

# Entropy Based Student's $t$ -Process Dynamical Model

Ayumu Nono <sup>1,\*</sup>, Yusuke Uchiyama <sup>2</sup> and Kei Nakagawa <sup>3</sup>

<sup>1</sup> Graduated School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

<sup>2</sup> MAZIN Inc., 3-29-14 Nishi-Asakusa, Tito City, Tokyo 111-0035, Japan; uchiyama@mazin.tech

<sup>3</sup> NOMURA Asset Management Co. Ltd., 2-2-1 Toyosu, Koto-ku, Tokyo 135-0061, Japan; k-nakagawa@nomura-am.co.jp

\* Correspondence: nono-ayumu303@g.ecc.u-tokyo.ac.jp

**Abstract:** Volatility, which represents the magnitude of fluctuating asset prices or returns, is used in the problems of finance to design optimal asset allocations and to calculate the price of derivatives. Since volatility is unobservable, it is identified and estimated by latent variable models known as volatility fluctuation models. Almost all conventional volatility fluctuation models are linear time-series models and thus are difficult to capture nonlinear and/or non-Gaussian properties of volatility dynamics. In this study, we propose an entropy based Student's  $t$ -process Dynamical model (ETPDM) as a volatility fluctuation model combined with both nonlinear dynamics and non-Gaussian noise. The ETPDM estimates its latent variables and intrinsic parameters by a robust particle filtering based on a generalized H-theorem for a relative entropy. To test the performance of the ETPDM, we implement numerical experiments for financial time-series and confirm the robustness for a small number of particles by comparing with the conventional particle filtering.

**Keywords:** finance; volatility fluctuation; Student's  $t$ -process; entropy based particle filter; relative entropy



**Citation:** Nono, A.; Uchiyama, Y.; Nakagawa, K. Entropy Based Student's  $t$ -Process Dynamical Model. *Entropy* **2021**, *23*, 560. <https://doi.org/10.3390/e23050560>

Academic Editor: Christian H. Weiss

Received: 30 March 2021

Accepted: 27 April 2021

Published: 30 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Asset allocation and pricing derivatives have been studied in both academia and industry as significant problems in financial engineering and quantitative finance. For these problems, various methodologies have been developed based on the variation of asset returns. In an idealized situation, the variation of returns has been assumed to follow the Gaussian distribution [1]. However, it is known that the variation of returns follows non-Gaussian distributions with fat tails [2]. To explain this observed fact, volatility, which quantifies the magnitude of fluctuating returns, has been introduced and utilized. Volatility, in particular, is often used as an indicator for constructing asset allocations that focus on macroeconomic fundamentals, and there are many studies related to them. Both researchers and investors have begun to attend to develop mathematical models of volatility fluctuations. For example, Yuhuang et al. investigated the impact of fundamental data on oil price volatility by focusing on time-varying skewness and kurtosis [3]. Hou et al. studied volatility spillovers between the Chinese fuel oil futures market and the stock index futures market, taking into account the time-varying characteristics of the markets [4].

In general, volatility is defined as the variance of a conditional Gaussian distribution for the variation of returns, namely, given as a latent variable in the literature of Bayesian statistical modeling. Based on this idea, various time-series models for the dynamics of asset returns have been developed and proposed. Such time-series models are generally called volatility fluctuation models, on which forecasting, state estimation and smoothing can be implemented.

In recent years, volatility fluctuation models with a machine learning technique have been proposed [5]. Since volatility is a latent variable, it is necessary for machine learning models to incorporate latent variables into their own methodology. The Gaussian process

is a candidate, such as a Bayesian learning model [6], and its applications for several problems in finance have been reported [7,8]. The Student's  $t$ -process is an extension of the Gaussian, for non-Gaussian distributed data such as asset returns. It has been proposed [9] and applied to the analysis of financial time-series and asset allocations, and it is confirmed for this model to estimate robustly [10].

This study extends the Student's  $t$ -process latent variable model to a dynamic latent variable model incorporating the structure of time-series. To estimate dynamic latent variables, we used the particle filter method [11]. The particle filter is used to estimate the latent variables. Conventional particle filters have problems called weight bias and the particle impoverishment problem (PIP), directly affecting the estimation accuracy [12]. Then, the merging particle method [13] and Monte Carlo filter particle filter [14] have been proposed. However, these methods are computationally expensive because they need a large number of particles. Therefore, we used an Entropy-based particle filter (EBPF), which constructs a parametric prior distribution on the generalized H-theorem for relative entropy [15]. It is expected to prevent the bias of particle weights and the loss of particle diversity while reducing the computational cost. Using EBPF in this experiment, and comparing it with conventional methods, we confirmed that it is effective for finance problems.

In summary, to estimate robustly and avoid the particle filter's problem, we combined  $t$ -process dynamical model and EBPF. We call the proposed model an entropy based Student's  $t$ -process dynamical model (ESTDM), in the following. We will verify this model's usefulness. The remains of this paper are summarized as follows—Section 2 introduces related statistical and machine learning models. In Section 3, we derive and propose ESTDM with its filtering method. In Section 4, we show the performance of volatility estimation using the proposed method and discuss the results. Section 5 is devoted to our conclusions and future perspectives.

