mathematics

MDPI

Article

TPLVM: Portfolio Construction by Student's *t*-Process Latent Variable Model

Yusuke Uchiyama [1,*] and Kei Nakagawa [2]

[1] MAZIN Inc., 3-29-14 Nishi-Asakusa, Taito City, Tokyo 111-0035, Japan
[2] NOMURA Asset Management Co. Ltd., 1-12-1 Nihonbashi, Chuo City, Tokyo 103-8260, Japan; kei.nak.0315@gmail.com
* Correspondence: uchiyama@mazin.tech

Received: 29 January 2020; Accepted: 12 March 2020; Published: 19 March 2020

check for updates

Abstract: Optimal asset allocation is a key topic in modern finance theory. To realize the optimal asset allocation on investor's risk aversion, various portfolio construction methods have been proposed. Recently, the applications of machine learning are rapidly growing in the area of finance. In this article, we propose the Student's *t*-process latent variable model (TPLVM) to describe non-Gaussian fluctuations of financial timeseries by lower dimensional latent variables. Subsequently, we apply the TPLVM to portfolio construction as an alternative of existing nonlinear factor models. To test the performance of the proposed method, we construct minimum-variance portfolios of global stock market indices based on the TPLVM or Gaussian process latent variable model. By comparing these portfolios, we confirm the proposed portfolio outperforms that of the existing Gaussian process latent variable model.

Keywords: student's *t*-process; latent variable model; factor model; Portfolio theory; global stock markets

---

1. Introduction

Estimation of the covariance matrix of timeseries plays a dominant role in applications of modern financial theory. The optimization of mean-variance portfolio, which is one of the pioneering works of the modern finance theory [1], is based on the covariance matrix of the multi-dimensional timeseries of return of assets. Since the return of assets are modelled by non-stationary stochastic processes, the covariance matrix should be estimated as a time-dependent symmetric matrix. In practice, we often estimate the covariance matrix by empirical time averaging, because of the lack of complete information of the corresponding probabilistic space. It is, however, pointed out that time averaging often causes serious estimation error of the covariance matrix in the case of larger assets [2,3]. To overcome this problem, several inference methods are proposed from the point of view of the random matrix theory [4,5].

With the aid of recently growing machine learning techniques, we can improve the accuracy of the estimation of the covariance matrix [6,7]. Furthermore, the applications of the machine learning techniques have been spreading in both theoretical and practical financial problems [8,9]. The prediction of the future price is implemented by the deep neural networks of various modeling [10,11]. In particular, the application of the machine learning techniques for the portfolio optimization has attracted the interest of both academia and industry [12,13]. The Gaussian process, which is known as a method of nonparametric Bayesian learning, is used as a model of dynamics of the covariance matrix of multi-dimensional timeseries. In the literature of option pricing theory, the model of the volatility of a risky asset is given by the Gaussian process [14].

In the field of mathematical finance, stochastic volatility models have been utilized in estimating dynamic covariance matrix of the return of assets. One of the most popular conditional volatility models is the generalized autoregressive conditional heteroscedasticity (GARCH) model [15], which describes the volatility clustering of the return of assets. To introduce a time-varying correlation structure to these conditional volatility models, the dynamic conditional correlation (DCC) GARCH model has been proposed [16]. The parameters of the GARCH and DCC GARCH can be estimated by the method of maximum-likelihood.

On the other hand, in the literature of the machine learning, some kinds of latent variable models can be utilized to infer the dynamics of the covariance matrix. Recently, the Gaussian process latent variable model (GPLVM) has been employed to the problem of the portfolio optimization, where latent variables are introduced as factors of return of the assets. Namely, this model can be interpreted as a latent variable factor model [17].

Despite these existing practical applications, we should reconsider the assumption and validation of the use of the GPLVM for financial problems because the GPLVM assumes that observed data follows the Gaussian distribution. In the most case of financial problems, the return of assets is regarded as an observed variable. It is well known that the fluctuations of the return of assets follow non-Gaussian distributions [18]. To describe such fluctuations, some fat-tailed distributions have been presented and applied to the financial timeseries. Thus, the GPLVM should be extended to fat-tailed distributions when we use it for the financial problems.

