Term Structure Models During the Global Financial Crisis: A Parsimonious Text Mining Approach

Kiyohiko G. Nishimura
The University of Tokyo

Seisho Sato
The University of Tokyo

Akihiko Takahashi
The University of Tokyo

Term Structure Models During the Global Financial Crisis: A Parsimonious Text Mining Approach[*]

Kiyohiko G. Nishimura

National Graduate Institute for Policy Studies (GRIPS) and CARF, University of Tokyo

Seisho Sato

Graduate School of Economics and CARF, University of Tokyo

Akihiko Takahashi

Graduate School of Economics and CARF, University of Tokyo

October 26, 2018

Abstract

This work develops and estimates a three-factor term structure model with explicit sentiment factors in a period including the global financial crisis, where market confidence was said to erode considerably. It utilizes a large text data of real time, relatively high-frequency market news and takes account of the difficulties in incorporating market sentiment into the models. To the best of our knowledge, this is the first attempt to use this category of data in term-structure models.

Although market sentiment or market confidence is often regarded as an important driver of asset markets, it is not explicitly incorporated in traditional empirical factor models for daily yield curve data because they are unobservable. To overcome this problem, we use a text mining approach to generate observable variables which are driven by otherwise unobservable sentiment factors. Then, applying the Monte Carlo filter as a filtering method in a state space Bayesian filtering approach, we estimate the dynamic stochastic structure of these latent factors from observable variables driven by these latent variables.

As a result, the three-factor model with text mining is able to distinguish (1) a spread-steepening factor which is driven by pessimists' view and explaining the spreads related to ultra-long term yields from (2) a spread-flattening factor which is driven by optimists' view and influencing the long and medium term spreads. Also, the three-factor model with text mining has better fitting to the observed yields than the model without text mining.

Moreover, we collect market participants' views about specific spreads in the term structure and find that the movement of the identified sentiment factors are consistent with the market participants' views, and thus market sentiment.

1   Introduction and Summary

Although "market sentiment" is often regarded as an important driver of asset markets,[1] it is not explicitly incorporated in traditional empirical factor models for the term structure of interest rates. This is because (1) it is not clear what sentiment factors mean, and moreover, (2) there are scant observations, if any, about these sentiment factors. This work formulates and estimates a factor model with explicit sentiment factors in the period including the global financial crisis, in which uncertainty was said to be heightened considerably. It utilizes a large text data of real-time, relatively high-frequency market news and takes account of the difficulties (1) and (2). To the best of our knowledge, this is the first attempt to use this category of data in term-structure models.

With respect to (1) [what sentiment means], we regard the "sentiment" of market participants as their attitude toward changes in economic conditions with which they are concerned. We consider two types: pessimists (taking the change as negative so as to lower the short rate) and optimists (taking it as positive to raise the rate). Pessimists and optimists are different with respect to their concerns. As time passes, initial sentiment impacts are eventually dying out. Thus, the pessimist factor initially reduces the short rate, and then increases it in the future. This implies the pessimist factor is spread-steepening. Symmetrically, the optimist factor is spread-flattening. This argument suggests the following three-factor model of term structure: A spread-steepening factor which is negative quadratic-Gaussian with zero-mean reversion (assuring initial negative impacts eventually dying out), a spread-flattening factor which is positive quadratic-Gaussian with zero mean-reversion (assuring initial positive impacts eventually dying out), and a level factor which is traditional Gaussian.[2] Appendix B contains an intuitive version of this theoretical model.

With respect to (2) [un-observability of sentiment factors], we use a text mining approach to generate "observable variables" which are driven by otherwise unobservable particular sentiment factors. To start with, we have pointed out that pessimists and optimists may have different concerns: Optimists may be concerned with the state of the current business cycle and its effect on interest rates, while pessimists may be more concerned with various long-term risks such as fiscal sustainability and uncertainty about the future of global economy and their effects on the rates. In this example, the pessimist spread-steepening factor is likely correlated with the frequency of news about what pessimists are concerned: fiscal sustainability and global risks. Similarly, the optimist spread-flattening factor and the frequency of news about business-cycle events may correlate to a high degree. The above argument indicates that the frequency of the relevant news can be considered as the observable variable driven by a particular sentiment factor.

To identify relevant news for a particular sentiment factor, we employ a parsimonious approach of finding one key-word group highly correlated with (the preliminary estimate of) the factor. Comprehensive text data of real-time financial news reports are collected from the beginning of 2008 to the end of 2011 (including Global Financial Crisis and

---

[1]For example, from the beginning of the Global Financial Crisis, some policymakers at central banks clearly identified the importance of sentiment factors or market confidence on asset markets' pricing and trading. See Nishimura (2008).

[2]In a companion paper now in progress, Nishimura, Saito and Takahashi develop theoretical foundation of this empirical model based on the theory of fundamental uncertainty called Knightian uncertainty. See Nishimura and Ozaki (2017)for the formulation of pessimism and optimism under Knightian uncertainty.

European Sovereign Debt Crisis) from Reuters JAPAN, which makes them available online for free access. Firstly, we identify words which appear frequently in the news and seem relevant to the bond market. Secondly, we calculate the frequency of a combination (group) of these key words. Finally, we select a key word group which is economically meaningful and strongly positively correlated to the preliminary estimate (without text mining) of the unobserved sentiment factor. Then,

- For the pessimist spread-steepening factor, the selected key word group is "fiscal conditions (zaisei)" and "foreign (gaikoku)".

- For the optimist spread–flattening factor: the selected key word group is "business conditions (keiki)" and "slowdown (gensoku)".

Thus, we have two new observable variables, which are the appearance frequency of the selected key words, in addition to traditional observable variables, which are observed interest rates.

The dynamic stochastic structure of these latent factors (pessimist spread-steepening, optimist spread-flattening and level) are estimated from traditional and new observable variables (traditional: observed interest rates and new: word frequencies), in a state space Bayesian filtering approach. Specifically, the Monte Carlo filter (as a particle filtering method) is used. [3]

Then, we obtain the following main results:

1. The three-factor model with text mining is able to distinguish a factor explaining the ultra-long[4] and long term spreads (e.g. 20-10 year) from a factor influencing the long and medium term spreads (e.g. 10-2 year), in contrast to the fact that they are hard to distinguish in the model without text mining. Moreover, RMSEs (square-root of the mean square errors) of the former (with text mining) are smaller than the latter (without it).

2. The factor influencing the ultra-long and long term spreads is a pessimist spread-steepening factor which concerns news containing "fiscal conditions" and "foreign", while the factor explaining the long - medium term spread is an optimist spread-flattening factor which concerns news containing "business conditions" and "slow down".

3. Thus, we are able to identify two spread-determining factors influencing different parts of the term structure of interest rates: optimist spread-flattening on the long and medium term spreads (e.g. 10-2 year) and pessimist spread-steepening on the ultra-long and long term spreads (e.g. 20-10 year).

To assess the results, we collect market participants' views about specific spreads in the term structure. We find that the movement of the identified sentiment factors are consistent with the market participants' views, and thus market sentiment.

---

[3]The Monte Carlo method is proposed by Kitagawa (1996). See Takahashi and Sato (2001) for its application to estimation of a term structure model. See also Fukui et al. (2017) for its application in other fields.

[4]In Japan, government bonds of maturity longer than ten years are called "super-long term" bonds, while in other countries they are often called "ultra-long term" bonds. In the following, we use "ultra-long", and put "super" in parentheses if necessary.

There are several studies in the related literature utilizing financial news.[5] In the behavioral finance literature, Gotthelf and Uhl (2018) uses a news-sentiment variable in the regression analysis, rather than a dynamic factor analysis of this paper in the term structure of interest rates. Their sentiment variable is based on Thomson Reuters News Analytics's sentiment measures about political, debt and monetary news, while the dependent variables are 1 & 3-month and 1,2,3,5,10,30-year U.S. treasury yields.[6] Thomson Reuters' variables are constructed from the assessment of particular news as positive, neutral or negative by their own text mining procedure, and do not focus on what set of economic conditions the market participants focus.

In the macro-finance literature, Bauer (2015) investigates impacts of macro economic news on yield curves with a simple no-arbitrage model by using daily interest rate futures and treasury yields data. He uses news data to classify each day to different "regimes" of the model according to the type of the news released on the day. He does not utilize news data to identify a dynamic (sentiment or non-sentiment) factor as we do.

The organization of the paper is as follows: After the next section (Section 2) presents an overview of our model, Section 3 explains a preliminary analysis that consists of estimation of two-factor term structure models, a text mining of financial and economic Japanese news provided by Reuters Japan from January 1st 2008 to December 31th 2011 (4 years), and selection of two-word sets relevant with steepening and flattening factors in the term structure models. Section 4 shows and discusses our main result of simultaneous estimation of a three-factor model together with two-word sets frequencies. Concluding remarks are in Section 5. Appendix A, B and C briefly explain a Monte Carlo filter algorithm used in the empirical analysis, interest rate models with market sentiment which may be considered as theoretical foundation of the models in this paper, and a new estimation method for trends and correlations embedded in noisy time series data utilized in our text mining analysis, respectively.

2 Outline of Model

This section explains the outline of our model. Following the argument in Introduction, we adopt as a term structure model a three-factor Gaussian Quadratic-Gaussian interest rate model,[7] where an instantaneous short rate is expressed as a sum of independent Gaussian (x_3) and ($\mp$) Quadratic-Gaussian (x_1^2 & x_2^2) factors:

$$r_t = c_1 x_{1,t}^2 + c_2 x_{2,t}^2 + x_{3,t};\ 0 > c_1 \geq -\left(\frac{\kappa_1}{\sigma_1}\right)^2,\ c_2 > 0, \tag{1}$$

Here, we assume that under a risk-neutral probability measure $\mathbb{Q}$, $x_{j,t}$ ($j = 1, 2$) follow Gaussian with zero mean-reversion processes, and $x_{3,t}$ does a Gaussian with no mean

---

[5]For surveys on applications of text mining in general to finance that include predictions of stock prices and foreign exchanges, see Kumar and Ravi (2016) and Nassirtoussi et al. (2014), for instance.

[6]They argue their news sentiment factor is distinct from the three yield curve factors such as level, slope, and curvature as well as from fundamental macroeconomic variables. However, although the t-statistics values for the news sentiment factor's coefficients look significant for the yields up to the 3-5 year maturities, the adjusted R^2s in their regression analyses are low: 0.02–0.17 with only the news sentiment factor as an explanatory variable, and 0.10–0.25 with macro variables (nonfarm payrolls, CPI, ISM Manufacturing Index) and the 10-2 year spread in addition to the news sentiment variable.

[7]See Appendix B.2 for a Quadratic-Gaussian model.

reversion process, which are the solutions of the stochastic differential equations (SDEs) as follows: given $x_{j,0}(j = 1, 2, 3)$,

$$dx_{j,t} = -\kappa_j x_{j,t} dt + \sigma_{x,j} dB_{j,t}^{\mathbb{Q}}; \ \kappa_j > 0, (j = 1, 2) \tag{2}$$
$$dx_{3,t} = \lambda dt + \sigma_{x,3} dB_{3,t}^{\mathbb{Q}},$$

where $\kappa_j(j = 1, 2)$, λ, $\sigma_{x,j}(j = 1, 2, 3)$ are positive constants, and $B_{j,t}^{\mathbb{Q}}(j = 1, 2, 3)$ are independent Brownian motions under a risk-neutral probability measure $\mathbb{Q}$. We remark that a condition $c_1 \geq -\left(\frac{\kappa_1}{\sigma_1}\right)^2$ $(\sigma_1 \equiv \sqrt{2}\sigma_{x,1})$ is necessary for a zero coupon bond price to be well-defined.

We also suppose that each risk-premium associated with $x_j(j = 1, 2)$ is zero, while the one with x_3 is λ. That is, under the physical probability measure $\mathbb{P}$ with independent Brownian motions $B_j^{\mathbb{P}}(j = 1, 2, 3)$, $x_j(j = 1, 2, 3)$ are the solutions to the SDEs: given $x_{j,0}(j = 1, 2, 3)$,

$$dx_{j,t} = -\kappa_j x_{j,t} dt + \sigma_{x,j} dB_{j,t}^{\mathbb{P}}; \ \kappa_j > 0, (j = 1, 2) \tag{3}$$
$$dx_{3,t} = \sigma_{x,3} dB_{3,t}^{\mathbb{P}},$$

which are the basis of system equations in a state space model used for our empirical analysis in the following sections.

Moreover, we note that it is well-known that the factor x_3 is expected to stand for an overall yield curve level.[8] So, we call x_3 the level factor. In contrast, we see that as $x_{j,t}^2$ $(j = 1, 2)$ have zero mean-reversion levels, $c_1 x_1^2$ with $c_1 < 0$ ($c_2 x_2^2$ with $c_2 > 0$) has a steepening (flattening) effect on a yield curve. Thus we call x_1 and x_1^2 the steepening factor and x_2 and x_2^2 the flattening factor. In particular, x_1^2 and x_2^2 will turn out to be crucial to select meaningful word sets from those obtained through text mining introduced in Section 3.2, since steepening and flattening of a yield curve reflect and are closely related to market participants' different views for the current and future economic conditions.

Then, we obtain time-t zero coupon bond price and zero yield with maturity τ, denoted respectively by $P_t(\tau)$ and $Y_t(\tau)$ as follows[9]:

$$
\begin{aligned}
P_t(\tau) &= E_t^Q\left[e^{-\int_t^{t+\tau} r_u du}\right] \tag{4} \\
&= \exp\left[-\tau\{X_{1,t}(\tau) + X_{2,t}(\tau) + X_{3,t}(\tau)\}\right], \\
Y_t(\tau) &= X_{1,t}(\tau) + X_{2,t}(\tau) + X_{3,t}(\tau), \tag{5}
\end{aligned}
$$

where

$$X_{j,t}(\tau) = \frac{-1}{\tau}\left[A_j(\tau) + C_j(\tau)x_{j,t}^2\right], \ (j = 1, 2) \tag{6}$$

$$X_{3,t}(\tau) = x_{3,t} + \frac{\lambda}{2}\tau - \frac{\sigma_3^2}{6}\tau^2, \ (\sigma_3 \equiv \sigma_{x,3}). \tag{7}$$

---

[8]For example, see Section 6 in Hull and White (1990) and Section 4.2 in Nakano et al.(2018).

[9]See Appendix B.2 in this paper and Appendix in Nakano et al.(2018) for outlines of the derivation of (6) and (7), respectively.

Here, with $c_1 \in [-\kappa_1^2/\sigma_1^2, 0)$, $c_2 > 0$ and $\sigma_j \equiv \sqrt{2}\sigma_{x,j}$ $(j = 1, 2)$, $C_j(\tau)$ $(j = 1, 2)$ are defined as

$$C_j(\tau) = C_{0j} + \frac{1}{z_j(\tau)}, \tag{8}$$

$$C_{0j} = \frac{\kappa_j + \sqrt{\kappa_j^2 + c_j\sigma_j^2}}{\sigma_j^2}, \tag{9}$$

$$z_j(\tau) = \frac{\sigma_j^2}{\alpha_j} - e^{\alpha_j \tau}\left(\frac{1}{C_{0j}} + \frac{\sigma_j^2}{\alpha_j}\right), \tag{10}$$

$$\alpha_j = 2(\kappa_j - \sigma_j^2 C_{0j}), \tag{11}$$

and $A_j(\tau), j = 1, 2$ as

$$A_j(\tau) = \frac{\sigma_j^2}{2}\int_0^\tau C_j(s)ds = \frac{1}{2}\left\{\left(\kappa_j + \frac{\alpha_j}{2}\right)\tau + \ln\frac{(1 + C_{2j})}{(1 + C_{2j}e^{\alpha_j \tau})}\right\} \tag{12}$$

with $C_{2j} = \frac{\alpha_j}{\sigma_j^2}z_j(0) - 1$ and $z_j(0) = \frac{-1}{C_{0j}}$.

Given the three-factor model, our aim is to obtain (two-)word sets meaningful in terms of economics and finance, which are able to stand for factors embedded in our interest rate model. To attain such an objective, we construct and estimate a state space model with state equations for the factors $x_j (j = 1, 2, 3)$ and observation equations for zero yields $Y_t(\tau)$ and two-word sets' frequencies relevant with steepening factor (x_1^2) and flattening factor (x_2^2), of which details will be given in Section 4.

Before implementing simultaneous estimation of the three factors and relevant two-word sets, we adopt a preliminary analysis that takes the following steps of which details will be shown in Section 3:

1. Estimate factors in the term structure model with observation equations only for the zero yields in our state space model (i.e., without text mining).