## 2. Related Work

### 2.1. Volatility Fluctuation Models

One of the most basic and utilized volatility fluctuation models is the GARCH model [16] given as follows:

$$x_t \sim \mathcal{N}(0, \sigma_t^2), \quad (1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j \sigma_{t-j}^2 + \sum_{i=1}^p \beta_i x_{t-i}^2, \quad (2)$$

where  $x_t$  is a time-dependent random variable sampled from a Gaussian distribution with mean 0 and variance  $\sigma_t^2$ , and the time evolution of the variance is given by Equation (2). The parameters  $\alpha_j$  and  $\beta_i$  take positive values, which can be estimated by observed data. Positive integers  $p$  and  $q$  are the order of the regression, respectively. Then this model is known as the GARCH( $p, q$ ) model. For the sake of simplicity, the order parameters are often fixed as  $p = q = 1$ . Various families of GARCH model have been developed and proposed in the area of econometrics and quantitative finance [17]. For instance, asymmetric effect has been introduced into a multivariate GARCH model [18,19].

### 2.2. Gaussian Process

For any finite number of vectors  $\{x_1, x_2, \dots, x_n\}$  and a stochastic process  $f(\cdot)$ , if the joint probability density function  $\{f(x_1), f(x_2), \dots, f(x_n)\}$  is a Gaussian distribution,  $f(\cdot)$  is called a Gaussian process [6]. Since the Gaussian process samples an infinite-dimensional vector, the mean value function  $m(\cdot)$  and the covariance function  $K(\cdot, \cdot)$  are introduced as follows:

$$m(x) = \mathbb{E}[f(x)], \quad (3)$$

$$K(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))^T]. \quad (4)$$

Then, given a matrix  $X = [x_1, x_2, \dots, x_n]^T$ ,  $p(f|X) = \mathcal{N}(m(X), K(X, X))$  is the probability density function of the Gaussian process. When we explicitly state that the stochastic process  $f$  is sampled from the Gaussian process, we write  $f \sim \mathcal{GP}(m, K)$ . Without loss of generality, the mean function of the Gaussian process is often assumed to be zero. The covariance function is represented by the kernel function  $k(\cdot, \cdot)$ , which is a positive symmetric bi-variate function, satisfying

$$K(x, x') = k(x, x'). \tag{5}$$

Hence,  $K(X, X)$  is a positive definite symmetric matrix. As a kernel function, for example, the radial basis function

$$k_{\text{RBF}}(x, x') = \alpha \exp(-l^{-2} \|x - x'\|^2) \tag{6}$$

is often used. Here,  $\alpha$  and  $l$  are hyper parameters.

For a pair of observed data  $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , let  $X = [x_1, x_2, \dots, x_n]^T$ ,  $Y = [y_1, y_2, \dots, y_n]^T$ . The hyper parameters of the Gaussian process can be estimated by gradient and Monte Carlo methods on  $\mathcal{D}$ . From the trained Gaussian process, the prediction  $Y^* = [y_1^*, y_2^*, \dots, y_m^*]^T$  for unknown input  $X^* = [x_1^*, x_2^*, \dots, x_m^*]^T$  is sampled from the conditional Gaussian distribution  $\mathcal{N}(f^*, K^*)$ . The mean function  $f^*$  and the covariance function  $K^*$  of the conditional Gaussian distribution are given by

$$f^* = m_X + K_{X^*, X} K_{X, X}^{-1} Y, \tag{7}$$

$$K^* = K_{X^*, X^*} - K_{X^*, X} K_{X, X}^{-1} K_{X, X^*}. \tag{8}$$

It is seen that the mean and covariance functions of the Gaussian process propagate the information of previously observed data to predicted values.

### 2.3. Student's $t$ -Process

In the Gaussian process, it is assumed for the probability density function to be the Gaussian distribution. Thus, when we apply the Gaussian process to data following a probability distribution with fat tails, such as financial time-series, it is impossible to perform an accurate estimation. A model that extends the Gaussian process to such data is the Student's  $t$ -process [9]. The Student's  $t$ -process is a stochastic process  $f(\cdot)$  with  $\nu$  degrees of freedom and a Student's  $t$ -distribution defined as follows:

$$\mathcal{T}(m, K, \nu) = \frac{\Gamma(\frac{\nu+n}{2})}{[(\nu-2)\pi]^{\frac{n}{2}} \Gamma(\frac{\nu}{2}) |K_{X, X}|^{\frac{1}{2}}} \left[ 1 + \frac{1}{\nu-2} (Y - m_X)^T K_{X, X}^{-1} (Y - m_X) \right]^{-\frac{\nu+n}{2}}. \tag{9}$$

Here,  $m(\cdot)$  and  $K(\cdot, \cdot)$  are the mean and covariance functions, respectively, and  $\Gamma(\cdot)$  is the gamma function. When the stochastic process  $f(\cdot)$  is a Student's  $t$ -process, it is denoted by  $f \sim \mathcal{TP}(m, K; \nu)$ . As with the Gaussian process, the mean function of the Student's  $t$ -process is often assumed to be zero without loss of generality.

