JPPIPA 9(12) (2023)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

The Implementation of Latent Gaussian Model in the Forecasting Process

Elyn Prina[1], Yusep Suparman[2*], Gumgum Darmawan[2]

[1] Post-Graduate Program in Applied Statistics. Faculty of Mathematics and Natural Sciences. Padjadjaran University. Sumedang. Indonesia
[2] Department of Statistics. Faculty of Mathematics and Natural Sciences. Padjadjaran University. Sumedang. Indonesia

Received: October 25, 2023
Revised: November 9, 2023
Accepted: December 20, 2023
Published: December 31, 2023

Corresponding Author:
Yusep Suparman
yusep.suparman@unpad.ac.id

DOI: 10.29303/jppipa.v9i12.5835

Abstract: This research aims to determine the implementation of the Latent Gaussian Model in the forecasting process. This research focuses on developing a forecasting model using the Multivariate Latent Gaussian Model (LGM) approach with shared components. which offers a more accurate representation without the assumption of stationarity and cointegration as it accommodates random components in the model. The forecasting results for the five KPPs are considered to have a very good level of accuracy with MAPE values < 10%. This shows that LGM can achieve reliable forecasting when applied to the real life problems. This condition supports forecasting and can be an effective and targeted benchmark. The Latent Gaussian Model using the Bayesian Approach in parameter estimation can be utilized in forecasting Personal Income Tax Article 25/29. This is supported by the highly accurate MAPE value of 0.01%. The implementation of the developed model is not limited to forecasting Personal Income Tax Article 25/29. but can also be used in various other fields. With its hierarchical structure. the Bayesian approach proves to be an effective method for addressing complex modeling challenges.

Keywords: Bayesian; Forecasting Process; Latent Gaussian Model; Tax

Introduction

The forecasting method employed in this research involved the implementation of Multivariate Latent Gaussian Model (LGM) as random components are accommodated in the model (Rue et al.. 2009) On the other hand, multivariate forecasting using the VAR approach is a classical forecasting and relatively complex because it requires the assumption of stationarity and the presence of cointegration among time series to ensure that the model can be used for forecasting future period (Wei. 2019). The latent process in LGM is related to the unknown parameters. with efficient calculations to perform Bayesian inference on LGM using the Integrated Nested Laplace approximations (INLA) approach (Opitz. 2016)The full Bayesian approach provides a suitable framework for dealing with complexity through hierarchical structures (Lee. 2011; Oleson et al.. 2021; Richardson et al.. 2006)One of the unique advantages of the Bayesian method providing robust results in interpreting posterior distributions and making inferences for the parameters in the model (Giacomini & Kitagawa. 2021; Wasserman. 1989)This research aims to obtain precise and accurate As a result. the accuracy of setting target figures is very important to encourage the realization of healthy and sustainable fiscal management in accordance with its potential (Badan Kebijakan Fiskal. 2016)

Latent Gaussian Models (LGM) is a forecasting method used in time series data analysis. LGM assumes that the observed time series data has a structure determined by latent (hidden) processes that follow a Gaussian distribution. These latent processes can explain variations in the data that cannot be explained by known observation processes. LGM consists of both observation and latent processes. The observation process is how data is observed. following a certain probability distribution. Meanwhile. the latent process describes how the data is related to a latent process that follows a Gaussian distribution.

How to Cite:
rina, E., Suparman, Y., & Prina, G. (2023). The Implementation of Latent Gaussian Model in the Forecasting Process. *Jurnal Penelitian Pendidikan IPA*, *9*(12), 12155–12165. https://doi.org/10.29303/jppipa.v9i12.5835

The implementation of LGM requires statistical techniques. such as Bayesian inference to determine the parameters of the model used and make inferences from the data (Rue. Martino. & Chopin. 2009). The LGM Bayesian approach integrates a likelihood function and a Gaussian prior distribution. resulting in the derivation of a Gaussian posterior distribution.

Shared component model is a type of statistical model used to include correlation information between variables into the model. The shared component model was first introduced by Knorr-Held in 2000). This model concerns to the correlation between variables that influence observations. In the Bayesian Shared Component Model. variables that belong to a common group modeled together to estimate parameters and collectively explain that group. Moreover. one variable can provide information about the parameters used to explain other variables in the similar group. Through shared component modeling. it is possible to accommodate the temporal correlation of tax revenue data of tax service offices (KPP) so that forecasting modeling can be carried out more accurately.

