to let higher moments differ between equivalent measures is tied to the modelling of variables over fixed intervals, $\Delta > 0$. For increasingly shorter intervals, $\Delta \to 0$, it also holds for the mixture model in the limit that only the expected rate of change for variables is affected by the change of measure. The effect of increasingly shorter time intervals on the Esscher transform used to get the components follows from results in Dai et al. (2006). Without loss of generality reformulate Λ^j_t in terms of another function $\mu^j(X_t)$ as

$$\Lambda^j(X_t) = (\Sigma S^c_t \Sigma')^{-1}\left(\mu^j(X_t) - \mu^{\mathbb{Q}}(X_t)\right), \tag{27}$$

where $\mu^{\mathbb{Q}}(X_t)$ is the drift and $\Sigma\sqrt{S^c_t}$ the volatility in the diffusion (6) that the $\mathbb{Q}$ model converges to as $\Delta \to 0$. Then a first order Taylor approximation of the mean and variance in the discrete model under $\mathbb{P}_j$ obtained with an Esscher transform defined by (27) gives

$$\begin{aligned} E^j_t(X_{t+\Delta}) &= X_t + \mu^j(X_t)\Delta + o(\Delta) \\ V^j_t(X_{t+\Delta}) &= \Sigma S^c_t \Sigma'\Delta + o(\Delta) = V^{\mathbb{Q}}_t(X_{t+\Delta}) + o(\Delta). \end{aligned} \tag{28}$$

This implies that the limiting continuous-time diffusion for the discrete model under $\mathbb{P}_j$ has drift $\mu^j(X_t)$ and volatility $\Sigma\sqrt{S^c_t}$. Thus in the limit the Esscher transform tends to the same effect as that achieved by a Girsanov transformation with kernel $\left(\Sigma\sqrt{S^c_t}\right)'\Lambda^j_t$ in the limiting diffusion model under $\mathbb{Q}$. As a result, Λ^j_t is approximately the market price of risk that gives the shift in mean per unit of variance,

$$E^j_t(X_{t+\Delta}) - E^{\mathbb{Q}}_t(X_{t+\Delta}) = V^j_t(X_{t+\Delta})\Lambda^j_t + o(\Delta). \tag{29}$$

Extending the results above to multiple components, a first order Taylor approximation to the mean and variance under $\mathbb{P}$ in the mixture model is found to give

$$\begin{aligned} E^{\mathbb{P}}_t(X_{t+\Delta}) &= X_t + \sum_{j=1}^{k} w^j_t \mu^j(X_t)\Delta + o(\Delta) \equiv X_t + \mu^{\mathbb{P}}(X_t)\Delta + o(\Delta) \\ V^{\mathbb{P}}_t(X_{t+\Delta}) &= \Sigma S^c_t \Sigma'\Delta + \sum_{j=1}^{k} w^j_t \left[\mu^j(X_t) - \mu^{\mathbb{P}}(X_t)\right]^2 \Delta^2 + o(\Delta) \\ &= \Sigma S^c_t \Sigma'\Delta + o(\Delta). \end{aligned} \tag{30}$$

Here the second equality for the variance assumes that the difference between the $\mu^j(X_t)$'s does not increase with Δ. Thus, it also holds for the mixture model as $\Delta \to 0$ that the effect

[4] The n'th centered moment under $\mathbb{Q}$ is $E^{\mathbb{Q}}_t\left(Z^n_{t+1}\right) = (n-1)!\nu c^n + n!c^{n-1}\rho Z_t$ and the Esscher transform make ρ and c depend on Z_t as given in (22).

of the measure to a first order approximation is only a change in the drift from $\mu^{\mathbb{Q}}(X_t)$ to the weighted average of the component drifts, $\mu^{\mathbb{P}}(X_t) = \sum_{j=1}^{k} w_t^j \mu^j(X_t)$. Consequently, the weighted average of Λ_t^j is approximately for small time intervals the market price of risk,

$$E_t^j(X_{t+\Delta}) - E_t^{\mathbb{Q}}(X_{t+\Delta}) = V_t^j(X_{t+\Delta}) \sum_{j=1}^{k} w_t^j \Lambda_t^j + o(\Delta),$$

measuring the change in mean per unit of variance.

That the measure change in the mixture model to a first order approximation for small time intervals only affects the mean requires that the difference $\mu^j - \mu^{\mathbb{P}}$ is not related to the interval length. For fixed interval length this difference may be large enough for the second term in (30) not to be negligible. This follows since $\mu^j - \mu^{\mathbb{P}}$ is not restricted by the wish to match a certain difference in the means and therefore may become large enough to matter for the variance at the interval length in the model. The variance in the single-component model, (28), may also have terms of order Δ^2, but the size of these cannot be changed without affecting the mean. Whether the possibility in the mixture model to let higher moments differ between the pricing model and the model for interest rate time series is relevant should therefore be determined empirically.

5 ML Estimation for Implicitly Observed Factors

The interest rate model is estimated from a data set of m zero-coupon bond yields with different maturities observed at fixed intervals. The period length in the discrete model is set equal to the observation frequency, such that yields are observed at times $t = 0, 1, ..., T$. Including an additive error term, ε_t^n, that captures the part of the observed yield unexplained by the model gives from (12) that observed yields are

$$y_t^n = \delta_0^n + \delta_X^{n\prime} X_t + \varepsilon_t^n. \tag{31}$$

Due to the error term in (31) the latent factors should generally be filtered out when the model is estimated. It is common to make the simplifying assumption that a number of yields equal to the number of factors are observed without error, such that factors can be calculated directly from the perfectly observed yields. This method is used in continuous-time models by Chen & Scott (1993) and Pearson & Sun (1994), and in discrete-time models by Dai et al. (2006). When factors under $\mathbb{P}$ follow a mixture distribution, this assumption allows the use of standard likelihood methods from mixture modelling such as the EM algorithm to estimate parameters. This section describes the estimation approach when factors are assumed implicitly observed from a subset of the yields.

To write down the likelihood function arrange yields into an N vector of those observed without error y_t^p and an $m-N$ vector of those with error y_t^e. Then factors are obtained from y_t^p using (31) without the error term to get

$$X_t = \left(\delta_X^p\right)^{-1}\left(y_t^p - \delta_0^p\right), \tag{32}$$

for δ_0^p the N-vector of δ_0^n's for maturities without error, and δ_X^p similarly the $N \times N$-matrix with rows $\delta_X^{n\prime}$ for maturities without error. Substituting the factors (32) in (31) for the remaining yields observed with error results in

$$y_t^e = \delta_0^e + \delta_X^e\left(\delta_X^p\right)^{-1}\left(y_t^p - \delta_0^p\right) + \varepsilon_t^e, \tag{33}$$

where the $m-N$-vector δ_0^e and the $(m-N)\times(m-N)$-matrix δ_X^e are obtained by collecting δ_0^n and $\delta_X^{n\prime}$ for maturities of yields observed with error. The error terms in y_t^e are assumed normal and independent across maturities, $\varepsilon_t^e \sim N\left(0, diag\left(\zeta_1^2, ..., \zeta_{m-N}^2\right)\right)$.

Collect all model parameters in the vector Ψ and write $y = \{y_t\}_{t=0}^T$ for yields. Then assuming y_0 is given, the likelihood function is

$$\mathcal{L}\left(\Psi|y\right) = f\left(y|\Psi\right) = \prod_{t=1}^T f\left(y_t|y_{t-1}, \ldots, y_0, \Psi\right) = \prod_{t=1}^T f\left(y_t^e|y_t^p, \Psi\right) f\left(y_t^p|y_{t-1}^p; \Psi\right),$$

where the last equation splits y_t into y_t^p and y_t^e, and uses that the Markovian structure for factors transfers to y_t^p. Use the change of variable formula to substitute y_t^p with factors given by (32) and get

$$\mathcal{L}\left(\Psi|y\right) = \prod_{t=1}^T f\left(y_t^e|y_t^p; \Psi\right) f^{\mathbb{P}}\left(X_t|X_{t-1}; \Psi\right)\left|\delta_X^p\left(\Psi\right)\right|^{-1}, \tag{34}$$

where superscript $\mathbb{P}$ indicates the physical factor dynamics. The risk-neutral dynamics of the factors, on the other hand, enter the likelihood function in the cross-sectional relation between yields through determination of vectors and matrices δ_0^p, δ_X^p, δ_0^e, and δ_X^e in the normal density for $y_t^e|y_t^p$, cf. (33). Finally, substitute the $\mathbb{P}$ density from the mixture model (17) and take logs to obtain the log-likelihood function

$$l\left(\Psi|y\right) = \sum_{t=1}^T\left\{\log f\left(y_t^e|y_t^p\right) + \log\sum_{j=1}^k w^j\left(X_{t-1}\right) f^j\left(X_t|X_{t-1}\right)\right\} - T\log\left|\delta_X^p\right|, \tag{35}$$

where dependence on the parameter vector Ψ is suppressed on the right hand side, and X must be calculated from (32).

Maximization of the log-likelihood function for mixture models can not be solved in close-form due to the log of a sum term that arises from the mixture. The classic solution to this in the literature is to use the expectation-maximization (EM) algorithm of Dempster, Laird & Rubin (1977). This method proceeds in two steps that are iterated and under weak regularity conditions monotone convergence to a local maximum is ensured. The dynamic structure in the suggested interest rate model is not standard in mixture models, but it is still possible to apply the EM algorithm to maximize (35). The interest rate model can be seen as missing the indicator variables $K_t = \left(K_t^1, ..., K_t^k\right)$ that determine to which component transition density X_t belongs, such that if $K_t^j = 1$ the density for X_t is $f^j\left(X_t|X_{t-1}\right)$. These variables are distributed $K_t|X_{t-1} \sim Mult_k\left(1, w_{t-1}\right)$ and K_t is independent of K_{t-1} when conditioning on X_{t-1}. The last property is important, since if the link between component indicators could not be broken by conditioning on factors, an extension of the algorithm would be necessary, such as the forward-backward algorithm used in hidden Markov models.

EM is especially useful when the M-step can be calculated in closed form, such as for normal component densities and constant membership probabilities. In these cases application of the algorithm avoids the use of standard numerical optimization routines which are then substituted with the iteration over two easy to calculate steps. The parameter estimates in component densities of the suggested interest rate model cannot be obtained in closed form and the weights are allowed to depend on the factors. Therefore the maximization step must be solved by numerical optimization. As the EM algorithm generally needs many iterations to converge, this numerical step will slow down the algorithm considerably. Indeed, in the empirical application later in the paper, it turns out that direct numerical optimization of the log-likelihood function (35) over the parameters Ψ is much faster than use of the EM algorithm, and therefore standard numerical optimization rutines are used to maximize the likelihood function instead.

There are a few special properties of the likelihood function for a model that includes a mixture element which should be considered. The likelihood function will generally have multiple local maxima. In models where components belong to the same family of distributions and differ only in parameter values, it is always possible to reorder component labels and obtain the same likelihood value. This label switching problem is solved by imposing some ordering of one of the parameters that differ over components. It is also possible that the likelihood function is unbounded. For instance in a mixture of normals with different mean and variance parameters in each component, the likelihood function goes to infinity if one component centers at a single observation and its variance goes to zero. Even in such cases regularity conditions satisfied by most parametric families ensure existence of a local maximum with the asymptotic properties equivalent to maximum likelihood estimators, see Redner & Walker (1984). Therefore when searching for parameter estimates, the procedure should be started several times from different initial points. Then after removing possible spurious solutions, the largest remaining local maximizer can be chosen as the maximum likelihood estimate.

It is desirable to be able to test for the number of components to include in the mixture distribution, and in particular whether a mixture is necessary at all. Removing one component density, i.e., going from k to $k-1$ components, can be achieved by appropriate parameter restrictions. Unfortunately regularity conditions are not met to test these restriction by a standard likelihood ration test. Reducing the number of components restricts the parameter vector to the boundary of the parameter space, or for components of the same parametric family to a non-identifiable subset of the parameter space. This implies that the LR statistic is not asymptotically chi-square distributed as in the standard setting. Instead models with different numbers of components can be compared using Akaike's information criterion and the Bayesian information criterion. For a model with d parameters and T observations AIC is $-2\log L\left(\hat{\Psi}\right)+2d$, while the BIC is $-2\log L\left(\hat{\Psi}\right)+d\log T$, and the model with the lowest statistic is preferred. The statistics measure lack of fit as twice the negative log-likelihood function and penalize models for their complexity as measured by the number of parameters. The BIC requires a larger gain in likelihood value to include an extra parameter when more data is observed, and thus it penalizes complex models stronger than AIC for all reasonably sized samples.

6 Three-Factor Models

It is standard in term structure literature that three factors are needed to adequately describe the cross section of interest rates. In the affine framework of Dai & Singleton (2000) a three-factor model with one factor affecting variance is often the preferred model. I will therefore focus on the discrete analog, the $DA_1^{\mathbb{Q}}(3)$ model, and estimate it together with a number of different specifications for the Radon-Nikodym derivative, $\xi_{t,t+1}$, to see whether the inclusion of a mixture element in the measure change for this model is justified empirically. Though the mixture extension is not related to the specific form of the market price of risk function Λ_t, I consider for comparability reasons both the effect of restricting this function to get linear $\mathbb{P}$ components and of allowing for a cubic formulation of Λ_Z, as suggested in Dai et al. (2006).

The risk-neutral part of the interest rate model to be estimated is presented in (2), (3) and (8). Thus the $DA_1^{\mathbb{Q}}(3)$ model has without restrictions 23 free parameters given by the set $\{\delta_0, \delta_Z, \delta_Y, \rho, \nu, c, \mu_0, \mu_Z, \mu_Y, \Sigma_Y, \alpha, \beta\}$. The restrictions imposed by definition on parameters ρ, ν, and c already ensure that Z_t is a well-defined autoregressive-gamma process that stays non-negative. To further avoid that Z_t can reach its lower bound of zero, the condition $\nu > 1$ is imposed. This sets $\Pr^{\mathbb{Q}}\{Z_{t+1}=0|Z_t\}=0$ for any conditioning Z_t, since the degrees of freedom, 2ν, in the non-central chi-square distribution then is larger than two. That Ω_{Yt} is positive definite can be obtained by imposing that $\alpha_i+\beta_i>0$ for all i in addition to the already given conditions, $\alpha \geq 0$, $\beta \geq 0$, and Σ_Y non-singular. There is no problem in either α_i or β_i being zero as long as this is not the case for both with the same index, in which case the corresponding error term would not enter the model. Further, mean-reverting Y factors is obtained by requiring that $eig\left(\mu_Y\right) < 1$. This already holds for Z, since $\rho < 1$ by definition.