The predictive distribution of the Student's  $t$ -process is also the Student's  $t$ -distribution  $\mathcal{T}(m^*, K^*, \nu^*)$ , where degrees of freedom, mean and covariance functions are then updated as follows:

$$\nu^* = \nu + n, \tag{10}$$

$$m^* = m_X + K_{X^*, X} K_{X, X}^{-1} (Y - m_X) \tag{11}$$

$$K^* = \frac{\nu - \beta - 2}{\nu - n - 2} [K_{X^*, X^*} - K_{X^*, X} K_{X, X}^{-1} K_{X, X^*}]$$

$$\beta = (Y - m_X)^T K_{X, X}^{-1} (Y - m_X). \tag{12}$$

Unlike the Gaussian process, in the Student's  $t$ -process, we can confirm that the effect of the number of data is reflected in the update equations of the degrees of freedom and the covariance function.

#### 2.4. Student's $t$ -Process Latent Variable Model

In the Student's  $t$ -process latent variable model, the input matrix  $X$  is given as a latent variable. Assume that the observed data  $y \in \mathbb{R}^D$  and the latent variable  $x \in \mathbb{R}^Q$  are related as  $y = f(x)$  by the Student's  $t$ -process  $f \sim \mathcal{TP}(m, K; \nu)$ . When we let  $Y \in \mathbb{R}^{D \times N}$  be the sequence of  $N$  observed data, and  $X \in \mathbb{R}^{Q \times N}$  be the sequence of  $N$  latent variables, we can define the following model as Student's  $t$ -process latent variable model [10]:

$$p(Y|X) = \frac{\Gamma\left(\frac{\nu+D}{2}\right)}{[(\nu-2)\pi]^{\frac{D}{2}} \Gamma\left(\frac{\nu}{2}\right) |K_{X,X}|^{\frac{1}{2}}} \left[1 + \frac{1}{\nu-2} (Y - m_X)^T K_{X,X}^{-1} (Y - m_X)\right]^{-\frac{\nu+D}{2}}. \quad (13)$$

Since the Student's  $t$ -distribution converges to the Gaussian distribution in the limit of  $\nu \rightarrow \infty$ , we can see that the Student's  $t$ -process latent variable model embraces the Gaussian process latent variable model [20].

### 3. Proposed Model

#### 3.1. Student's $t$ -Process Dynamical Model

Since volatility fluctuations cannot be observed directly, they are modeled as dynamic latent variables, such as the family of GARCH models, most of which are given by linear time-series models [18]. To describe nonlinear dynamics of volatility fluctuations, we extend the Student's  $t$ -process latent variable model to dynamic latent variables, namely, Student's  $t$ -process dynamical model (TPDM), which is expected to be robust for both observable and unobservable with outliers.

Suppose  $p_t$  represents an asset price at time  $t$ , the log-return is given by  $r_t = \log(p_t/p_{t-1})$ . Let  $\sigma_t^2$  denote the volatility of  $r_t$ . Here, for an observable  $r_t$  and a latent variable  $\sigma_t^2$ , we provide a volatility fluctuation model by a TPDM as follows:

$$r_t \sim \mathcal{T}(0, \sigma_t^2; \nu), \quad (14)$$

$$v_t \equiv \log \sigma_t^2, \quad (15)$$

$$v_t = f(v_{t-1}, r_{t-1}; \nu) + \epsilon_t \quad (16)$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma_n^2), \quad (17)$$

where the observable  $r_t$  as centered at 0 and following a Student's  $t$ -distribution with  $\nu$  degrees of freedom, whose parameter is given by  $\sigma_t^2$ . The dynamic latent variable  $v_t$  is defined by Equation (15) to take the whole real number as its range of value. The time evolution of the dynamic latent variable  $v_t$  is given by Equation (16) with a Gaussian white noise whose variance is  $\sigma_n$ . The stochastic process  $f$  on the right-hand side of Equation (16) follows a Student's  $t$ -process given by

$$f \sim \mathcal{TP}(m, K; \nu), \quad (18)$$

$$m(\xi_{t-1}) = av_{t-1} + bx_{t-1}, \quad (19)$$

$$k(\xi_{t-1}, \xi'_{t-1}) = \gamma \exp(-l^{-2} \|\xi_{t-1} - \xi'_{t-1}\|^2), \quad (20)$$

where  $\xi_t = (x_t, v_t)$ , and the hyper parameters are  $\theta = (\nu, \sigma_n, a, b, \gamma, l)$ . Given a series of observed data  $r_{1:T} = \{r_1, r_2, \dots, r_T\}$ , it is possible to obtain the volatility fluctuations by estimating a series of dynamic latent variables  $v_{1:T} = \{v_1, v_2, \dots, v_T\}$ .

#### 3.2. Particle Filter

Particle filter is a method of state estimation by Monte Carlo sampling, where a large number of particles approximates posterior distributions. Hence, it can be applied to

nonlinear systems, where posterior distributions are intractable [21]. For  $N$  particles, let  $\{v_{1:t-1}^i\}_{i=1}^N$  and  $W_{t-1}^i$  ( $i = 1, 2, \dots, N$ ) be the realizations of the dynamic latent variables and their associated weights up to time  $t - 1$ , respectively. The weights are normalized to  $\sum_{i=1}^N W_{t-1}^i = 1$ . With these values, the posterior distribution  $p(v_{1:t-1}|x_{1:t-1})$  at time  $t - 1$  can be approximated as follows [22]:

$$\hat{p}(v_{1:t-1}|x_{1:t-1}) = \sum_{i=1}^N W_{t-1}^i \delta(v_{1:t-1}), \quad (21)$$

where  $\delta(\cdot)$  is the Dirac's delta function. In other words, the posterior distribution is approximated by a mixture of the delta functions.