Method

This research employed a Bayesian approach through the Latent Gaussian Model (LGM) to estimate parameters in a multivariate model as in equation (3). Bayesian method is a data analysis approach based on Bayes' theorem. where the available knowledge about the parameters in a statistical model is updated with information in the observed data (Van De Schoot et al.. 2021)The Bayesian approach views parameters as random variables that have a distribution. namely a prior distribution (Wagenmakers et al.. 2008)From the prior and likelihood. the posterior distribution can be determined to obtain a Bayesian estimator from the posterior distribution. The subject of this research is tax revenue data for Personal Income Tax (PPh) Article 25/29 at five Pratama Tax Service Offices (KPP) in the South Jakarta I Regional Office of the Directorate General of Taxes. These offices are KPP Pratama Jakarta Setiabudi Satu. KPP Pratama Mampang Prapatan. KPP Pratama Jakarta Tebet. KPP Pratama Jakarta Setiabudi Dua. and KPP Pratama Jakarta Pancoran. The data used are annual data from 2009 to 2022 obtained from the Directorate of Data and Information (DIP) of the Directorate General of Taxes.

Statistical Modelling

Tax revenue data is continuous data with area of (i = 1) represents tax revenue of KPP Pratama Jakarta Setiabudi Satu; (i = 2) represents tax revenue of KPP Pratama Mampang Prapatan; (i = 3) represents tax revenue of KPP Pratama Jakarta Tebet; (i = 4) represents tax revenue of KPP Pratama Jakarta Setiabudi Dua; and (i = 5) represents tax revenue of KPP Pratama Jakarta Pancoran and time (t). we describe a multivariate model to model tax revenue data. KPP tax revenue data i at time t is modeled as a Gaussian random variable.

$$y_{it}|\mu_{it}.\sigma^2{}_y \sim Gaussian\ (\mu_{it}.\sigma^2{}_y) \quad (1)$$

with $i = 1.2.\ldots n$ and $t = 1.2.\ldots T$

$$y_{it} = \alpha_i + \beta_i\zeta_t + \delta_i\lambda_t + \varepsilon_{it} \quad (2)$$
$$\varepsilon_{it} \sim N(0.\sigma^2)$$

$$E[y_{it}] = \mu_{it} = \alpha_i + \beta_i\zeta_t + \delta_i\lambda_t + u_t + v_t \quad (3)$$

with μ_{it}: expectation of a random variable y_{it}
α_i: average tax revenue data of KPP i;
β_i: temporal coefficient for the tax revenue data of KPP i;
ζ_t : temporal effect component (using Random Walk order 2);
δ_i : shared temporal component coefficient;
λ_t : shared temporal components;
u_t: unstructured temporal random effects (temporal heterogeneity)
v_t : random component

There are three main stages in the Bayesian workflow. including the initial distribution is expressed as a prior distribution determined before using the data. determining the likelihood function using information about the parameters available in the observed data. and combining the prior distribution with the likelihood function using Bayes' theorem in determining the posterior distribution. This distribution reflect updated knowledge. balancing prior knowledge with observed data. and used to make predictions about future events (Van De Schoot et al.. 2021)

The Bayesian approach distinguishes between observations and unknown quantities. Observation refers to data observed during research while unknown quantities are random variables that include the parameters to be estimated (Blangiardo & Cameletti. 2015)In the Bayesian paradigm. unknown parameters in the model are treated as random variables and aims to calculate (or estimate). With Bayes' theorem the posterior probability distribution is defined as follows (Gelman et al.. 2013)

$$p(\theta|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\theta)p(\theta)}{p(\boldsymbol{y})} \quad (4)$$

with:
$p\ (\theta\ |\ \boldsymbol{y})$: posterior distribution

$p\ (\mathbf{y}\mid\theta)$: likelihood function for the joint distribution function $\boldsymbol{y}$ conditional on the parameters θin the model
$p\ (\theta)$: prior distribution for parameters θ
$p\ (\boldsymbol{y})$: marginal probability distribution $\boldsymbol{y}$ obtained from $p\ (\boldsymbol{y})=\int \mathrm{p}(\mathbf{y}|\theta)\mathrm{p}(\theta)\mathrm{d}\theta$

In order to obtain the posterior probability distribution of $p\ (\theta\mid\boldsymbol{y})$. which states the uncertainty of the parameters to be estimated after the observation data. so that it depends on the data $\boldsymbol{y}$. The probability distribution function of $p\ (\boldsymbol{y})$ expresses the marginal probability distribution of data $\boldsymbol{y}$ and is considered as a constant normalizing factor that does not depend on the parameters θ. Therefore. equation (4) can be expressed as (Jaya & Andriyana. 2020)

$$\mathrm{p}(\theta|y) \propto \mathrm{p}(y|\theta)\mathrm{p}(\theta) \tag{5}$$

To obtain parameter estimates θ. it can be obtained by looking for the posterior mean value θ. namely $E\ (\theta)$ as follows (Lawson. 2013).

$$\hat{\theta} = E(\theta|y) = \int \theta\ p(\theta|y)\ d\theta \tag{6}$$

Estimating the parameters of the posterior distribution is often difficult because the posterior function is not in the form of a distribution function that is commonly known. To overcome this condition the INLA approach can be used.