Analogous to the continuous-time affine models it is necessary to set some of the 23 parameters constant to get an identified model. This is an issue, since factors in the linear model can be rotated without affecting the model for the one-period rate and thus without changing pricing implications. In the discrete model the main transformations are a rescaling of Z with a positive constant l, an affine transformation $LY + v$ of the Y factors, and a reparametrization of Y-factor variance using the transformation D. The effect of these transformations on the parameter vector is given in the appendix. Different identifying restrictions can be chosen to avoid that such transformations are possible. It is natural to consider the restrictions that would place the approximating continuous-time $A_1^{\mathbb{Q}}(3)$ model in its canonical form. For the parametrization (6)-(7) of the $A_1^{\mathbb{Q}}(3)$ model the canonical form imposes the identifying restrictions $\sigma_Z = 1$, $\theta_Y = 0$, $\Sigma = I$, $\alpha^c = \iota$, and $\delta_Y \geq 0$. Therefore by the relation (5), if the restrictions

$$c = \Delta/2, \quad \mu_0 = -\mu_Z \frac{\nu\Delta/2}{1-\rho}, \quad \alpha = \Delta\iota, \quad \Sigma_Y = I, \quad \delta_Y \geq 0 \tag{36}$$

are imposed in the discrete model, the approximating $A_1^{\mathbb{Q}}(3)$ model will be on canonical form. This sets 9 parameters constant and the risk-neutral model therefore has 14 free parameters left. The sign restrictions on some of the free parameters,

$$0 < \rho < 1, \quad v > 1, \quad eig\left(\mu_Y\right) < 1, \quad \beta \geq 0, \tag{37}$$

also ensure that admissibility restrictions in the continuous-time model are satisfied. In the appendix it is argued that the restrictions in (36) are indeed identifying in the discrete model, and that they are unrestrictive for any model with all α_i strictly positive. This follows since the conditions can be imposed in any discrete model, with $\alpha > 0$, via transformations by l, L, v, and D, and that no further transformations are possible when the restrictions are imposed.

6.1 Empirical Identification Issues

The canonical identification restrictions (36) can be imposed in any model with all α strictly positive. For the yield data used to estimate the 3-factor model, this turns out to be a problematic assumption. When the model is estimated with canonical identifying restrictions, this results in a Y_2 factor with very large scale, such that in the variance term $\alpha_2 + \beta_2 Z_t$, β_2 is of the order $10^7 - 10^9$, while the scale-identifying restriction for Y_2 is $\alpha_2 = \Delta = 1/52$. For such a difference in magnitude α_2 is in practical terms equal to zero, and setting α_2 equal to zero in the model removes theoretical identification of the Y_2 scale. The effect in the estimated model is that $L = diag\,(1, L_{22})$ rotations can be performed that change parameters δ_2, μ_{2Z}, μ_{12}, μ_{21}, and β_2 while α_2 is kept at its fixed value with very limited impact on the likelihood value. Thus the likelihood is flat in direction of L_{22} rotations, and the standard errors of estimates for affected parameters will be very large. When data requires an α_2 close to zero and this parameter is fixed for identification, a small constant term in the Y_2 variance is instead achieved by scaling up Y_2. Equivalently, if the free parameter estimate of α_2 is close to zero, then the transformations needed to get to (36) include an L rotation with $L_{22} = \Delta/\alpha_2$, which scales Y_2 up by a large number. This issue is not specific to discrete-time affine models, but would also arise in continuous-time models if data requires one of the non-variance factors to have no constant term in its volatility.

The issue can be solved by imposing another scale-identifying restriction for Y_2 than setting α_2 constant. Any other free parameter affected by L_{22}, i.e., δ_2, μ_{2Z}, μ_{12}, μ_{21}, and β_2, can be chosen. In general, any of these may be zero for some data, so there is no universal choice. In the specific case where a restriction of α_2 gives problems, the natural choice is to set β_2 constant instead, since for $\alpha_2 = 0$ it must hold that $\beta_2 > 0$. I choose $\beta_2 = \Delta$, such that $\beta_2^c = 1$ in the continuous approximation. The rotation that achieves $\beta_2 = \Delta$ is $L_{22} = \sqrt{\Delta/\beta_2}$ and relative to (36) this makes α_2 a free parameter, such that it can go to zero if the data requires this. Thus the 14 free parameters to be estimated in the risk-neutral model, including sign restrictions, are

$$\delta_0, \quad \delta_Z, \quad \delta_Y \geq 0, \quad 0 < \rho < 1, \quad \nu > 1, \quad \mu_Z, \quad eig\,(\mu_Y) < 1, \quad \alpha_2 \geq 0, \quad \beta_1 \geq 0. \tag{38}$$

6.2 Radon-Nikodym Derivative Specifications

A number of different specifications for the time-series dynamics of yields are estimated. These differ on their degree of mixture in the RN derivative and in the formulation of the market price of risk function. Besides a single-component model, I estimate models with two components, $k = 2$ in (16), that either have constant weights, $w_t = w$, or weights that depend on the factors. The relation between weights and factors is chosen to be a logistic function

$$w\,(X_t) = \left(1 + \exp\left(-\eta_0 - \eta_X' X_t\right)\right)^{-1}, \tag{39}$$

for η_0 scalar and η_X a 3×1 vector. The function is convenient since its range is $(0, 1)$ and it is analogous to a logistic regression for the unobserved variables that indicate to which component of the mixture each data point belongs.

For each degrees of mixture two types of market price of risk functions are investigated. The linear specification implies that component models under $\mathbb{P}$ are affine with constant parameters, so $\Theta\left(\Lambda_t^j\right) = \Theta^j$ in (21) and (22).[5] The other specification allows for cubic terms in the Z factor, similar to the model preferred by Dai et al. (2006), and thus results in some parameters of the $\mathbb{P}$ components varying over time. The cubic model sets the market price of risk function for Z in each component to

$$\Lambda_{Zt}^j = Z_t^{-1}\left(\lambda_{Z0}^j + \lambda_{Z1}^j Z_t + \lambda_{Z2}^j Z_t^2 + \lambda_{Z3}^j Z_t^3\right), \tag{40}$$

for scalars λ_{Zi}^j, $i = 0, ..., 3$. The impact of Λ_{Zt}^j on c and ρ is given in (22), and from (4) it follows that the component conditional distribution will only be affine if Λ_Z is constant. Thus the linear specification restricts λ_{Z0}^j, λ_{Z2}^j, and λ_{Z3}^j to zero. For all estimated models the market price of risk functions for the Y factors in each component are defined as

$$\Lambda_{Yt}^j = S_{Yt}^{-1}\left(\lambda_{Y0}^j + \lambda_{YZ}^j Z_t + \lambda_{YY}^j X_t\right), \tag{41}$$

for λ_{Y0}^j and λ_{YZ}^j both 2×1 vectors and λ_{YY}^j a 2×2 matrix. For general Λ_Y^j the μ_0 and μ_Z parameters in the j'th component become time varying under $\mathbb{P}$ as given in (22). The specific formulation in (41) affects the a and b functions in (4), see appendix, such that the j'th component remains an affine model with constant parameters,

$$\mu_0^{\mathbb{P}_j} = \mu_0 + \lambda_{Y0}^j, \qquad \mu_Z^{\mathbb{P}_j} = \mu_Z + \lambda_{YZ}^j, \qquad \mu_Y^{\mathbb{P}_j} = \mu_Y + \lambda_{YY}^j.$$

6.2.1 Restrictions for well-defined Λ_t

Conditions on the component functions Λ_t^j for the RN derivative to be well-defined given in (19) are now translated to conditions on parameters in specifications (40) and (41). To ensure $\Lambda_{Zt}^j < 1/c$, it must hold that $\lambda_{Z0}^j \leq 0$ and $\lambda_{Z3}^j \leq 0$ in the cubic model, since Λ_{Zt}^j otherwise increases without bound for $Z \to 0$ or $Z \to \infty$.[6] In the linear model $\Lambda_{Zt}^j = \lambda_{Z1}^j$, so the requirement is just $\lambda_{Z1}^j < 1/c$. It is automatically avoided that Z can reach its lower boundary of zero in the component models under $\mathbb{P}$, since $\nu > 1$ is unchanged by the Esscher transform, cf. (22). That Λ_{Zt}^j is finite $\mathbb{Q}$-a.s. then follows from $0 < Z < \infty$ with probability one under $\mathbb{Q}$. Λ_{Yt} is finite as well since Y is mean reverting under $\mathbb{Q}$ due to $eig\left(\mu_Y\right) < 1$.

It is desirable to put structure on the model under $\mathbb{P}$ such that variables do not become explosive. This can be done component-wise in the mixture model. The Y factors have constant parameters in each component under $\mathbb{P}$, so Y is mean reverting when $eig\left(\mu_Y^j\right) < 1$. For Z in the linear models this is achieved by $\rho^{\mathbb{P}_j} < 1$ which requires $\lambda_{Z1}^j < \left(1 - \sqrt{\rho}\right)/c$. In the cubic models $\rho^{\mathbb{P}_j}$ is time varying, so the requirement is instead that this parameter is less than one for all large values of Z, i.e., that $\exists K > 0 : \rho\left(\Lambda_Z^j\right) < 1,\ \forall Z > K$. This is

[5] This definition of a linear Λ_t is slightly different from what is called a linear Λ_t by Dai et al. (2006), who instead define this in terms of the impact on the drift in the approximating $A_M\left(N\right)$ model.

[6] Parameters may also be such that the polynomial Λ_{Zt}^j exceeds $1/c$ in the interior, but this is less likely as c is small, and it is therefore checked for each model individually.

satisfied when $\lambda^j_{Z3} < 0$.

By the limiting arguments in (27) and (28) a continuous-time diffusion approximating the dynamics of factors in the j'th $\mathbb{P}$ component can be found to be on the form $dX_t = \mu^{\mathbb{P}_j}(X_t)\,dt + \Sigma\sqrt{S_t}dW_t^{\mathbb{P}_j}$. Here the drift for the Z factor is equal to

$$\mu_Z^{\mathbb{P}_j}(X_t) = \kappa_{ZZ}(\theta_Z - Z_t) + \left(\lambda^j_{Z0} + \lambda^j_{Z1}Z_t + \lambda^j_{Z2}Z_t^2 + \lambda^j_{Z3}Z_t^3\right), \tag{42}$$

with κ_{ZZ} and θ_Z from the approximation under $\mathbb{Q}$ given by (5). Although boundary non-attainment for Z_t is automatically satisfied under $\mathbb{P}$ in the discrete models, since ν in the components is unchanged by the Esscher transform, this property need not be satisfied in the approximating continuous model. Boundary non-attainment for drift (42) in the continuous model requires that $2\left(\kappa_{ZZ}\theta_Z + \lambda^j_{Z0}\right) > 1$ in the parametrization with $\Sigma = I$, such that the drift is positive and large enough as $Z \to 0$. Since $\nu = 2\kappa_{ZZ}\theta_Z$ from (5), the condition in terms of the discrete parameters is that $\lambda^j_{Z0} > 1/2 - \nu$. This need not be satisfied in the discrete model for restrictions already imposed, though it may be, since the interval where λ^j_{Z0} satisfies both conditions, $(1/2 - \nu, 0]$, is non-empty for $\nu > 1$. The reason that the discrete model can be well-defined without $\lambda^j_{Z0} < 1/2 - \nu$ is that the continuous model approximation becomes worse as $Z \to 0$.

In estimation of the discrete model I choose not to impose restrictions for boundary non-attainment of Z in the approximating continuous model. I focus on the discrete model and impose only restrictions to ensure that this is well-defined. Dai et al. (2006) present their estimation results of the discrete model with focus on results for the corresponding continuous-time model and they therefore impose $\lambda_{Z0} < 1/2 - \nu$ on the cubic model. They focus less on the discrete model in the background and thus do not impose $\Lambda_{Zt} < 1/c$, as seen from $\lambda_{Z0} > 0$ in some of their estimated models. In the mapping to the continuous model $c = \sigma_Z^2\Delta/2$, and thus for $\Delta \to 0$ the constraint reduces to $\Lambda_{Zt} < \infty$ almost surely. Therefore one can arrive at a well-defined continuous model, although the estimated discrete model has $\Lambda_{Zt} > 1/c$ for some set of Z with positive probability.

6.3 Results

The models are estimated using constant maturity treasury rates from the Federal Reserve Board's database on Selected Interest Rates (H.15). The constant maturity rates in this data set are based on a yield curve estimated each period from on-the-run treasury securities with a cubic spline model. I use weekly observations by selecting all Wednesday yields from the daily data with maturities 3 and 6 months and 1, 2, 3, 5, 7, and 10 years. The sample period is 1988 : 1 to 2009 : 1, giving a data set of $8 \cdot 1097$ yields. For the estimation method described in Section 5, I assume that the yields with maturities 3 month, 2 and 7 years are assumed observed without error, while the remaining ones include a Gaussian error term, cf.

(33). Direct numerical optimization of the log-likelihood function (35) turns out to converge and be much faster than using the EM algorithm. Standard errors are estimated using the negative inverse of the Hessian at maximum likelihood estimates.

The 3-factor models with parameters given in (38)-(41) are over-parameterized, so I follow Dai et al. (2006) and reduce the maximal models to preferred parsimonious models by removing insignificant parameters. Table 1 shows the results of this procedure. For instance, 9 parameters were removed from the maximal single-component linear model at a loss of 2.75 log-likelihood points. The method used for this elimination was to successively remove the parameter that resulted in the least drop in likelihood value as long as the corresponding likelihood-ratio statistic was above 5%. In a couple of instances a parameter with p-value slightly below 5% was removed when this allowed for subsequent removal of a parameter with high p-value. In all models a considerable number of parameters were removed at low cost in likelihood value. Table 1 shows the joint LR statistics of the restrictions imposed by the preferred models and these do in all cases have p-values above 40%.

The likelihood ratio statistics of restrictions imposed in the preferred linear models relative to the preferred cubic models are shown in Table 1. For each degree of mixture the linear and cubic preferred models are nested and only differ by the preferred cubic models including 2-3 of the λ_{0Z}^j, λ_{Z2}^j and λ_{Z3}^j parameters not allowed in linear models. In the single-component model the cubic elements are insignificant and should have been removed in the reduction to the preferred model, but where kept for comparison. In the two-component models the cubic terms are very significant with p-values of the joint LR statistic close to zero.

The performance of the two-component models relative to that of the single-component models is compared using Akaike's information criterion and the Bayesian information criterion also shown in Table 1. Formulating the RN derivative as a weighted sum of two Esscher transforms, such that the $\mathbb{P}$ distribution of factors and interest rates become a mixture gives a considerable increase in the log-likelihood value. When accounting for the increased complexity of models by using either information criteria, the models with two components are preferred. For instance, between the linear single-component and the linear constant mixture models there is a difference in the two information criteria of respectively 165.4 and 145.4 points. Among all models the time-varying cubic mixture model is the preferred model in terms of both the AIC and BIC statistics.

The mixture models with constant weights impose the restriction $\eta_X = 0$ relative to the models with time-varying weights. Testing this restriction is somewhat complicated by the preferred models not being nested since λ_{1Z}^2 is removed in the preferred time-varying mixture models, but not in the preferred constant mixture models. The models can instead be tested up against an encompassing model that includes all parameters in either model. λ_{1Z}^2 is trivially insignificant in this encompassing model since it is the only addition relative to the time-varying preferred model, where it was removed. The constant mixture models set the 4 parameters $\{\lambda_{10}^2, \eta_X\}$ to zero relative to the encompassing model, and this gives LR statistics of 23.6 and 24.2 in the cubic and linear models, respectively, which in both cases have p-values below 10^{-4}. There is thus evidence in favor of letting weights depend on factors in the two-component models, although the gain in model fit this provides is less than that of including an extra component relative to a single, as seen from the information criteria.

6.3.1 Risk-neutral Models

Estimates of free parameters in the risk-neutral part of the preferred models, (38), are shown in Table 2 together with standard errors estimated via the numerical Hessian at maximum likelihood estimates. Of the 14 free parameters δ_2 and α_2 are set to zero in all preferred models, while β_1 is also removed in all two-component models. Thus the second Y factor does not affect the one-period rate directly, but only through its effect on Y_1 via μ_{12}, while the first Y factor has constant variance under $\mathbb{Q}$ in all mixture models. That α_2 is insignificant and removed in all preferred models is the reason that the canonical identification restrictions do not work for the given data, since models with $\alpha_2 = 0$ cannot be rotated to satisfy (36).[7] Parameter values in the approximating continuous-time $A_1^{\mathbb{Q}}(3)$ models are calculated by (5) and shown in Table 3 with standard errors obtained by the delta method.