In this article, we propose Student's-t process latent variable model (TPLVM) as an extension of the GPLVM. This model is developed based on the Student's t-distribution, which is a symmetric fat-tailed distribution. Since the Student's t-distribution converges to the Gaussian distribution with the limit of a parameter, degree of freedom, the TPLVM includes the GPLVM as a special case. To use the TPLVM in practice, as with the GPLVM, we derive its predictive distribution of closed form and an estimator of hyper parameters by the variational inference in Bayesian sense.

The reminder of this article is organized as follows. Section 2 gives a brief introduction of the GPLVM including the Gaussian process with the concept of kernel functions. In Section 3, we introduce the formula of TPLVM, which consists of the kernel functions, predictive distribution and variational inference for estimating hyper parameters. As a preliminary preparation of finance, we explain the basis of factor model and portfolio optimization in Section 4. Section 5 implements portfolio optimization, where we compare the performance of the GPLVM and TPLVM. Section 6 is dedicated to conclusions and future works.

2. Short Review of Gaussian Process

2.1. Gaussian Process

The Gaussian process, a kind of stochastic processes, is a non-parametric method of machine learning [19,20]. This has been firstly introduced to describe random dynamics such as a fluctuating pollen on water surface known as Brownian motion [21]. Without loss of generality, the argument of the Gaussian process can be extended from one-dimensional time to multi-dimensional feature space. In this chapter, we provide a short review of the Gaussian process for multi-dimensional features as the preliminary preparation of the proposed model.

For a sequence of input features $\{x_1, x_2, \cdots, x_n\}$, a stochastic process $f(\cdot)$ is the Gaussian process when the sequence of random variables $\{f(x_1), f(x_2), \cdots, f(x_n)\}$ is sampled from a multivariate Gaussian distribution. In general, the form of the multivariate Gaussian distribution is determined by the mean vector and covariance matrix. Likewise, the Gaussian process are specified by the mean and covariance functions. Thus, the Gaussian process is regarded as a representation of the infinite dimensional Gaussian distribution.

The mean and covariance functions are defined as follows:

$$m(x) = \mathbb{E}[f(x)], \tag{1}$$

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))], \tag{2}$$

where the operator $\mathbb{E}[\cdot]$ denotes expectation operator, $m(\cdot)$ and $k(\cdot, \cdot)$ are respective mean and covariance functions. The mean vector and covariance matrix of the Gaussian process for given dataset are represented by

$$m_i = m(x_i) \ \ (1 \leq i \leq n), \tag{3}$$

$$K_{i,j} = k(x_i, x_j) \ \ (1 \leq i, j \leq n). \tag{4}$$

On these settings, the stochastic process $f(\cdot)$ is sampled from the Gaussian distribution $\mathcal{N}(m(\cdot), K(\cdot, \cdot))$. In this situation, the stochastic process $f(\cdot)$ is the Gaussian process expressed as $f \sim \mathcal{GP}(m, K)$. The covariance function satisfies to be symmetric and positive definite, and thus is also called as a kernel function. In the literature of the Gaussian process, the covariance matrix is often called as a kernel matrix. The mathematical characteristics of the kernel functions are explained in [22].

Given an additional input dataset $\mathcal{D}^* = \{x_1^*, x_2^*, \cdots, x_{n^*}^*\}$, the corresponding outputs $\{y_1^*, y_2^*, \cdots, y_{n^*}^*\}$ can be predicted by the conditional Gaussian process with prior dataset $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$. With notations that $X = [x_1, x_2, \cdots, x_n]^T$, $X^* = [x_1^*, x_2^*, \cdots, x_{n^*}]^T$ and $Y = [y_1, y_2, \cdots, y_n]$, the predictive distribution of the conditional Gaussian process is also given by the Gaussian process $\mathcal{GP}(f^*, K^*)$, where

$$f^* = m_X + K_{X^*,X} K_{X,X}^{-1} Y, \tag{5}$$

$$K^* = K_{X^*,X^*} - K_{X^*,X} K_{X,X}^{-1} K_{X,X^*}. \tag{6}$$

In Equations (5) and (6), it is seen that the covariance function propagates the information about $\mathcal{D}$ to $\mathcal{D}^*$. Hence, the covariance functions play the dominant role in the use of the Gaussian process.