It is however known that an insufficient number of particles fails to approximate the posterior distribution by the degeneracy of ensemble. Indeed, each particle's weights become unbalanced and biased toward a tiny number of particles as the time step progresses [11,12]. To overcome this problem, a huge amount of particles is used for filtering processes in practice.

### 3.3. Entropy-Based Particle Filter

In the use of the conventional particle filter, it is necessary to sample a large number of particles for accuracy. That leads to the growth of estimation time. In the case of online estimation, reducing run time is desired. For this purpose, we introduce a robust particle filter for a small number of particles.

Let us reconsider approximating the probability density function for the dynamic latent variable,  $Q(v, t)$ , called a background distribution. In the conventional particle filter, the background distribution is approximated by the mixture of delta functions. This approximation works well only when the background distribution exhibits an extensively sharp peak. Nevertheless, the delta function has no width, and the distribution peaks only at a single point.

To improve the accuracy for the approximation of the background distribution, we replace the mixture of the delta functions with that of Gaussian distributions as

$$\hat{Q}(v, t) = \sum_{i=1}^M W_t^i \mathcal{N}(\mu_t^i, \sigma_t^{2,i}), \quad (22)$$

where  $\mu_t^i, \sigma_t^{2,i}$  ( $1 \leq i \leq M$ ) are the mean and variance of the Gaussian distributions at  $t$ . Unlike the delta function, the Gaussian distribution has a certain width in its distribution. Hence, the mixture of the Gaussian distributions is capable of fitting properly to data with large variance and fat tails.

With the use of finite samples from the background distribution  $Q(v, t)$ , the posterior/filter distribution  $P(v, t)$  is inferred by the minimum principal for relative entropy, which is known as an entropy based particle filter [15]. The relative entropy (Kullback-Liebler divergence) between the filter distribution  $P(v, t)$  and the background distribution  $Q(v, t)$  are defined as follows [23–25]:

$$H[P|Q] = \int_{\Omega_v} P(v, t) \log \left( \frac{P(v, t)}{Q(v, t)} \right) dv, \quad (23)$$

where  $\Omega_v$  is the domain of the dynamic latent variable  $v_t$ . On the properties of the relative entropy as a quasi-distance for probability density functions, the filter distribution is obtained as the minimizer for the relative entropy in Equation (23). Combined with the entropy based particle filter, the state estimation of the ESTDM is implemented. An overview of its algorithm is explained in the following Algorithm 1.

**Algorithm 1** Entropy Based Student's  $t$ -Process Dynamical Model (ETPDM)**Require:** Initial particles  $\mathbf{X}_0 = [\mathbf{X}_0^0, \dots, \mathbf{X}_0^M]$ , Initial particles' weights  $W_0^i = 1/M$ **Ensure:**  $\sum_{i=1}^M W_t^i = 1.0$  at any time  $t$ 1: **while** There are observations to be assimilated **do**2: Compute importance weights proportional to the likelihood with observation  $x_t$ 

$$W_t^i \propto p(y_t | \mathbf{X}_t) \quad (24)$$

According to weights  $\{W_t^i\}$ , resample  $M$  particles  $\{\mathbf{X}_t^j\}_{j=1}^M$ .Then we can compute filter distribution  $Q'(x)$  at time  $t$ 

$$Q'_t = \sum_{i=1}^M W_t^i \mathcal{N}(\mathbf{X}_t^i). \quad (25)$$

At the same time, we're also able to estimate expected status  $v_t$ , extracting any finite number of samples  $\{v^k\}$  from background density  $Q_t$ 

$$v_t = \mathbb{E}(v^k). \quad (26)$$

With stochastic process  $f \sim \mathcal{T}(m, k; v)$ , generate new particles

$$\{\mathbf{X}_{t+1}^i\} = f(\mathbf{X}_t^j, x_t) \quad (27)$$

Then we can predict distribution  $\hat{Q}(x)$  at time  $t + 1$ 

$$\hat{Q}_{t+1} = \sum_{i=1}^M W_t^i \mathcal{N}(\mathbf{X}_{t+1}^i). \quad (28)$$

3: **return** Log likelihood for estimation  $p(y_t | v_t)$ .4: **end while****4. Numerical Experiments**

In this section, we implement numerical experiments to validate the ETPDM for the time-series of a foreign exchange rate. As a dataset, we use USD/JPY exchange rate in 2010—every 1-min sampled, 30-min sampled and 1-h sampled. Figures 1–3 show the time-series of the log-return of the USD/JPY exchange rate  $r_t$  and volatility fluctuations estimated by respective the ETPDM, the conventional particle filter for the GARCH model (cp-GARCH) and the conventional particle filter for the Student's  $t$ -process dynamical model (cp-TPDM). Warm up period of the estimations is  $0 \leq t \leq 20$ , where the values of volatility show zeros. In Figure 1a, intermittent jumps are observed, which are evidence of the non-Gaussian behavior of  $r_t$ . Indeed, the estimated volatility fluctuations show higher peaks at the same time point of the intermittent jumps in Figure 1b–d. That means all of the models capture the nature of volatility fluctuations of the USD/JPY exchange rate effectively. Besides, the same can be said for other types of data sets—30-min and 1-h—in Figure 2 or Figure 3, which means that these models can be applied to data of any sampling rate. Previous volatility estimation studies used the GARCH model with various estimation methods. A typical example is the particle filter [26], or the Markov chain Monte Carlo simulation [27]. In all of these studies, including this experiment, the GARCH model has been implemented well.