2. Find meaningful two-word sets based on a text mining and a regression analysis, whose frequencies have high correlations with the estimated factors in Step 1.

We firstly remark that, in Step 1 above, we are unable to use our three-factor model, because the estimation of the three-factor model without text mining turns out to be unstable, (i.e. different seeds in Monte Carlo simulations produce quite different state estimates of x_j (j=1,2)), mainly due to high negative correlation between the estimated steepening and flattening factors. In fact, we will see in Section 3.1 that the estimated correlation is close to -1 in the two factor models introduced below. Thus, instead of the three-factor, we consider two-factor models with steepening & level factors $(x_1 \& x_3)$ and flattening & level factors $(x_2 \& x_3)$ for a preliminary analysis in the next section. More concretely, setting $c_1 \neq 0$ and $c_2 = 0$ or $c_1 = 0$ and $c_2 \neq 0$ in the equation (1) provide two-factor models with

$$r_t = c_1 x_{1,t}^2 + x_{3,t};\ 0 > c_1 \geq -\left(\frac{\kappa_1}{\sigma_1}\right)^2, \tag{13}$$

6

or

$$r_t = c_2 x_{2,t}^2 + x_{3,t}; \ c_2 > 0, \tag{14}$$

respectively.

Hereafter, let us call the model with instantaneous short rate (13) as "two-factor Gaussian minus Quadratic-Gaussian" that stands for steepening & level factors in the term structure of interest rates, and the model with (14) as "two-factor Gaussian plus Quadratic-Gaussian" that represents flattening & level factors. The corresponding zero yield with term τ is given by

$$Y_t(\tau) = X_{1,t}(\tau) + X_{3,t}(\tau) \tag{15}$$

for "two-factor Gaussian minus Quadratic-Gaussian", and

$$Y_t(\tau) = X_{2,t}(\tau) + X_{3,t}(\tau) \tag{16}$$

for "two-factor Gaussian plus Quadratic-Gaussian", where $X_{j,t}(\tau)(j = 1, 2)$ and $X_{3,t}(\tau)$ are defined as the equations (6) and (7), respectively.

3 Preliminary Analysis

3.1 Estimation of Two-factor Term Structure Model

This subsection shows the result for estimation of our two-factor term structure model by using daily data of the Japanese Government Bond (JGB) zero yields during January 4th 2008 and December 30th 2011, where the zero yields are estimated from par rate data available on the website of Ministry of Finance (MOF) Japan, with a cubic spline and a bootstrap method.[10] Particularly, we apply a state space model, whose system and observation equations are described as follows:

(State space model for two-factor term structure model)

- ("two-factor Gaussian minus Quadratic-Gaussian")
  (System equation)

$$x_{1,t} = e^{-\kappa_1 \Delta t} x_{1,t-\Delta t} + \frac{\sigma_1}{\sqrt{2}} \sqrt{\frac{1 - e^{-2\kappa_1 \Delta t}}{2\kappa_1}} \epsilon_{1,t}, \tag{17}$$

$$x_{3,t} = x_{3,t-\Delta t} + \sigma_3 \sqrt{\Delta t} \epsilon_{3,t}, \ (\Delta t = 1/250), \tag{18}$$

(Observation equation)

$$Y_t(\tau) = X_{1,t}(\tau) + X_{3,t}(\tau) + e_{t,\tau}(\tau = 2, 5, 10, 20, 30 \ (\text{year})) \tag{19}$$

---

[10]See, for example, Nakano et al. (2018) for the details.

- ("two-factor Gaussian plus Quadratic-Gaussian")
  (System equation)

$$x_{2,t} = e^{-\kappa_2 \Delta t} x_{2,t-\Delta t} + \frac{\sigma_2}{\sqrt{2}} \sqrt{\frac{1 - e^{-2\kappa_2 \Delta t}}{2\kappa_2}} \epsilon_{2,t}, \tag{20}$$

$$x_{3,t} = x_{3,t-\Delta t} + \sigma_3 \sqrt{\Delta t} \epsilon_{3,t}, \ (\Delta t = 1/250), \tag{21}$$

(Observation equation)

$$Y_t(\tau) = X_{2,t}(\tau) + X_{3,t}(\tau) + e_{t,\tau} (\tau = 2, 5, 10, 20, 30 \text{ (year)}) \tag{22}$$

Here, with $A_j(\tau)$ and $C_j(\tau)$ defined by (8) and (12),

$$X_{j,t}(\tau) = \frac{-1}{\tau} \left[A_j(\tau) + C_j(\tau) x_{j,t}^2 \right], (j = 1, 2) \tag{23}$$

$$X_{3,t}(\tau) = x_{3,t} + \frac{\lambda}{2} \tau - \frac{\sigma_3^2}{6} \tau^2. \tag{24}$$

Also, $t = 0$ stands for Jan. 4th 2008, and the initial $x_{j,-\Delta t}$ $j = 1, 2, 3$ are determined by realization of random variables following normal distributions whose means and variances are determined so that the likelihood in estimation becomes improved. Moreover, we assume $\epsilon_{i,t} \sim$ i.i.d. $N(0,1), i = 1, 2, 3$ and $e_{t,j} \sim$ i.i.d. $N(0, \gamma_j^2), (j = 2, 5, 10, 20, 30)$. See Section 4.1 for the details of our estimation procedure including how to set γ_j^2.

The estimation result is listed below: Table 1 gives the estimated parameters. Figure 1 shows the time series of the estimated factors. Figure 2 presents the time series of the observed & estimated zero yields, and Table 2 provides RMSEs (square-root of mean squared errors) of our zero yields' estimates.

	"two-factor Gaussian minus Quadratic-Gaussian"	"two-factor Gaussian plus Quadratic-Gaussian"
κ_1	0.1290	–
σ_1	0.0402	–
κ_2	–	0.1290
σ_2	–	0.0253
λ	0.0042	0.0040
σ_3	0.0149	0.0147
c_1	-4.230	–
c_2	–	2.120
log-likelihood	16631.05	17077.11

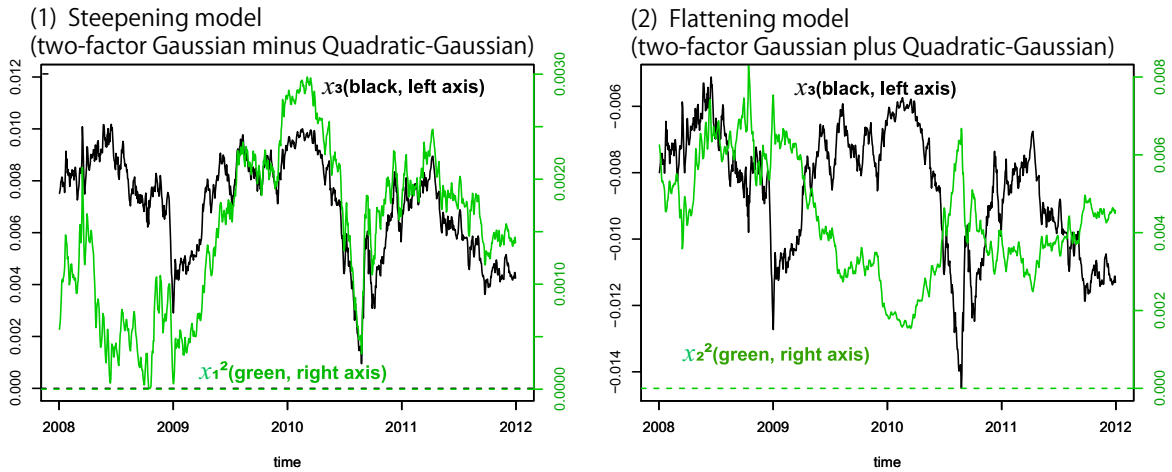Table 1: Estimates of parameters: two factor model

(1) Steepening model
(two-factor Gaussian minus Quadratic-Gaussian)

(2) Flattening model
(two-factor Gaussian plus Quadratic-Gaussian)

Figure 1: Estimated factors of two factor models

(1) Steepening model
(two-factor Gaussian minus Quadratic-Gaussian)

(2) Flattening model
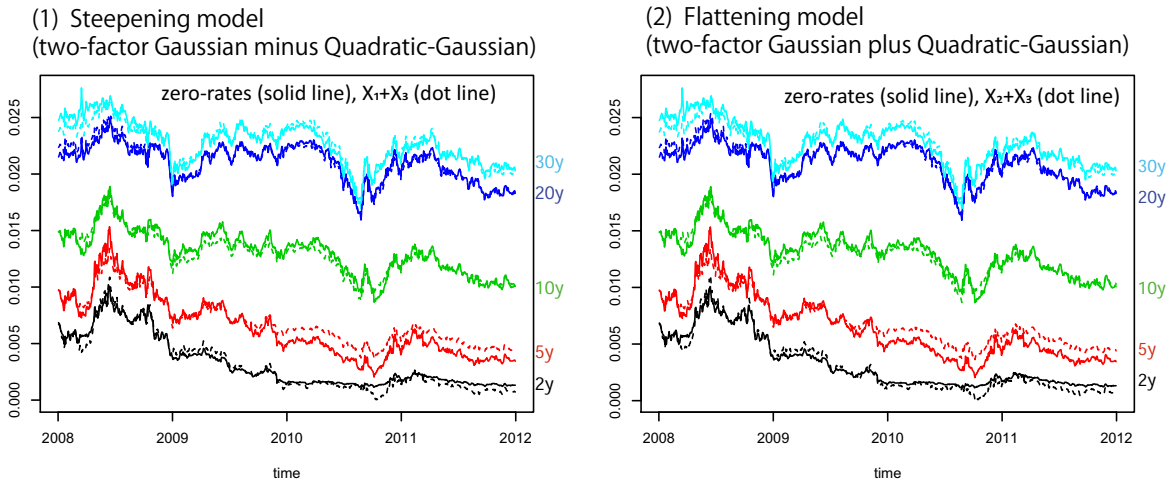(two-factor Gaussian plus Quadratic-Gaussian)

Figure 2: Observed zero rates (solid lines) and estimated zero rates (dot lines) by two-factor models

	2y	5y	10y	20y	30y	total average
"two-factor Gaussian minus Quadratic-Gaussian"	4.52	9.00	5.35	3.57	5.31	5.85
"two-factor Gaussian plus Quadratic-Gaussian"	4.54	8.94	5.27	3.54	5.31	5.82

Table 2: Square-root of mean squared errors (RMSEs) of estimated yields(basis points(bps)): two factor model

In Figure 1, we observe that the estimates of the level factor x_3 have very similar shapes in both "two-factor Gaussian minus Quadratic-Gaussian" and "two-factor Gaussian plus Quadratic-Gaussian" models. (In fact, the correlation is close to one (more than 0.99).) In each model, the risk premium (λ) is positive, that is a term premium exists, which is about 40 basis points (bps) per year, and the volatility (σ_3) is around 150 bps.

9

However, the mean level of x_3 is positive for "two-factor Gaussian minus Quadratic-Gaussian" and negative for "two-factor Gaussian plus Quadratic-Gaussian", because given the same observed yields, x_1^2 (x_2^2) affect the interest rates, negatively (positively). (i.e. $c_1 < 0$ in (13) and $c_2 > 0$ in (14)).

As for steepening and flattening factors ((x_1 & x_2 respectively), we first note that the x_1^2's level is about a third of the x_2^2's level, while the absolute value of c_1 ($|c_1| = 4.23$) is around twice that of c_2 ($|c_2| = c_2 = 2.12$). Hence, the flattening factor x_2^2 has larger (about 1.4 times absolute) effects on the interest rates. (See the short rate equations (13) and (14).) It is also observed that the mean reversion speeds ($\kappa_j (j = 1, 2)$) are almost the same (around 0.129), and that the estimated steepening factor x_1^2 has a quite high negative correlation (close to -1) with the the estimated flattening factor x_2^2, as expected.

In contrast, the volatility ($\sigma_1 = 402$ bps) for the steepening factor is around 1.6 times as large as the one ($\sigma_2 = 253$ bps) for the flattening factor, possibly because it is easier for the model with flattening factor ("two-factor Gaussian plus Quadratic-Gaussian") to be fitted to the observed yields than the one with steepening factor ("two-factor Gaussian minus Quadratic-Gaussian") for this estimation period. In other words, steepening factor needs a higher volatility for well fitting to the observations.

Also, the difference of the estimated volatilities is partially reflected in the difference of the log-likelihoods between two models, which is not negligible ($16,631$ vs $17,077$), since standard deviations of Monte Carlo simulations used for calculation of these log-likelihoods fall within 10.

Finally, we observe in Figure 2 and Table 2 that the fittings to the observed yields in "two-factor Gaussian minus Quadratic-Gaussian" & "two-factor Gaussian plus Quadratic-Gaussian" are almost the same, but the "two-factor Gaussian plus Quadratic-Gaussian" is slightly better, where the average of RMSE (square-root of mean square errors) is 5.82 vs 5.85 bps, and the fitting to the 20 year yield is the best (3.54 vs 3.57 bps), while the fitting to the 5 year yield is the worst (8.94 vs 9.00 bps).

As seen before, the correlation between the estimated steepening and flattening factors (x_1^2, x_2^2, respectively) is extreme: the correlation is very close to -1. This suggests that a statistical (three-factor) model with the two factors at the same time in addition to the level factor x_3, is not likely to be stably estimated. However, the volatility differs considerably between the two (as seen before in Table 1), suggesting they are not the same (i.e., the steepening factor is not equal to the inverted flattening factor). Thus, it may be possible to distinguish between the two by including additional information, which is news reports that may correlate with the two factors differently. This line of reasoning leads us to the next part of a text mining.

3.2 Text Minning

This subsection analyzes financial and economic Japanese news provided by Reuters Japan from January 1st 2008 to December 31th 2011 (4 years), where the total number of news is 265,466. First, we construct a word group by the following procedure.

1. We decompose each news into words by using MeCab software[11] for each year, and select only nouns, though they include the ones with adjectival usage. Then, we remove only-one-character words and numbers.

---

[11]See MeCab (2006) for the description of the software.

2. For each word, we count the number of news where the word appears.

3. In each year we extract top 2,000 words in terms of the number of news.

4. We select 2,474 words which appear in top 2,000 words for more than or equal to 2 years during the 4 years.

5. Moreover, from above 2,474 words we select 645 words which seem relevant for representing steepening and flattening factors in our term structure model.

6. We select these 645 words as components of our word groups in the subsequent research.

The final goal of this text mining is to find a meaningful two-word[12] combination whose frequency is likely to be linked to a steepening factor and flattening one, respectively. To achieve this goal, we calculate the frequencies of word sets which are combinations of above 645 words described above. First, let $F_{i,t}(A)$ stand for appearance frequency of one word A in the i-th news of the day-t, where day-t is set to be from 15:00 of the previous business day to 14:59 of the day t. Then, let us define two words (A, B) appearance frequency, $F_{i,t}(A$ and $B)$ as $F_{i,t}(A$ and $B) = \min(F_{i,t}(A), F_{i,t}(B))$. Then, define $F_t(S)$, the frequency of the day-t of a word set S (e.g. $S = A$ and B) as $F_t(S) = \sum_{i=1}^{n_t} F_{i,t}(S)$, where n_t denotes the total number of the i-th news appearing in the day-t.

3.3 Selection of Two-word Sets relevant with Steepening and Flattening Factors in Term Structure Model

Next, to find two word sets (A and B) which have high correlations with the steepening factor x_1^2 and the flattening factor x_2^2 estimated by the two-factor interest rate models in Section 3.1, we implement a regression analysis, and calculate standard sample correlations for the levels of those series. Additionally, we calculate SIML correlations proposed by Kunitomo et al. (2018) in order to remove possible spurious correlations that might be caused by non-stationarity found in daily data. [13]

Our regression analysis is based on the following equations:

$$\log(F_t(A \text{ and } B) + 1) = a_1 + b_1 x_{1,t}^2 + e_{1,t} \tag{25}$$

or

$$\log(F_t(C \text{ and } D) + 1) = a_1 + b_2 x_{2,t}^2 + e_{2,t} \tag{26}$$

where $e_{1,t}, e_{2,t}$ are i.i.d. noises.

Figure 3-4 show that the time series of actual and estimated $\log(F_t(\cdot \text{ and } \cdot) + 1)$ for economically/financially meaningful two-word sets, which have relatively high correlations with x_j^2 ($j = 1, 2$) estimated by our two-factor term structure models in Section 3.1. In these figures, we denote standard sample correlations by "cor1" and SIML correlations

---

[12]We have also examined one-word groups. It has turned out that one-word groups usually lack sufficient detail. So we use two-word groups.