2.2. Gaussian Process Latent Variable Model

In the literature of big data analysis, it is often expected that observed variables can be explained by lower dimensional latent variables. For this purpose, various methods of dimension reduction have been developed. One of the most popular methods is the principal component analysis (PCA), which extracts latent variables by the singular value decomposition. To extend the PCA for nonlinear and random data, the Gaussian process latent variable model (GPLVM) has been proposed [23]. The GPLVM expresses nonlinear relationships between observed and latent variables by the covariance function. The randomness is assumed to be originate from the Gaussian distribution.

To describe an observed variable $y \in \mathbb{R}^D$, we introduce a latent variable $x \in \mathbb{R}^Q$ $(Q < D)$, and a nonlinear map $f : \mathbb{R}^Q \to \mathbb{R}^D$ with a Q-dimensional noise $\epsilon \sim \mathcal{N}(0, \sigma_0 I)$ as

$$y = f(x) + \varepsilon. \tag{7}$$

For this latent variable model, we assume that the nonlinear map $f(\cdot)$ is sampled from the Gaussian process as $f \sim \mathcal{GP}(0, K)$. This model is known as the GPLVM. For the sake of brevity, we introduce notations for the set of latent and observed variables as $X = [x_1, x_2, \cdots, x_N]^T$ and $Y = [y_1, y_2, \cdots, y_N]^T$. Assume that the columns of the observed matrix $Y \in \mathbb{R}^{N \times D}$ are samples from the independently identical distributed Gaussian distributions which have the covariance functions with respect to the latent variable matrix $X \in \mathbb{R}^{N \times Q}$, the probability density function of the GPLVM is introduced as follows:

$$p(Y|X) = \frac{1}{(2\pi)^{ND/2}|K_{X,X}|^{D/2}} \exp\left(-\frac{1}{2}Y^T K_{X,X}^{-1} Y\right). \tag{8}$$

In the GPLVM, latent variables and hyperparameters of the covariance functions are inferred by several existing methods such as gradient methods, variational inference and Markov Chain Monte Carlo methods.

3. Proposed Model: Student's *t*-Process Latent Variable Model

3.1. Introduction of the Student's t-Process

The Gaussian process has diverse applications in the fields of computer science, robotics and others. However, it seems not to be applicable to financial problems because the fluctuations of the financial data follow non-Gaussian distributions with fat-tails. It is thus necessary to extend the existing methods of the Gaussian process to those of non-Gaussian stochastic processes with fat-tails.

For this purpose, the Student's *t*-process has been proposed as a generalization of the Gaussian process [24]. This stochastic process follows the Student's *t*-distribution, of which tails show power-law behaviours. As with the Gaussian process, the Student's *t*-process is specified by the mean and covariance functions. Given the mean and covariance functions, the probability density function of the Student's *t*-process is defined as

$$\mathcal{T}(m, K, \nu) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{[(\nu-2)\pi]^{\frac{N}{2}}\Gamma\left(\frac{\nu}{2}\right)|K|^{\frac{1}{2}}}\left[1 + \frac{1}{\nu-2}(y-m)^T K^{-1}(y-m)\right]^{-\frac{\nu+N}{2}}, \tag{9}$$

where $\Gamma(\cdot)$ is the multivariate gamma function and the positive real parameter ν is degrees of freedom. In this setting, the stochastic process $f(\cdot)$ is the Student's *t*-process expressed as $f \sim \mathcal{TP}(m, K; \nu)$. Note that the Student's *t*-process converges to the Gaussian process at the limit of $\nu \to \infty$.