Laplace Approximation

Laplace approximation becomes a definite integral approximation expressed through the Taylor series. Laplace approximation is used to find the marginal posterior distribution of each parameter with the following formula (Blangiardo & Cameletti. 2015).

$$\int f(x)dx = \int exp\ (log(f(x))\ dx \tag{7}$$

with $f(x)$ representing the density function of the random variable X. The function of log $f\ (x)$ can be expressed in the Taylor series and evaluate log $f\ (x)$ on $x = x_0\ as$ follows:

$$log\ f(x) \approx logf(x_0) + (x - x_0)\frac{\partial log(f(x)}{\partial x}\Big|_{x=x_0} + \frac{(x-x_0)^2}{2}\frac{\partial^2 log(f(x)}{\partial x^2}\Big|_{x=x_0} \tag{8}$$

If x_0 expressed as mode $x^*= argmax_x$ log $f\ (x)$. then $\frac{\partial log(f(x)}{\partial x}\Big|_{x=x^*} = 0$ and $logf(x)$ in equation (8) can be estimated as follows:

$$log\ f(x) \approx logf(x^*) + \frac{(x - x^*)^2}{2}\frac{\partial^2 log\ f(x^*)}{\partial x^2}\Big|_{x=x^*} \tag{9}$$

Then. the integral can be solved as follows:

$$\int f(x)dx \approx \int exp\ \left(log(f(x^*) + \frac{(x-x^*)^2}{2}\frac{\partial^2 log\ f(x^*)}{\partial x^2}\Big|_{x=x^*}\right)\ dx$$
$$= exp(log\ f(x^*)) \int exp\left(\frac{(x-x^*)^2}{2}\frac{\partial^2 log\ f(x^*)}{\partial x^2}\Big|_{x=x^*}\right)\ dx \tag{10}$$

where the integral in equation (10) can be approximated through the density of the normal distribution by taking the form $\sigma^{2*} = {-1}\Big/{\frac{\partial^2 log(f(x^*)}{\partial x^2}\Big|_{x=x^*}}$ to obtain:

$$\int f(x)dx = exp(log(f(x^*)) \int exp\ \left(-\frac{(x-x^*)^2}{2\sigma^{2*}}\right)\ dx \tag{11}$$

The integral in Equation (11) is identical to integratie the normal density function with mean $x^*$and variance σ^{2*}. For the interval $(a.b)$ it can be written as follows:

$$\int_a^b f(x)dx \approx f(x^*)\sqrt{2\pi\sigma^{2^*}}\int_a^b \frac{1}{\sqrt{2\pi\sigma^{2*}}} exp\left(\frac{(x-x^*)^2}{2\sigma^{2*}}\right)\ dx \approx f(x^*)\sqrt{2\pi\sigma^{2^*}}\big(\Phi(b) - \Phi(a)\big) \tag{12}$$

with $\Phi(.)$expressing the cumulative density function of the normal distribution $N\ (x^*.\sigma^{2*})$. This concept will help to solve the problems contained in the Bayesian method.

Integrated Nested Laplace Approximations (INLA)

Bayesian theory is always looking for new methods to recover the underlying posterior density assumptions of a model. When using an approach. such as *Markov Chain Monte Carlo* (MCMC). the issue often encountered is slow convergence to achieve accurate results (Robert & Casella. 2011)However. this can be overcome with LGM. which has a latent field structure containing unobserved parameters assumed to follow a Gaussian distribution. The additive structure combined with the Gaussian prior assumption on the latent field offers a natural way to encode data information into a precision matrix. The Integrated Nested Laplace Approximation (INLA) (Chiuchiolo. 2022)

The INLA implementation relies on a combination of analytical approximation and an efficient numerical integration scheme to achieve accurate deterministic

approximation for the posterior (Martino & Riebler. 2020)LGM for INLA-based inference as in equation (3) can be defined in three stages: likelihood function. latent Gaussian field. and hyperprior model (Morrison et al.. 2016)

First stage: likelihood function with the following equation:

$$(y|\Omega.\tau) = \prod_{i=1}^{n} p(y_i|\Omega_i.\tau) = \prod_{i=1}^{n}\prod_{t=1}^{T} p(y_{it}|\Omega_{it}.\tau) \quad (13)$$

where $\boldsymbol{y}$ is a vector containing observation values. vector Ω is a Gaussian field containing all latent model components $\Omega = \{ \alpha_1 . \ldots . \alpha_5 . \beta_1 . \ldots . \beta_5 . \zeta_1 . \ldots . \zeta_{14} . \delta_1 . \ldots . \delta_5 . \lambda_1 . \ldots . \lambda_{14} . u_1 . \ldots . u_{14} . v_1 \ldots . v_{14} \}$ and $\tau = \{\sigma^2\}$ is a hyperparameter of Ω.