The speed of mean reversion for each factor is determined by ρ, μ_{11}, and μ_{22}. These parameters are close to one, so Table 2 shows $1-\rho$, $1-\mu_{11}$, and $1-\mu_{22}$ instead for convenience. The same pattern emerges in all estimated risk-neutral models. The Z factor reverses faster to its mean than the Y factors of which Y_1 is slightly more persistent than Y_2. The loading functions, δ_X^n, giving the impact of each factor on yields with different maturities, are shown in Figure 1 for the linear constant mixture model and the shape is similar in the other models. Y_1 is the factor with most even impact over maturities. The impact of Y_2 is increasing from zero for the one-period rate, $\delta_2 = 0$, to be strongest (negatively) for yields with long time to maturity. Z impacts yields with about 2 years to maturity most (negatively) and short and longer yields less or in the opposite direction. These observations give the natural interpretation of Y_1, Y_2, and Z as level-, (negative) slope- and (negative) curvature factors. Their successively increasing speed of mean reversion is consistent with standard findings. Comparing models, the difference between persistence of the Y factors is smallest in the time-varying mixture models and largest in the constant mixture models. The time series of implied factors from the yields observed without error obtained by (32) are drawn in Figure 2 for the linear constant mixture model. Abstracting

[7] δ_2 identifies the sign of Y_2, so when it is set to zero this becomes undetermined. Thus the estimated models could be transformed with $L_{22} = -1$ without any real effects. This might be solved by assuming that $\mu_{21} \leq 0$.

from possible differences by the change to $\mathbb{P}$, the graphs further illustrate the differences in the persistence of the three factors.

The parameters that govern the relation between factors, μ_{1Z}, μ_{12}, μ_{2Z}, and μ_{21}, are all estimated to be negative. Given the above mentioned interpretation of factors as level, (negative) slope, and (negative) curvature the parameter estimates have the following implications. More positive curvature or equivalently high medium term yields (Z low) tend to push the general level of yields higher and make the curve less steep. A high general level of yields (Y_1 high) will tend to tilt the curve steeper, a relation which also goes the other way, such that a steeper curve (Y_2 low) generally tends to push the yield level higher.

The estimated risk-neutral models are fairly similar across different specifications of the $\mathbb{P}$ model for interest rates. There are some notable differences between the single-component and mixture models mainly related to the variance of the level factor no longer increasing with Z under $\mathbb{Q}$ in the mixture models, i.e., that β_1 becomes insignificant. The explanation for this is discussed after the presentation of estimates for the $\mathbb{P}$ models.

6.3.2 Single-Component Model

Table 4 displays estimates of parameters in the Radon-Nikodym derivatives for the preferred models of the 6 different specifications that are investigated. Those of the implied $\mathbb{P}$ parameter that are changed relative to their values under $\mathbb{Q}$ are shown in Table 5. Omitted are the parameters in the cubic models that become time varying under $\mathbb{P}$. These parameters are instead illustrated in Figure 3.

Starting with the single-component linear model, the preferred model only keeps the two market price of risk parameters that impact how the level factor, Y_1, is affected by the other two factors, λ_{1Z} and λ_{12}. Thus the negative impact of Z on Y_1 is increased under $\mathbb{P}$, while the impact of the slope factor, Y_2, is strongly reduced. Since the unconditional mean of Z is positive, $\theta_Z = 5.1$, the increase in μ_{1Z} under $\mathbb{P}$ causes the unconditional mean of Y_1 to fall. Through the negative μ_{21}, it also causes the mean of Y_2 to increase. Thus $\theta_Y^{\mathbb{P}} = (-24.3, 7.3)'$, which was zero under $\mathbb{Q}$. This is the standard effect that the level of interest rates are expected to be higher under the pricing measure, or equivalently that bond prices are expected to be lower, to compensate for risk.

The effect of the measure change on the mean of the one-period rate is illustrated in the first row of Figure 4. This figure shows the expected change of the short rate over the next period under $\mathbb{Q}$ and $\mathbb{P}$ as a function of each of the factors. As a function of Y_1 there is a parallel shift downward from $\mathbb{Q}$ to $\mathbb{P}$, whereas the increase in μ_{1Z} causes the graph to tilt steeper as a function of Z. Since the slope factor only affects the short rate through its effect on the level factor, $\delta_2 = 0$, the expected change as a function of this factor almost become zero due to the fall in μ_{12}.

When the single-component model allows for a cubic specification in Λ_Z, the estimate of this function is illustrated in Figure 3 together with implied values of c and ρ as a function of Z. The leading coefficient for high Z, λ_{Z3}, is estimated to be negative, consistent with findings in Dai et al. (2006), thus inducing larger mean reversion for large Z under $\mathbb{P}$ than in the linear model. The Λ_Z function is also negative for small Z, implying a less positive expected change in Z up towards the unconditional mean, $E^{\mathbb{P}}(Z)$. The shape of $E_t^{\mathbb{P}}(Z_{t+1})$ as a function of Z_t is shown in Figure 5. This figure also illustrates that $E_t^{\mathbb{P}}(Z_{t+1})$ would be positive and induce stronger mean reversion for small Z if λ_{Z0} was estimated without restriction, i.e., an "S on the side" shape is obtained. Though, to avoid that Λ_Z increases above $1/c$ as Z goes to zero, it must be imposed that $\lambda_{Z0} \leq 0$. In terms of the expected change in the one-period rate, illustrated in first row of Figure 4, the cubic model produces results similar relative to those of the linear model, and $E_t^{\mathbb{P}}(r_{t+1}) - r_t$ can only just be discerned to be slightly curved as a function of Z_t, the dash-dot line in the upper left panel.

There may be different reasons why a cubic element in Λ_Z turns out to be less significant in this study than in Dai et al. (2006). It is possible that a longer time scale in the discrete model, 1 month instead of 1 week, increases the need for nonlinearities in the expected change of the one-period rate under $\mathbb{P}$. Such nonlinearity may also mainly be necessary as a function of the level, whereas the Z factor that allows for nonlinearities comes out as the curvature factor in the present study. Other models yet to be presented do find clear nonlinear patterns in the expected r change as a function of both the level and the curvature factor at the weekly timescale. The main reason that the cubic elements are less significnt seems instead to be that I impose $\lambda_{Z0} \leq 0$ to get a well-defined Λ_Z in the discrete model, which is a binding restriction. Had $\lambda_{Z0} > 0$ been allowed, the LR statistic for the restrictions imposed by the linear model would have been 9.5 instead with a p-value of 5.1%. Thus the cubic element would be substantially more significant in the single-component model if it was also allowed to increase the mean reversion for low Z through a positive λ_{Z0}.

6.3.3 Constant Mixture Model

Models with a RN derivative that imply a two-component mixture distribution with constant weights under $\mathbb{P}$ estimate the weight on the second component at respectively 2.2% and 2.5% in the linear and cubic models, see Table 4. Due to the high weight on the first component in the mixture its distribution must be quite similar to the $\mathbb{P}$ distribution in the single-component models. Thus the preferred constant mixture models also include a positive λ_{12}^1, which reduces $\mu_{12}^{\mathbb{P}_1}$ almost to zero and makes the expected change in the short rate almost independent of the second Y factor, see the left panel of the middle row in Figure 4. The level of rates is again lowered relative to $\mathbb{Q}$ by the measure change, but now with λ_{10}^1 instead of λ_{1Z} in the single-component model. The effect is still to reduce the level and increase the negative slope factor, $\theta_Y^{\mathbb{P}_1} = (-11.1, 5.5)'$, but $E_t^{\mathbb{Q}}(r_{t+1}) - r_t$ is not tilted as a function of Z, compare the left panels in the first two rows of Figure 4.

The shift in distribution from $\mathbb{Q}$ to the second component under $\mathbb{P}$ determined by Λ_t^2 is much larger than that to the first component, but to a large extend still in the same direction. The level of rates is lowered relative to $\mathbb{Q}$, but much stronger and with λ_{1Z}^2. Thus the tilt relative to Z, which was seen in the single-component models, but not in the main component of the mixture, is delegated to the secondary component. Changes in the Z variable from $\mathbb{Q}$ to $\mathbb{P}_2$ shift c and ρ down in the linear model implying stronger mean reversion, lower level and lower variance as well. The estimated cubic function, see second row of Figure 3, has the same shape as that in the cubic single-component model, but again with greatly magnified effect due to the low weight on this component. As in the single-component case the condition $\lambda_{Z0} \leq 0$ imposed to avoid $\Lambda_Z^2 > 1/c$ is binding. The second component mainly differs from the first or the single-component model in relation to changes for the second Y factor. Thus mean reversion for this factor is increased under $\mathbb{P}_2$ by a negative λ_{22}^2, an effect not seen elsewhere, and the impact of this factor on Y_1, and thus the one-period rate, is increased by a negative λ_{12}^2, which is opposite to that in the first component. This seems to allow $\mu_{12}^{\mathbb{P}_1}$ to be even lower in the first component relative to the single-component model, but with approximately the same combined effect, compare the mean graphs for the one-period rate as a function of Y_2 under $\mathbb{P}$. In general the estimated first moment of the one-period rate as a function of the different factors, row one and two of Figure 4, is not very different in the constant-weight mixture and single-component models. The main differences between these instead occur in higher moments to be discussed later.

6.3.4 Time-varying Mixture Model

When weights on the two Esscher transforms that define the measure change to $\mathbb{P}$ are allowed to depend on factors, this gives estimates of coefficients in the logistic weight function shown in Table 4. Figure 6 plots $w(X_t)$ as a function of each factor with the other two fixed at their unconditional $\mathbb{P}$ means. At $X_t = E^{\mathbb{P}}(X)$ the weights on the second component are 1.6% and 1.8% in each of the linear and cubic models. The weight on the second component is higher when medium term yields are comparably low, Z high, when the general level of rates is high, Y_1 high, and when the curve is steeper, Y_2 low. Thus for instance in the cubic model the weight increases to 6.7% or 5.4% when Z or Y_1, respectively, is one standard deviation above its mean. When Y_2 is one standard deviation below its mean the weight increases to 3.1%. The combined effect is much stronger. If the (Z, Y_1, Y_2) factors simultaneously are $(1, 1, -1)$ standard deviation from their means, the weight on the second component would be 27.1%. The numbers for the linear model are similar, but slightly lower.

As it can be seen from Table 4, parameters of the estimated market price of risk functions for each of the two components, $\Lambda^j(X_t)$, are very similar to those in the constant mixture models. The main difference is that λ_{10}^2 is included in the preferred models instead of λ_{1Z}^2, such that the negative shift in yield level relative to $\mathbb{Q}$ is constant with respect to Z in both components. The reason for this is that a steeper relation under $\mathbb{P}$ between Z and the mean change in the one-period rate instead is captured by the increase in weight on the second component as Z becomes large. Thus looking at the first panel in the bottom row of Figure 4, $E_t^{\mathbb{P}}(r_{t+1}) - r_t$ as a function of Z obtains a clear nonlinear pattern with higher negative changes for large Z. The single-component model includes a tilt via λ_{1Z} to partly capture this, and the constant mixture has a negative λ_{1Z}^2 in the secondary component. Also the negative estimate of λ_{Z3} in the cubic representation of these models implies effects in this direction, but gives nonlinear effects that are barely discernible in the leftmost panels of the top and middle rows of Figure 4. The regularity condition $\lambda_{Z0} \leq 0$ is not binding in the model with time-varying weights, although it is binding in the single-component model and in the second component of the constant mixture model. This is due to the factor dependent weight on the second component being almost zero for small Z where a positive λ_{Z0} would have been advantageous.

The time-varying mixture also generates a clear nonlinear pattern in the expected one-period rate change as a function of the level factor. Thus the short rate is expected to move much more negatively when the yield level in general is high, an effect not seen in any of the other models. Since the nonlinearities that show up in the mean-change functions are due to the dependence of weights on factors and not due to the form of the component market price of risk functions, there are only small differences between the linear and cubic formulations of $\Lambda^j(X_t)$.

The time series of weights calculated from estimated factors in the model are shown in Figure 7 together with the factor time series. The fast moving Z factor capturing changes in medium term rates drives most of the variation in the weights. The slow moving level factor determines the base level for weights around which they vary mainly due to Z. The slope factor moves medium fast and has lower impact on weights than the two other factors as seen in Figure 6. It is interesting to notice how weights on the second component tend to rise when the general level of yields is about to fall. Examples are found at the end of 1990, the beginning of 2001, and the period from late 2007 to early 2008.

6.3.5 Changes in Conditional Variance

Aside from the nonlinearities in the time-varying mixture models there are only relatively small differences in expected short rate change between mixture and single-component models, and under $\mathbb{Q}$ the mean functions are virtually identical in all models. This is different with respect to the second moment of the one-period rate. Figure 8 displays, in a fashion similar to that of Figure 4, the conditional standard deviation of the one-period rate in each model under $\mathbb{Q}$ and $\mathbb{P}$ as a function of each factor, while the other two are at their unconditional $\mathbb{P}$ mean, i.e., the function $f(x) = Var^{\upsilon}\left[r_{t+1}|X_{it} = x, X_{j\neq i,t} = E^{\mathbb{P}}(X_{j\neq i,t})\right]^{1/2}$. As seen in the first row of this figure, the conditional standard deviation for the single-component models barely changes between measures. The conditional variance of the Gaussian factors is unaffected for given Z_t by the Esscher transform, while it changes slightly for the autoregressive-gamma factor when $\Lambda_Z \neq 0$. Thus in the cubic model the change in standard deviation is less than 0.3% for all conditioning values of Z_t in the displayed range.

For the mixture models, in contrast, the conditional variance is estimated to be quite different under the two measures. This is possible as seen from (26), since the different means in component densities under $\mathbb{P}$ increase the variance of the mixture distribution. Though this effect by (30) goes to zero with the square of the discrete model period length, it is evident from Figure 8 that the component means can be sufficiently different for this effect to be large at weekly interval lengths. The possibility of a pronounced difference in the conditional second moment in the mixture models between measures implies that estimated values differ from those in the single-component models not only under $\mathbb{P}$ but also under $\mathbb{Q}$. The risk-neutral part of the mixture models thus estimates both a lower level for the conditional variance of the one-period rate and a lower dependence of this on the variance factor, Z, than in the single-component models. The higher level and stronger dependence on Z, akin to that in the single-component models, instead occur only under $\mathbb{P}$ in the mixture models. In addition, the constant mixture has conditional variance under $\mathbb{P}$ increasing as a function of the slope factor, due to the different values of μ_{12} and μ_{22} in the two components.

In the time-varying mixture models conditional standard deviation under $\mathbb{P}$ increases faster than linearly as a function of Z and Y_1 over the displayed range. This is due to the increasing weight on the second component for large values of Z and Y_1 in these models. This result for the time series of interest rates is qualitatively similar to results of both Chan et al. (1992) and Aït-Sahalia (1996*a*). These papers find that the volatility of the short rate increases faster than linearly with the short rate, and that a non-parametric estimator of the short rate diffusion increases fast in a range above the mean of the short rate.