The conditional distribution of the Student's *t*-process can be also derived analytically and given as the conditional Student's *t*-distribution. Namely, we can update the mean and covariance functions and the degrees of freedom from the conditional distribution. Through cumbersome calculations, the update formulas of the mean and covariance functions and the degrees of freedom are derived as follows:

$$m^* = m + K_{X^*,X} K_{X,X}^{-1} Y, \tag{10}$$

$$K^* = \frac{\nu - \beta - 2}{\nu - N - 2}\left[K_{X^*,X^*} - K_{X^*,X} K_{X,X}^{-1} K_{X,X^*}\right], \tag{11}$$

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

$$\nu^* = \nu + N. \tag{13}$$

It is seen that the update formula of the covariance function in Equation (11) explicitly depends on the number of observed variables, which property does not appear in the case of the Gaussian process. Hence, the Student's *t*-process is regarded to utilize prior information more effectively than the Gaussian process.

3.2. Student's t-Process Latent Variable Model

To extend the GPLVM to stochastic processes following non-Gaussian distributions, we propose the TPLVM. Suppose an observed variable $y \in \mathbb{R}^D$ is explained by a lower dimensional latent variable $x \in \mathbb{R}^Q$ $(Q < D)$ by a nonlinear map $f : \mathbb{R}^D \rightarrow \mathbb{R}^Q,\ f \sim \mathcal{TP}(m, K; \nu)$, the TPLVM is introduced as follows:

$$p(Y|X) = \frac{\Gamma\left(\frac{\nu+D}{2}\right)}{[(\nu-2)\pi]^{\frac{D}{2}}\Gamma\left(\frac{\nu}{2}\right)|K_{X,X}|^{\frac{1}{2}}}\left[1 + \frac{1}{\nu-2}(Y - m_X)^T K_{X,X}^{-1}(Y - m_X)\right]^{-\frac{\nu+D}{2}}. \tag{14}$$

The nonlinear dependency of the latent variable matrix $X \in \mathbb{R}^{N \times Q}$ is given by the covariance matrix. It is expected that the TPLVM provides a robust estimation especially for observed data with large fluctuations because the Student's t-distribution can capture large deviated data from the Gaussian distribution in its sampling.

As with the GPLVM, the latent variables and hyperparameters of the TPLVM can be estimated from its likelihood. The logarithmic likelihood of the TPLVM is given as

$$\begin{aligned} \log p(Y|X) = \log\Gamma\left(\frac{\nu+D}{2}\right) - \frac{D}{2}\log[(\nu-2)\pi] - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log|K_{X,X}| \\ - \frac{\nu+D}{2}\log\left[1 + \frac{1}{\nu-2}(Y - m_X)^T K_{X,X}^{-1}(Y - m_X)\right], \end{aligned} \tag{15}$$

By means of existing optimization methods, we can estimate the latent variables and hyperparameters of the covariance function and the degrees of freedom. However, it is known that the optimization of the covariance function with respect to the latent variables often induces numerical instability because of its complexity. Hence, we should carefully select the initial values of optimization procedures and repeat with diverse seeds of the initial values to refuse dropping in local minima.

3.3. Variational Inference

To overcome the shortcomings of the method of maximum-likelihood, we utilize the method of variational inference [25]. Instead of optimizing the logarithmic likelihood in Equation (15), we consider that of posterior $p(X|Y) = p(Y|X)p(X)/p(Y)$ in the Bayesian sense. In solving the optimization problem with respect to the posterior, we approximate $p(X|Y)$ by $q(X)$. As a measure of the difference between two probability density functions, we introduce the Kullback-Leibler (KL) divergence as follows:

$$\mathrm{KL}[q(X)||p(X|Y)] = \int \log\frac{q(X)}{p(X|Y)}q(X)\mathrm{d}X. \tag{16}$$

With the use of the Bayes theorem, the KL divergence is alternatively represented as

$$\mathrm{KL}[q(X)||p(X|Y)] = -\int \log\frac{p(Y|X)p(X)}{q(X)}q(X)\mathrm{d}X + \log p(Y). \tag{17}$$

Since the second term in the right hand side in Equation (17) does not depend on $q(\cdot)$, we just have to maximize the first term in the right hand side, which is known as the evidence lower bound (ELBO), to minimize the KL divergence. The ELBO provides the lower bound of the evidence $\log p(Y)$ because the KL divergence is non-negative. Therefore, this procedure realizes the sufficient fitting of the observed data at the same time. Indeed, the maximization of the ELBO serves the best explanation of the reduced dimension Q of the latent variables.