Second stage: Latent Gaussian Field Distribution Ω conditional vector hyperparameters τ. $p(\Omega|\tau)$ follows the Multivariate normal. The prior density function of is Ω denoted as follows:

$$p(\Omega|\tau) = |Q_\tau|^{\frac{1}{2}} exp\left(-\frac{1}{2}\Omega' Q_\tau \Omega\right) \quad (14)$$

The latent variables are assumed to be multivariate Gaussians with a conditional independence structure that produces a sparse precision matrix (Rue & Held. 2005).

Third stage: The prior distribution of the hyperparameters is a hyperprior. The use of hyperprior in the model will produce robust parameter estimates (Aguerrebere et al.. 2014). $p(\boldsymbol{\tau})$ is the hyperpior of $\boldsymbol{\tau}$. INLA provides a simple way to define priors. For computational reasons. INLA works with an internal representation of the parameters and not with random effect precision parameters τ. but rather with $\theta = \log(\tau)$. Therefore. the prior must be determined on θ (Gómez-Rubio et al.. 2019).

This study uses the Half–Cauchy prior with the scale parameter γ defined σas:

$$p(\sigma|\gamma) = \frac{2}{\pi\gamma(1+(\sigma/\gamma)^2} \quad (15)$$

with scale parameter equal to 25 (Jaya & Folmer. 2020). From the three stages explained. the combined posterior distribution is as follows:

$$p(\Omega.\tau|y) = \frac{p(y|\Omega.\tau)p(\Omega|\tau)p(\tau)}{p(y|\tau)} \propto p(\tau)p(\Omega|\tau)p(y|\Omega.\tau) \quad (16)$$

The INLA procedure does not consider the full posterior distribution of Ω and $\boldsymbol{\tau}$ but rather is based on a marginal posterior distribution approach. namely $p(\Omega_i \mid \boldsymbol{y})$ and $p(\boldsymbol{\tau_k} \mid \boldsymbol{y})$. The marginal posterior distribution of Ω_i is defined as follows:

$$p(\Omega_i|y) = \int p(\Omega_i.\tau|y)d\tau = \int p(\Omega_i|\tau.y)p(\tau|y)d\tau . i = 1. \ldots . n \quad (17)$$

while the marginal posterior distribution of $\tau_k$is defined as follows:

$$p(\tau_k|y) = \int p(\tau|y)d\tau_{-k} \quad (18)$$

with τ_{-k} denoting all elements in $\boldsymbol{\tau}$ except element $-\ k$.

The first step is to estimate the marginal posterior distribution of the hyperparameter. $p(\boldsymbol{\tau} \mid \boldsymbol{y})$. using the Laplace approach (Jaya et al.. 2022; Jaya & Folmer. 2020)

$$p(\tau|y) = \frac{p(\Omega.\tau|y)}{p(\Omega|\tau.y)} = \frac{p(y|\Omega.\tau)p(\Omega|\tau)p(\tau)}{p(y|\tau)} \frac{1}{p(\Omega|\tau.y)} \propto \frac{p(y|\Omega.\tau)p(\Omega|\tau)p(\tau)}{p(\Omega|\tau.y)} \approx \left.\frac{p(y|\Omega.\tau)p(\Omega|\tau)p(\tau)}{p_G(\Omega|\tau.y)}\right|_{\Omega=\Omega^*(\tau)} =: \tilde{p}(\boldsymbol{\tau} \mid \boldsymbol{y}) \quad (19)$$

where $p(\Omega \mid \boldsymbol{\tau}.\boldsymbol{y})$ is the conditional distribution of the Ω approximation. which $p_G(\Omega|\tau.y)$ is a Gaussian approximation based on the Laplace transform. $\Omega^*(\boldsymbol{\tau})$ is the posterior mode of $p(\Omega \mid \boldsymbol{\tau}.\ \boldsymbol{y})$ for the hyperparameter $\boldsymbol{\tau}$. and $\tilde{p}(\boldsymbol{\tau} \mid \boldsymbol{y})$ are Laplace approximations for $p(\boldsymbol{\tau} \mid \boldsymbol{y})$. The second step is to calculate the marginal posterior conditional distribution $p(\Omega_i \mid \boldsymbol{\tau}.\boldsymbol{y})$. In the process. it is rewritten Ω as $\Omega = (\Omega_i . \Omega_{-i})$ so the equation becomes:

$$p(\Omega|\tau.y) = p\big((\Omega_i.\Omega_{-i})\big|\tau.y\big) = p(\Omega_{-i}|\Omega_i.\tau.y)p(\Omega_i|\tau.y) \quad (20)$$