Going back to the estimated parameters in the pricing models shown in Table 2, the main differences between mixture and single-component models can be explained by the difference in level and dependence on Z of the conditional variance under $\mathbb{Q}$. Thus β_1, the impact of Z on variance of the level factor, is included in the single-component models, but becomes insignificant in the mixture models. The reason must be that this parameter also describes variance under $\mathbb{P}$ in the single-component models, which for the mixtures is also determined by differences in the component means. When β_1 is removed from the mixture models, the constant term of the level factor's variance, α_1, needs to increase somewhat to compensate, although conditional variance still remains lower at $X_t = E^{\mathbb{P}}(X)$ in the mixture, see Figure 8. Since α_1 is fixed to identify the scale of Y_1, this increase is instead achieved by scaling Y_1 down, which explains why μ_{1Z} and μ_{12} are lower, and why δ_1 and μ_{21} are higher in the mixture models.

6.3.6 Higher moments and the Second Component Location

The mixture formulation of the $\mathbb{P}$ distribution for yields implies considerable variation in the conditional higher moments of the one-period rate as shown in Figures 9 and 10. In the single-component models under both measures and in the risk-neutral part of the mixture models, the skewness and excess kurtosis of $r_{t+1}|X_t$ are very close to zero. The mixture models under $\mathbb{P}$ in general estimate negative skewness and positive excess kurtosis, and for $X_t = E^{\mathbb{P}}(X)$ skewness is about -1 while excess kurtosis is about 3-4.

The scaled higher moments vary with the conditioning factors, as these affect the distance between the component means and their relative weights. They also vary due to changes in the second moment used for scaling. In the constant mixture models both skewness and excess kurtosis increase numerically with both lower medium term rates and lower slope of the yield curve, i.e., higher Z or Y_2. With respect to the level they are constant. In the time-varying models the quickly increasing second moment makes the negative skewness and excess kurtosis highest at about +1 standard deviation for both the negative curvature factor, Z, and the level factor Y_1. It may be noted that the shapes of the conditional standard deviation, skewness, and excess kurtosis are closely related. As the distribution for factors in the single-component model for a given X_t is only shifted by a single value Λ_t, the distribution in the two-component mixtures is shifted from $\mathbb{Q}$ only in two directions described by the mix over Λ_t^1 and Λ_t^2.

The impact of the mixture distribution can alternatively be illustrated by the conditional density of the factors, $f(X_{t+1}|X_t)$, in the direction of the mean of the second component. Thus for the vector representing the (negative) direction to the second component mean, $v = X_t - E^{\mathbb{P}_2}(X_{t+1}|X_t)$, Figure 11 draws the density function $g(\alpha) = f^{\mathbb{P}}(X_{t+1} = X_t + \alpha v|X_t)$, such that $\alpha = -1$ will be the conditional $\mathbb{P}$ density for $X_{t+1} = E_t^{\mathbb{P}_2}(X_{t+1})$. The density is drawn for conditioning X_t values at $E^{\mathbb{P}}(X)$ and points where each factor, Z, Y_1, and Y_2, in turn are shifted ± 1 standard deviation. A joint shift of $(1/2, 1/2, -1/2)$ standard deviations from the factors' unconditional $\mathbb{P}$ means is also considered. Table 6 gives the corresponding values of X_t and v used to draw Figure 11.

From Figure 11 and Table 6 it follows that the second component adds density to the conditional distribution of factors under $\mathbb{P}$ for X_{t+1}'s that represent relatively large movements away from X_t. At $X_t = E^{\mathbb{P}}(X)$ the distance to the conditional mean of the second component is $-v = -(.3, .7, .2)$, i.e., in direction of a lower level of yields, a steeper curve and more positive curvature. The distance almost does not vary with the conditioning value of the level factor, while for a less steep curve, Y_{2t} at +1 std. dev., the location is further towards a steeper curve and a lower level, and for less positive curvature, Z_t at +1 std. dev., it is further towards more positive curvature. The variation in the impact of the second component in the varying weights model is clear from Figure 11. Thus, the height of the second component in the density is noticeably larger for high values of Z_t and Y_{1t}, as well as for the joint case where all factors are shifted in the directions that give a higher second component weight.

The introduction of a mixture distribution under $\mathbb{P}$ allows for the possibility of specific large factor moves that have almost zero probability in single-component models. This has similarities to adding jumps in continuous-time models in the sense that it adds the possibility of infrequent large moves in variables. In the mixture model the weight on the second component can thus be seen as the intensity of the specific unconventional moves described by this component. For instance in the linear constant weight model this is estimated at 2.2% per week equal to 68% probability per year of at least one such move. The time-varying weights models can have periods of higher and lower intensity of these larger moves dependent on how weights change with factors over time. The particular infrequent move that the models estimate should be added under $\mathbb{P}$ is a relatively large move in direction of simultaneously shifting the yield curve lower, steeper, and more positively curved.

7 Conclusion

Affine risk-neutral factor models in the sense of Duffie & Kan (1996) are attractive because they lead to relatively simple expressions for the yield curve and still are observed to capture the cross-section of interst rates well when about three factors are included. Affine models are simultaneously found to provide a poor description of the time-series dynamics of interest rates such as shown for instance by Duffee (2002). It is well documented that the mean and variance of the short rate depend nonlinearly on its own level, and that interest rate changes depart from the normal distribution such as captured in models that include jumps. This paper suggests a method that in a dynamic term structure model with an affine pricing relation allows for nonlinear relations between factors and moments of the short rate in the time-series dimension, as well as departures from the normal distribution. This is possible, since the model is set in discrete time which offers substantial flexibility in formulating how the one-period ahead distributions for interest rates differ between measures. Specifically, the model exploits a weighted average of Radon-Nikodym derivatives, a combination which results in a mixture model for the time-series of rates and thus allows for the requested properties. Empirically it is indeed found that the distribution of interest rates over time departs from normality, and that the one-period rate has nonlinear dependencies on the factors. The preferred model with factor-dependent weights estimates for instance large increases in mean-reversion and variance of the one-period rate for high levels of the yield curve and finds substantial negative skewness and positive excess kurtosis. These properties are the consequences of the type of yield curve changes that the second component allows for in form of simultaneous shifts toward lower level, more steepness, and more curvature. This can be seen as capturing some of the same effects that jumps do in other types of models.

In the model suggested here, though, the form of the infrequent large shifts that are added can be given a quite specific interpretation since it relates to the factors that impact the level, slope, and curvature of the yield curve.

Appendix

Zero-coupon Bond Price Since $\mathbb{Q}$ is defined as the risk-neutral measure, discounted prices must be martingales under $\mathbb{Q}$ to avoid arbitrage,

$$P_t^n = E_t^{\mathbb{Q}}\left[e^{-\Delta r_t} P_{t+1}^{n-1}\right] \tag{43}$$

where it has been emphasized that the one-period rate is measured in per annum terms by the scaling with Δ. Suppose that prices are on the form $P_t^n = e^{-A_n - B_n' X_t}$ and check whether this satisfies the no arbitrage condition. Insert the price to get

$$\begin{aligned} e^{-A_n - B_n' X_t} &= e^{-\Delta r_t} E_t^{\mathbb{Q}}\left[e^{-A_{n-1} - B_{n-1}' X_{t+1}}\right] \\ -A_n - B_n' X_t &= -\Delta\delta_0 - \Delta\delta_X' X_t - A_{n-1} + a\left(-B_{n-1}\right) + b\left(-B_{n-1}\right)' X_t. \end{aligned}$$

The second equation is obtained by taking logs and using (1) and (8). Then by equating constants and coefficients of X_t, this equation is satisfied when A_n and B_n solve the recursions

$$\begin{aligned} A_n &= A_{n-1} + \Delta\delta_0 - a\left(-B_{n-1}\right) \\ B_n &= \Delta\delta_X - b\left(-B_{n-1}\right), \end{aligned}$$

with boundary conditions $A_0 = 0$ and $B_0 = 0$ following from $P_t^0 = 1$. Thus an exponential-affine price does indeed satisfy (43). For yields measured in per annum terms, $P_t^n = \exp\left(-n\Delta y_t^n\right)$, so

$$y_t^n = \frac{A_n}{n\Delta} + \frac{B_n'}{n\Delta} X_t \equiv \delta_0^n + \delta_X^{n\prime} X_t$$

Multi-period RN derivative By iterated expectations and from (14) for each period it follows that

$$\begin{aligned} E_t^{\mathbb{P}}\left(X_{t+k}\right) &= E_t^{\mathbb{P}}\left(E_{t+1}^{\mathbb{P}}\left(X_{t+k}\right)\right) = E_t^{\mathbb{Q}}\left(\xi_{t,t+1} E_{t+1}^{\mathbb{Q}}\left(\xi_{t+1,t+k} X_{t+2}\right)\right) \\ &= E_t^{\mathbb{Q}}\left(E_{t+1}^{\mathbb{Q}}\left(\xi_{t,t+1}\xi_{t+1,t+k} X_{t+2}\right)\right) = E_t^{\mathbb{Q}}\left(\xi_{t,t+1}\xi_{t+1,t+k} X_{t+2}\right), \end{aligned}$$

such that $\xi_{t,t+k} = \xi_{t,t+1}\xi_{t+1,t+k}$ and then (15) follows by induction. Similarly, if $E_{t+1}^{\mathbb{Q}}\left(\xi_{t+1,t+k}\right) = 1$, it holds that

$$\begin{aligned} E_t^{\mathbb{Q}}\left(\xi_{t,t+k}\right) &= E_t^{\mathbb{Q}}\left(\xi_{t,t+1}\xi_{t+1,t+k}\right) = E_t^{\mathbb{Q}}\left(E_{t+1}^{\mathbb{Q}}\left(\xi_{t,t+1}\xi_{t+1,t+k}\right)\right) \\ &= E_t^{\mathbb{Q}}\left(\xi_{t,t+1} E_{t+1}^{\mathbb{Q}}\left(\xi_{t+1,t+k}\right)\right) = E_t^{\mathbb{Q}}\left(\xi_{t,t+1}\right) = 1, \end{aligned}$$

and then again by induction, all multi-period RN derivatives have expected value under $\mathbb{Q}$ equal to 1, and they thus define probability measures. Equivalence, i.e., $\xi_{t,t+k} > 0$ for $\mathbb{Q}_{t+k}$-a.s., follows since all one-period ξ are strictly positive $\mathbb{Q}_{t+k}$-a.s.

Centered Moments in Mixture The centered moments in a mixture distribution can be obtained by

$$
\begin{aligned}
& E_t^{\mathbb{P}}\left[\left(X_{t+1}-E_t^{\mathbb{P}}\left(X_{t+1}\right)\right)^n\right] \\
= & \sum_{j=1}^k \sum_{i=1}^n \binom{n}{i}\left(E_t^j\left(X_{t+1}\right)-E_t^{\mathbb{P}}\left(X_{t+1}\right)\right)^{n-i} w_t^j E_t^j\left[\left(X_{t+1}-E_t^j\left(X_{t+1}\right)\right)^i\right].
\end{aligned}
$$

Alternatively, the uncentered moments of the mixture are linear in the component uncentered moments, $E_t^{\mathbb{P}}\left(X_{t+1}^n\right)=\sum_j w^j E_t^j\left(X_{t+1}^n\right)$. Then the centered moments of the mixture follow by standard relations.

Invariant Transformations in the Discrete Model Analogous to continuous-time affine models different transformations of the risk-neutral $DA_1^{\mathbb{Q}}(N)$ model are possible without changing the dynamics of the one-period rate r_t. Instead of performing invariant transformations in the continuous-time model and then considering the implications for the discrete model via the approximation (5), the rotations can be formulated directly in the discrete model. Let the parameter vector be

$$
\{\delta_0, \delta_Z, \delta_Y, \rho, \nu, c, \mu_0, \mu_Z, \mu_Y, \Sigma_Y, \alpha, \beta\}
$$

as defined in (2), (3), and (8). Then consider the following transformations:

Rescaling of Z: Positive scalar l, which transforms the variance factor to lZ and the parameters to

$$
\{\delta_0, \delta_Z/l, \delta_Y, \rho, \nu, lc, \mu_0, \mu_Z/l, \mu_Y, \Sigma_Y, \alpha, \beta/l\}.
$$

A location shift of Z is not considered, since the discrete model does not allow Z to be a location-shifted $\chi^2(k,\lambda)$ variable.

Affine transformation of Y: A $(N-1)\times(N-1)$ non-singular matrix L and $(N-1)$ vector v, which transform the Y factors to $LY+v$ with new parameters given by

$$
\left\{\delta_0-\delta_Y' L^{-1} v, \delta_Z, L'^{-1}\delta_Y, \rho, \nu, c, L\mu_0+\left(I-L\mu_Y L^{-1}\right) v, L\mu_Z, L\mu_Y L^{-1}, L\Sigma_Y, \alpha, \beta\right\}.
$$

Linear transformations of Z and Y, i.e., lZ and LY, are separated in the discrete model, whereas the continuous models allow for general $L_X X$ scaling. This separation is necessary in the discrete model, since Z and Y are defined to be instantaneously independent.

Reparametrization of Y variance: A diagonal matrix D that does not affect factors but changes the parametrization of Ω_{Yt}, such that the new parameter vector is

$$
\left\{\delta_0, \delta_Z, \delta_Y, \rho, \nu, c, \mu_0, \mu_Z, \mu_Y, \Sigma_Y D^{-1}, D^2\alpha, D^2\beta\right\}.
$$

When the alternative representation of the discrete model variance, $\Omega_{Yt} = h_0 + h_Z Z_t$, is used, the h_0 and h_Z matrices are unaffected by a D transformation. An affine transformation of Y has the effect $\{Lh_0L', Lh_ZL'\}$, while a rescaling of Z with l gives $\{h_0, h_Z/l\}$.

Further possible transformations are: A rotation of the Gaussian error terms, i.e., an orthogonal matrix O that commutes with S_{Yt} and rotates only Σ_Y to $\Sigma_Y O'$. A permutation of the Y factors.

Effect of Invariant Transformations on Parameters in Λ An affine transformation of Y has no effect on Λ_Z, whereas it changes $\{\lambda_{Y0}, \lambda_{YZ}, \lambda_{YY}\}$ to

$$\left\{L\lambda_{Y0} + \left(I - L\lambda_{YY}L^{-1}\right)v, L\lambda_{YZ}, L\lambda_{YY}L^{-1}\right\},$$

and $\{\eta_0, \eta_Z, \eta_Y\}$ to

$$\left\{\eta_0 - \eta_Y' L^{-1} v, \eta_Z, L'^{-1}\eta_Y\right\}.$$

A rescaling of Z implies: $\{\lambda_{Z0}, \lambda_{Z1}/l, \lambda_{Z2}/l^2, \lambda_{Z3}/l^3\}$, $\{\lambda_{Y0}, \lambda_{YZ}/l, \lambda_{YY}\}$, and $\{\eta_0, \eta_Z/l, \eta_Y\}$. Reparameterizing the variance with D or a rotation of the Gaussian error terms with O has no effect.