4. Problem Formulation in Finance

4.1. Factor Model

Arbitrage pricing theory [26] assumes that the D-days expected return of an asset $r_n \in \mathbb{R}^N$ is explained by the factor model as

$$r_n = \alpha_n + F\beta_n + \epsilon, \tag{18}$$

where $\alpha_n \in \mathbb{R}^D$ is an excess return, $\beta_n \in \mathbb{R}^Q$ is weight coefficients, $F \in \mathbb{R}^{D \times Q}$ is a factor matrix, and $\epsilon \in \mathbb{R}^D$ is an error term with zero mean and a finite covariance. The factor model manifests that the return of the asset is originated from the returns of Q-factors. In fact, without the excess return α_n, the expected return of the factor model is derived as follows:

$$\mathbb{E}[r_n] = \mathbb{E}[F]\beta_n. \tag{19}$$

The special case of this formula with only one factor is known as the model of the capital asset pricing model, which is a cornerstone of the modern finance theory [27].

The weight coefficients β_n in the factor model in Equation (18) can be interpreted as latent variables which explain the return of the asset. Based on this idea, we introduce a nonlinear factor model as

$$r_n = f(\beta_n). \tag{20}$$

This model is regarded as a latent variable counterpart of nonlinear factor model [10]. Here, we employ the Student's t-process as the model of nonlinear mapping $f : \mathbb{R}^Q \rightarrow \mathbb{R}^D$. In other words, the nonlinear factor model in Equation (20) is given by the TPLVM. The nonlinear correlation of the latent variable factors depends on the specific form of the covariance function of the TPLVM, and the predicted return of the asset can be inferred by the predicted distribution. Furthermore, the nonlinear factor model can be interpreted as a dimension reduction model when $Q < D$. Hence we can expect to obtain the essential lower dimensional variable which explains the dynamics of the return of the asset.

4.2. Portfolio Theory

Markowitz established the modern portfolio theory on the mean-variance portfolio. In this theory, a portfolio consists of multi assets classes such as stock, bond, currency and commodity with their optimal allocations based on both individual and entangled risk of assets.

The mean-variance portfolio is designed by the constrained quadratic programming problem with respect to the objective function as

$$w^T K w - \lambda(\mathbb{E}[r] - \mu), \tag{21}$$

where $w \in \mathbb{R}^D$ is the weight coefficients of the portfolio, $K \in \mathbb{R}^{D \times D}$ is the covariance matrix of the returns, λ is a Lagrangian multiplier, r is the return of the portfolio and μ is the expected return of the portfolio. In practical use, the return of the portfolio is quite hard to be estimated, whereby, without the constraint condition of the expected return, the mean-variance portfolio is often replaced by the minimum-variance portfolio with empirically estimated covariance matrix.

5. Experiment

In this section, we test the performance of the minimum-variance portfolio with the TPLVM by comparing with the counterpart of the GPLVM. Before proceeding, we explain the experimental dataset of our performance test.

As the experimental data, we use the following global stock market indices: S&P 500 (US), S&P/TSX 60 (Canada), FTSE 100 (UK), CAC 40 (France), DAX (Germany), IBEX 35 (Spain), FTSE MIB (Italy), AEX (the Netherlands), OMX 30 (Sweden), SMI (Switzerland), Nikkei 225 (Japan), HKHSI

(Hong Kong), ASX 200 (Australia), KOSPI (Korea), OBX (Norway), MSCI (Singapore). These stock indices are sampled every month between Jun 1998 to Jun 2019 from the Bloomberg's data platform. The statistics of the return of the stock indices are shown in Table 1. In this table, mean (Mean), standard deviation (Std.), the ratio of mean and standard deviation (R/R), skewness (Skew) and kurtosis (Kurtosis) of returns of the stock indices are presented.