[13]See Appendix C for a new estimation method for trends and correlations in non-stationary noisy time series data.

cor_{SIML} by "cor2". The black bars and the red solid line show log-frequencies and the regressed estimates by each factor, respectively. We select two-word sets, "fiscal conditions (zaisei) & foreign (gaikoku)" for the factor x_1^2 (steepening factor) and "business conditions (keiki) & slowdown (gensoku)" for the factor x_2^2 (flattening factor), which explicitly contain more financial and economic meaning than the other two-word sets. Moreover, both two-word sets "fiscal conditions (zaisei) & foreign (gaikoku)" and "business conditions (keiki) & slowdown (gensoku)" have the highest standard and SIML correlations among three kinds of two-word sets shown in Figure 3-4.
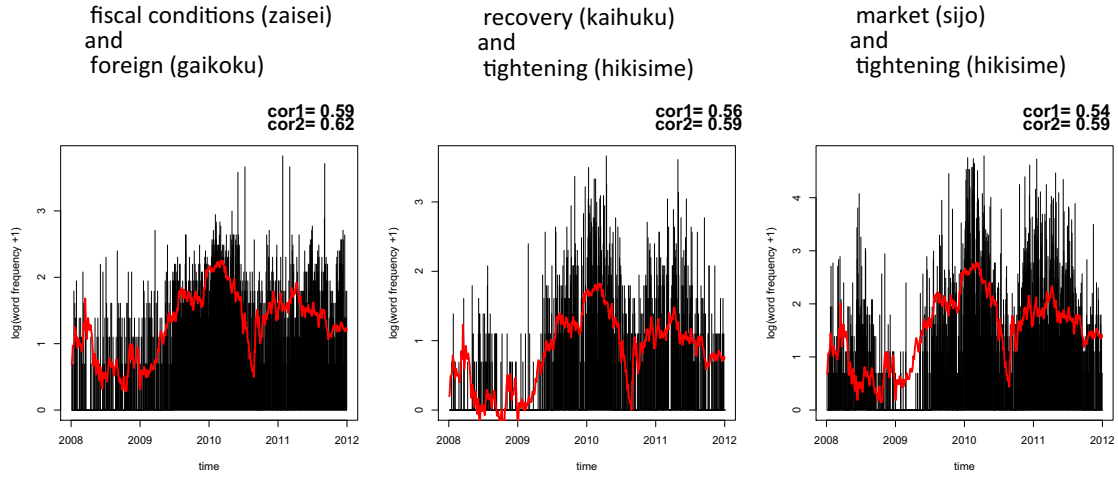

Figure 3: Actual(black) and estimated(red) two-word set's log-frequencies using "two-factor Gaussian minus Quadratic-Gaussian model" (level factor + steepening factor)
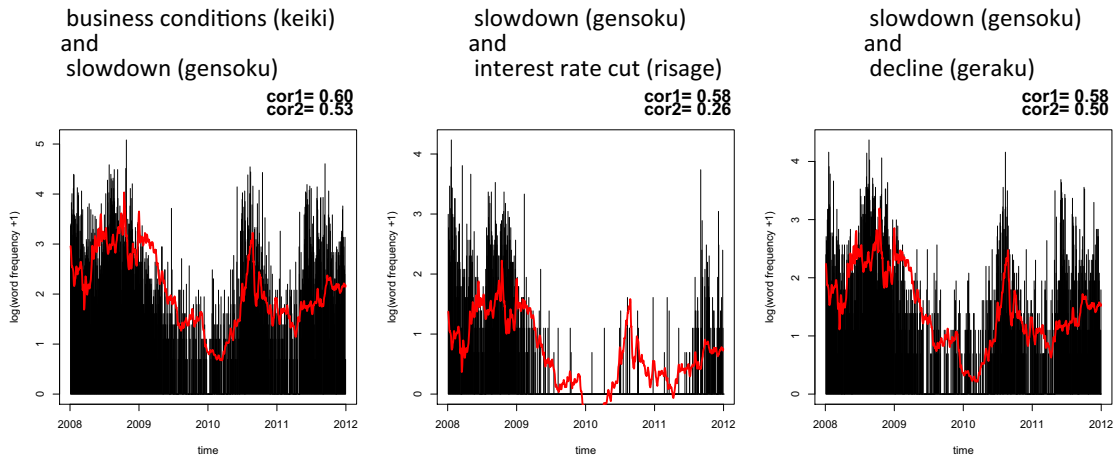

Figure 4: Actual(black) and estimated(red) two-word set's log-frequencies using "two-factor Gaussian plus Quadratic-Gaussian model" (level factor + flattening factor)

4 Simultaneous Estimation of Three-factor Term Structure Model and Two-word Set Frequency

This section implements estimation of the three-factor ("three-factor Gaussian Quadratic-Gaussian") term structure model (whose short rate is defined as (1)) with frequencies of the two-word sets selected in Section 3.3. Again, we apply a state space model whose system and observation equations are described as follows:

(State space model for three-factor term structure model and two-word set frequency)

(System equation)

$$x_{i,t} = e^{-\kappa_i \Delta t} x_{i,t-\Delta t} + \frac{\sigma_i}{\sqrt{2}} \sqrt{\frac{1 - e^{-2\kappa_i \Delta t}}{2\kappa_i}} \epsilon_{i,t} \ (i = 1, 2) \tag{27}$$

$$x_{3,t} = x_{3,t-\Delta t} + \sigma_3 \sqrt{\Delta t} \epsilon_{3,t} \tag{28}$$

where $\epsilon_{i,t} \sim i.i.d.N(0,1), i = 1, 2, 3$, and we set Δt as $1/250$.

(Observation equation)

$$Y_t(\tau) = X_{1,t}(\tau) + X_{2,t}(\tau) + X_{3,t}(\tau) + e_{t,\tau}, \tag{29}$$

$$X_{i,t}(\tau) = \frac{-1}{\tau} \left[A_i(\tau) + C_i(\tau) x_{i,t}^2 \right], (i = 1, 2) \tag{30}$$

$$X_{3,t}(\tau) = x_{3,t} + \frac{\lambda}{2}\tau - \frac{\sigma_3^2}{6}\tau^2 \tag{31}$$

$$\log(F_t(A \text{ and } B) + 1) = \xi_{c,1} + \xi_1 x_{1,t}^2 + e_{t,w1} \tag{32}$$

$$\log(F_t(C \text{ and } D) + 1) = \xi_{c,2} + \xi_2 x_{2,t}^2 + e_{t,w2}, \tag{33}$$

where $C_i(\tau)$ and $A_i(\tau)$ are defined by (8) and (12), respectively; The two-word sets are given as ($A =$ "fiscal conditions (zaisei)", $B =$ "foreign (gaikoku)") and ($C =$ "business conditions (keiki)", $D =$ "slowdown (gensoku)"). Here, we assume $e_{t,j} \sim i.i.d.N(0, \gamma_j^2), (j = 2, 5, 10, 20, 30, w1, w2)$. Moreover, we fix $c_j(j = 1, 2)$ as $c_1 = -4.23, c_2 = 2.12$, the estimates obtained by the two-factor model in Section 3.1, because it is quite difficult to estimate $c_j(j = 1, 2)$ at the same time with the other parameters in the three-factor model. We also note that the regression equations in the observation model are equivalent to those in (25) and (26).

4.1 Procedure for Parameter Estimation

This subsection shows the detailed procedure in estimation of our model. In sum, our estimation consists of the following three parts.

- Apply Monte Carlo filter with 10,000 particles to estimate state variables x_1, x_2, x_3. (See Appendix A for a summary of our Monte Carlo filter algorithm.)

- Use a maximum likelihood (ML) method with a *stochastic grid search* to estimate parameters other than variances of observation noises.

- Set variances of observation noises.

Particularly, we note that applying maximum likelihood (ML) method to estimation of the variances of observation noises suffers substantial computational burden to obtain reasonable estimates, since they have large impact on the likelihood as well as on the other parameters' estimates. Thus, we set the variances of observation noises separately and use ML method for estimation of the other parameters.

Next, let us describe a *stochastic grid search* for estimation of parameters, λ, $\sigma_i (i = 1, 2, 3)$, $\kappa_j, \xi_{c,j}, \xi_j (j = 1, 2)$, as well as $c_j (j = 1, 2)$ for the two-factor model in Section 3.1. (In the three-factor model in Section 4, they are fixed as $c_1 = -4.23$ and $c_2 = 2.12$.)

First, we define a parameter vector $\theta = (\theta_i)_{i=1}^n (n = 12)$ as

$$\theta = (\hat{\lambda}, \hat{\kappa}_1, \hat{\kappa}_2, \sigma_1, \sigma_2, \sigma_3, \xi_{c,1}, \xi_{c,2}, \xi_1, \xi_2, c_1, c_2), \tag{34}$$

where $\hat{\lambda} := \lambda/\sigma_3^2$, $\hat{\kappa}_2 := \kappa_2/\sigma_2^2$, and $\hat{\kappa}_1 := \sqrt{\kappa_1^2/(|c_1|\sigma_1^2) - 1}$ with a restriction $\kappa_1^2/(|c_1|\sigma_1^2) - 1 > 0$ due to a condition $c_1 \in (-(\frac{\kappa_1}{\sigma_1})^2, 0)$ in the interest rate models (1) and (13).

1. Set the initial values $\theta_i^{(0)}$, $(i = 1, \cdots, n)$ of $\theta = (\theta_i)_{i=1}^n$ by using those of λ, $\sigma_i (i = 1, 2, 3)$, $\kappa_j, \xi_{c,j}, \xi_j (j = 1, 2)$ and $c_j (j = 1, 2)$: As initial estimates of λ and σ_3, we use estimates λ and σ_3 obtained by a one-factor Gaussian term structure model with $r_t = x_{3,t}$. We randomly select initial estimates of κ_j, σ_j around $\kappa_j = 0.25$ and $\sigma_j = 0.016$ $(j = 1, 2)$ that are obtained by a two-factor Gaussian term structure model with $r_t = x_{1,t} + x_{3,t}(= x_{2,t} + x_{3,t})$. For the three-factor model, as initial estimates of ξ_j and $\xi_{c,j}$ $(j = 1, 2)$, we use estimates a_j and b_j $(j = 1, 2)$ in the regression analysis (25) and (26), respectively. For the two-factor models, as initial estimates of c_1 and c_2, we set $c_1 = -2$ and $c_2 = 2$. (In the three-factor model, we do not estimate c_1 nor c_2, and fix those as $c_1 = -4.23$ and $c_2 = 2.12$.)

2. Generate n particles from a random variable r_i following a uniform distribution taking its values in $(-\delta, \delta)$ with some $\delta > 0$.

   Initially set δ as around 2 and gradually decrease the value to be about 0.2. (However, if the initial value seems close to be the optimal one, we set it as a small value. (e.g. λ, σ_3)).

3. Compute a likelihood using $\theta_i = \theta_i^{(0)} \exp(r_i)$ $(i = 1, \cdots, n)$. (As the parameters $(\xi_{c,j} (j = 1, 2))$ change the signs, we change their signs randomly.)

4. Repeat Step 2-3 by m times to set a parameter giving the highest likelihood as θ_i^* $(i = 1, \cdots, n)$.

   While generally setting m as around $100 \sim 5000$, we stop the repetition if the likelihood is substantially improved. Also, in each repetition we use different seeds of random numbers in Monte Carlo simulations.

5. Set $\theta_i^{(0)} = \theta_i^*$ $(i = 1, \cdots, n)$ and repeat Step 2-4.

6. When observing no improvement of the likelihood, set a parameter giving the best likelihood to be our estimate θ_i^* $(i = 1, \cdots, n)$.

In the following, we show how to set variances of observation noises:

1. With regard to the observation noise variance γ_τ^2 for the zero yield with term τ, set it proportionally to the variance of the corresponding yield level $Y_t(\tau)$, that is $\gamma_\tau^2 = K \times Var(Y_t(\tau))$ with a common constant $K > 0$ for all $\tau = 2, 5, 10, 20, 30$.

   The rationale of setting a common constant K is to avoid particular preferences or biases for fitting the model to specific terms of the yield curve. Moreover, it also avoids computational inefficiency. Concretely, in our algorithm of the Monte Carlo filter method, setting very small variances of the observation noises makes computation of the resampling probability difficult/infeasible or causes inadequate resampling probabilities. (See Appendix A, especially Step (d) for the details.) In particular, changing values of K for different τ with too small K for specific terms make the estimation quite inefficient.

2. To decide the constant K, decrease the value of K gradually from $K = 0.1$ until appropriate computation becomes impossible. (See 1. above.) In the two-factor model, we obtain $K = 0.01$, the lowest value that seems to make our computation possible. This value gives us well fitting to observed yields in the two factor model. Thus, we keep the same value $K = 0.01$ in estimation of the three-factor model.

3. With regard to variances of observation noises for two-word sets' frequencies $\gamma_{wj}^2 (j = 1, 2)$, we start with 0.56 and 0.93, respectively, since our preliminary regression analysis provides around 0.56 and 0.93 as the variances of the regression errors. Then, set $\gamma_{wj}^2 = L \times$ (variances of regression errors) with a constant $L > 0$ to be determined.

4. We first note that given $K = 0.01$, $L \geq 1$ gives higher likelihoods than $L < 1$, and around $L = 3$ does the highest in the three-factor model. However, in this case we have the following undesirable properties: (a) factors are estimated so as to fit zero yields solely and the two-word sets' frequencies become irrelevant. (b) estimates of x_1 and x_2 are unstable (i.e. different seeds in Monte Carlo simulations produce quite different estimates), and the obtained estimates of x_1 and x_2 are indistinguishable (almost the same).

   In contrast, given $K = 0.01$, setting $L = 0.1$ improves fitting to the two-word sets' frequencies, and at the same time it enables better fitting to the observed zero yields than the two-factor model. However, $L = 0.05$ makes the fitting to two-word sets' frequencies better, but to the zero yields worse. Moreover, $L = 0.01$ causes our computation in the Monte Carlo filter quite difficult. Since we do not need to closely replicate sharp fluctuations of the two-word sets' frequencies with sacrificing the fitting to the yields, we choose $L = 0.1$.

We remark that for each value of K and L, we estimate the parameters other than variances of observation noises by applying ML method with stochastic grid search explained above (Step 1-6).

4.2 Estimation Result

The estimation result is shown below in Table 3-4 and Figure 5-8. We first note that although it is shown in Table 3 that the log-likelihood is lower than the one in the two-factor models, those are not comparable, because the current observation data are not the same as in Section 3.1, that is, the observation data associated with two-word sets' frequencies are added in the estimation. Nonetheless, we obtain better fitting to the observed zero-yields together with well replication of the trends of the two-word sets' frequencies, as will be seen below.

Next, we see in Table 3 and Figure 5 with Figure 1 that the term premium λ and the volatility σ_3 for the level factor x_3 are similar to the ones in the two-factor models, and the estimated x_3 has a quite similar shape and a high correlation (about 0.9) with those in the two-factor models.

Third, with regard to the steepening factor x_1, the mean-reversion speed κ_1 and the volatility parameter σ_1 become lower than the ones in the corresponding two-factor model ("two-factor Gaussian minus Quadratic-Gaussian"), while the resulting correlation of x_1^2 between three and two factor models is still high(0.79).