Mapping to Continuous-time $A_1^{\mathbb{Q}}(N)$ ***Approximation*** The reverse relation between the discrete model and the continuous-time approximation (5) is

$$\begin{aligned}
&\kappa_{ZZ} = (1-\rho)/\Delta, && \theta_Z = c\nu/(1-\rho), && \sigma_Z^2 = 2c/\Delta, \\
&\theta_Y = (I - \mu_Y)^{-1}\left(\mu_0 + \mu_Z \tfrac{c\nu}{1-\rho}\right), && \kappa_{YZ} = -\mu_Z/\Delta, && \kappa_{YY} = (I-\mu_Y)/\Delta, \\
&\alpha^c = \alpha/\Delta && \beta^c = \beta/\Delta &&
\end{aligned} \tag{44}$$

Identification in the Discrete Model When the restrictions (36) and (37) are imposed, it follows from (44) and (45) that the approximating continuous-time model has the fixed parameters

$$\theta_Y = 0, \quad \sigma_Z = 1, \quad \Sigma_Y = I, \quad \alpha^c = \iota,$$

while $\{\delta_0, \delta_Z, \delta_Y, \kappa_{ZZ}, \kappa_{YZ}, \kappa_{YY}, \theta_Z, \beta^c\}$ are free parameters that satisfy

$$\delta_Y \geq 0, \quad \kappa_{ZZ} > 0, \quad \theta_Z > (2\kappa_{ZZ})^{-1}, \quad eig(\kappa_{YY}) > 0, \quad \beta^c \geq 0.$$

These are the same fixed parameters and restrictions on the free ones as in canonical representation.

Whether the conditions (36) indeed impose no limitations and are identifying in the discrete model is checked by seeing whether these restrictions can be imposed on any discrete model via invariant transformations and that no further rotations are possible from this point. Starting from a general $DA_1^{\mathbb{Q}}(N)$ model with the full parameter vector $\Theta = \{\delta_0, \delta_Z, \delta_Y, \rho, \nu, c, \mu_0, \mu_Z, \mu_Y, \Sigma_Y, \alpha, \beta\}$ free, the following rotations can be performed, which progressively do not affect parameters already fixed in previous steps:

I) $l = \Delta/(2c)$ sets $c = \Delta/2$.

II) $D = \Delta diag(\alpha)^{-1}$ sets $\alpha = \Delta\iota$. Requires α strictly positive.

III) $L = \Sigma_Y^{-1}$ sets $\Sigma_Y = I$.

III) $v = -(I - \mu_Y)^{-1}\left[\mu_Z \frac{c\nu}{1-\rho} + \mu_0\right]$ sets $\mu_0 = -\mu_Z \frac{c\nu}{1-\rho}$.

IV) $L = diag(sign(\delta_Y))$ ensures that $\delta_Y \geq 0$.

Thus a model that satisfies (36) can be reached from any discrete model with α strictly positive, so besides this, the set of conditions are unrestrictive. Also, any further transformations by either l, L, v, D or O where these are not identity operators will change the value of either c, μ_0, Σ_Y, or α, or affect the sign of δ_Y. Since no further transformations are possible, the conditions (36) are identifying.

As discussed, this rotation turns out to be problematic for the data set used in this paper and it is better to fix the scale of Y_2 by setting μ_{21} constant instead of α_2. Thus for some constant γ the rotation $L = diag(1, \gamma/\mu_{21})$ sets $\mu_{21} = \gamma$ and changes α_2 to a free parameter, while it also releases the sign of δ_2.

Λ_{Yt}^j in (41) imply constant $\mathbb{P}_j$ parameters In the moment generating function (21) for the j'th component of the $\mathbb{P}$ model set $u_Z = 0$ to focus on the Y factors and then take logs to get

$$\log\phi_t^j(0, u_Y) = u_Y'\mu_0\left(\Lambda_{Yt}^j\right) + \frac{1}{2}u_Y'h_0u_Y + \left[\frac{1}{2}u_Y'h_Zu_Y + u_Y'\mu_Z\left(\Lambda_{Yt}^j\right)\ ,\ u_Y'\mu_Y\right]X_t.$$

Substitute that by (22) $\mu_0\left(\Lambda_{Yt}^j\right) = \mu_0 + h_0\Lambda_{Yt}^j$ and $\mu_Z\left(\Lambda_{Yt}^j\right) = \mu_Z + h_Z\Lambda_{Yt}^j$ to get that

$$\log\phi_t^j(0, u_Y) = u_Y'\mu_0 + \frac{1}{2}u_Y'h_0u_Y + \left[\frac{1}{2}u_Y'h_Zu_Y + u_Y'\mu_Z\ ,\ u_Y'\mu_Y\right]X_t + u_Y'(h_0 + h_ZZ_t)\Lambda_{Yt}^j.$$

For $\Sigma_Y = I$, the variance is $\Omega_{Yt} = S_{Yt} = h_0 + h_ZZ_t$, and then for the j'th component market price of risk function for the Y factors in *(41)*, $\Lambda_{Yt}^j = S_{Yt}^{-1}\left(\lambda_{Y0}^j + \lambda_{YZ}^jZ_t + \lambda_{YY}^jX_t\right)$, it follows that

$$\begin{aligned}\log\phi_t^j(0, u_Y) &= u_Y'\mu_0 + \frac{1}{2}u_Y'h_0u_Y + \left[\frac{1}{2}u_Y'h_Zu_Y + u_Y'\mu_Z\ ,\ u_Y'\mu_Y\right]X_t \\ &\quad + u_Y'\left(\lambda_{Y0}^j + \lambda_{YZ}^jZ_t + \lambda_{YY}^jX_t\right) \\ &= u_Y'\mu_0^j + \frac{1}{2}u_Y'h_0u_Y + \left[\frac{1}{2}u_Y'h_Zu_Y + u_Y'\mu_Z^j\ ,\ u_Y'\mu_Y^j\right]X_t,\end{aligned}$$

for constant j-component parameters $\mu_0^j = \mu_0 + \lambda_{Y0}^j$, $\mu_Z^j = \mu_Z + \lambda_{YZ}^j$, and $\mu_Y^j = \mu_Y + \lambda_{YY}^j$.

α, β, and Σ_Y from h_0 and h_Z In the alternative representations of the variance for the Y factors, $\Omega_{Yt} = \Sigma_Y diag\left(\alpha + \beta Z_t\right) \Sigma_Y' = h_0 + h_Z Z_t$, it may be relevant to be able to go from h_0 and h_Z to the α, β, and Σ_Y parameters. Of course, since D rotations can be performed on α, β, and Σ_Y, these are not identified and different solutions are possible. Suppose h_0 is positive definite, then one solution would be the following

$$\Sigma_Y = U'V^{-1}, \qquad \alpha = \iota, \qquad \beta = diag\left(D\right), \tag{45}$$

where V is the Cholesky decomposition of $h_0 = VV'$, and U and D are the orthogonal and diagonal matrices in the diagonalization of the symmetric matrix $V^{-1}h_Z V'^{-1} = UDU'$. Any other solutions for α, β, and Σ_Y can then be obtained by further rotations by D. The equivalent method can be used with h_0 and h_Z switched if h_Z is positive definite. As discussed in the main text, both h_0 and h_Z may be singular, since only $\alpha_i + \beta_i > 0$ is required and therefore $\alpha_i = 0$ and $\beta_{j \neq i} = 0$ is possible. In that case the suggested solution would not work, but with only two Y factors a solution is obtained by $\sigma_{11} = \sigma_{22} = 1$, $\alpha_1 = h_0^{11}$, $\beta_2 = h_Z^{22}$, $\sigma_{12} = h_Z^{12}/h_Z^{22}$, and $\sigma_{21} = h_0^{12}/h_0^{11}$.

References

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

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

Ang, A. & Bekaert, G. (2002), 'Short rate nonlinearities and regime switches', *Journal of Economic Dynamics and Control* **26**, 1243–1274.

Bansal, R. & Zhou, H. (2002), 'Term structure of interest rates with regime shifts', *The Journal of Finance* **57**(5), 1997–2043.

Boudoukh, J., Richardson, M., Stanton, R. & Whitelaw, R. F. (1999), 'A multifactor, nonlinear, continuous-time model of interest rate volatility', *Working Paper* .

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

Chen, R.-R. & Scott, L. (1993), 'Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates', *The Journal of Fixed Income* pp. 14–31.

Cheridito, P., Filipovic, D. & Kimmel, R. L. (2007), 'Market price of risk specifications for affine models: Theory and evidence', *Journal of Financial Economics* **83**, 123–170.

Dai, Q., Le, A. & Singleton, K. J. (2006), 'Discrete-time term structure models with generalized market prices of risk', *Working Paper* .

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

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society* **39**(1), 1–38.

Duarte, J. (2004), 'Evaluating an alternative risk preference in affine term structure models', *The Review of Financial Studies* **17**(2), 379–404.

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

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

Hamilton, J. D. (1988), 'Rational-expectations econometric analysis of changes in regime', *Journal of Economic Dynamics and Control* **12**, 385–423.

Johannes, M. (2004), 'The statistical and economic role of jumps in continuous-time interest rate models', *The Journal of Finance* **59**(1), 227–260.

Litterman, R. & Scheinkman, J. (1991), 'Common factors affecting bond returns', *The Journal of Fixed Income* pp. 54–61.

Pearson, N. D. & Sun, T.-S. (1994), 'Exploiting the conditional density in estimating the term structure: An application to the cox, ingersoll, and ross model', *The Journal of Finance* **49**(4), 1279–1304.

Piazzesi, M. (2001), 'An econometric model of the yield curve with macroeconomic jump effects', *NBER Working Paper* **8246**.

Redner, R. A. & Walker, H. F. (1984), 'Mixture densities, maximum likelihood and the EM algorithm', *SIAM Review* **26**(2), 195–239.

Relative Performance of 3-factor Models

Single Component	*Linear*	*Cubic*
maximal[df]	2.75 [28]	4.26 [31]
preferred [df]	0.00 [19]	1.58 [22]
LR(pref.) [pval]	5.50 [0.79]	5.36 [0.80]
LR(linear) [pval]	3.16 [0.368]	
AIC	38.0	40.8
BIC	133.0	150.8

Constant Mixture	*Linear*	*Cubic*
maximal [df]	92.61 [38]	100.66 [44]
preferred [df]	86.68 [23]	95.08 [25]
LR(pref.) [pval]	11.86 [0.69]	11.16 [0.92]
LR(linear) [pval]	16.81 [0.000]	
AIC	-127.4	-140.2
BIC	-12.4	-15.2

Time-varying Mixture	*Linear*	*Cubic*
maximal [df]	105.8 [41]	114.8 [47]
preferred [df]	98.47 [26]	107.1 [28]
LR(pref.) [pval]	14.61 [0.48]	15.39 [0.70]
LR(linear) [pval]	17.29 [0.000]	
AIC	-144.9	-158.2
BIC	-15.0	-18.2

Table 1: The table compares the performance of $DA_1^{\mathbb{Q}}(3)$ models with different specifications for interest rates dynamics under the physical measure. For each of the 6 models the table shows log-likelihood values of the maximal and preferred models. The degrees of freedom for each model is given brackets. Likelihood values are normalized relative to the preferred single component linear model. Then follows first the likelihood-ratio test of restrictions imposed by the preferred model relative to the maximal, and then the LR test of restrictions imposed by the preferred linear models relative to allowing for cubic terms. Finally, Akaike's and the Bayesian information criteria are shown.

Parameters in Risk-Neutral Models

Linear	*Single Component*		*Constant Mixture*		*Time-varying Mixture*		10^x
$1-\rho$	28.23	(0.94)	28.16	(0.35)	28.16	(0.33)	(-3)
ν	15.02	(0.32)	16.24	(0.39)	15.81	(0.34)	
μ_{1Z}	-47.37	(12.78)	-27.32	(0.93)	-28.05	(0.94)	(-3)
$1-\mu_{11}$	4.776	(0.489)	4.437	(0.016)	4.819	(0.018)	(-3)
μ_{12}	–12.87	(3.47)	-7.193	(0.301)	-7.368	(0.276)	(-3)
μ_{2Z}	-18.29	(1.95)	-19.25	(1.08)	-17.86	(0.99)	(-3)
μ_{21}	-1.495	(0.409)	-2.691	(0.125)	-2.645	(0.113)	(-3)
$1-\mu_{22}$	5.010	(0.484)	5.369	(0.198)	4.994	(0.020)	(-3)
β_1	11.88	(8.47)	-		-		(-3)
δ_0	106.5	(5.4)	110.5	(5.4)	108.7	(5.1)	(-3)
δ_Z	2.599	(0.168)	2.097	(0.119)	2.072	(0.114)	(-3)
δ_1	4.307	(1.150)	7.378	(0.173)	7.297	(0.172)	(-3)

Cubic	*Single Component*		*Constant Mixture*		*Time-varying Mixture*		10^x
$1-\rho$	28.22	(0.94)	28.25	(0.34)	28.16	(0.32)	(-3)
ν	15.13	(0.31)	14.58	(0.29)	15.96	(0.27)	
μ_{1Z}	-47.26	(12.83)	-28.74	(0.99)	-27.78	(0.93)	(-3)
$1-\mu_{11}$	4.770	(0.488)	4.405	(0.016)	4.804	(0.018)	(-3)
μ_{12}	-12.85	(3.49)	-7.698	(0.323)	-7.407	(0.307)	(-3)
μ_{2Z}	-18.26	(1.95)	-19.19	(1.06)	-17.84	(1.04)	(-3)
μ_{21}	-1.493	(0.412)	-2.492	(0.117)	-2.621	(0.126)	(-3)
$1-\mu_{22}$	5.003	(0.484)	5.381	(0.197)	4.993	(0.196)	(-3)
β_1	11.89	(8.53)	-		-		(-3)
δ_0	106.3	(5.3)	110.0	(5.3)	107.7	(5.1)	(-3)
δ_Z	2.586	(0.167)	2.279	(0.120)	2.158	(0.112)	(-3)
δ_1	4.294	(1.156)	7.315	(0.174)	7.215	(0.169)	(-3)

Table 2: The tables show maximum likelihood estimates of parameters in the risk-neutral part of models. Estimates are for preferred models where insignificant parameters have been removed as indicated by '-'. δ_2 and α_2 that were set to zero in all preferred models are not shown and neither are parameters whose values are fixed due to identification or admissibility. Figures in parenthesis are estimated standard errors, while the last column indicates that all numbers in the row are to the order 10^x.

Parameters in Continuous-time
Approximation to Risk-Neutral Models

Linear	*Single Component*		*Constant Mixture*		*Time-varying Mixture*	
κ_{ZZ}	1.468	(0.049)	1.465	(0.018)	1.464	(0.017)
κ_{1Z}	2.464	(0.664)	1.420	(0.049)	1.458	(0.049)
κ_{11}	0.248	(0.025)	0.231	(0.001)	0.251	(0.001)
κ_{12}	0.669	(0.180)	0.374	(0.016)	0.383	(0.014)
κ_{2Z}	0.951	(0.101)	1.001	(0.056)	0.929	(0.051)
κ_{21}	0.077	(0.021)	0.140	(0.007)	0.138	(0.006)
κ_{22}	0.260	(0.025)	0.279	(0.010)	0.260	(0.001)
θ_Z	5.117	(0.202)	5.546	(0.150)	5.397	(0.131)
β_1^c	0.618	(0.441)	-		-	

Cubic	*Single Component*		*Constant Mixture*		*Time-varying Mixture*	
κ_{ZZ}	1.468	(0.049)	1.469	(0.018)	1.464	(0.017)
κ_{1Z}	2.458	(0.668)	1.494	(0.052)	1.445	(0.048)
κ_{11}	0.248	(0.025)	0.229	(0.001)	0.250	(0.001)
κ_{12}	0.667	(0.181)	0.400	(0.017)	0.385	(0.016)
κ_{2Z}	0.949	(0.101)	0.998	(0.055)	0.928	(0.054)
κ_{21}	0.078	(0.021)	0.130	(0.006)	0.136	(0.007)
κ_{22}	0.260	(0.025)	0.280	(0.010)	0.260	(0.010)
θ_Z	5.153	(0.202)	4.963	(0.116)	5.451	(0.112)
β_1^c	0.618	(0.444)	-		-	

Table 3: The tables show parameter values in the continuous-time approximation to the risk-neutral part of estimated discrete-time models. Values are for preferred models in which insignificant parameters have been removed as indicated by '-'. δ_2 and α_2^c that were set to zero in all preferred models are not shown and neither are parameters whose values are fixed due to identification or admissibility. Figures in parenthesis are standard errors calculated from the discrete model by the delta method.