Overview of results in 1-minute chart

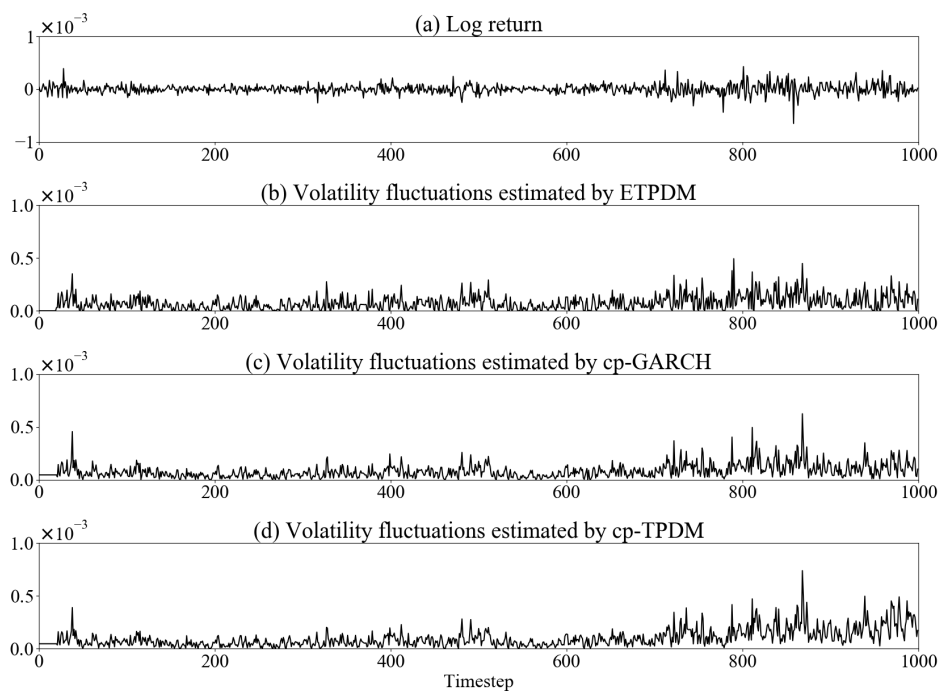


Figure 1. Overview of estimation results in 1-min chart.

Overview of results in 30-minute chart

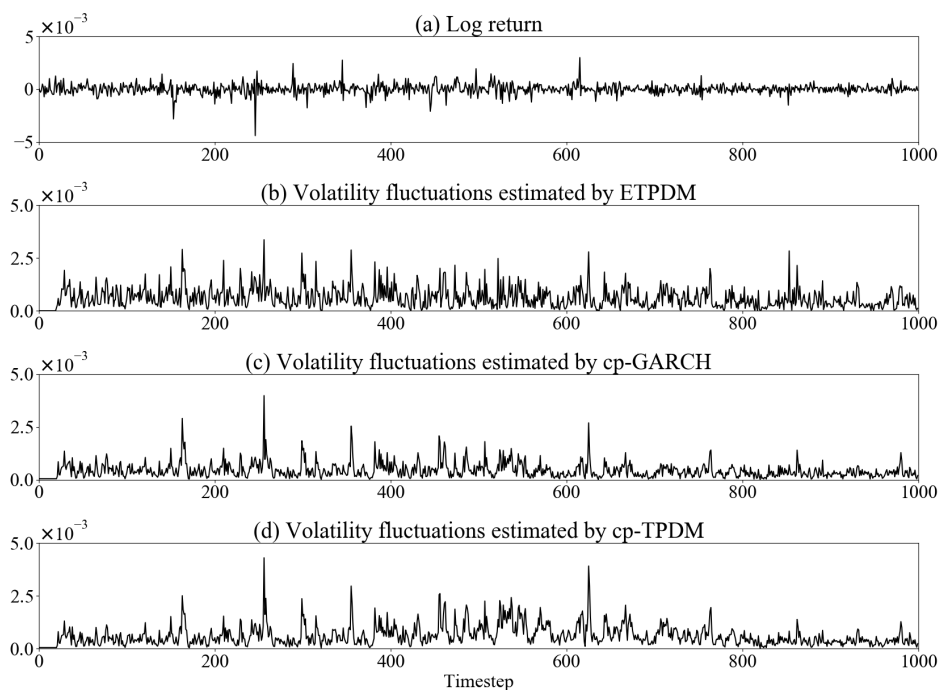
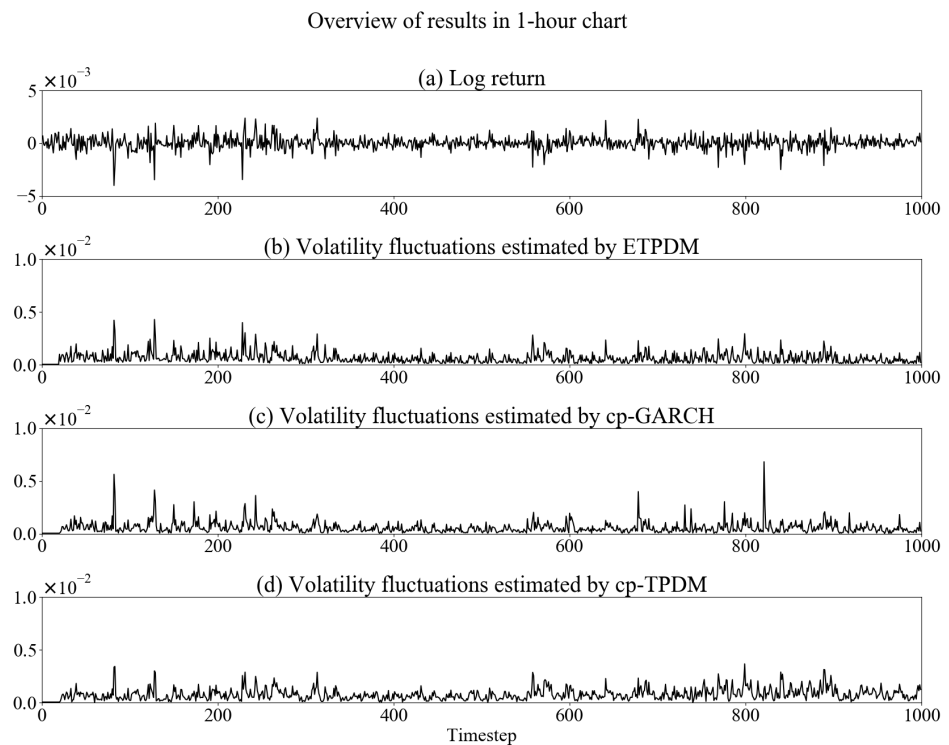