	3 factor	2 factor(steepening)	2 factor(flattening)
κ_1	0.0599	0.1290	–
σ_1	0.0189	0.0402	–
κ_2	1.0523	–	0.1290
σ_2	0.0562	–	0.0253
λ	0.0033	0.0042	0.0040
σ_3	0.0136	0.0149	0.0147
ξ_1	612.0	–	–
$\xi_{c,1}$	-0.535	–	–
ξ_2	311.1	–	–
$\xi_{c,2}$	-0.036	–	–
log-likelihood	15,392.91	16,631.05	17,077.11

Table 3: Estimates of parameters: ("3 factor" means three-factor Gaussian Quadratic-Gaussian model, and "2 factor(steepening/flattening)" means two-factor Gaussian minus/plus Quadratic-Gaussian model, respectively.)
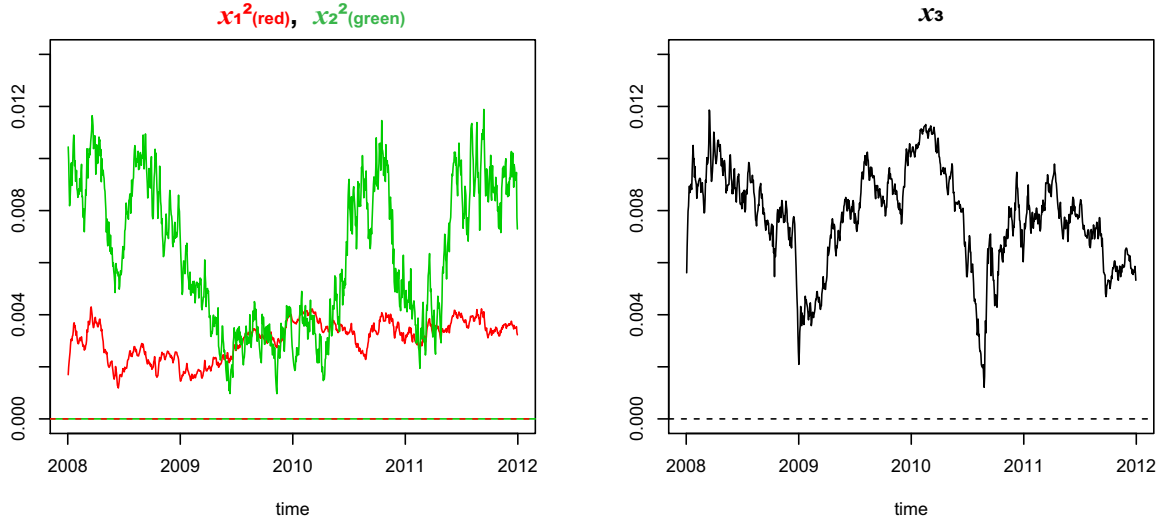
Figure 5: Estimated factors for three-factor Gaussian Quadratic-Gaussian model with word frequencies

As for the flattening factor x_2, the mean-reversion speed κ_2 and the volatility σ_2 become more than 8 times and twice as large as the ones in the corresponding two-factor ("two-factor Gaussian plus Quadratic-Gaussian") model, respectively, presumably because the best κ_2's estimate is obtained so that this factor reflects a strong mean reversion with a high volatility of the two-word set frequency ("business conditions (keiki) & slowdown (gensoku)"), which is observed in Figure 8. Thus, the correlation of x_2^2 between three and two factor models is lower than that of x_1^2(0.52 vs 0.79).

	2y	5y	10y	20y	30y	total average
"three-factor Gaussian Quadratic-Gaussian"	3.28	7.40	3.66	3.68	4.89	4.82
"two-factor Gaussian minus Quadratic-Gaussian"	4.52	9.00	5.35	3.57	5.31	5.85
"two-factor Gaussian plus Quadratic-Gaussian"	4.54	8.94	5.27	3.54	5.31	5.82

Table 4: Square root of mean squared errors (RMSEs) of estimated yields: three-factor Gaussian Quadratic-Gaussian model with word frequency (basis points (bps))

Table 4 provides RMSEs(square root of mean squared errors) of zero yields' estimates, and Figure 6 shows the time series of the observed & estimated zero yields' estimates, respectively. In Figure 6 the fitting to the observed five and two year zero yields in the three-factor model looks better than the one in the two-factor models examined in Section 3.1. We also see in Table 4 that the total average of RMSEs in the three-factor model is less than 5 basis points(bps), which is smaller than those (around 6 bps) in the two-factor models. Also, it is observed that the RMSEs for all the maturities except the 20 year yield in the three-factor model are smaller than those in the two-factor models. Considering

17

that two observation equations related to the two-word sets' frequencies are added, our three-factor model seems to work well.
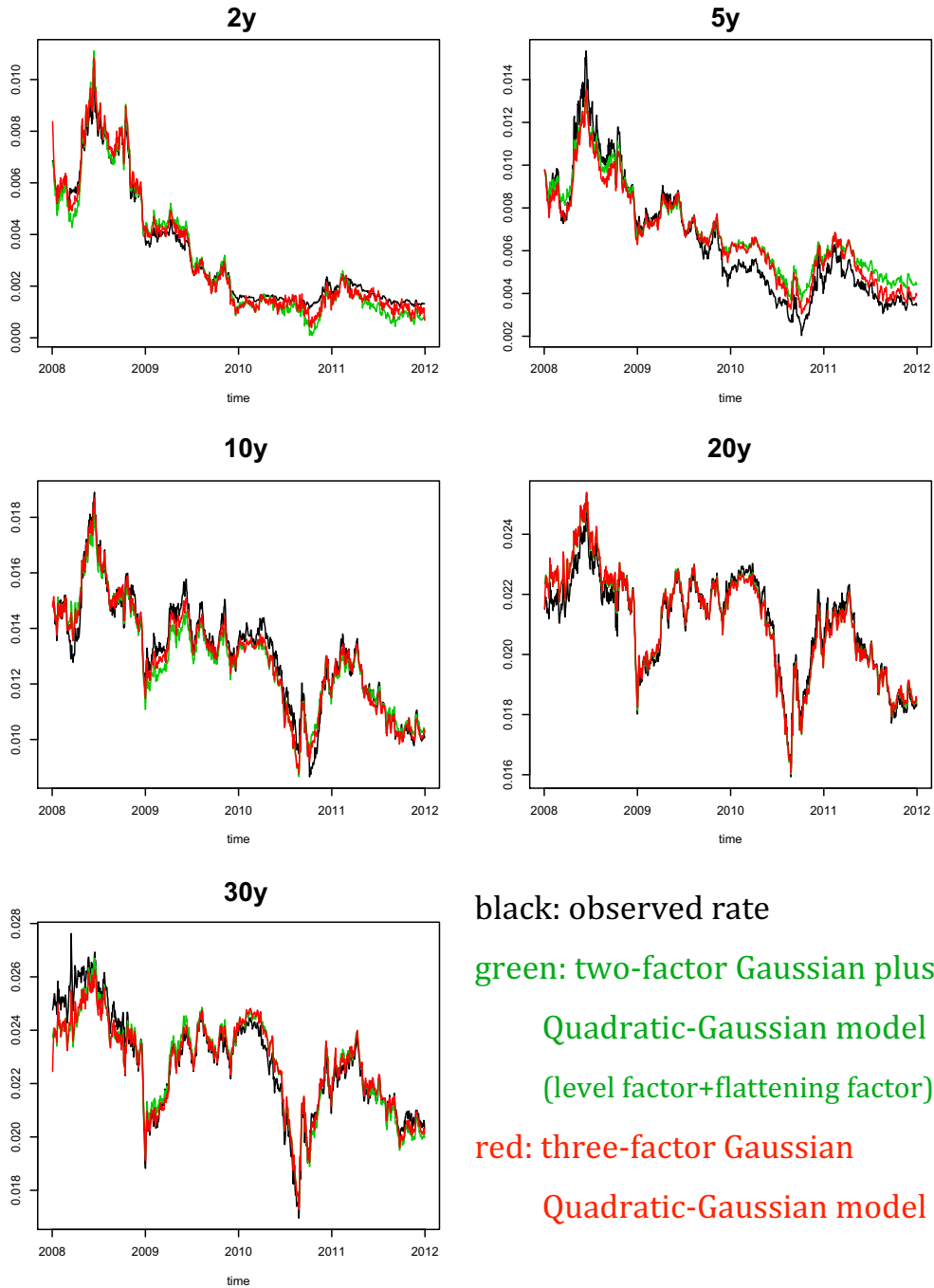

Figure 6: Observed rates, estimated rates of two-factor Gaussian plus Quadratic-Gaussian model without text mining, and estimated rates of three-factor Gaussian Quadratic-Gaussian model with text mining

Finally, Figure 7 shows the time series of actual and estimated log-frequencies $[\log(F_t(\cdot \text{ and } \cdot)+$

1)] of "fiscal conditions (zaisei) & foreign (gaikoku)" and Figure 8 shows that of "business conditions (keiki) & slowdown (gensoku)". Comparing those figures, we observe that the flattening factor (x_2^2) captures the trend of the corresponding two-word set's log-frequency better than the steepening one (x_1^2). In fact, the correlation between the log-frequency of "business conditions (keiki) & slowdown (gensoku)" and the factor x_2^2 (flattening factor) is 0.78, while that of the log-frequency of "fiscal conditions (zaisei) & foreign (gaikoku)" and the factor x_1^2 (steepening factor) is 0.54.
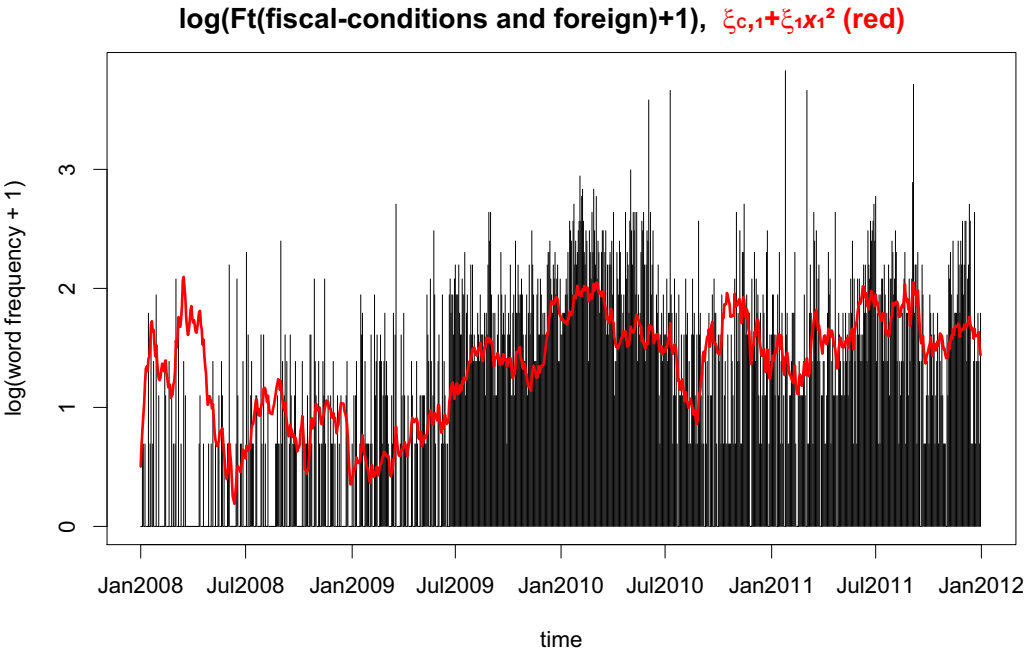

Figure 7: Actual(black) and estimated(red) log-frequencies of "fiscal conditions (zaisei) & foreign (gaikoku)" by three-factor Gaussian Quadratic-Gaussian model with word frequencies
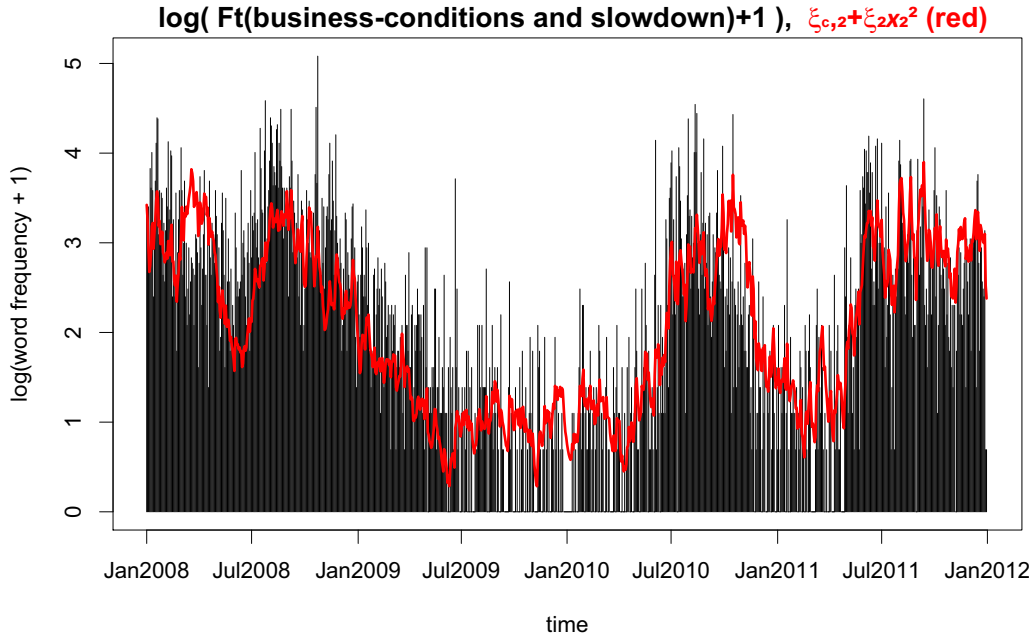
Figure 8: Actual(black) and estimated(red) log-frequencies of "business conditions (keiki) & slowdown (gensoku)" by three-factor Gaussian Quadratic-Gaussian model with word frequencies

4.3 Discussion

This subsection focuses on the relation between the estimated steepening & flattening factors and the movement of spreads in the term structure. Moreover, we would like to interpret "sentiment factors" in the context of spreads among medium-term, long-term and ultra(super)-long term yields. We also informally collect market participants' views about what some specific spreads mean. We then examine whether each of the estimated optimist spread-flattening and pessimist spread-steepening factors has reasonable interpretations.

First, we show correlations between yields/spreads/butterflies and estimated factors in Table 5 below, where the steepening factor x_1^2 is highly correlated with 30-5y and 20-10y spreads (0.94 and 0.90, respectively), while the flattening factor x_2^2 has a high negative correlation (-0.85) with the 10-2y spread. These are observed in Figure 11 and 10. We also see the level factor x_3 has high correlations with 20y and 30y yields (0.83 and 0.82, respectively).

		x_1^2	x_2^2	x_3
yield	2y	-0.66	0.24	0.18
yield	5y	-0.70	-0.01	0.28
yield	10y	-0.55	-0.33	0.52
yield	20y	-0.15	-0.45	0.83
yield	30y	-0.12	-0.17	0.82
spread	5-2y	-0.49	-0.66	0.42
spread	10-2y	0.31	-0.85	0.43
spread	20-2y	0.65	-0.66	0.49
spread	30-2y	0.75	-0.51	0.68
spread	30-10y	0.78	0.31	0.52
spread	20-10y	0.90	-0.09	0.38
spread	30-20y	0.04	0.55	0.29
spread	10-5y	0.68	-0.48	0.18
spread	20-5y	0.83	-0.35	0.29
spread	30-5y	0.94	-0.16	0.44
butterfly	10-5-2y	0.77	-0.02	-0.08
butterfly	30-20-5y	-0.68	0.46	-0.15
butterfly	5-10-20y	-0.06	0.59	0.12

Table 5: Correlations between yields and estimated factors by three-factor Gaussian Quadratic-Gaussian model with word frequencies

Next, let us discuss on the market participants' remarks collected from portfolio managers of hedge funds and others who are actively participating in Japanese government bond markets. There are two kinds of remarks: one about long and medium term spreads and the other about ultra(super)-long and long-term spreads.