Parameters in the Radon-Nikodym Derivative

	Single Component				
	Linear		*Cubic*		10^x
λ_{Z1}	-		-1.905	(1.715)	
λ_{Z2}			0.828	(0.657)	
λ_{Z3}			-85.38	(61.24)	(-3)
λ_{1Z}	-19.62	(6.23)	-19.54	(6.23)	(-3)
λ_{12}	10.70	(4.55)	10.72	(4.57)	(-3)

	Constant Mixture				*Time-varying Mixture*				
	Linear		*Cubic*		*Linear*		*Cubic*		10^x
λ^1_{10}	-43.35	(8.60)	-41.82	(8.40)	-44.04	(9.30)	-42.38	(9.35)	(-3)
λ^1_{12}	6.153	(1.825)	6.576	(1.941)	6.154	(1.881)	6.127	(1.894)	(-3)
λ^2_{Z1}	-3.402	(0.943)	-52.00	(11.01)	-3.439	(0.863)	-57.69	(15.56)	
λ^2_{Z2}			18.29	(3.906)			18.78	(4.85)	
λ^2_{Z3}			-1.600	(0.340)			-1.519	(0.372)	
λ^2_{10}	-		-		-384.6	(70.0)	-397.1	(72.2)	(-3)
λ^2_{1Z}	-75.03	(12.90)	-86.09	(12.15)	-		-		(-3)
λ^2_{12}	-52.84	(17.14)	-43.53	(16.31)	-62.99	(14.95)	-51.46	(15.14)	(-3)
λ^2_{22}	-54.60	(20.52)	-49.49	(19.06)	-37.48	(18.12)	-35.75	(16.52)	(-3)
w_0	0.978	(0.005)	0.975	(0.006)					
η_0					5.789	(1.365)	5.958	(1.335)	
η_Z					-1.043	(0.260)	-1.084	(0.243)	
η_1					-0.321	(0.084)	-0.315	(0.080)	
η_2					0.156	(0.098)	0.158	(0.098)	

Table 4: The tables show estimates of parameters in different specifications for the Radon-Nikodym derivative. The estimates are for the preferred models in which insignificant parameters have been set to zero as indicated by '-'. Parameters set to zero in all models are left out of the table. Entries for parameters that does not enter a given model are left blank. Figures in parenthesis are estimated standard errors, while the last column indicates that all numbers in the row are to the order 10^x.

Changed Parameters in the $\mathbb{P}$-Models

	Single Component		
	Linear	*Cubic*	10^x
$1-\rho^{\mathbb{P}}$		$\star$	
$c^{\mathbb{P}}$		$\star$	
$\mu^{\mathbb{P}}_{1Z}$	-67.00	-66.80	(-3)
$\mu^{\mathbb{P}}_{12}$	–2.170	–2.124	(-3)

	Constant Mixture		*Time-varying Mixture*		
	Linear	*Cubic*	*Linear*	*Cubic*	10^x
$\mu^{\mathbb{P}_1}_{10}$	108.1	100.8	107.3	109.0	(-3)
$\mu^{\mathbb{P}_1}_{12}$	-1.041	-1.123	-1.214	-1.280	(-3)
$1-\rho^{\mathbb{P}_2}$	88.76	$\star$	89.37	$\star$	(-3)
$c^{\mathbb{P}_2}$	9.311	$\star$	9.308	$\star$	(-3)
$\mu^{\mathbb{P}_2}_{10}$			-233.2	-245.7	(-3)
$\mu^{\mathbb{P}_2}_{1Z}$	-102.4	-114.8			(-3)
$\mu^{\mathbb{P}_2}_{12}$	-60.03	-51.22	-70.36	-58.87	(-3)
$1-\mu^{\mathbb{P}_2}_{22}$	59.97	54.87	42.48	40.74	(-3)

Table 5: The tables show parameters in the models under $\mathbb{P}$. Those that are not shown or where entries are left blank have the same values as under $\mathbb{Q}$. Parameters with a '$\star$' in the cubic models are time varying and their values as a function of Z are shown in figure 3. The last column indicates that all numbers in the row are to the order 10^x.

Values of X_t and Direction v in Figure 11

Constant Mixture

X_t	Z	Y$_1$	Y$_2$	v	Z	Y$_1$	Y$_2$
+1 std.dev.	6.61 -9.54 5.75	5.27 -6.29 5.75	5.27 -9.54 9.21	+1 std.dev.	0.44 0.83 0.34	0.32 0.71 0.32	0.32 0.90 0.52
$E^{\mathbb{P}}(X)$	-	5.27 -9.54 5.75	-	$E^{\mathbb{P}}(X)$	-	0.32 0.69 0.31	-
-1 std.dev.	3.94 -9.54 5.75	5.27 -12.8 5.75	5.27 -9.54 2.28	-1 std.dev.	0.20 0.55 0.29	0.32 0.68 0.30	0.32 0.48 0.10
$\left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$	-	5.94 -7.92 4.01	-	$\left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$	-	0.38 0.66 0.23	-

Time-varying Mixture

X_t	Z	Y$_1$	Y$_2$	v	Z	Y$_1$	Y$_2$
+1 std.dev.	6.11 -8.06 5.82	4.95 -4.65 5.82	4.95 -8.06 9.07	+1 std.dev.	0.40 0.78 0.24	0.30 0.76 0.23	0.30 0.97 0.36
$E^{\mathbb{P}}(X)$	-	4.95 -8.06 5.82	-	$E^{\mathbb{P}}(X)$	-	0.30 0.74 0.22	-
-1 std.dev.	3.78 -8.06 5.82	4.95 -11.5 5.82	4.95 -8.06 2.57	-1 std.dev.	0.19 0.71 0.20	0.30 0.73 0.21	0.30 0.51 0.08
$\left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$	-	5.53 -6.36 4.20	-	$\left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$	-	0.35 0.65 0.16	-

Table 6: Tables show the direction, v, of the center of the second component at different conditioning factor values, X_t, which are used to draw the density graphs in Figure 11. To the left are the 8 conditioning factor vectors considered: the unconditional mean and one in which each factor is shifted either $+1$ or -1 standard deviation, as well as the point where factors jointly are shifted $(1/2, 1/2, -1/2)$ std. dev. To the right are the values of $v = X_t - E^{\mathbb{P}_2}(X_{t+1}|X_t)$, i.e., the negative direction to the mean of the second component at each of the X_t points to the left.

Uncondition mean and standard deviation of factors under $\mathbb{P}$

	Linear		*Cubic*	
Z	*Mean*	*S.D.*	*Mean*	*S.D.*
Single Component	5.11	1.33	5.03	1.18
Constant Mixture	5.27	1.34	4.74	1.33
Time-varying Mixture	4.95	1.17	5.16	1.24

	Linear		*Cubic*	
Y_1	*Mean*	*S.D.*	*Mean*	*S.D.*
Single Component	-24.28	7.51	-23.02	6.39
Constant Mixture	-9.54	3.24	-9.31	3.46
Time-varying Mixture	-8.06	3.41	-8.64	3.51

	Linear		*Cubic*	
Y_2	*Mean*	*S.D.*	*Mean*	*S.D.*
Single Component	7.27	3.51	7.23	3.48
Constant Mixture	5.75	3.47	5.11	3.32
Time-varying Mixture	5.82	3.25	5.60	3.39

Table 7: Unconditional mean and standard deviation of factors under $\mathbb{P}$ were estimated by Monte Carlo simulation. A draw from the unconditional $\mathbb{P}$ distribution of factors was obtained as the final factor vector after simulating 3000 periods in the estimated model. Though the initial X has little influence over this many periods, the value that solves $E_t^{\mathbb{P}}(X_{t+1}) = X_t$ was used. 10000 draws were made for each model, from which mean and standard deviation were calculated.

Linear Constant Mixture Model

Loading Functions

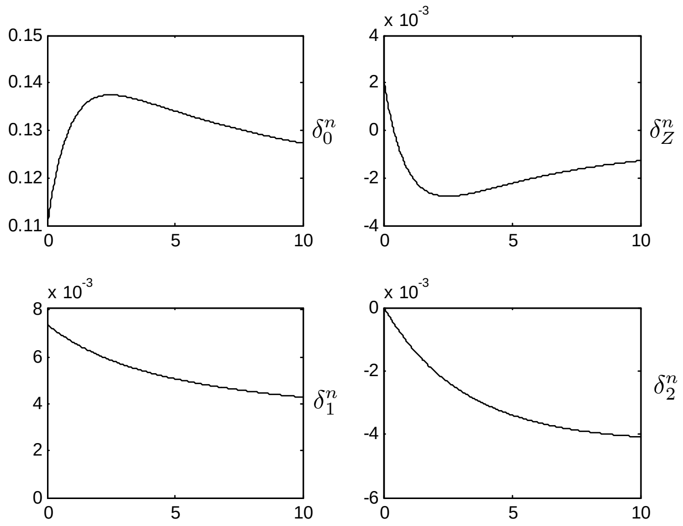

Figure 1: Draws loading functions, δ_X^n, and the constant part, δ_0^n, in the yield to factor relation, $y_t^n = \delta_0^n + \delta_X^{n\prime} X_t$, for the linear constant mixture model. The x-axis is measured in years, i.e., $n/52$.

Implied Factor Time Series

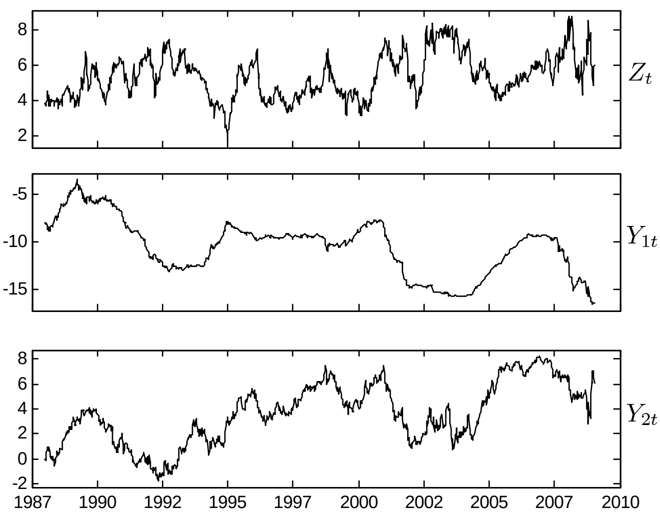

Figure 2: Displays time series of factors implied by yields observed without error in the linear constant mixture model.

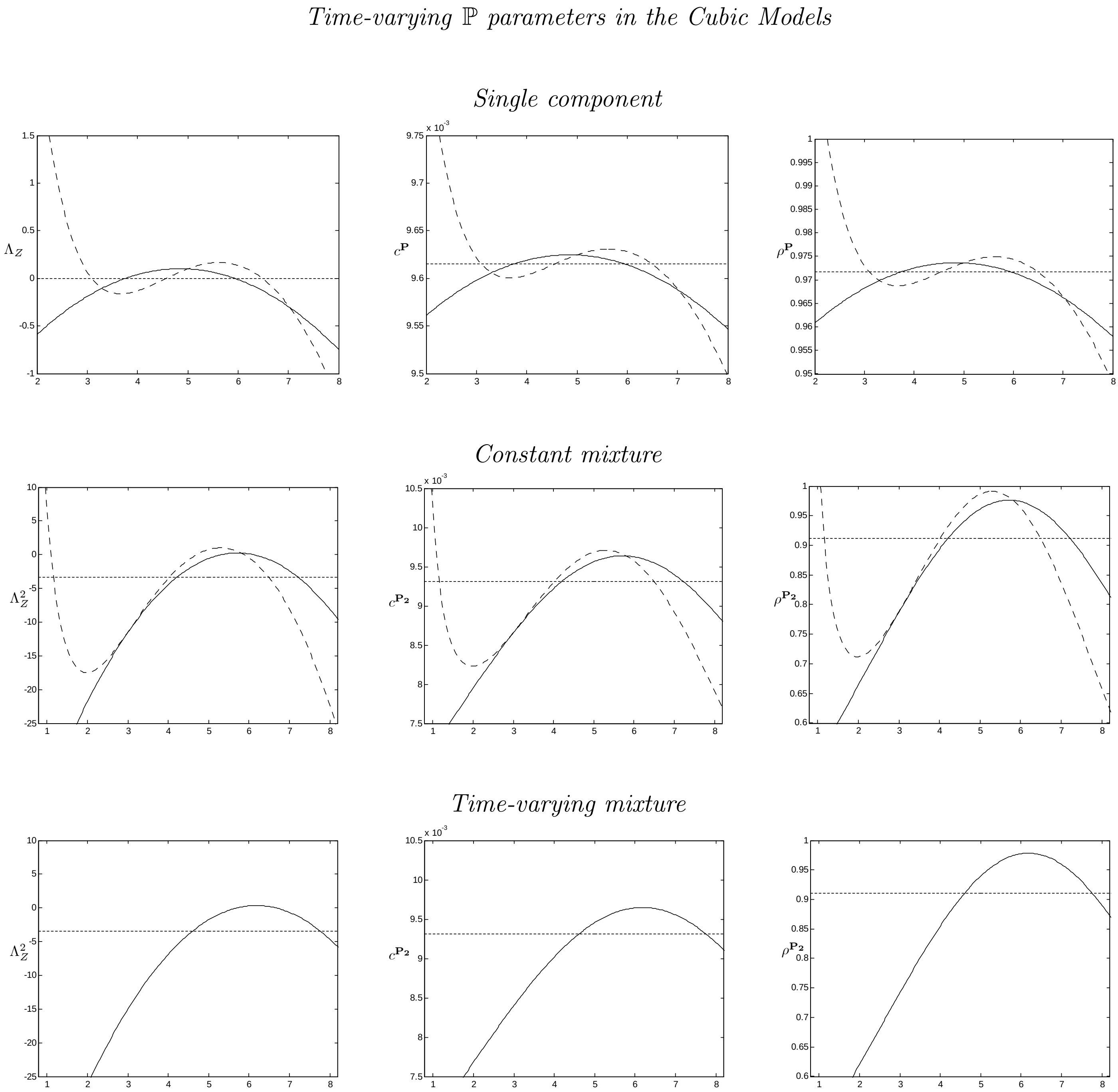

Figure 3: Figures draw Z-factor parameters c and ρ that become time varying under $\mathbb{P}_j$ when the market price of risk function Λ_Z^j, drawn as well, is not constant. This is the case in the cubic models, though for the mixture models $\Lambda_Z^1 = 0$ in the preferred models, so only the second component is shown. In all graphs the solid line is the cubic model, while the horizontal dotted line illustrates the value in the corresponding linear model, also given in table 5. The dashed lines are the values that arose in the first two cubic models, when the $\lambda_{Z0} \leq 0$ restriction, necessary for the $\mathbb{P}$ model to be well-defined, was not imposed. This implies $\rho > 1$ for low Z and also that $\Lambda_Z > 1/c$ for $Z \to 0$.