With the use of the historical returns of the stock indices, we construct the minimum-variance portfolios based on the GPLVM (Port_G) and TPLVM (Port_t). The covariance matrix of each portfolio is estimated by the covariance function with 120 past samples. As the kernel function, we utilize the exponential kernel defined as

$$k_{\mathrm{Exp}}(x, x') = \theta_1 \exp\left(-\theta_2^{-2}||x - x'||\right) \tag{22}$$

with θ_l $(l = 1, 2)$ being hyper parameters. For the sake of brevity, the dimension of the latent variable are fixed $Q = 1$. Under these conditions, we compare the performance of the Port_G and Port_t by annualized return (Return), annualized risk as the standard deviation of return (Risk), risk/return (R/R) as return divided by risk, which are defined as follows:

$$\mathbf{Return} = \frac{12}{T}\sum_{t=1}^{T} R_t^P, \tag{23}$$

$$\mathbf{Risk} = \sqrt{\frac{12}{T-1} \times (R_t^P - \mu^P)^2}, \tag{24}$$

$$\mathbf{R/R} = \mathbf{Return/RISK}. \tag{25}$$

Here, R_t^P indicates return of GPLVM or TPLVM portfolio at time t, and $\mu^P = (1/T)\sum_{t=1}^{T} R_t^P$ denotes the average return of the GPLVM or TPLVM portfolio. All our experiments were implemented by a laptop PC with Intel(R) Core(TM) i7-76660U CPU@2.50 GHz and 16GB RAM. We used PyStan in variational inference procedures.

Table 2 shows the performances of the portfolios by comparing annual return, risk and return-risk ratio. The sample period is separated into anterior half period (Jun 2008–Jun 2013) and posterior half period (Jul 2013–Jun 2019). Note that the anterior half period contains the global financial crisis 2007–2008. As is seen in this table, the Port_t outperforms the Port_G in the both half periods. In particular, the difference of the annual return in the anterior half period is larger than that in the posterior half period. It is said that the market volatility during the global financial crisis intensively fluctuated whereby non-Gaussian nature clearly emerged in the global stock market. In such situation, the TPLVM is a consistent model to describe the intermittent volatility fluctuations. As is well known, the performance of the minimum-variance portfolio depends on the accuracy of estimated covariance matrix. In other words, accurately estimated covariance matrix, which is given by the kernel matrix with respect to the latent variable, is expected to make a better profit. Thus, we can construct a robust portfolio by the TPLVM based minimum-variance portfolio.

Table 1. Statistics of global stock indices.

	US	**Canada**	**UK**	**France**	**Germany**	**Spain**	**Italy**	**Netherlands**
Mean [%]	6.00	5.41	2.39	4.08	6.87	3.20	1.35	2.96
Std. [%]	14.93	14.92	13.62	18.12	21.13	20.66	21.71	19.13
R/R	0.40	0.36	0.18	0.23	0.33	0.15	0.06	0.15
Skew	−0.66	−0.92	−0.55	−0.38	−0.50	−0.17	0.03	−0.74
Kurtosis	5.23	7.36	4.53	4.52	6.12	4.96	4.80	5.88
	Sweden	**Switzerland**	**Japan**	**HongKong**	**Australia**	**Korea**	**Norway**	**Singapore**
Mean [%]	6.32	2.80	3.35	7.27	4.70	12.98	10.72	5.05
Std. [%]	19.51	14.68	19.24	23.46	12.40	28.80	21.49	21.71
R/R	0.32	0.19	0.17	0.31	0.38	0.45	0.50	0.23
Skew	−0.19	−0.73	−0.54	0.28	−0.69	1.39	−0.93	−0.26
Kurtosis	5.29	6.11	4.75	5.78	4.54	11.63	6.84	6.81

Table 2. Performance of Port_G and Port_t.