From equation (20) obtained:

$$p(\Omega_i|\tau.y) = \frac{p\big((\Omega_i.\Omega_{-i})\big|\tau.y\big)}{p(\Omega_{-i}|\Omega_i.\tau.y)} = p(\Omega|\tau.y)\frac{1}{p(\Omega_{-i}|\Omega_i.\tau.y)} = \frac{p(\Omega.\tau|y)}{p(\tau|y)p(y)}\frac{1}{p(\Omega_{-i}|\Omega_i.\tau.y)} \propto \frac{p(\Omega.\tau|y)}{p(\tau|y)}\frac{1}{p(\Omega_{-i}|\Omega_i.\tau.y)} \propto \frac{p(\Omega.\tau|y)}{p(\Omega_{-i}|\Omega_i.\tau.y)}$$

$$\approx \frac{p(\Omega.\tau|y)}{p_G(\Omega_{-i}|\Omega_i.\tau.y)}\Big|_{\Omega_{-i}=\Omega^*{}_{-i}(\Omega_{-i}.\tau)} =: \tilde{p}(\Omega_i|\tau.y) \quad (21)$$

by Ω_{-i} showing all the inner elements $\Omega.$ except the element i-th. $p_G(\Omega_{-i}|\Omega_i.\tau.y)$ showing the Gaussian approximation of $p\ (\Omega_{-i}\,|\,\Omega_i.\ \boldsymbol{\tau}.\ \boldsymbol{y})$. then $\Omega^*{}_{-i}(\Omega_{-i}.\ \boldsymbol{\tau})$ showing the mode of $p\ (\Omega_{-i}\,|\,\mu_i.\ \boldsymbol{\tau}.\ \boldsymbol{y})$ and $\tilde{p}(\ \Omega_i\,|\,\boldsymbol{\tau}.\boldsymbol{y})$ is Laplace approximation for $p\ (\Omega_i\,|\,\boldsymbol{\tau}.\ \boldsymbol{y})$.

The third stage after getting $\tilde{p}(\Omega_i|\tau.y)$and $\tilde{p}(\boldsymbol{\tau}\,|\,\boldsymbol{y})$ then the marginal posterior distribution of the parameter $p\ (\Omega_i|y)$ can be calculated with the following equation:

$$p(\Omega_i|y)\ \approx \int \tilde{p}\ (\Omega_i|\tau.y)\tilde{p}(\tau|y)d\tau \quad (22)$$

Multivariate Forecasting

To obtain the multivariate forecasting value for each KPP. the researchers use the multivariate posterior prediction distribution. which is defined as follows:

$$p(\hat{y}_{T+h}|\mathrm{y}.\theta) = \int p(\hat{y}_{T+h}|\Omega.\theta)\ p(\Omega|y.\theta)d\Omega \quad (23)$$

Where $\hat{y}_{T+h} = (\hat{y}_{1(T+h)}.\hat{y}_{2(T+h)}.\hat{y}_{3(T+h)}.\hat{y}_{4(T+h)}.\hat{y}_{5(T+h)})$ represents a vector of forecasting values for each KPP at time t. In INLA. forecasting is implemented by entering 'Not Available (NA)' for the period T + h in which the forecast is made (Morrison et al.. 2016).

Evaluation of Forecasting Models

A model evaluation is performed to determine when a forecasting model has a high level of accuracy. Evaluate forecasting models generally aims to minimize out-of-sample prediction errors. Mean Absolute Percentage Error (MAPE) is one of the most commonly used measures to calculate forecasting accuracy. MAPE is the average of Absolute Percentage Errors (APE) (Kim & Kim. 2016). The MAPE value is defined as follows:

$$MAPE\ = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|\frac{\hat{Y}_{it}-Y_{it}}{Y_{it}}\right| \text{ x } 100\% \quad (24)$$

with :

n : number of KPP units

T: amount of data

$\hat{Y}_{it}$:Forecasting value of the i-th tax revenue at the t-th time

Y_{it}: Actual value of the i-th tax revenue at time t

Results and Discussion

South Jakarta I Regional Tax Office is one of 34 regional offices in Indonesia (pajak.go.id). In this study. we utilized tax revenue data at the South Jakarta I Regional Office of DJP with several KPPs. including KPP Pratama Jakarta Setiabudi Satu. KPP Pratama Jakarta Mampang Prapatan. KPP Pratama Jakarta Tebet. KPP Pratama Jakarta Setiabudi Dua. and KPP Pratama Jakarta Pancoran as response variables Multivariate. Before carrying out further analysis. we carried out descriptive analysis of each research variable to better understand the characteristics of each variable.

Table 1. Descriptive Analysis

Office	Minimal	Average	Median	Maximum
Jakarta Setiabudi Satu	10.05	20.57	16.52	55.50
Jakarta Mampang Prapatan	20.66	39.93	30.99	101.68
Jakarta Tebet	16.91	38.08	30.09	84.13
Jakarta Setiabudi Dua	8.81	25.65	19.93	58.50
Jakarta Pancoran	6.98	26.49	20.35	54.63

The table 1 presents tax revenue data for 2009 - 2022 with minimum tax revenue ranging from IDR 6.976.642.060 to IDR 20.665.236.039 and maximum tax revenue ranging from IDR 54.638.175.452 to IDR 101.683.887.079.