Figure 2. Overview of estimation results in 30-min chart.



**Figure 3.** Overview of estimation results in 1-h chart.

Figure 4 show the estimated log-likelihoods of the ETPDM, cp-GARCH and the cp-TPDM. The likelihood tends to be higher for the ETPDM, the cp-TPDM and cp-GARCH in that order. As mentioned in Section 2, TPDM is a superordinate model that encompasses Gaussian process dynamical model, the GPDM, and the fact that the likelihood of the GARCH was at a lower level than that of TPDM is consistent with the results of previous studies comparing the GARCH and the GPDM [26]. Likelihood is the most reliable indicator to quantify the model performance, and the EPTDM had the best performance of three models. Besides, in the case of cp-TPDM, the log-likelihoods scatter around  $-0.7$  in the range of particle numbers from 10 to 500 without convergence. This means the performance of the conventional particle filter is insufficient for the given data. On the other hand, the log-likelihoods of the ETPDM exhibit good convergence for the particle numbers larger than 100 in Figure 4b, which indicates the ETPDM is expected to be robust for fewer sampling.

To investigate the effectiveness of particle filtering, we introduce an effective particles rate

$$R_{eff} = \frac{1}{N \sum_{i=1}^N (W_t^i)^2} \quad (29)$$

as a measure for evaluating the bias of sampled particles. This value gives the maximum value  $R_{eff} = 1$  when the weights are uniformly distributed as  $W^i = 1/N$  ( $i = 1, 2, \dots, N$ ). In Figure 5a, the effective particle rates for the cp-TPDM scatter for whole particle numbers from 10 to 500. This kind of worse performance for the effective particle rates stems from the weight bias problem of the particle filtering. In other words, the conventional particle filter is hard to overcome the particle impoverishment problem, even if, by increasing particle numbers. On the contrary, for the case of the ETPDM as shown in Figure 5b, we can see that the effective particle rate converges beyond 50%. This is the expected advantage of the ETPDM, which stems from the finite band of each Gaussian distribution as a component of the prior distribution. Thus, the ETPDM serves as an accurate estimation for lower particle numbers and then would contribute to effective online estimation. Focusing on another comparison of the ETPDM, the cp-GARCH, also looks good in the view of the



effective particles rate. However, when we also consider the likelihood in Figure 4a, we can say that the practical ensembles didn't affect to performance because cp-GARCH was less suitable for this problem than the ETPDM. This is another evidence that the ETPDM have better potential.

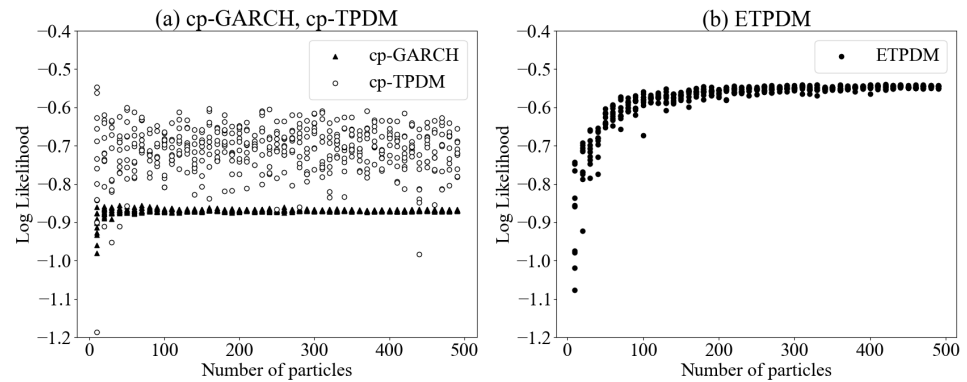


Figure 4. Estimated log-likelihoods of (a) the cp-GARCH, the cp-TPDM and (b) the ETPDM.