	Port$_G$	**Port$_t$**	**Difference**
Anterior half (Jun 2008–Jun 2013)			
Return	−4.89%	−2.63%	2.25%
Risk	19.57%	18.33%	−1.24%
R/R	−0.25	−0.14	0.11
Posterior half (Jul 2013–Jun 2019)			
Return	6.08%	6.30%	0.22%
Risk	11.16%	10.56%	−0.60%
R/R	0.54	0.60	0.05
Whole period (Jun 2008–Jun 2019)			
Return	0.64%	1.87%	1.23%
Risk	15.92%	14.93%	−0.99%
R/R	0.04	0.12	0.09

6. Conclusions

In the literature of Bayesian machine learning, the Gaussian process has been developed and utilized to the diverse area including finance. It is, however, well known that the historical financial data follows non-Gaussian distributions. The Student's t-process is proposed, as the generalization of the Gaussian process, to model the observed data following the non-Gaussian distributions with fat-tails.

In this article, we proposed the TPLVM by incorporating the latent variables into the Student's t-process. The TPLVM can be used to reduce the number of explanation variable following the non-Gaussian distributions with fat-tails. The nonlinear correlation of the TPLVM is modelled by prescribed kernel functions. The hyperparameters of the TPLVM can be determined by the method of maximum-likelihood. As a robust parameter optimization, we presented the method of variational inference of the TPLVM, which utilize the information of prior distribution of latent variables.

The problem of the portfolio optimization has been studied in both academia and industry. We applied the TPLVM into the portfolio optimization with the use of the minimum-variance portfolio. To test the performance of the proposed portfolio, we implemented the empirical analysis for the global stock market data and compared the Port_G with Port_t. It was shown that the Port_t outperforms the Port_G in the whole test periods because Port_t can capture the non-Gaussian nature of the global stock market especially in the period of the global financial crisis.

The TPLVM can be applied other risk-based portfolios such as risk parity [28], maximum risk diversification [29], and complex valued risk diversification [30], in which Value at Risk (VaR), instead of standard deviation, is often used as an appropriate risk measure. These applications are expected to show higher performance compared with conventional ones. In addition, the TPLVM can be modified to a latent variable dynamical model to catch the nature of historical volatility fluctuations. These ways of research are our future works.

Author Contributions: Conceptualization, Y.U.; Methodology, Y.U. and K.N.; Software, K.N.; Validation, K.N.; Formal analysis, Y.U.; Ddata curation, K.N.; Writing–original draft preparation, Y.U.; Writing–review and editing, Y.U. and K.N.; Project administration, Y.U. and K.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Abbreviations

The following abbreviations are used in this manuscript:

GARCH	Generalized AutoRegressive Conditional Heteroscedasticity
DCC	Dynamic Conditional Correlation
GPLVM	Gaussian Process Latent Variable Model
TPLVM	Student's t-Process Latent Variable Model

References

1. Markowitz, H. Portfolio selection. *J. Financ.* **1952**, *7*, 77–91.
2. Nakagawa, K.; Imamura, M.; Yoshida, K. Risk-based portfolios with large dynamic covariance matrices. *Int. J. Financ. Stud.* **2018**, *6*, 52. [CrossRef]
3. Engle, R.F.; Ledoit, O.; Wolf, M. Large dynamic covariance matrices. *J. Bus. Econ. Stat.* **2019**, *37*, 363–375. [CrossRef]
4. Ledoit, O.; Wolf, M. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Stat.* **2012**, *40*, 1024–1060. [CrossRef]
5. Ledoit, O.; Wolf, M. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financ. Stud.* **2017**, *30*, 4349–4388. [CrossRef]
6. Chen, X.; Lyu, M.R.; King, I. Toward Efficient and Accurate Covariance Matrix Estimation on Compressed Data. In Proceedings of the 34th International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 767–776.
7. Wu, Y.; Lobato, J.M.H.; Ghahramani, Z. Dynamic Covariance Models for Multivariate Financial Time Series. In Proceedings of the 30th International Conference on International Conference on Machine Learning—Volume 28, Atlanta, GA, USA, 17–19 June 2013; pp. 558–566.
8. Atsalakis, G.S.; Valavanis, K.P. Surveying stock market forecasting techniques—Part II: Soft computing methods. *Expert Syst. Appl.* **2009**, *36*, 5932–5941. [CrossRef]
9. Cavalcante, R.C.; Brasileiro, R.C.; Souza, V.L.; Nobrega, J.P.; Oliveira, A.L. Computational intelligence and financial markets: A survey and future directions. *Expert Syst. Appl.* **2016**, *55*, 194–211. [CrossRef]
10. Nakagawa, K.; Uchida, T.; Aoshima, T. Deep factor model. In Proceedings of the ECML PKDD 2018 Workshops, Dublin, Ireland, 10–14 September 2018; Springer: Berlin, Germany, 2018; pp. 37–50.
11. Nakagawa, K.; Ito, T.; Abe, M.; Izumi, K. Deep recurrent factor model: Interpretable non-linear and time-varying multi-factor Model. *arXiv* **2019**, arXiv:1901.11493.
12. Shen, W.; Wang, J.; Jiang, Y.G.; Zha, H. Portfolio Choices with Orthogonal Bandit Learning. In Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015; AAAI Press: Palo Alto, CA, USA, 2015; pp. 974–980.