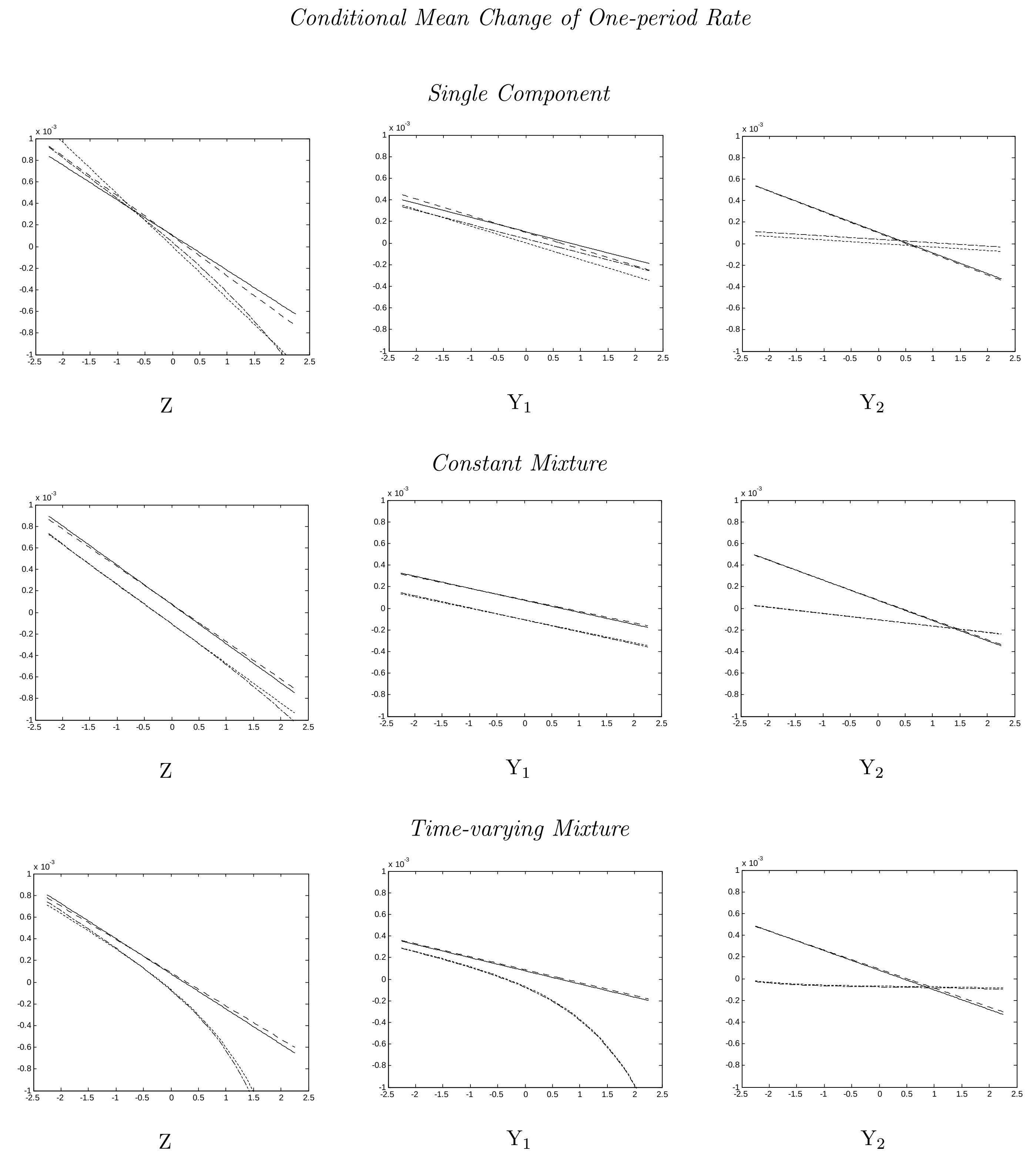

Figure 4: The graphs show how the expected change in the one-period rate (in per annum terms) over the next period varies as one of the conditioning factors changes and the other two are kept at their unconditional $\mathbb{P}$ mean. The function that is drawn is thus $f(x) = E\left[r_{t+1} - r_t | X_{it} = x, X_{h \neq i,t} = E^{\mathbb{P}}(X_{h \neq i,t})\right]$, and the factor, X_{it}, that varies is respectively Z_t, Y_{1t}, and Y_{2t} in each column from left to right . The x-axes are measured in standard deviations from the $\mathbb{P}$ mean. In each figure values are shown for the linear model under $\mathbb{Q}$ (dash) and $\mathbb{P}$ (dot), and for the cubic model under $\mathbb{Q}$ (solid) and $\mathbb{P}$ (dash-dot).

Expected change in Z factor - Cubic Single-Component model

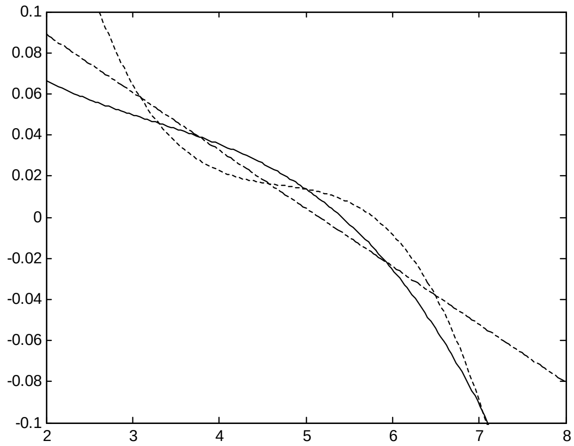

Figure 5: This shows the conditional expected change in the Z factor over different conditioning values, $f(z) = E(Z_{t+1}|Z_t = z) - Z_t$, for the cubic single-component model. This is calculated under $\mathbb{Q}$ (dash-dot) and under $\mathbb{P}$ (solid), as well as under $\mathbb{P}$ (dot) in the cubic model when the restriction $\lambda_{Z0} \leq 0$ is not imposed.

Weights as Function of Factors

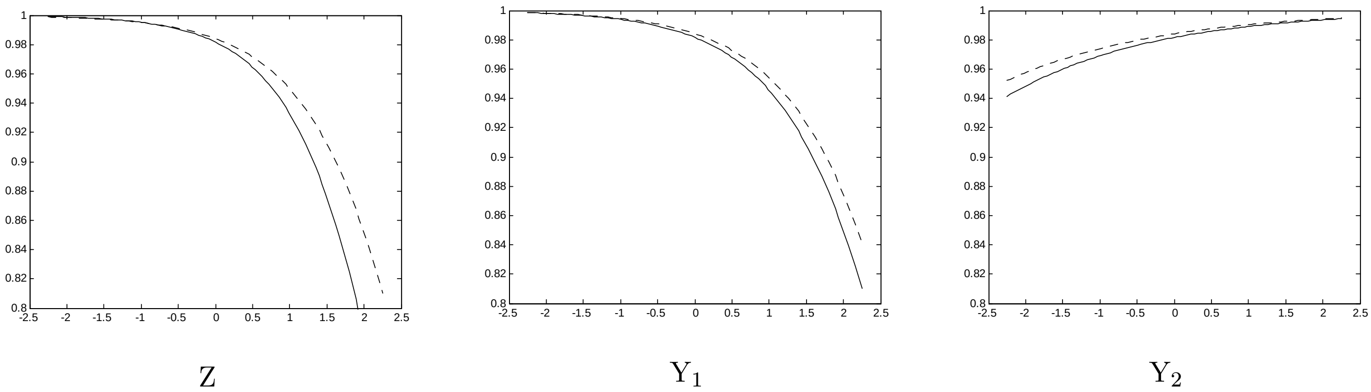

Figure 6: Figures illustrate how weights depend on factors in the time-varying mixture models by drawing the function $w(X_t) = (1 + \exp(-\eta_0 - \eta_X' X_t))$. From left to right respectively Z_t, Y_{1t}, and Y_{2t} varies, while the other factors are at their unconditional $\mathbb{P}$ mean. The linear models are the dashed lines and the cubic models are the solid lines. The x-axes are measured in standard deviations from the $\mathbb{P}$ mean.

Time-varying Weights - Cubic Model

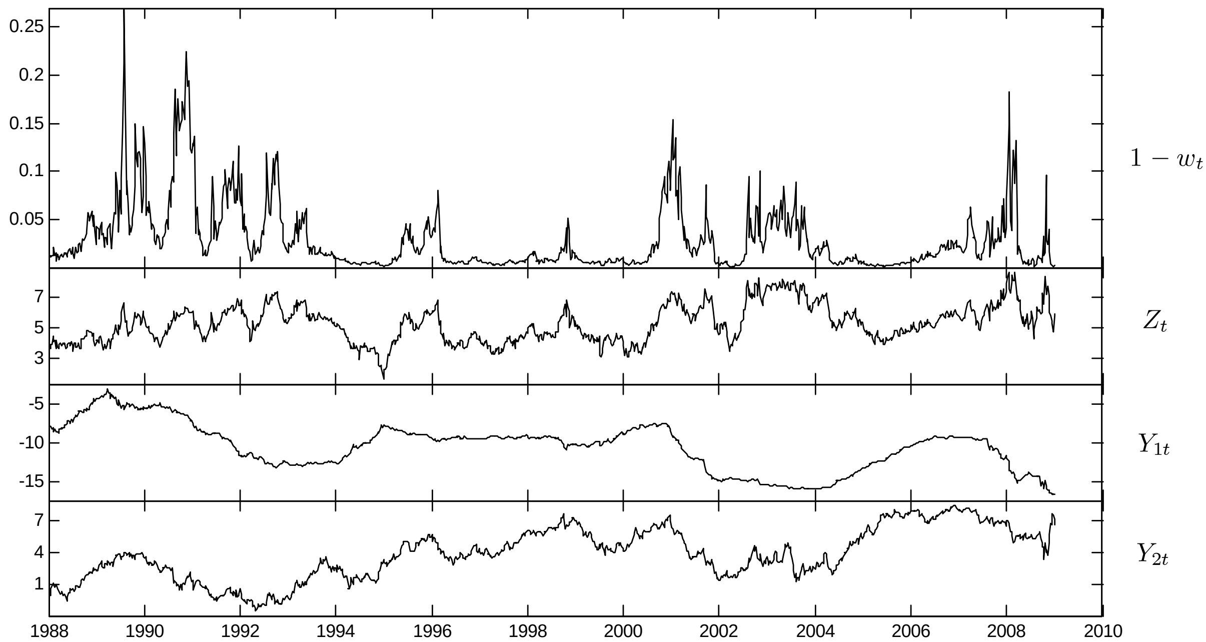

Figure 7: Displays time series of weights on the second $\mathbb{P}$ component, $1 - w_t$, in the cubic mixture model, for which weights depend on factors, as shown in Figure 6. Also shown are the series of each of the three factors.

Conditional Standard Deviation of One-period Rate

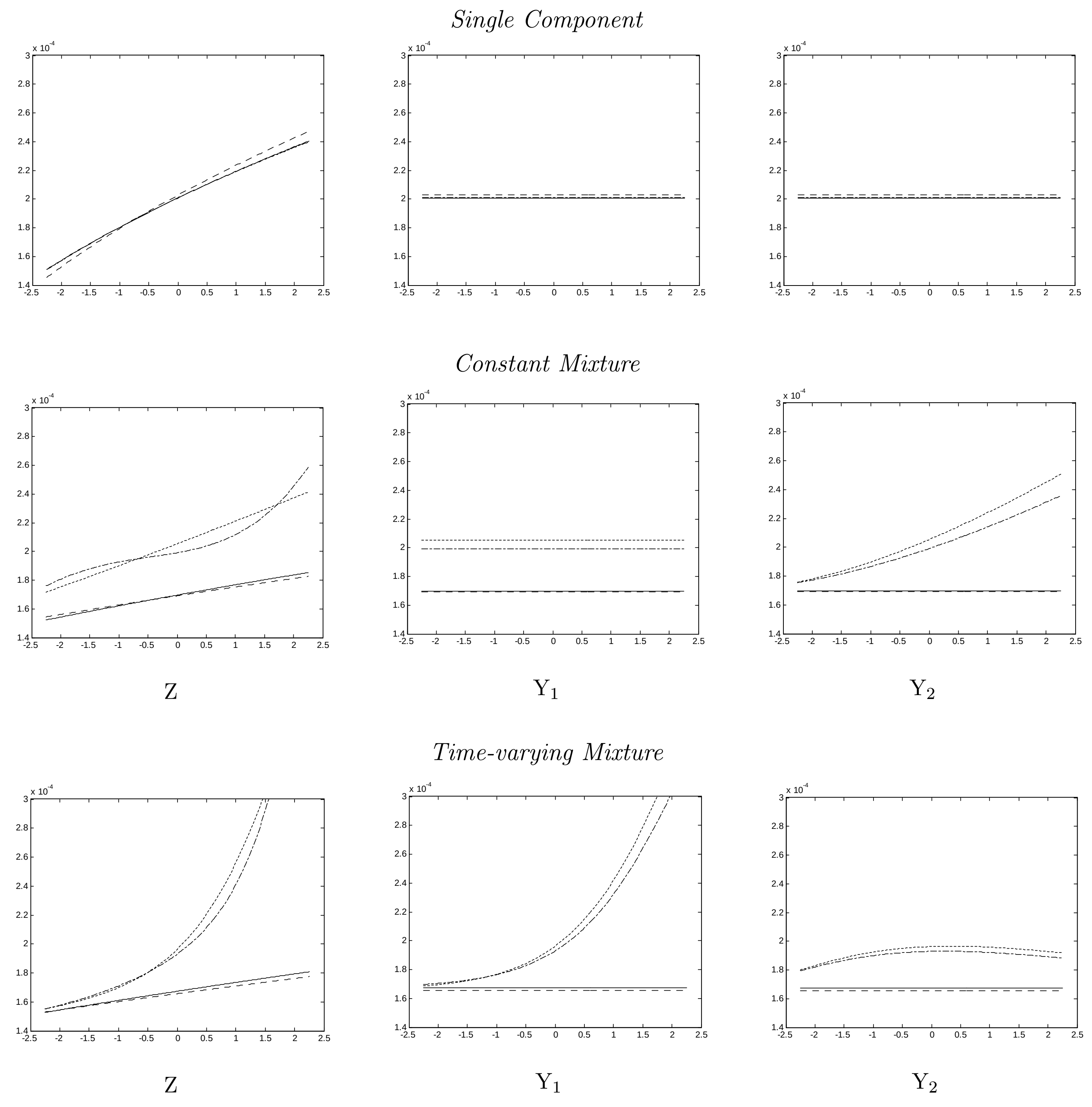

Figure 8: The graphs show the variation in standard deviation of the one-period rate (in per annum terms) as one of the conditioning factors changes and the other two are kept at their unconditional $\mathbb{P}$ mean. The factor that varies is respectively Z_t, Y_{1t}, and Y_{2t} in each column from left to right . The x-axes are measured in standard deviations from the $\mathbb{P}$ mean. In each figure values are shown for the linear model under $\mathbb{Q}$ (dash) and $\mathbb{P}$ (dot), and for the cubic model under $\mathbb{Q}$ (solid) and $\mathbb{P}$ (dash-dot).