Table 2. Correlation Analysis

Office	Jakarta Setiabudi Satu	Jakarta Mampang Prapatan	Jakarta Tebet	Jakarta Setiabudi Dua	Jakarta Pancoran
Jakarta Setiabudi Satu	1				
Jakarta Mampang Prapatan	0.65	1			
Jakarta Tebet	0.67	0.88	1		
Jakarta Setiabudi Dua	0.59	0.66	0.80	1	
Jakarta Pancoran	0.30	0.53	0.63	0.66	1

Table 2 presents the positive correlation between each KPP which ranges from 0.3087 to 0.8853. which shows that there is a fairly strong correlation between each KPP. These results indicate that KPPs have relatively similar characteristics regarding tax revenues. thus supporting the suitability of using shared components in LGM multivariate model modeling.

Bayesian hierarchical temporal modeling

Table 3 shows the intercept value for the Jakarta Setiabudi Satu KPP is 23.820. which can be concluded that the posterior mean of the KPP tax revenue data is exp (23.820) = IDR 22.125.574.641 when there are no random effects from the temporal component and shared component. Likewise for other KPP intercepts have almost the same intercept values.

Based on the Table 4. it is found that the temporal patterns are similar between each KPP. which is indicated by the relatively similar shared component coefficient values which are in the range 0.927 – 1.03. This is considered normal because the five KPPs are located in the same regional office.

Table 3. Summary statistics for the fixed effects

KPP	Posterior Mean	Standard Deviation	95% Credible Interval	
			Lower bound	Upper bound
Intercept Jakarta Setiabudi Satu	23820	0.06	23703	23937
Intercept Jakarta Tebet	24326	0.05	24225	24427
Intercept Jakarta Setiabudi Dua	23924	0.05	23812	24036
Intercept Jakarta Mampang Prapatan	24357	0.07	24216	24499
Intercept Jakarta Pancoran	23800	0.09	23620	23980

Table 4. Shared Component Model Estimation

KPP	Posterior Mean	Standard Deviation	95% Credible Interval	
			Lower bound	Upper bound
Jakarta Setiabudi Satu	1.00	-	-	-
Jakarta Tebet	0.927	0.309	0.321	1.54
Jakarta Setiabudi Dua	1.03	0.301	0.440	1.63
Jakarta Mampang Prapatan	1.00	-	-	-
Jakarta Pancoran	0.958	0.313	0.348	1.58

Table 5 presents the posterior mean value of the statistics for standard deviation (SD) in the temporal unstructured effect of tax revenue data for the five KPPs. The standard deviation value for KPP Jakarta Setiabudi Satu. KPP Jakarta Tebet and KPP Setiabudi Dua has almost the similar value. namely in the range 0.0622 – 0.0762 which indicates that the three KPPs have the same characteristics in terms of tax revenue contribution. as well as for KPP Jakarta Mampang Prapatan and KPP Jakarta Pancoran. which has standard deviation values of 0.1131 and 0.1009. Apart from that. there is also a 95% credible interval. Hyperparameters provide information regarding the contribution of each fitted random effect in the model.

Based on the shared component plot above. it can be seen that the five KPPs have similar fluctuations. This means that the model is significant for the parameter estimates in the forecasting model.

It can be seen that KPP Pratama Jakarta Setiabudi Satu. KPP Pratama Jakarta Tebet. and KPP Pratama Jakarta Setiabudi Dua have relatively the similarr temporal pattern. Likewise with KPP Pratama Mampang Prapatan and KPP Pratama Jakarta Pancoran also have relatively the same pattern. This shows that the forecasting model can be used with information based on KPP temporal patterns. which have relatively similar characteristics and patterns.