When asking particularly about the relation between short-run business conditions and yield spreads, we customarily get the following kind of reply: When economy picks up (recovery), the 10-2 year spread often increases. Or conversely, when business conditions slow down, the 10-2 year spread decreases. In fact, this statement is consistent with data which show a high correlation of the 10-2 year spread with the leading diffusion index published monthly by the Cabinet Office, which is a good predictor of business conditions. Figure 9 below shows that the 10-2 year spread has a relatively high correlation(0.72) with the leading diffusion index (DI).
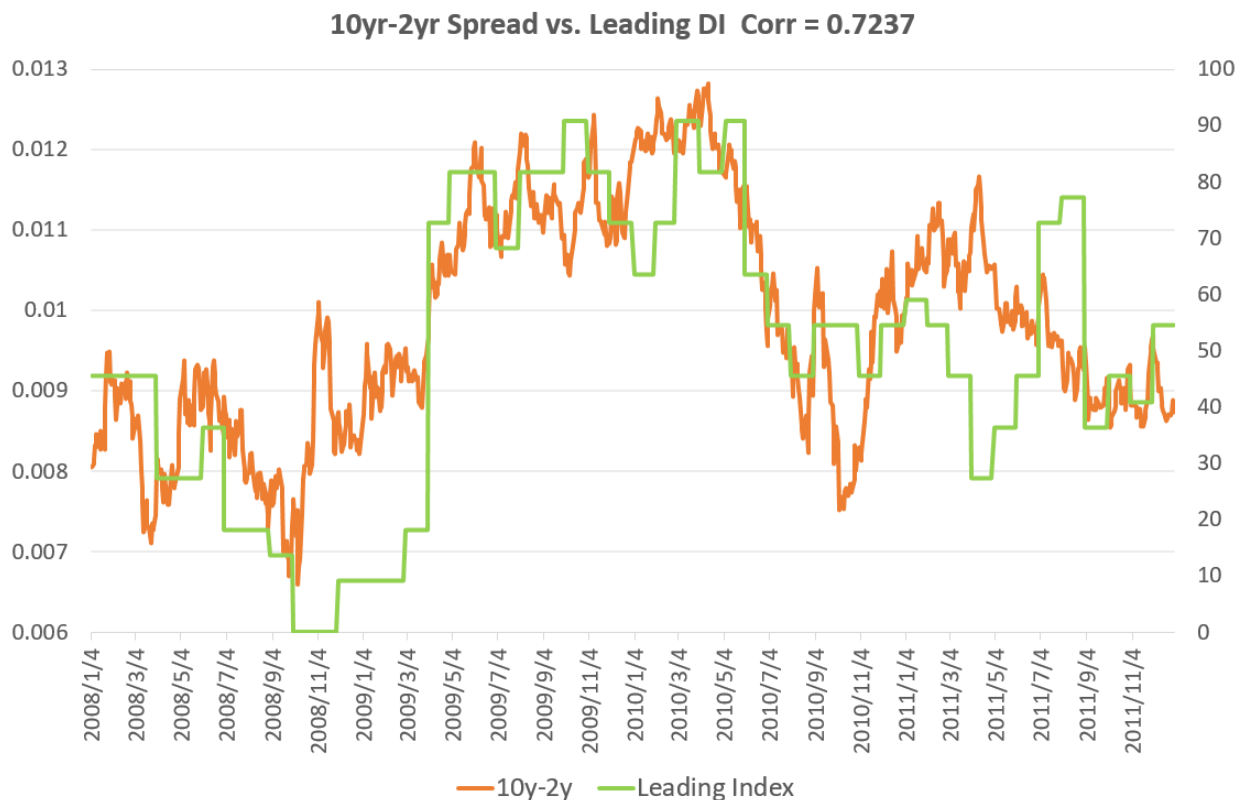
Figure 9: Plot of 10-2year spread and leading diffusion index (DI)

Then, as seen in Table 5, we observe in Figure 10 below with an inverted scale for the 10-2 year spread that the flattening factor x_2^2 moves in a highly negatively correlated way with the 10-2 year spread, and thus with short run business conditions. Also, in Figure 8, the flattening factor x_2^2 explains the two-word set's frequency of "business conditions" and "slowdown" (= inverted "recovery") very well.

Consequently, the spread-flattening factor x_2^2 has a reasonable interpretation of a factor representing short-run business conditions' influence on market sentiment. The market participants are closely following news about business conditions, especially those suggesting slowdown.
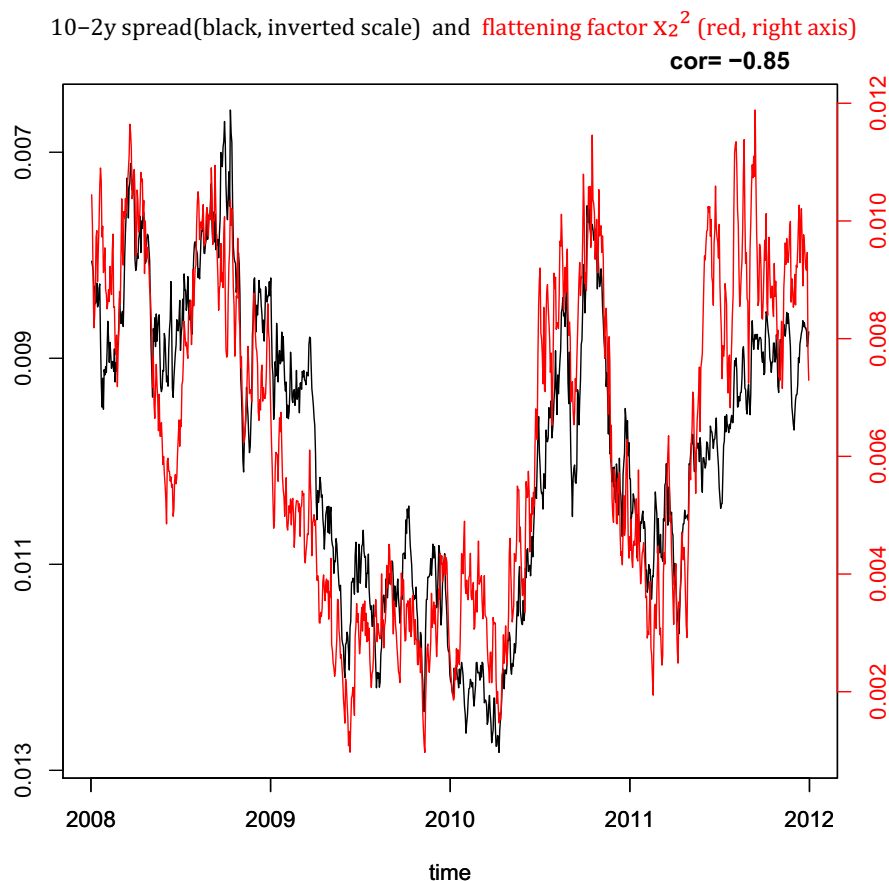
Figure 10: Plots of estimated factor for x_2^2 in "three-factor Gaussian Quadratic-Gaussian" model and 10-2year spread

Finally, let us consider a market participants' remark regarding long-run concerns and yield spreads. They often note that bad news about fiscal conditions (e.g., budget deficits) and substantial increase in uncertainty about future (long-run trends) (e.g., negative structural change after financial crisis) have more effects on ultra-long (20 year, 30 year) rates than shorter (equal to and less than 10 year) rates, making spreads between ultra-long term yields and the others increase substantially. Our results show the estimated steepening factor has a very high correlation with the ultra-long term related spreads (i.e. the 30-5 year, 20-10 year spreads), and at the same time explains the word frequency of market participants' long-run concerns.

Figure 11 below demonstrates a close relation between the steepening factor x_1^2 and the 30-5 & 20-10 year spreads. (Let us recall Table 5 shows their high correlations, 0.94 for the 30-5y and 0.90 for the 20-10y spreads.) Moreover, as seen in Figure 7, the steepening factor x_1^2 explains the frequency of the two-word group "fiscal conditions" and "foreign" very well. "Fiscal conditions" have been a long run concern of the market during our research period. News about "foreign" reflects the market's concern about heightened uncertainty triggered by the global financial crisis and European sovereign debt crisis, both of which are explained in detail in the following analysis. Thus, as an overall assessment, the spread-steepening factor has a reasonable interpretation of a factor representing the market's long run concerns.
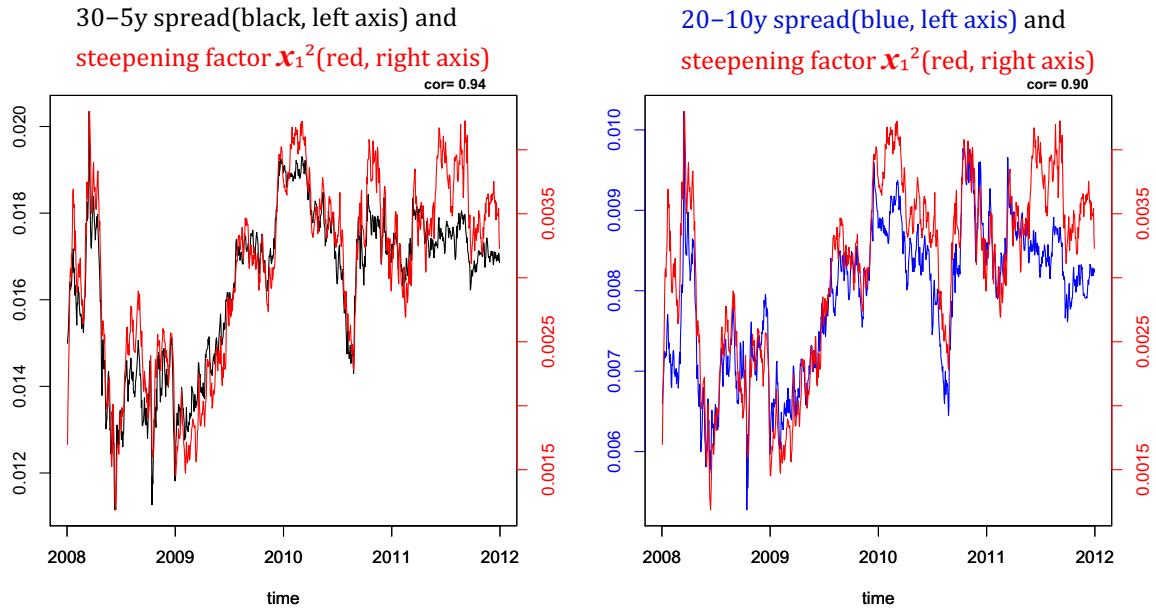
Figure 11: Plot of estimated factor for x_1^2 in "three-factor Gaussian Quadratic-Gaussian" model and 30-5year or 20-10year spread

Regarding Japanese fiscal conditions, we see the word, "deterioration (akka)" or "deficit (akaji)" or "supplementary [budget] (hosei [yosan])" in 34% (1,202 out of 3,495) of the news including the two-word set "fiscal conditions (zaisei) & foreign (gaikoku)", indicating worsening fiscal condition. In fact, in Figure 12 below, the general government gross debt to GDP ratio implies a quite serious situation in Japan especially during our research period 2008-2011, compared with in Norway, Sweden, Germany, France, U.S. and Italy.
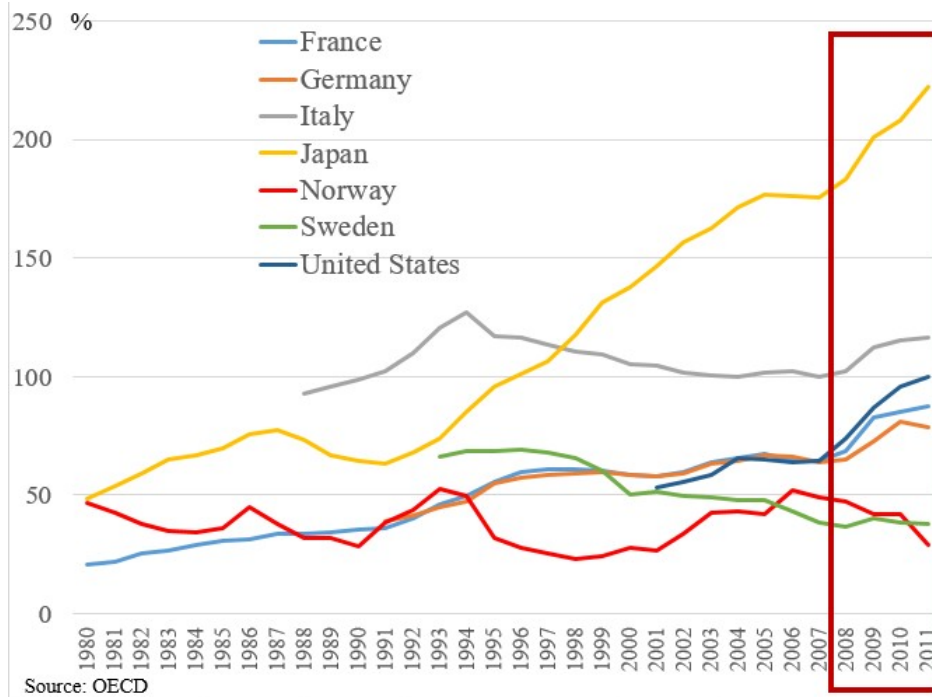
Figure 12: Fiscal conditions: general government gross debt to GDP ratio

Concerning heightened uncertainty, it seems caused by events in "foreign": In fact, our research period includes the financial crisis in U.S. & Europe and the sovereign debt crisis in Europe, both of which shook the market confidence and elevated uncertainty. In contrast, Japanese financial markets were relatively calm at that time and there was no domestic incidence to negatively affect Japanese market confidence. Thus, it seems reasonable that an erosion of market confidence, if any, was originated in foreign countries.

This heightened uncertainty also causes drastic appreciation of Japanese yen (JPY) against U.S. dollar (USD), from JPY 108.60 in January 4th, 2008 to JPY 76.91 in December 30th, 2011. See Figure 13 that shows the price of the unit amount of USD in terms of JPY (USDJPY).
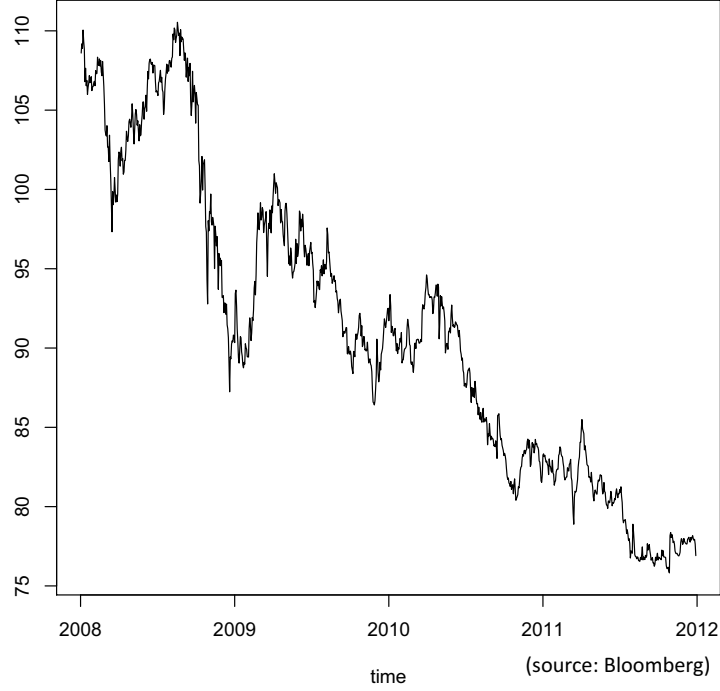
Figure 13: Spot rates of USDJPY

At the same time, the word, "foreign exchange (gaikoku kawase)" appears in 83% (2,885 out of 3,495) of the news including the two-word set "fiscal conditions (zaisei) & foreign (gaikoku)".

Furthermore, the correlations between the USDJPY rate and the estimated x_1^2, the 30-5 year and 20-10 year spreads are given by -0.56, -0.55 and -0.61, respectively, that is, JPY appreciation against USD has positive correlations with the steepening factor and ultra-long term related spreads.

These phenomena reflect that a globally increasing risk-aversion (risk-off) generally leads to long JPY, and also JGB, especially the JGB futures and then cash bonds with maturities closer to the futures' such as 10 and 5 years, which results in the steepening of the 30-5 and 20-10 year spreads.

Finally, we examine word sets related to the word "foreign exchange" more carefully.

JPY appreciation	JPY depreciation	long JPY	short JPY
552	132	379	413
USD depreciation	USD appreciation	short USD	long USD
337	177	549	527
JPY appreciation&USD depreciation	JPY depreciation & USD appreciation	long JPY& short USD	short JPY& long USD
209	84	157	158
EUR depreciation	EUR appreciation	short EUR	long EUR
167	55	213	160
JPY appreciation&EUR depreciation	JPY depreciation & EUR appreciation	long JPY& short EUR	short JPY& long EUR
65	6	45	12

Table 6: Words & their frequencies in the news where the word "foreign exchange" appears

Table 6 above shows meaningful words concerning currencies' appreciation/depreciation, and their frequencies in the news where "foreign exchange" appears.

First, the left hand side in Table 6 shows that frequencies of the words including currency appreciation/depreciation clearly implies the JPY appreciation: we see 552 news containing the word "JPY appreciation (yendaka)" against 132 news including "JPY depreciation (yenyasu)"; 337 news "USD depreciation (doruyasu)" against 177 news "USD appreciation (dorudaka)"; 167 news "EUR depreciation(yuroyasu)" against 55 news "EUR appreciation(yurodaka)"; 209 and 84 news containing the words "JPY appreciation & USD depreciation (yendaka& doruyasu)" and "JPY depreciation & USD appreciation (yenyasu& dorudaka)", respectively; 65 and 6 news containing "JPY appreciation & EUR depreciation (yendaka& yuroyasu)" and "JPY depreciation & EUR appreciation (yenyasu& yurodaka)", respectively.

However, the right hand side in Table 6 shows that frequencies of the words with "JPY(USD) long/short" do not indicate which currency is strong/weak, i.e. they do not tell whether JPY appreciates against USD or not: 379 news containing the word "long JPY (yengai)" against 413 news including "short JPY (yenuri)"; 549 news "short USD (doruuri)" against 527 news "long USD (dorugai)"; in particular, the words "long JPY & short USD (yengai & doruuri)" and "short JPY & long USD (yenuri & dorugai)" appear in almost the same number of the news (157 and 158, respectively). In contrast, the number of news regarding long/short EUR may imply appreciation of JPY against EUR: The word "long JPY & short EUR (yengai & yurouri)" appears in 45 news, while "short JPY & long EUR (yenuri & yurogai)" does in 12 news.