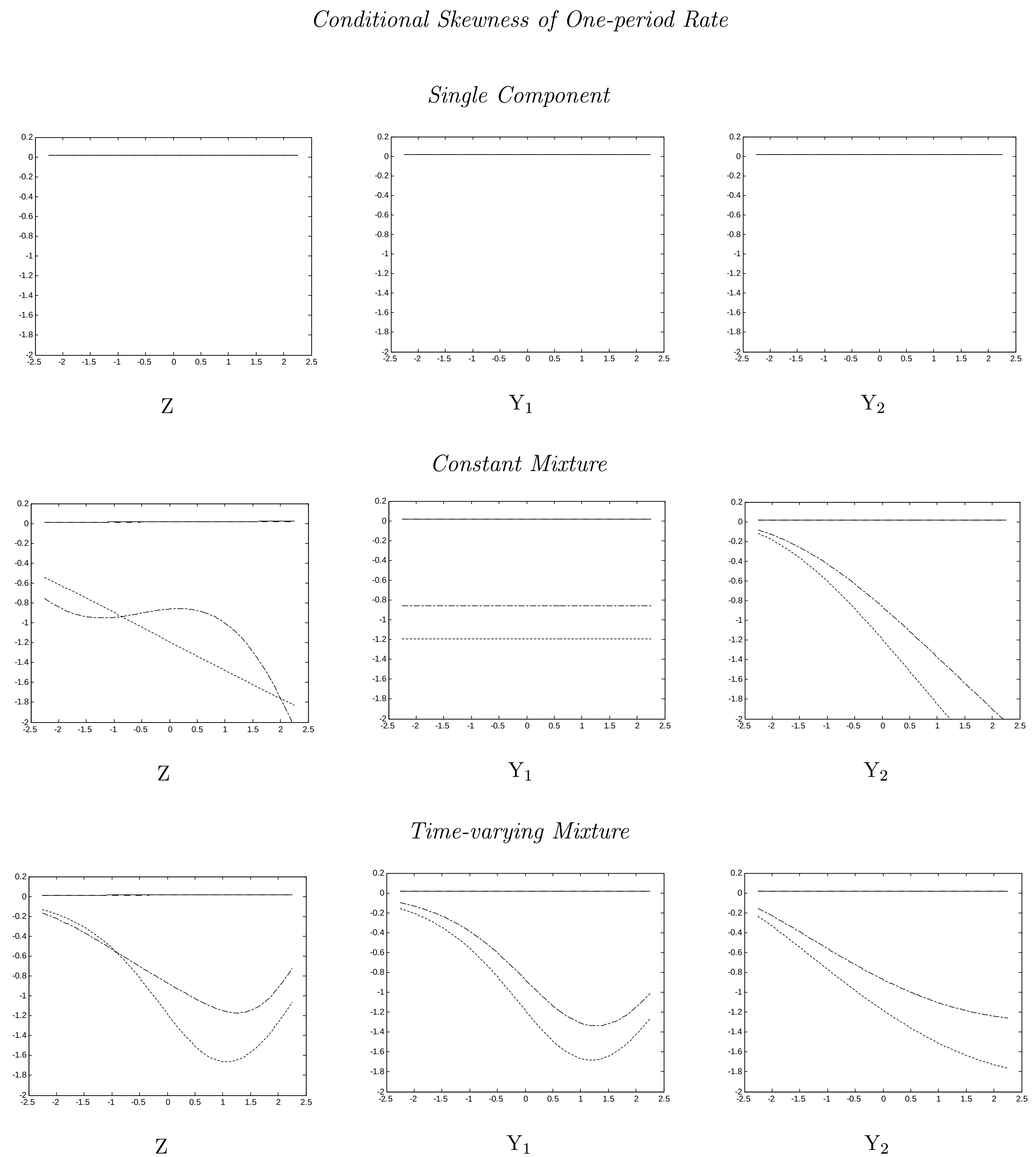

Figure 9: The graphs show the variation in skewness of the one-period rate as one of the conditioning factors changes and the other two are kept at their unconditional $\mathbb{P}$ mean. The factor that varies is respectively Z_t, Y_{1t}, and Y_{2t} in each column from left to right . The x-axes are measured in standard deviations from the $\mathbb{P}$ mean. In each figure values are shown for the linear model under $\mathbb{Q}$ (dash) and $\mathbb{P}$ (dot), and for the cubic model under $\mathbb{Q}$ (solid) and $\mathbb{P}$ (dash-dot).

Conditional Excess Kurtosis of One-period Rate

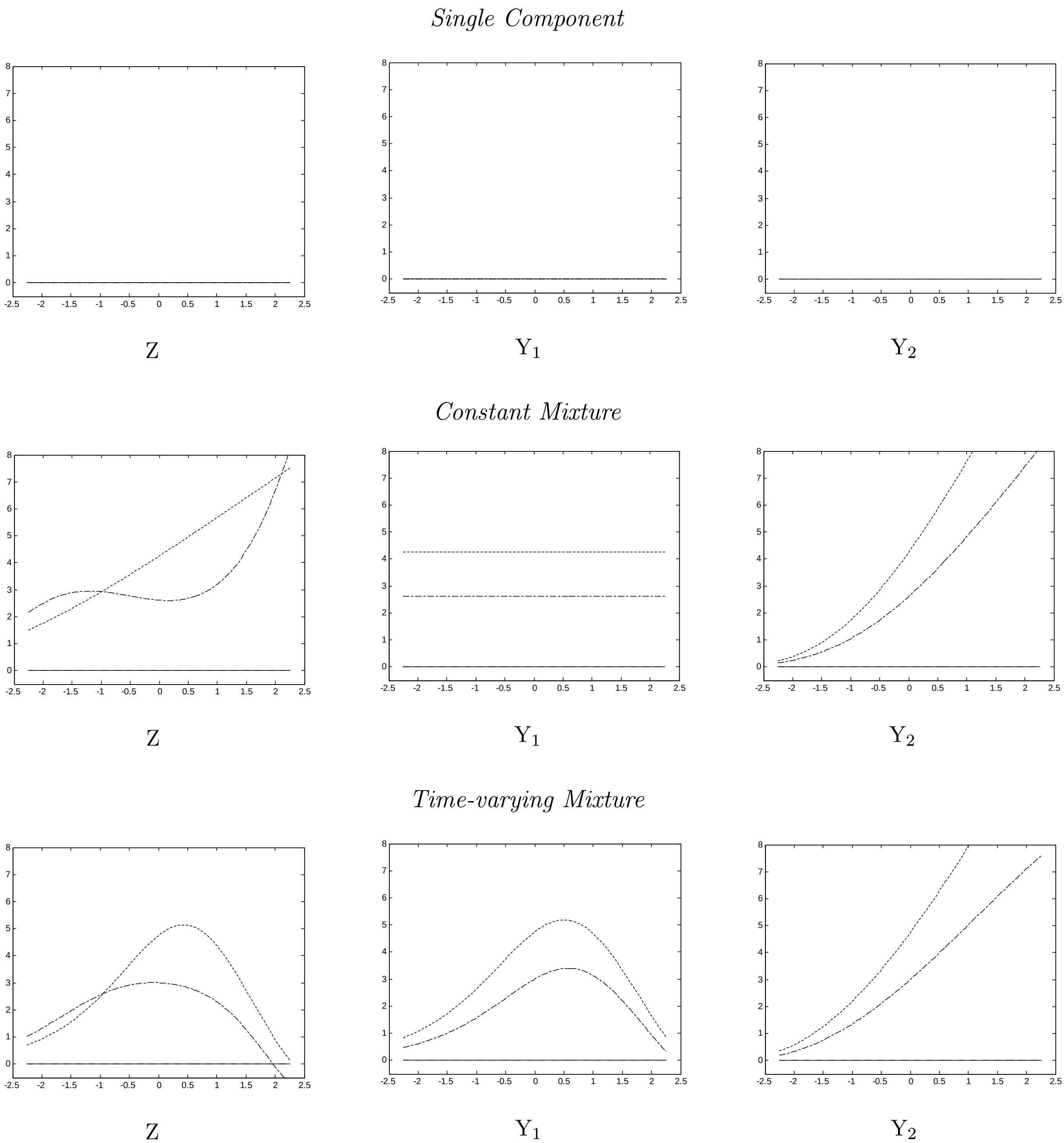

Figure 10: The graphs show the variation in excess kurtosis of the one-period rate as one of the conditioning factors changes and the other two are kept at their unconditional $\mathbb{P}$ mean. The factor that varies is respectively Z_t, Y_{1t}, and Y_{2t} in each column from left to right . The x-axes are measured in standard deviations from the $\mathbb{P}$ mean. In each figure values are shown for the linear model under $\mathbb{Q}$ (dash) and $\mathbb{P}$ (dot), and for the cubic model under $\mathbb{Q}$ (solid) and $\mathbb{P}$ (dash-dot).

Density in Direction of Second Component - Linear Mixture Models

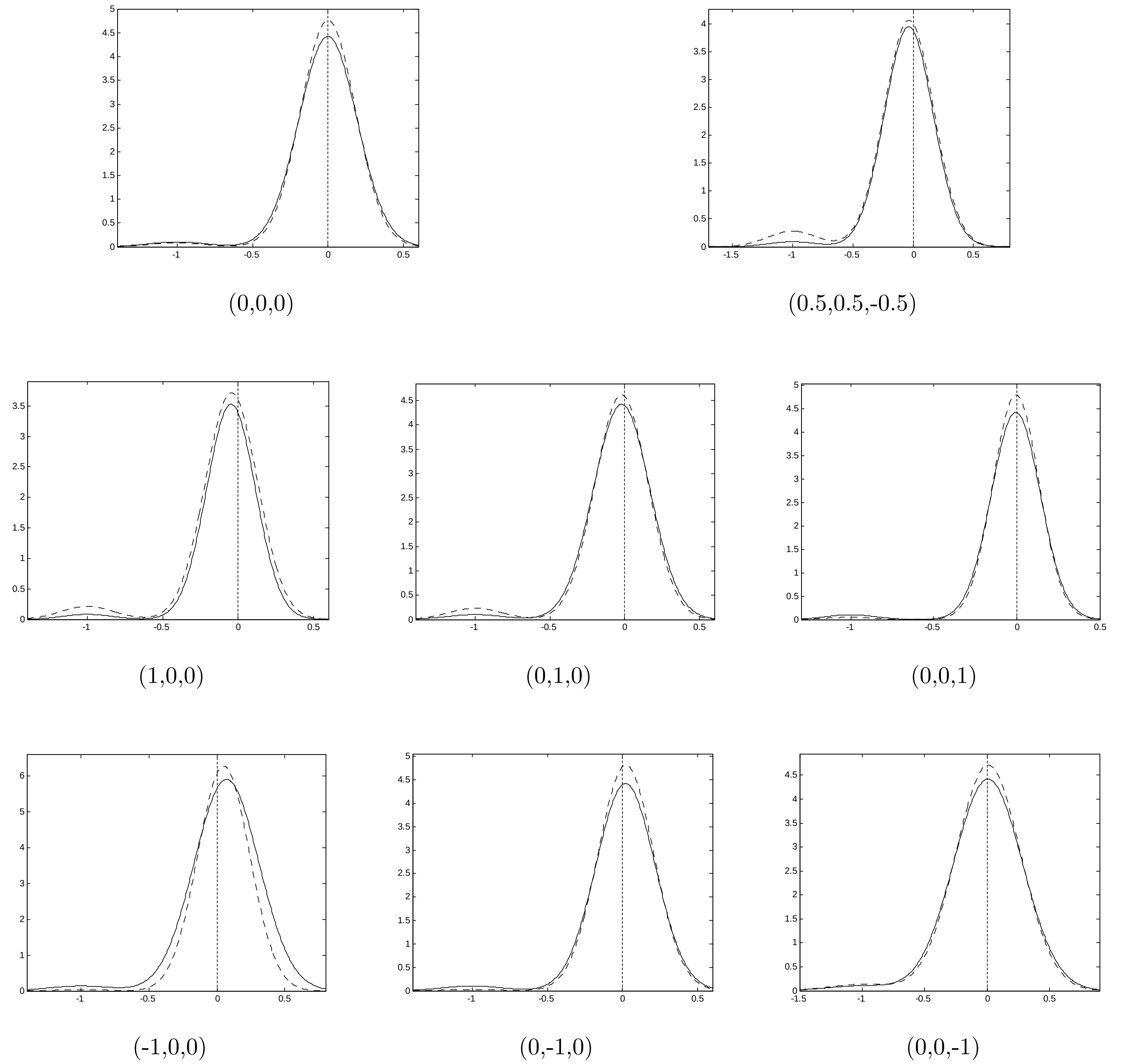

Figure 11: The figures illustrate how the second component shows up in the $\mathbb{P}$ density of factors. Thus the graphs draw the density of $X_{t+1}|X_t$ under $\mathbb{P}$ in direction of $E^{\mathbb{P}_2}(X_{t+1}|X_t)$. That is, for the vector $v = X_t - E^{\mathbb{P}_2}(X_{t+1}|X_t)$, the function $g(\alpha) = f^{\mathbb{P}}_{X_{t+1}|X_t}(X_{t+1} = X_t + \alpha v|X_t)$ is drawn, such that $\alpha = -1$ is the $\mathbb{P}$ density at the mean of the second component. This is done for the linear constant mixture model (solid) and the linear time-varying mixture model (dash). The 8 figures are for different values of the conditioning X_t, and the parenthesis below each figure indicate the position of each factor (Z, Y_1, Y_2) measured in standard deviations from its unconditional $\mathbb{P}$ mean. The values of X_t and the corresponding negative distance to the second component, v, are shown in Table 6.

Research Papers 2010

2009-57: Jesper Rangvid, Maik Schmeling and Andreas Schrimpf: Global Asset Pricing: Is There a Role for Long-run Consumption Risk?

2009-58: Olaf Posch: Risk premia in general equilibrium

2009-59: Christian M. Dahl and Emma M. Iglesias: Modelling the Volatility-Return Trade-off when Volatility may be Nonstationary

2009-60: Ole E. Barndorff-Nielsen, José Manuel Corcuera and Mark Podolskij: Limit theorems for functionals of higher order differences of Brownian semi-stationary processes

2010-01: Anders Bredahl Kock and Timo Teräsvirta: Forecasting with nonlinear time series models

2010-02: Gunnar Bårdsen, Stan Hurn and Zoë McHugh: Asymmetric unemployment rate dynamics in Australia

2010-03: Jesper Rangvid, Maik Schmeling and Andreas Schrimpf: Cash Flow-Predictability: Still Going Strong

2010-04: Helle Bunzel and Walter Enders: The Taylor Rule and "Opportunistic" Monetary Policy

2010-05: Martin M. Andreasen: Non-linear DSGE Models and The Optimized Particle Filter

2010-06: Søren Johansen and Bent Nielsen: Discussion of The Forward Search: Theory and Data Analysis by Anthony C. Atkinson, Marco Riani, and Andrea Ceroli

2010-07: Giuseppe Cavaliere, Anders Rahbek and A.M.Robert Taylor: Bootstrap Sequential Determination of the Co-integration Rank in VAR Models

2010-08: Peter R. Hansen and Asger Lunde: Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error

2010-09: Tom Engsted, Thomas Q. Pedersen and Carsten Tanggaard: Pitfalls in VAR based return decompositions: A clarification

2010-10: Torben G. Andersen and Luca Benzoni: Stochastic Volatility

2010-11: Torben B. Rasmussen: Affine Bond Pricing with a Mixture Distribution for Interest Rate Time-Series Dynamics