13. Song, Q.; Liu, A.; Yang, S.Y. Stock portfolio selection using learning-to-rank algorithms with news sentiment. *Neurocomputing* **2017**, *264*, 20–28. [CrossRef]
14. Wu, Y.; Hernández-Lobato, J.M.; Ghahramani, Z. Gaussian Process Volatility Model. In *Advances in Neural Information Processing Systems 27*; Curran Associates, Inc.: Red Hook, NY, USA, 2014; pp. 1044–1052.
15. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327. [CrossRef]
16. Engle, R. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econ. Stat.* **2002**, *20*, 339–350. [CrossRef]
17. Nirwan, R.S.; Bertschinger, N. Applications of Gaussian Process Latent Variable Models in Finance. In Proceedings of the SAI Intelligent Systems Conference, London, UK, 5–6 September 2019; Springer: Berlin, Germany, 2019; pp. 1209–1221.
18. Mandelbrot, B.B. The variation of certain speculative prices. In *Fractals and Scaling in Finance*; Springer: Berlin, Germany, 1997; pp. 371–418.
19. Rasmussen, C.E. Gaussian processes in machine learning. In *Summer School on Machine Learning*; Springer: Berlin, Germany, 2003; pp. 63–71.
20. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 2.
21. Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Der Phys.* **1905**, *322*, 549–560. [CrossRef]
22. Hofmann, T.; Schölkopf, B.; Smola, A.J. Kernel methods in machine learning. *Ann. Stat.* **2008**, *36*, 1171–1220. [CrossRef]
23. Lawrence, N.D. Gaussian process latent variable models for visualisation of high dimensional data. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2004; pp. 329–336.
24. Shah, A.; Wilson, A.; Ghahramani, Z. Student-t processes as alternatives to Gaussian processes. In *Artificial Intelligence and Statistics*; PMLR Press: New York, NY, USA, 2014; pp. 877–885.
25. Damianou, A.C.; Titsias, M.K.; Lawrence, N.D. Variational inference for latent variables and uncertain inputs in Gaussian processes. *J. Mach. Learn. Res.* **2016**, *17*, 1425–1486.
26. Ross, S.A. The arbitrage theory of capital asset pricing. In *Handbook of the Fundamentals of Financial Decision Making: Part I*; World Scientific: Singapore, 2013; pp. 11–30.
27. Harvey, C.R.; Liu, Y.; Zhu, H. ... and the cross-section of expected returns. *Rev. Financ. Stud.* **2016**, *29*, 5–68. [CrossRef]
28. Qian, E. Risk parity and diversification. *J. Invest.* **2011**, *20*, 119–127. [CrossRef]
29. Choueifaty, Y.; Coignard, Y. Toward maximum diversification. *J. Portf. Manag.* **2008**, *35*, 40–51. [CrossRef]
30. Uchiyama, Y.; Kadoya, T.; Nakagawa, K. Complex Valued Risk Diversification. *Entropy* **2019**, *21*, 119. [CrossRef]