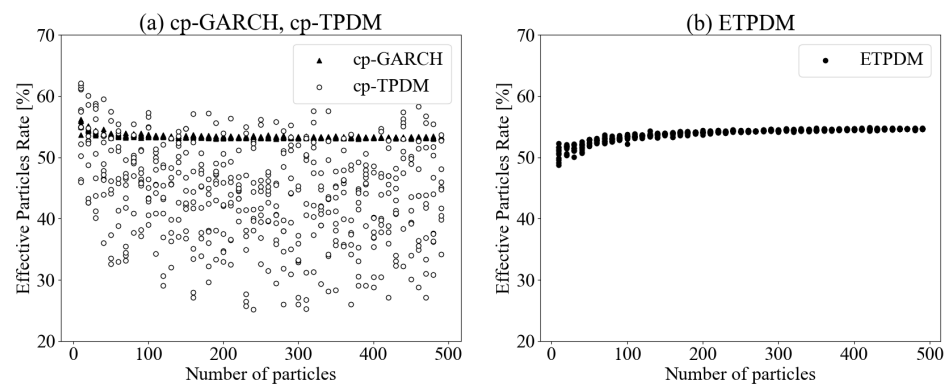
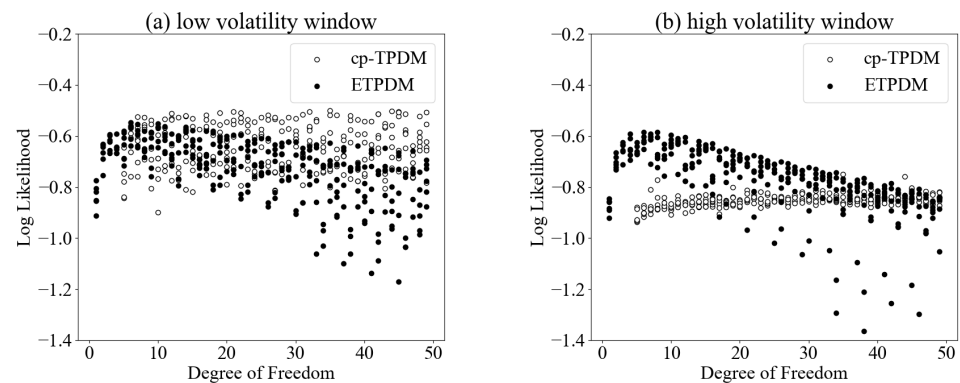


Figure 5. Effective particle rates of (a) cp-GARCH, the cp-TPDM and (b) the ETPDM.

In order to validate the robustness for estimating intermittent return dynamics, we investigate the degree of freedom  $\nu$  of the Student's  $t$ -process dynamical models. For this purpose, we split time window of return fluctuations; one is low volatility window ( $50 \leq t \leq 250$ ) and the other is high volatility counterpart ( $360 \leq t \leq 560$ ). The descriptive statistics of the return fluctuations in the two windows are shown in Table 1. As is seen in the table, kurtoses in both time windows are larger than 3, namely, corresponding return fluctuations follow non-Gaussian statistics. Prior research has confirmed that when the data set follows a Gaussian distribution, the strengths of models that excel at robust estimation do not come into play [28]. Therefore, such a data set that follows a non-Gaussian distribution is appropriate for the purpose of this experiment. Figure 6 exhibit the log-likelihoods of the cp-TPDM and the ETPDM in (a) low volatility window and (b) high volatility one. In these figures, the log-likelihoods of the ETPDM in both time windows have maxima in  $6 \leq \nu \leq 7$  though the estimations by the the cp-TPDM are unstable. This result evidences the robustness of the ETPDM.



**Figure 6.** Log-likelihoods of the cp-TPDM and the ETPDM in (a) low volatility window and (b) high volatility one.

**Table 1.** Two types of window.

Window Type	Mean	Variance	Skewness	Kurtosis
high volatility window	$-1.0 \times 10^{-6}$	$8.31 \times 10^{-4}$	-0.974	5.22
low volatility window	$1.0 \times 10^{-6}$	$4.41 \times 10^{-4}$	-0.531	3.56

## 5. Conclusions

In this study, we proposed the ETPDM to implement robust estimation for dynamical latent variables of nonlinear and non-Gaussian fluctuations. In estimating the dynamic latent variables and hyper parameters, the entropy based particle filter with the Gaussian mixture distribution was adopted. To validate the performance of the ETPDM, we carried out the numerical experiment for the return fluctuations of a foreign exchange rate compared with the cp-GARCH and the cp-TPDM. As a result, we confirmed the advantages of the ETPDM; (i) good convergence property, (ii) high effective particle rate and (iii) robustness for a small number of particles.

Based on its advantages, the ETPDM is applicable for online volatility estimation for the problem of asset allocation and derivative pricing in a short time span. As a basis distribution for background distribution, we employed the Gaussian distribution in our numerical experiments. Nevertheless, the framework of the entropy based particle filter is able to be extended to other probability density functions. Additionally, we can adapt this research to any other time-series data, not just asset data. It has the potential to be applied to control engineering, such as the self-positioning estimation problem. These are our future works.