Moreover, the total number of the news (without double counting) containing "JPY appreciation & USD depreciation (yendaka & doruyasu)" or "longJPY & shortUSD (yengai & doruuri)" is 336, while that of JPY depreciation & USD appreciation (yenyasu & dorudaka)" or "shortJPY & longUSD (yenuri & dorugai)" is 230, respectively. Obviously, the former is more, but its difference is not so much.

Also, the correlation between the USDJPY forex rate and the frequency of the two-word sets "fiscal conditions (zaisei) & foreign (gaikoku)" is -0.46, which is not high.

As such our analysis of the word sets' frequencies does not clearly reflect substantial JPY appreciation against USD during the research period. It seems partly because our simple text mining method is not able to distinguish whether those words are used in the context of JPY appreciation or not. Thus, one of our future research topics is to incorporating a more advanced and elaborate analysis of the news (e.g. through a natural language processing) into our term structure modeling.

5   Concluding Remarks

Market sentiment is an elusive qualitative concept in empirically-oriented asset market analyses. Almost everybody in the market (regardless of portfolio managers, researchers, and central bankers) agrees that it is important, and yet it is not well incorporated into quantitative tool kits of asset market modeling, since the sentiment is not observable and thus considered to be not quantifiable. In contrast, quantitative dynamic factor models are successfully applied in many countries to "explain" the actual term structure

of interest rates, though it is not clear what drives these factors.[14] They are a good stochastic description of the market interest rates, but they are short of discerning the underlying economic determinants.

The contribution of this paper is to solve both problems at the same time, first by formulating market sentiment as quantifiable factors in the term structure models, and then by finding observable variables (frequencies of relevant news reports) linked to the sentiment factors, so that they can be estimated quantitatively. Along this way, what drives sentiment is also revealed.

Specifically, this paper has formulated and successfully estimated a three-factor term structure model with explicit sentiment factors in a period including the Global Financial Crisis and the European Sovereign Debt Crisis, the most challenging period of volatile economic activities and heightened uncertainty. The key of our analysis is a text mining of a large text data of real time, relatively high-frequency market news, to find a key word set about market participants' major concerns, that is, determinants of their sentiment.

In this procedure, we have been able to distinguish (1) a spread-steepening factor which is driven by pessimists' concerns and explaining the ultra-long and long term spreads (e.g. 20-10 year) from (2) a spread-flattening factor which is driven by optimists' concerns and influencing the long and medium term spreads (e.g. 10-2 year). Moreover, a pessimist spread-steepening factor is mostly concerned with news about long-run fiscal conditions and global risks, while an optimist spread-flattening factor is mostly influenced by short-run business cycle activities. In this way, we have been able to identify two spread-determining factors influencing different parts of the term structure of interest rates.

Finally, there is one caveat, related to the discussion in Section 4.3, especially about the word "foreign (exchange)". The current work takes a parsimonious approach of text mining, namely to find a key word group and to use its frequency as an observable variable of an otherwise unobservable sentiment factor. This approach has been quite successful in distinguishing different spread factors. However, beyond this point, the link between key words and events influencing sentiment is still loose, especially with respect to the context key words are used. Thus, we still need to improve our analysis to identify in what context words are used and what connotations they have in financial news. More essentially, it is of utmost importance to develop a method that is able to select meaningful combination or sequence of word groups in the news from a specific viewpoint of research objectives we have, which remains as a main topic in our future research.

---

[14]There are attempts (for example, Rudebusch and Wu (2008)) in the macro-finance literature that try to explain dynamic factors in the term structure models by macro economic variables including monetary policy. However, the attempts are not very successful so far. The reason is that level, slope and curvature (volatility) components in the term structure of interest rates happen to be poorly correlated with macro variables in the first place. See Adrian (2017)).

A   Algorithm for Monte Carlo Filter

This appendix describes the outline of an algorithm of Monte Carlo filter used in this work, which is an adaptation of Fukui et al. (2017).

We introduce a state space model that consists of the following system and observation models:

$$\begin{cases} x_t = & f(x_{t-\Delta t}, \epsilon_t) \quad \text{(system equation)} \\ y_t = & h(x_t) + e_t \quad \text{(observation equation)}, \end{cases} \tag{35}$$

where x_t, y_t and Δt denote a N dimensional state vector, a M dimensional observation vector at time t and the time interval of observational data, respectively while ϵ_t and e_t denote the system noise and the observational noise whose density functions are given by $q(v)$ and $\psi(u)$ respectively. The functions f and h are generally non-linear maps, $R^N \times R^N \mapsto R^N$ and $R^N \mapsto R^M$, and the initial state vector x_0 is assumed to be a random variable whose density function is given by $p_0(x)$.

Next, we summarize the notations: $p(x_t|y_{t-\Delta t})$, called "one step ahead prediction" denotes the conditional density function of a state vector x_t given an observation vector $y_{t-\Delta t}$ where Δt is the interval of time series data. $p(x_t|y_t)$, called "filter" denotes the conditional density function of x_t given y_t. $\{p_t^{[1]}, \cdots, p_t^{[m]}\}$ and $\{s_t^{[1]}, \cdots, s_t^{[m]}\}$ represent the vectors of realization of m trials of Monte Carlo simulations from $p(x_t|y_{t-\Delta t})$ and $p(x_t|y_t)$, respectively. Then, when we choose $\{s_0^{[1]}, \cdots, s_0^{[m]}\}$ from the density function $p_0(x)$ of the initial state vector x_0, as realization of Monte Carlo simulations, an algorithm of Monte Carlo filter is as follows.

[**Summary of Algorithm for Monte Carlo filter**]

1. Apply the following steps (a)$\sim$(d) to each time $t = 0, \Delta t, 2\Delta t, \cdots, (T_* - \Delta t), T_*$ where T_* denotes the final time point of the data.

   - (a) Generate the system noise $\epsilon_t^{[k]}$, $k = 1, \cdots, m$ according to the density function $q(\epsilon)$.

   - (b) Compute for each $k = 1, \cdots, m$

     $$p_t^{[k]} = f(s_{t-\Delta t}^{[k]}, \epsilon_t^{[k]}).$$

     We note that $f(\cdot, \cdot)$ is linear in our model, such as $Fs_{t-\Delta t}^{[k]} + G\epsilon_t^{[k]}$. (See the system equations in "State space model for two-factor term structure model" in Section 3.1 and "State space model for three-factor term structure model and two-word set frequency" in Section 4.)

   - (c) Evaluate the density function of $\psi(u)$ at $u = y_t - h(p_t^{[k]})$, $k = 1, \cdots, m$ and define the evaluated densities as $\alpha_t^{[k]}$, $k = 1, \cdots, m$. In our models, $\alpha_t^{[k]}$ is given by:

     $$\alpha_t^{[k]} = \Pi_{l=1}^M \frac{1}{\sqrt{2\pi\gamma_l^2}} \exp\left(-\frac{[y_{l,t} - h_l(p_t^{[k]})]^2}{2\gamma_l^2}\right), \tag{36}$$

     where $h(\cdot)$ in our models is expressed as

     $$h_l(p_t^{[k]}) = a_l + b_l(p_{1,t}^{[k]})^2 + c_l(p_{2,t}^{[k]})^2 + d_l p_{3,t}^{[k]},$$

here a_l, b_l, c_l and d_l are explicitly given as the constant terms, the coefficients of $x_{j,t}^2$ $(j = 1, 2)$ and $x_{3,t}$ in the observation equations in "State space model for two-factor term structure model" in Section 3.1 and "State space model for three-factor term structure model and two-word sets frequency" in Section 4.

- (d) Resample $\{s_t^{[1]}, \cdots, s_t^{[m]}\}$ from $\{p_t^{[1]}, \cdots, p_t^{[m]}\}$. More precisely, resample each $s_t^{[k]}$, $k = 1, \cdots, m$ from $\{p_t^{[1]}, \cdots, p_t^{[m]}\}$ with the probability given by

$$\text{Prob.}(s_t^{[k]} = p_t^{[i]}|y_t) = \frac{\alpha_t^{[i]}}{\sum_{k=1}^{m} \alpha_t^{[k]}}, \quad k = 1, \cdots, m, \ i = 1, \cdots, m.$$

We note that when the variances of the observation noises, γ_l^2 are very small, $\alpha_t^{[k]}$ tends to take values close to zero (cf. (36)), which makes computation of this resampling probability difficult (infeasible) or causes inadequate resampling probabilities. (e.g. The support of a resampling distribution concentrates on a few particular values.)

- (e) We obtain the filtered estimates $\hat{x}_t$ by

$$\hat{x}_t = \frac{1}{m} \sum_{i=1}^{m} s_t^{[i]}.$$

Particularly, in our models, we calculate

$$\hat{x}_{j,t}^2 = \frac{1}{m} \sum_{i=1}^{m} (s_{j,t}^{[i]})^2, j = 1, 2,$$

$$\hat{x}_{3,t}^2 = \frac{1}{m} \sum_{i=1}^{m} s_{3,t}^{[i]}.$$

The estimation of unknown parameters is based on the maximum likelihood method. If θ denotes the vector representing whole unknown parameters, the likelihood $L(\theta)$ is given by

$$L(\theta) = g(y_{\Delta t}, \cdots, y_{T_*}|\theta) = \Pi_{i=1}^{\frac{T_*}{\Delta t}} g_i(Y_{i\Delta t}|y_{\Delta t}, \cdots, y_{(i-1)\Delta t}, \theta); \quad (37)$$

$$g_1(y_{\Delta t}|y_0) = p_0(y_{\Delta t}), \quad (38)$$

where $g(y_{\Delta t}, \cdots, y_{T_*}|\theta)$ and $g_i(y_{i\Delta t}|y_{\Delta t}, \cdots, y_{(i-1)\Delta t}, \theta)$ denote the joint density function of $y_{\Delta t}, \cdots, y_{T_*}$ with parameter vector θ and the conditional density function of $y_{i\Delta t}$ given $y_{\Delta t}, \cdots, y_{(i-1)\Delta t}$ with θ, respectively. The log-likelihood $l(\theta)$ is computed approximately within the framework of the Monte Carlo filter by:

$$l(\theta) = \sum_{i=1}^{\frac{T_*}{\Delta t}} \left(\log \sum_{k=1}^{m} \alpha_{i\Delta t}^{[k]} \right) - \frac{T_*}{\Delta t} \log m.$$

Then, maximize $l(\theta)$ with respect to θ to obtain the maximum likelihood estimator θ^*.

B   Interest Rate Models with Market Sentiment

This appendix briefly explains an equilibrium model with a representative agent, which supports interest rate models with market sentiment. The model is motivated by the argument of Nishimura and Ozaki (2017 Chapter 12), which suggests the economic agent may find himself in three different states with respect to the (part or whole) of stochastic model he faces: he may be confident about the stochastic model, not confident and pessimistic (and thus maximize his utility assuming the worst case), or optimistic (assuming the best case). Particularly, we assume in a consumption-portfolio allocation a representative agent is confident about some risks (Brownian motions in this paper), but not confident about the others, of which he may be pessimistic assuming the worst case for some and optimistic assuming the best case for the rest. We present the model in an intuitive way in this appendix. A rigorous argument is found in Nishimura, Saito and Takahashi (work in progress).

We start with an appropriate filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$ with d-dimensional Brownian motion $B = (B_1, \cdots, B_d)$. Then, we specify fundamental uncertainty as uncertainty about the stochastic model in the following way. We work with some d-dimensional progressively measurable process $\lambda = \{\lambda_t; 0 \leq t < \infty\}$, a martingale $Z_t(\lambda)$ defined by

$$Z_t(\lambda) := \exp\left\{\sum_{j=1}^d \int_0^t \lambda_{j,s} dB_{j,s} - \sum_{j=1}^d \frac{1}{2} \int_0^t \lambda_{j,s}^2 ds\right\} \tag{39}$$

and probability measure P^λ,

$$P^\lambda(A) := E[Z_T(\lambda) 1_A]; \ A \in \mathcal{F}_T, \text{ for any } T \in [0, \infty). \tag{40}$$

We assume that the Maruyama-Girsanov theorem can be applied; that is, a process $B^\lambda = \{(B_{1,t}^\lambda, \cdots, B_{d,t}^\lambda); 0 \leq t < \infty\}$, $B_{j,t}^\lambda := B_{j,t} - \int_0^t \lambda_{j,s} ds$ $(j = 1, 2, \cdots, d)$ is a d-dimensional Brownian motion on $(\Omega, \mathcal{F}, P^\lambda)$. (For a rigorous argument, see Chapter 1.7 and Chapter 3.9 in Karatzas & Shreve (1998) and Chapter 3.5 in Karatzas & Shreve (1991). Here λ_j represents fundamental uncertainty about the j-th risk (Brownian motion B_j). If $\lambda_j \equiv 0$ (i.e. $B_j^\lambda = B_j$), there is no fundamental uncertainty about the j-th risk. When there is fundamental uncertainty about the the j-th risk, we only know the true j-th risk is one of $\{B_j^\lambda; \lambda_j \in \Lambda_j\}$ with $B_{j,t}^\lambda := B_{j,t} - \int_0^t \lambda_{j,s} ds, 0 \leq t < \infty$ for some set Λ_j, and we cannot tell which is the true one.

In a consumption-portfolio allocation for a single risky asset and an risk-free asset, a representative agent takes his/her own views for uncertainties (risks) associated with Brownian motions into account. Specifically, a representative agent who has a pessimistic (optimistic) view on Brownian motion $B_1(B_2)$ assumes the worst (best) case. Thus, he/she implements optimization in a consumption-portfolio allocation with respect to λ_j $(j = 1, 2)$, that is, minimize (maximize) the expected utility with respect to $\lambda_1(\lambda_2)$, in addition to standard maximization for an allocation of consumption(c), a risk-free asset and a risky asset whose proportion of the agent's wealth (W) is denoted by α. In contrast, for $j = 3, \cdots, d$, the economic agent has perfect confidence, so that we have $\lambda_j \equiv 0$. Then, $B_{1,t}^\lambda = B_{1,t} - \int_0^t \lambda_{1,s} ds$, $B_{2,t}^\lambda = B_{2,t} - \int_0^t \lambda_{2,s} ds$ and $B_j^\lambda = B_j$ for $j = 3, \cdots, d$ are Brownian motions under the probability measure P^λ generated by a martingale $Z(\lambda)$ with $\lambda = (\lambda_1, \lambda_2, 0, \cdots, 0)$.

More concretely, the agent with a time-separable expected utility specified by a strictly increasing and concave function solves the following problem:

$$\sup_{(\alpha,c)\in\Pi}\ \sup_{\lambda_2\in\Lambda_2}\ \inf_{\lambda_1\in\Lambda_1}\ E^{P^\lambda}\left[\int_0^\infty e^{-\beta t}u(c_t)dt\right],\ (\beta>0),\tag{41}$$

where *pessimism and optimism* are expressed by $\inf_{\lambda_1}$ and $\sup_{\lambda_2}$, respectively. Here, Π and Λ_j $(j=1,2)$ denote appropriate admissible control sets. In particular, the conditions specifying Λ_j contain that λ_j $(j=1,2)$ are progressively measurable processes such that $\lambda_{j,t}^2\le\bar\lambda_j(x_t)^2$ with a state vector process $x=\{x_t;t\ge0\}$, and $Z(\lambda)$ is a martingale under P given $\lambda_i(i\ne j)$.

Moreover, we exogenously specify $\bar\lambda_j(x)$, (a function of x, $\mathcal{R}^n\to\mathcal{R}$) so that $Z(\bar\lambda)$ with a progressively measurable process $\bar\lambda(x_t)$ is a martingale under P.

Further, in what follows, we will specify $u(c)$ as $u(c)=\log c$ for $c>0$.