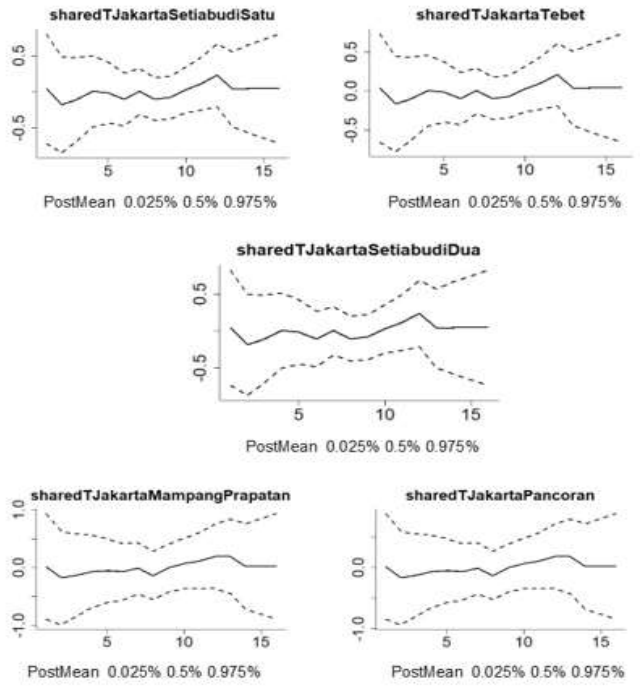

Figure 1. Plot shared components

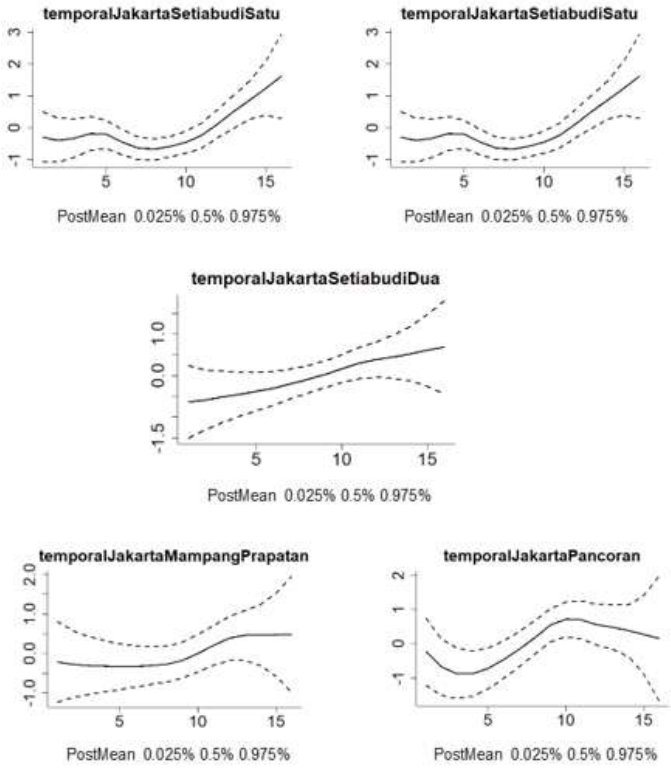

Figure 2. Temporal plot

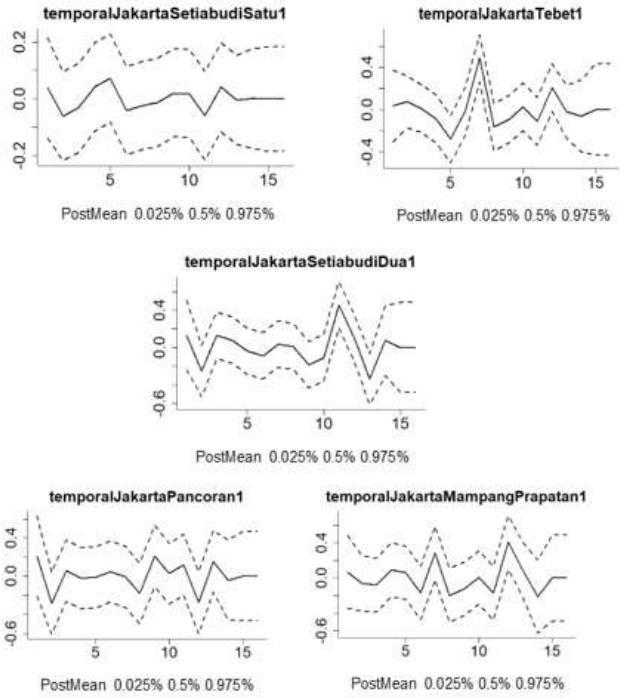

Figure 3. Temporal heterogeneity plot

Figure 3. shows in detail the existence of variations in tax revenues that are relatively the same from each KPP. This can occur because each KPP has the same tax policy but different number of potential taxpayers in each region. Therefore. a KPP with a larger of potential taxpayers will result in greater tax revenues.

Forecasting Result using LGM

Based on Table 6. almost the similar MAPE values are obtained for all five KPPs. under 10% using training data. This indicates that the forecast using LGM with the INLA approach produces very accurate MAPE values.

Table 7 shows the forecast results for tax revenue data in the five KPPs in the South Jakarta I Regional Office of the Directorate General of Taxes using the INLA approach for the years 2023 and 2024. The visual representation of actual and forecasted tax revenue values for each KPP can be seen in Figure 4.

Figure 4 represents the plot of actual data and forecasted revenue for the personal income tax Article 25/29 in the five KPPs in the South Jakarta I Regional Office of theDirectorate General of Taxes using the INLA approach. Based on the above figure. it can be observed that the forcasted data patern aligns well. following the increase or decrease observed in the data pattern. It can be seen that among the five KPPs. KPP Jakarta Pancoran experiences a decrease revenue. mainly due to the distinctive temporal data pattern of Jakarta Pancoran compared to the other four KPPs. Forecasting using the Multivariate Latent Gaussian Model tends to perform better when there is a stronger correlation between

variables. In a ddition. Bayesian forecasting tends to provide better results by avoiding constant outcomes.