**Author Contributions:** Conceptualization, Y.U.; methodology, A.N.; software, A.N.; validation, A.N.; formal analysis, A.N.; investigation, A.N.; resources, A.N.; data curation, K.N.; writing—original draft preparation, A.N.; writing—review and editing, Y.U., and K.N.; visualization, A.N.; project administration, Y.U.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author, Nono, A., upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cochrane, J.H. *Asset Pricing: Revised Edition*; Princeton University Press: Princeton, NJ, USA, 2009.
2. Mandelbrot, B.B. The variation of certain speculative prices. In *Fractals and Scaling in Finance*; Springer: Berlin/Heidelberg, Germany, 1997; pp. 371–418.
3. Shang, Y.; Dong, Q. Oil volatility forecasting and risk allocation: Evidence from an extended mixed-frequency volatility model. In *Applied Economics*; Routledge: London, UK, 2021; pp. 1127–1142.

4. Hou, Y.; Li, S.; Wen, F. Time-varying volatility spillover between Chinese fuel oil and stock index futures markets based on a DCC-GARCH model with a semi-nonparametric approach. *Energy Econ.* **2019**, *83*, 119–143. [[CrossRef](#)]
5. Horvath, B.; Muguruza, A.; Tomas, M. Deep Learning Volatility. *arXiv* **2019**, arXiv:1901.09647.
6. Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; Adaptive computation and machine learning series; The MIT Press: Cambridge, MA, USA, 2006.
7. Gonzalez, J.; Lezmi, E.; Roncalli, T.; Xu, J. Financial Applications of Gaussian Processes and Bayesian Optimization. *arXiv* **2019**, arXiv:1903.04841.
8. Nirwan, R.S.; Bertschinger, N. Applications of Gaussian process latent variable models in finance. In *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 1209–1221.
9. Shah, A.; Wilson, A.; Ghahramani, Z. Student-t processes as alternatives to Gaussian processes. In *Artificial Intelligence and Statistics*; 2014; pp. 877–885. Available online: <http://proceedings.mlr.press/v33/shah14.html> (accessed on 30 March 2021)
10. Uchiyama, Y.; Nakagawa, K. TPLVM: Portfolio Construction by Student's t-Process Latent Variable Model. *Mathematics* **2020**, *8*, 449. [[CrossRef](#)]
11. Sarkka, S. *Bayesian Filtering and Smoothing*; Cambridge University Press: Cambridge, UK, 2013.
12. Murata, M.; Hiramatsu, K. On Ensemble kalman Filter, Particle Filter, and Gaussian Particle Filter. *Trans. Inst. Syst. Control Inf. Eng.* **2016**, *29*, 448–462.
13. Nakano, S.; Ueno, G.; Haguchi, T. Merging particle filter for sequential data assimilation. *Nonlinear Process. Geophys.* **2007**, *14*, 395–408. [[CrossRef](#)]
14. Murata, M.; Nagano, H.; Kashino, K. Monte Carlo filter particle filter. In Proceedings of the 2015 European Control Conference (ECC), Linz, Austria, 15–17 July 2015; Volume 14, pp. 2836–2841.
15. Eyink, G.L.; Kim, S. A Maximum Entropy Method for Particle Filtering. *J. Stat. Phys.* **2006**, *123*, 1071–1128. [[CrossRef](#)]
16. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econ.* **1986**, *31*, 307–327. [[CrossRef](#)]
17. Bollerslev, T. Glossary to arch (garch). *CREATES Res. Pap.* **2008**, *49*, 1–46. [[CrossRef](#)]
18. Teräsvirta, T. An introduction to univariate GARCH models. In *Handbook of Financial Time Series*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 17–42.
19. Nakagawa, K.; Uchiyama, Y. GO-GJRSK Model with Application to Higher Order Risk-Based Portfolio. *Mathematics* **2020**, *8*, 1990. [[CrossRef](#)]
20. Lawrence, N.D. Gaussian process latent variable models for visualisation of high dimensional data. *Nips* **2003**, *2*, 5.
21. Kato, T. Introduction of Particle Filter and Its Implementation. *IPSJ SIG Tech. Rep.* **2007**, *157*, 161.
22. Wu, Y.; Hernández-Lobato, J.M.; Ghahramani, Z. Gaussian Process Volatility Model. In *Advances in Neural Information Processing Systems 27*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; pp. 1044–1052.
23. Vințe, C.; Ausloos, M.; Furtună, T.F. A Volatility Estimator of Stock Market Indices Based on the Intrinsic Entropy Model. *Entropy* **2021**, *23*, 484. [[CrossRef](#)]
24. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 1991.
25. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
26. Wu, Y.; Hernandez-Lobato, J.M.; Ghahramani, Z. Gaussian process volatility model. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2014.
27. Fleming, J.; Kirby, C. A closer look at the Relation between GARCH and Stochastic Autoregressive volatility. *J. Financ. Econ.* **2003**, *1*, 365–419.
28. Jacquier, E.; Polson, N.G.; Rossi, P.E. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *J. Econ.* **2004**, *122*, 185–212. [[CrossRef](#)]