Next, let us suppose that a n-dimensional state vector process x, a dividend process D and a price process η of a risky asset receiving the dividend stream $\{D_t:t\ge0\}$ are obtained by the following system of stochastic differential equations:

$$dx_t\ =\ \mu_x(x_t)dt+\sum_{j=1}^d\sigma_{x,j}(x_t)dB_{j,t}\tag{42}$$

$$=\ [\mu_x(x_t)+\sum_{j=1}^2\lambda_{j,t}\sigma_{x,j}(x_t)]dt+\sum_{j=1}^d\sigma_{x,j}(x_t)dB_{j,t}^\lambda;\ x_0\text{ given},\tag{43}$$

$$dD_t\ =\ D_t[\mu_D(x_t)dt+\sum_{j=1}^d\sigma_{D,j}(x_t)dB_{j,t}]\tag{44}$$

$$=\ D_t[(\mu_D(x_t)+\sum_{j=1}^2\lambda_{j,t}\sigma_{D,j}(x_t))dt+\sum_{j=1}^d\sigma_{D,j}(x_t)dB_{j,t}^\lambda];\ D_0>0\text{ given},\tag{45}$$

$$d\eta_t\ =\ \eta_t[\{\mu(x_t)-\frac{D_t}{\eta_t}\}dt+\sum_{j=1}^d\sigma_j(x_t)dB_{j,t}]\tag{46}$$

$$=\ \eta_t[(\{\mu(x_t)-\frac{D_t}{\eta_t}\}+\sum_{j=1}^2\lambda_{j,t}\sigma_j(x_t))dt+\sum_{j=1}^d\sigma_j(x_t)dB_{j,t}^\lambda];\ \eta_0>0\text{ given},\tag{47}$$

with $\mu_x,\sigma_{x,j}:\mathcal{R}^n\to\mathcal{R}^n$ and $\mu_D,\sigma_{D,j},\mu,\sigma_j:\mathcal{R}^n\to\mathcal{R}$, $j=1,\cdots,d$.

We also suppose that while a state vector process x and a dividend process D are exogenously given, an interest rate r of an risk-free asset, and the expected rate of return μ and volatilities σ_j of a price process η are endogenously determined in equilibrium.

For simplicity, for the first and second element of x, x_i $(i=1,2)$, we assume $\sigma_{x,j,i}=0$, $j\ne i$ $(i=1,2)$, where $\sigma_{x,j,i}$ denotes the i-th element of $\sigma_{x,j}$.

Then, the representative agent's wealth process W is described as follows: given stochastic processes of a risk-free interest rate r, a consumption c and proportion α of the

wealth invested in a risky asset,

$$
\begin{aligned}
dW_t &= [r_tW_t + \alpha_tW_t(\mu_t - r_t) - c_t]dt + \alpha_tW_t\left[\sum_{j=1}^{d}\sigma_{j,t}dB_{j,t}\right] \tag{48}\\
&= \left[r_tW_t + \alpha_tW_t(\mu_t - r_t) - c_t + \alpha_tW_t\sum_{j=1}^{2}\sigma_{j,t}\lambda_{j,t}\right]dt + \alpha_tW_t\left[\sum_{j=1}^{d}\sigma_{j,t}dB_{j,t}^{\lambda}\right] ; \quad W_0 = w. \tag{49}
\end{aligned}
$$

Then, solving the associated HJB equation, we obtain candidates for optimal controls are given by

$$
\lambda_{1,t} = -\bar{\lambda}_1(x_t); \ \lambda_{2,t} = +\bar{\lambda}_2(x_t), \tag{50}
$$

$$
c_t = \beta W_t; \quad \alpha_t = \frac{\mu(x_t) - r(x_t) - \bar{\lambda}_1(x_t)\sigma_1(x_t) + \bar{\lambda}_2(x_t)\sigma_2(x_t)}{|\sigma(x_t)|^2}.
$$

We remark that if $\bar{\lambda}_j(x_t)\sigma_j(x_t) > 0(j = 1, 2)$, a pessimistic (optimistic) view reduce (increase) α_t, investment proportion of an risky asset. Hereafter, we suppose that the candidates (50) of optimal controls attain (41).

(**Equilibrium**) An equilibrium in this economy is characterized as follows: it holds that $c = D$ and $\alpha = 1$. Given $r = \{r_t; t \geq 0\}$, (43), (45), (47) and (49), the representative agent solves (41).

We also note that the equilibrium conditions above imply that $W_t = \frac{1}{\beta}D_t(= \eta_t)$, and hence $\mu(x_t) = \beta + \mu_D(x_t)$ and $\sigma_j(x_t) = \sigma_{D,j}(x_t)$ in equilibrium. Thus, the condition $\alpha = 1$ provides an equilibrium interest rate as

$$
r_t = \beta + \mu_D(x_t) - |\sigma_D(x_t)|^2 - \bar{\lambda}_1(x_t)\sigma_{D,1}(x_t) + \bar{\lambda}_2(x_t)\sigma_{D,2}(x_t), \tag{51}
$$

in which the term $\beta + \mu_D(x_t) - |\sigma_D(x_t)|^2$ is a well-known equilibrium interest rate process *without pessimism and optimism* for the log-utility. We remark that if $\bar{\lambda}_j(x_t)\sigma_{D,j}(x_t) > 0$ $(j = 1, 2)$, the expression agrees with our intuition, that is pessimistic (optimistic) views reduces (increases) the equilibrium interest rate.

Next, we note that x is expressed under Q^{λ}, a risk-neutral probability measure induced by P^{λ} as

$$
dx_t = [\mu_x(x_t) + \sum_{j=1}^{2}\lambda_j\sigma_{x,j}(x_t)]dt + \sum_{j=1}^{d}\sigma_{x,j}(x_t)[dB_{j,t}^{Q^{\lambda}} - \sigma_{D,j}(x_t)dt], \tag{52}
$$

where $Q^{\lambda}(A) = E^{P^{\lambda}}[Z_T(\sigma_D)1_A]$ for any $T > 0$, $A \in \mathcal{F}_T$,

$$
Z_T(\sigma_D) = \exp\left(-\int_0^T \sigma_D(x_s)dB_s - \frac{1}{2}\int_0^T |\sigma_D(x_s)|^2 ds\right), \tag{53}
$$

and $B^{Q^{\lambda}}$ is a d-dimensional Brownian motion under Q^{λ}.

Hence, equilibrium zero coupon bond price $P(t,T)$ and zero yield $Y(t,T)$ at time t with maturity T are expressed respectively by

$$P(t,T) = E^{Q^{\bar{\lambda}}}\left[e^{-\int_t^T r_s ds}|\mathcal{F}_t\right] \tag{54}$$

$$Y(t,T) = \frac{-1}{(T-t)}\log P(t,T), \tag{55}$$

with $\bar{\lambda} = (-\bar{\lambda}_1, \bar{\lambda}_2, 0, \cdots, 0)$ and an equilibrium interest rate (51).

B.1 Example: Three-factor Gaussian Quadratic-Gaussian Model

When we specify

$$\mu_D(x_t), \sigma_{D,j}(x_t); \mu_x(x_t), \sigma_{x,j}(x_t); \bar{\lambda}_j(x_t) \ (j = 1, 2) \tag{56}$$

associated with the equation (51), we obtain a concrete expression of equilibrium interest rate r_t and zero yields $Y(t,T)$, $0 \le t \le T < \infty$. In what follows, we present such an example for yield curve models with new factors (proxies of *pessimism and optimism*) that follow Quadratic-Gaussian processes, which is a simplified version of the previous section's models.

First, as an example of the state variable process (42), let us consider the following model with three-dimensional Brownian motion $B = (B_1, B_2, B_3)$, and constants $a_j \ge 0$, $b_j > 0$ $(j = 1, 2)$, $b_3 \ge 0$, a_3, $\sigma_{x,j}$, $j = 1, 2, 3$:

$$dx_{j,t} = (a_j - b_j x_{j,t})dt + \sigma_{x,j}dB_{j,t}, \ x_{j,0} \text{ given}, \ j = 1, 2 \tag{57}$$

$$dx_{3,t} = (a_3 - b_3 x_{3,t})dt + \sigma_{x,3}dB_{3,t}, \ x_{3,0} \text{ given}. \tag{58}$$

Moreover, as an example of the dividend process (44), we suppose

$$dD_t = D_t[\mu_D(x_t)dt + \sum_{j=1}^3 \sigma_{D,j}(x_t)dB_{j,t}], \ D_0 > 0, \tag{59}$$

in which

$$\mu_D(x_t) = \mu_0 + \mu_{1,1}x_{1,t} + \mu_{2,1}x_{2,t} + \mu_1 x_{1,t}^2 + \mu_2 x_{2,t}^2 + \mu_3 x_{3,t},$$
$$\text{with constants } \mu_{1,1}, \mu_{2,1}, \mu_j(j = 0, 1, 2, 3), \tag{60}$$
$$\sigma_{D,j}(x_t) = \sigma_{D,j}x_j, \ \sigma_{D,j} > 0 \ (j = 1, 2); \ \sigma_{D,3}(x_t) = \sigma_{D,3},$$
$$\text{with constants } \sigma_{D,j}(j = 1, 2, 3), \tag{61}$$

with $\mu_{1,1} = 0$ when $a_1 = 0$ and $\mu_{2,1} = 0$ when $a_2 = 0$.

In addition, we specify

$$\bar{\lambda}_j(x_t) = \phi_j x_{j,t} \text{ with constants } \phi_j > 0 \ (j = 1, 2). \tag{62}$$

Thus, we have $\bar{\lambda}_j(x_t)\sigma_{D,j}(x_t) = \sigma_{D,j}\phi_j x_{j,t}^2$ $(j = 1, 2)$ in (51), and the optimal $\lambda_{j,t}$ $(j = 1, 2)$ are given by $\lambda_{1,t} = -\bar{\lambda}_1(x_t) = -\phi_1 x_{1,t}$ and $\lambda_{2,t} = \bar{\lambda}_2(x_t) = \phi_2 x_{2,t}$. ($\lambda_{3,t} \equiv 0$ by our standing assumption.)

Then, each of $B^{\bar{\lambda}}_{1,t} = B_{1,t} + \int_0^t \bar{\lambda}_{1,s} ds$, $B^{\bar{\lambda}}_{2,t} = B_{2,t} - \int_0^t \bar{\lambda}_{2,s} ds$ and $B^{\bar{\lambda}}_{3,t} = B_{3,t}$ is a Brownian motion under $P^{\bar{\lambda}}$ with $\bar{\lambda} = (-\bar{\lambda}_1, \bar{\lambda}_2, 0)$. (e.g. Lemma 3.9 and Exercise 3-11-i) in Chapter 3 of Bain and Crisan (2008).)

Thus, the equilibrium interest rate (51) is given by

$$r_t = \beta + \mu_D(x_t) - \sum_{j=1}^{2} \sigma_{D,j}^2 x_{j,t}^2 - \sigma_{D,3}^2 - \sigma_{D,1}\phi_1 x_{1,t}^2 + \sigma_{D,2}\phi_2 x_{2,t}^2,$$

$$(\beta > 0, \mu_j \in \mathcal{R}, \ \sigma_{D,j} > 0, \ \phi_j > 0 \ (j = 1, 2); \ \mu_3, \sigma_{D,3} \in \mathcal{R}), \tag{63}$$

in which $\mu_D(x_t) = \mu_0 + \mu_{1,1} x_{1,t} + \mu_{2,1} x_{2,t} + \mu_1 x_{1,t}^2 + \mu_2 x_{2,t}^2 + \mu_3 x_{3,t}$. ($\mu_{1,1} = 0$ when $a_1 = 0$ and $\mu_{2,1} = 0$ when $a_2 = 0$.)

Moreover, we obtain the SDEs of $x_{j,t}$, $j = 1, 2, 3$ under $P^{\bar{\lambda}}$ as

$$dx_{1,t} = [a_1 - (b_1 + \phi_1\sigma_{x,1})x_{1,t}]dt + \sigma_{x,1}dB^{\bar{\lambda}}_{1,t}, \tag{64}$$

$$dx_{2,t} = [a_2 - (b_2 - \phi_2\sigma_{x,2})x_{2,t}]dt + \sigma_{x,2}dB^{\bar{\lambda}}_{2,t}, \tag{65}$$

$$dx_{3,t} = (a_3 - b_3 x_{3,t})dt + \sigma_{x,3}dB^{\bar{\lambda}}_{3,t}, \tag{66}$$

where we suppose $b_1 + \phi_1\sigma_{x,1} > 0$ and $b_2 - \phi_2\sigma_{x,2} > 0$.

In addition, we have each SDE of $x_{i,t}$, $i = 1, 2, 3$ under $Q^{\bar{\lambda}}$ as

$$dx_{1,t} = [a_1 - (b_1 + (\phi_1 + \sigma_{D,1})\sigma_{x,1})x_{1,t}]dt + \sigma_{x,1}dB^{Q^{\bar{\lambda}}}_{1,t}, \tag{67}$$

$$dx_{2,t} = [a_2 - (b_2 - (\phi_2 - \sigma_{D,2})\sigma_{x,2})x_{2,t}]dt + \sigma_{x,2}dB^{Q^{\bar{\lambda}}}_{2,t}, \tag{68}$$

$$dx_{3,t} = (a_3 - \sigma_{D,3}\sigma_{x,3} - b_3 x_{3,t})dt + \sigma_{x,3}dB^{Q^{\bar{\lambda}}}_{3,t}, \tag{69}$$

where $B^{Q^{\bar{\lambda}}} = (B^{Q^{\bar{\lambda}}}_j)_{j=1}^3$ is a three dimensional Brownian motion under $Q^{\bar{\lambda}}$, and we assume $b_1 + (\phi_1 + \sigma_{D,1})\sigma_{x,1} > 0$ and $b_2 - (\phi_2 - \sigma_{D,2})\sigma_{x,2} > 0$.

Finally, as for the term structure of interest rates (55), we omit the expression of Gaussian part (x_3), which is well-known. (e.g. Appendix in Nakano et al. (2018)) The expression corresponding to quadratic-Gaussian factors (x_i^2, $i = 1, 2$) will be given in the next section.

We remark that if we set $a_j = 0$ ($j = 1, 2$) in (57) as a special case, $y_j := x_j^2$ is the solution to the following SDE: with $a_{y,j} = \sigma_{x,j}^2$, $b_{y,j} = 2b_j$ and $\sigma_{y,j} = 2\sigma_{x,j}$,

$$dy_{j,t} = (a_{y,j} - b_{y,j}y_{j,t})dt + \sigma_{y,j}\sqrt{y_{j,t}}dB_{j,t}. \tag{70}$$

That is, x_j^2 becomes the so called squared-root (CIR[15]) process. (See Theorem 3.2 in Shirakawa (2002) for a measure change associated with $\bar{\lambda}_j(y_{j,t}) = \phi_j\sqrt{y_{j,t}}$ with positive constants ϕ_j ($j = 1, 2$).)

B.2 Expression of Term Structure in Quadratic-Gaussian Model

This appendix shows an expression of a term structure of interest rates corresponding to a Quadratic-Gaussian factor in the previous section. Without loss of generality, we use a notation $\mathbb{Q}$ instead of $\mathbb{Q}^\lambda$ for a risk-neutral probability measure.

---

[15]Cox-Ingersoll-Ross(1985)

First, let us introduce the following SDE: with constants $\kappa, \hat{\theta}, \sigma$,

$$dy_t = \kappa(\hat{\theta} - y_t)dt + \sigma dB_t^{\mathbb{Q}}; \ y_0 \text{ given}, \ \kappa > 0, \tag{71}$$

and consider a quadratic function of y_t with constants $\hat{a}, b, c$:

$$\hat{a} + by_t + \frac{1}{2}cy_t^2 = \frac{c}{2}\left(y_t + \frac{b}{c}\right)^2 + \left(\hat{a} - \frac{b^2}{2c}\right) = \frac{c}{2}x_t^2 + a, \tag{72}$$

in which $x_t = y_t + \frac{b}{c}$, $a = \hat{a} - \frac{b^2}{2c}$, and

$$dx_t = \kappa(\theta - x_t)dt + \sigma dB_t^{\mathbb{Q}}; \ \theta = \hat{\theta} + \frac{b}{c}. \tag{73}$$

Thus, an interest rate model, determined by a quadratic function (72) of factor y which follows a mean-reverting Gaussian process (71), is represented by

$$dx_t = \kappa(\theta - x_t)dt + \sigma dB_t^{\mathbb{Q}}; \ x_0 \text{ given}, \ \kappa > 0, \tag{74}$$

$$r_t = a + \frac{1}{2}cx_t^2; \ a, c \in \mathbf{R}; \ c \geq -\left(\frac{\kappa}{\sigma}\right)^2. \tag{75}$$

This is the quadratic-Gaussian part of the interest rate model ((1) and (2)) in Section 2 with appropriate modification. Hereafter, we consider the above model for a term structure of interest rates.

Let $P(t, T)$ denote a zero coupon bond price at time t with maturity $T(t \leq T)$, and define $P_t(\tau) := P(t, T)$ with $\tau = T - t$. Then, we would like to obtain

$$P_t(\tau) = E^{\mathbb{Q}}\left[e^{-\int_t^{t+\tau} r_u du} | \mathcal{F}_t\right], \tag{76}$$

$$Y_t(\tau) = \frac{-1}{\tau} \log P_t(\tau). \tag{77}$$

Next, we conjecture that

$$P(t, T) = P(x_t, t, T) = \exp\left[A(T - t) + B(T - t)x_t + C(T - t)\frac{x_t^2}{2}\right]. \tag{78}$$

Since $P(x_t, t; T)e^{-\int_0^t r_u du}$ is a martingale under $\mathbb{Q}$, after applying Ito's formula to $P(x_t, t; T)e^{-\int_0^t r_u du}$ and setting the drift term to be zero, it is enough to solve the following ODEs to obtain $A(\tau), B(\tau), C(\tau)$:

$$C'(\tau) + 2\kappa C(\tau) - \sigma^2 C^2(\tau) + c = 0; \ C(0) = 0, \tag{79}$$

$$B'(\tau) + \kappa B(\tau) - \kappa\theta C(\tau) - \sigma^2 B(\tau)C(\tau) = 0; \ B(0) = 0, \tag{80}$$

$$A'(\tau) - \kappa\theta B(\tau) - \frac{1}{2}\sigma^2 B^2(\tau) - \frac{1}{2}\sigma^2 C(\tau) + a = 0; \ A(0) = 0 \tag{81}$$

Then, $C(\tau)$ is given as follows:

$$C(\tau) = C_0 + \frac{1}{z(\tau)}; \ C_0 = \frac{\kappa + \sqrt{\kappa^2 + c\sigma^2}}{\sigma^2}, \tag{82}$$

$$z(\tau) = \frac{\sigma^2}{\alpha} - e^{\alpha\tau}\left(\frac{1}{C_0} + \frac{\sigma^2}{\alpha}\right); \ \alpha = 2(\kappa - \sigma^2 C_0) = -2\sqrt{\kappa^2 + c\sigma^2}. \tag{83}$$

Given $C(\tau)$, we have $B(\tau)$ as

$$B(\tau) = \frac{\kappa\theta}{\sigma^2}\left[\frac{e^{\frac{\alpha\tau}{2}}(1+C_2)}{1+C_2e^{\alpha\tau}} - 1 + \kappa\frac{2e^{\frac{\alpha\tau}{2}}\{1 - e^{\frac{-\alpha\tau}{2}} + C_2(e^{\frac{\alpha\tau}{2}} - 1)\}}{\alpha(1+C_2e^{\alpha\tau})}\right], \qquad (84)$$

$$C_2 = \frac{\alpha}{\sigma^2}z(0) - 1; \ z(0) = \frac{-1}{C_0}. \qquad (85)$$

In particular, when $\theta = 0$, $B(\tau) \equiv 0$.

Then, given $C(\tau)$ and $B(\tau)$, we obtain $A(\tau)$ as follows:

$$A(\tau) = \kappa\theta\int_0^\tau B(s)ds + \frac{\sigma^2}{2}\int_0^\tau B^2(s)ds + \frac{\sigma^2}{2}\int_0^\tau C(s)ds - a\tau, \qquad (86)$$

where

$$\frac{\sigma^2}{2}\int_0^\tau C(s)ds = \frac{1}{2}\left\{\left(\kappa + \frac{\alpha}{2}\right)\tau + \ln\frac{(1+C_2)}{(1+C_2e^{\alpha\tau})}\right\}. \qquad (87)$$

Here, we note that since $\alpha = -2\sqrt{\kappa^2 + c\sigma^2} < 0$ with $e^{\alpha\tau} < 1$ and $C_0 > 0$, we have $1 + C_2 = \frac{\alpha}{\sigma^2}z(0) = \frac{2\sqrt{\kappa^2+c\sigma^2}}{C_0\sigma^2} > 0$, and $1 + C_2e^{\alpha\tau} = (1 - e^{\alpha\tau}) + \frac{2\sqrt{\kappa^2+c\sigma^2}}{\sigma^2 C_0}e^{\alpha\tau} > 0$.

Finally, we remark that $a = 0$ and $\theta = 0$ (i.e. $B(\tau) \equiv 0$) in our empirical analysis of the main text. Moreover, setting $\hat{x}_t := \frac{x_t}{\sqrt{2}}$, we have

$$d\hat{x}_t = -\kappa\hat{x}_t dt + \frac{\sigma}{\sqrt{2}}dB_t^{\mathbb{Q}}, \qquad (88)$$

and then redefining $\hat{x}_t$ as x_t with $\sigma_x \equiv \frac{\sigma}{\sqrt{2}}$ provide the equations (2) and (6) in Section 2. (Here, the subindex j is omitted for simplicity.)

C   Estimation of Trend and Correlation in Noisy Time Series Data

This appendix introduces a new method for estimation of trends in time series and their correlations, which is based on the Separating Information Maximum Likelihood (SIML) method applied in Section 3.3.

First, we briefly explains the SIML method by using a simple case. (See Kunitomo et. al. (2018) for more general cases.) Let us suppose that observed time series $(x_t)_{t=1}^n$ and $(y_t)_{t=1}^n$ are generated by the following stochastic processes, each of which is a sum of a stationary process e_x (e_y) and a non-stationary process μ_x (μ_y):

$$x_t = \mu_{x,t} + e_{x,t}, \qquad (89)$$
$$y_t = \mu_{y,t} + e_{y,t}, \qquad (90)$$
$$\mu_{x,t} = \mu_{x,t-1} + \eta_{x,t}, \qquad (91)$$
$$\mu_{y,t} = \mu_{y,t-1} + \eta_{y,t}, \qquad (92)$$

where each of $\eta_{x,t}, \eta_{y,t}$ is i.i.d. normal with mean 0 across time t, and the correlation between $\eta_{x,t}$ and $\eta_{y,t}$ is given as a constant ρ_{xy} for all t. Then, we define a SIML correlation

estimator $cor_{SIML}(x, y)$ for ρ_{xy} as follows:

$$cor_{SIML}(x, y) = \frac{\sum_{i=1}^{m} z_{x,i} z_{y,i}}{\sqrt{(\sum_{i=1}^{m} z_{x,i}^2)(\sum_{i=1}^{m} z_{y,i}^2)}}, \tag{93}$$

where with $x = (x_1, x_2, \cdots, x_n)^\top$, $y = (y_1, y_2, \cdots, y_n)^\top$, $z_x = (z_{x,1}, z_{x,2}, \cdots, z_{x,n})^\top$ and $z_y = (z_{y,1}, z_{y,2}, \cdots, z_{y,n})^\top$ are defined as

$$z_x = P_n C_n^{-1}(x - x_0) \tag{94}$$
$$z_y = P_n C_n^{-1}(y - y_0), \tag{95}$$

and $m = \lfloor n^\alpha \rfloor$ with $\alpha = 0.45$.[16] Here, C_n^{-1} is an $n \times n$-matrix representing the first-order difference, and P_n is an $n \times n$-orthogonal matrix such that $C_n^{-1} C_n^{\top -1} = P_n D_n P_n^\top$ with a diagonal matrix D_n. That is,

$$C_n^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \tag{96}$$

$$P_n = (p_{ij})_{1 \le i,j \le n} \tag{97}$$

$$p_{ij} = \sqrt{\frac{2}{n + 1/2}} \cos\left[\frac{2\pi}{2n+1}(i - \frac{1}{2})(j - \frac{1}{2})\right]. \tag{98}$$

To detect a trend component in each x and y, we use $z_{x,i}, z_{y,i}$ only with $i = 1, \cdots, m(< n)$ (e.g. $m = 22$ out of $n = 977$ in an example below) for an estimation, because z_i with smaller i contains information about a longer cyclic component.

Next, we introduce a new method to estimate trends and their correlations. Concretely, we estimate trend series by T_x for x and T_y for y that are defined as as follows:

$$T_x = C_n P_{m,n}^\top P_{m,n} C_n^{-1} x, \tag{99}$$
$$T_y = C_n P_{m,n}^\top P_{m,n} C_n^{-1} y, \tag{100}$$

in which

$$P_{m,n} = (p_{ij}) \ (i = 1, \cdots, m, j = 1, \cdots, n). \tag{101}$$

Then, it is easily seen that a standard sample correlation between the first-order differences of these trend series is same as a SIML correlation estimator, i.e.

$$cor(C_n^{-1} T_x, C_n^{-1} T_y) = cor_{SIML}(x, y). \tag{102}$$

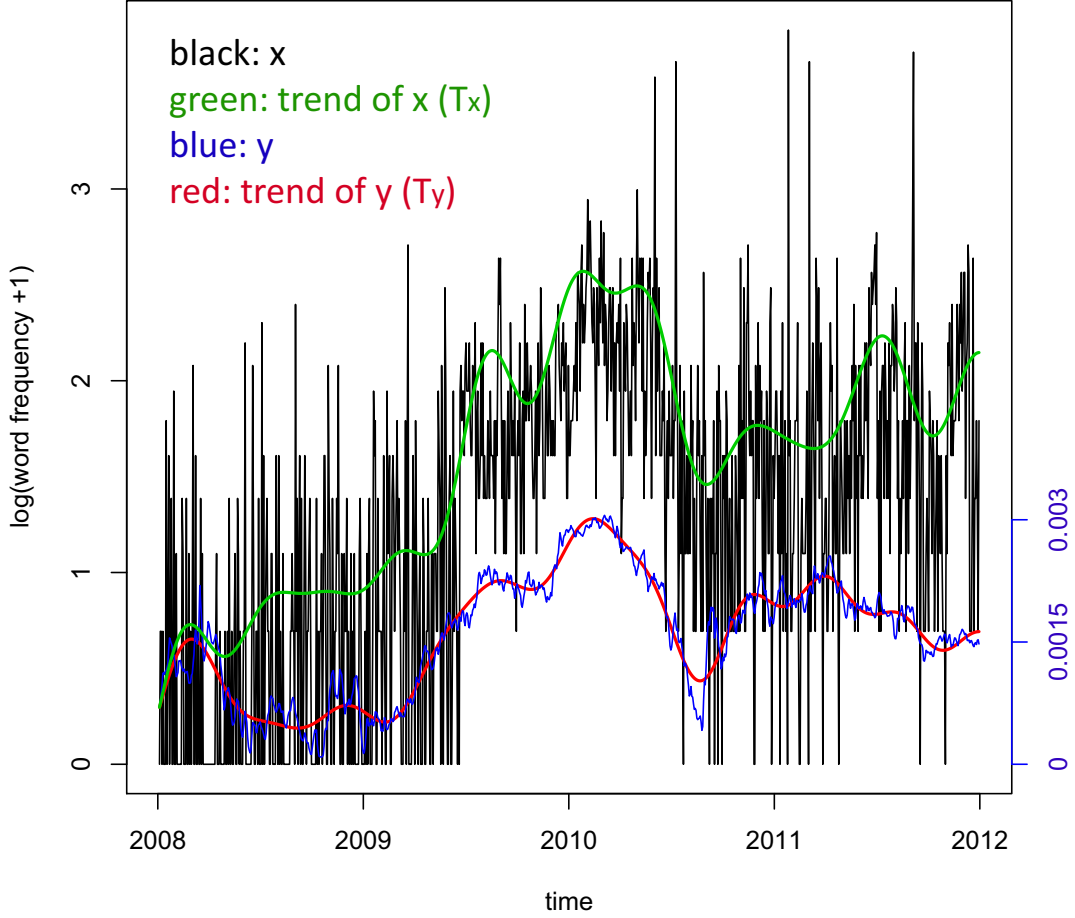[16]$\alpha \in (0, 0.5)$ gives asymptotic consistency for the SIML correlation estimator.

Figure 14: time series plot for log-frequencies of "fiscal conditions (zaisei) & foreign (gaikoku)" (x) and the steepening factor x_1^2 estimated by "two-factor Gaussian minus Quadratic-Gaussian" model (y, right axis)

Next, we apply our method to a text mining analysis in Section 3.3, particularly, the observed log-frequency of two-word set "fiscal conditions (zaisei) & foreign (gaikoku)" as x, and the steepening factor x_1^2 estimated in Section 3.1 as y. We remark that since our estimate of the mean-reversion speed in the system equation (17) is $\kappa_1 = 0.129$, which implies that the autoregressive coefficients for $x_{1,t}$ and $y_t = x_{1,t}^2$ (i.e. $e^{-\kappa_1 \Delta t}$ and $e^{-2\kappa_1 \Delta t}$ with $\Delta t = 1/250$) are close to 1, we may regard the time series $(x_{1,t})_{t=1,\cdots,n}$ and $(y_t)_{t=1,\cdots,n}$ with $n = 977$ as non-stationary processes. In fact, given the estimated time series $(y_t)_{t=1,\cdots,n}$, a null hypothesis, $\beta = 0$ is not rejected at 10% significant level in a simple regression equation: $y_t - y_{t-1} = \alpha + \beta y_{t-1} + \epsilon_t$ ($\epsilon_t \sim$ i.i.d. $N(0, \sigma_\epsilon^2)$).

Figure 14 shows the result. It is observed that the estimated trends T_x and T_y show similar movements, which is captured by the SIML correlation, 0.62 as reported in Figure 3. In contrast, because the observed two-word set's frequency suffers from large noises, the standard sample correlation between the first-order differences of x and y is only 0.02. Moreover, the correlation between levels of those trends, i.e. T_x and T_y becomes very high (0.86), while the standard sample correlation is 0.59 as reported in Figure 3. Further, since the estimated x_1^2 has relatively small noises, the correlation between T_x and y is also high (0.83).

We finally remark that the method described in this section is expected to be effective

with relatively little computational burden in detecting trends and long-term relationships embedded in very noisy time series data such as word frequencies.

References

[1] Adrian, T. (2017). The term structure of interest rates and macrofinancial dynamics, Speech at Bank of Canada Conference on Advances in Fixed Income and Macro-Finance Research, August 17, 2017.

[2] Bauer, M.D.(2015). Nominal interest rates and the news, Journal of Money, Credit and Banking, Vol. 47, No. 2–3 (March–April 2015).

[3] Cox, J.C., Ingersoll J.E., and Ross, S.A.(1985). A theory of the term structure of interest rates. Econometrica, 53, 385–407.

[4] Fukui, T., Sato, S., and Takahashi, A. (2017). Style analysis with particle filtering and generalized simulated annealing. International Journal of Financial Engineering, 4(02n03), 1750037.

[5] Gotthelf, N., and Uhl, M.W. (2018). News Sentiment: A new yield curve factor, Journal of Behavioral Finance, DOI: 10.1080/15427560.2018.1432620.

[6] Hull, J., and White, A. (1990). Pricing interest-rate-derivative securities. The Review of Financial Studies, 3(4), 573-592.

[7] Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of Computational and Graphical Statistics, **5**, 1, 1–25.

[8] Kumar B.S. and Vadlamani Ravi, V. (2016). A survey of the applications of text mining in financial domain, Knowledge-Based Systems, 114, 128–147.

[9] Kunitomo, N., Sato, S., and Kurisu, D. (2018). Separating Information Maximum Likelihood Method for High-Frequency Financial Data. Springer.

[10] MeCab (2006). Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge. jp/

[11] Nakano, M., Takahashi, A., Takahashi, S., and Tokioka, T. (2018). On the effect of Bank of Japan's outright purchase on the JGB yield curve. Asia-Pacific Financial Markets, 25(1), 47-70.

[12] Nassirtoussi A.K., Aghabozorgi, S., Waha, T.Y. Ngo, D.C.L. (2014). Text mining for market prediction: A systematic review, Expert Systems with Applications, 41, 7653–7670.

[13] Nishimura, K. G. (2008). Recent Economic and Financial Developments and the Conduct of Monetary Policy, Speech at the Foreign Correspondents' Club of Japan, September 29, 2008, Bank of Japan.

[14] Nishimura, K. G., and H. Ozaki. (2017) Economics of Pessimism and Optimism: Theory of Knightian Uncertainty and Its Applications, Springer.

[15] Rudebusch, G. D., and Wu, T. (2008). A macro- finance model of the term structure,no Monetary Policy and the Economy, Economic Journal 118(530), 906-926.

[16] Takahashi, A. and Sato, S. (2001). Monte Carlo Filtering Approach for Estimating the Term Structure of Interest Rates, *Annals of The Institute of Statistical Mathematics*, **53**, No.1, 50-62.

[17] Bain, A. and Crisan, D. (2008). Fundamentals of Stochastic Filtering, Springer.

[18] Karatzas and Shreve(1998). Methods of Mathematical Finance, Springer.

[19] Karatzas and Shreve (1991). Brownian Motion and Stochastic Calculus, Springer.

[20] Shirakawa,H. (2002). Squared Bessel Processes and their Applications to the Squared Root Interest Rate Model, Asia-Pacific Financial Markets 9, pp. 169-190.

[21] Shreve, S.E. (2004). Stochastic Calculus for Finance II: Continuous-Time Models, Springer.