Table 5. Statistics of the posterior means for temporally unstructured effect

Hyperparameters	Posterior Mean	Standard Deviation	95% Credible Interval	
			Lower bound	Upper bound
SD for Temporally unstructured effects Jakarta Setiabudi Satu	0.09	0.06	0.02	0.08
SD for Temporally unstructured effects Jakarta Tebet	0.24	0.06	0.13	0.23
SD for Temporally unstructured effects Jakarta Setiabudi Dua	0.26	0.07	0.14	0.25
SD for Temporally unstructured effect Jakarta Mampang Prapatan	0.30	0.11	0.15	0.27
SD for Temporally unstructured effect Jakarta Pancoran	0.2651	0.1009	0.1200	0.2474

Tabel 6. Result of MAPE Value

KPP	MAPE
Jakarta Setiabudi Satu	0.09
Jakarta Tebet	0.01
Jakarta Setiabudi Dua	0.01
Jakarta Mampang Prapatan	0.01
Jakarta Pancoran	0.01

Table 7. Forecast result

KPP	Year	Prediction Value	95% Credible Interval	
			Lower bound	Upper bound
Jakarta Setiabudi Satu	2023	80.28	40308042404	159886883816
	2024	116.27	33629925851	401961947660
Jakarta Tebet	2023	76.77	35719557703	165014315011
	2024	84.34	31179144.405	228169107527
Jakarta Setiabudi Dua	2023	47.23	20244331242	110215131963
	2024	51.08	16896827161	154448629887
Jakarta Mampang Prapatan	2023	62.45	23007058.942	169557903185
	2024	62.93	14968974.261	264579745579
Jakarta Pancoran	2023	28.93	9167531.592	91329740606
	2024	25.84	3945604.016	169271486430

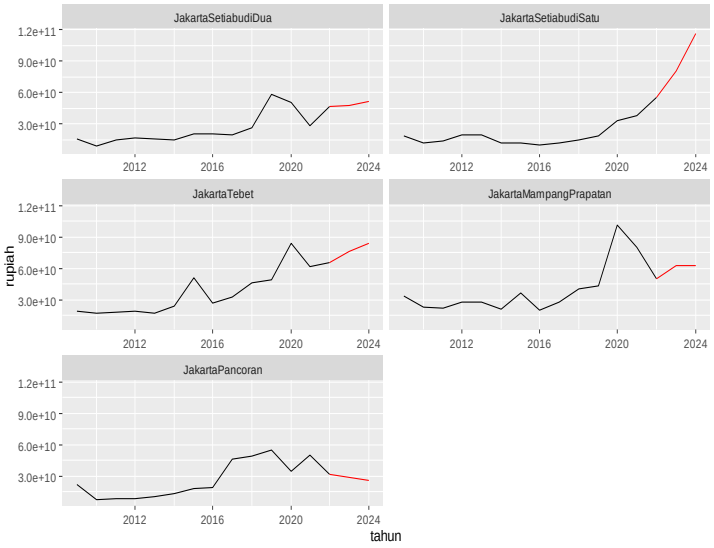

Figure 4. Plot forecasts the personal income tax revenue Article 25/29 in the Five KPPs

Tax reform in Indonesia was started in 1993. transitioning from the official assessment to self-assessment. Since tax reform. taxes have played a crucial role in government operations and have become a backbone of the country. representing one of the largest sources of revenue in the state budget (Bawazier, 2018). The state utilizes taxes as a promary instrument to fund central and regional government expenditures for the welfare of the community. Expenditures on development or state operations. such as providing healthcare. education. infrastructure. and other public services. can only be achieved if tax revenue is effectively mobilized. Taxation shares a concept similar to mutual cooperation *(gotong royong)* as it requires contributions from all citizens.

The government's efforts through tax reform aimed not only at improving taxpayer compliance but also to increase tax revenue (Kemenkeu. 2023). Personal Income Tax Article 25/29 is one form of tax that is a practice of the self-assessment system and affects state revenues . With the high revenue from Personal Income Tax. it can be interpreted that there is a positive increase in the income of the population. Tax revenue plays a dominant role as a source of state income.

Tax revenues are the largest contributor to state income. More than 70% of Indonesia's state income comes from tax revenues . The role of tax revenues in financing the State Budget is getting increase every year. Considering the important role of taxes in financing state development. a series of efforts which include analyzing tax potential. forecasting tax revenues. and monitoring the realization of tax revenues must be implemented properly to ensure the stability of tax and expenditure policies (Jenkins et al.. n.d.)

Several researchers have forecasted tax revenues using different models and methods. including comparing tax forecasting models using the Random Walk. SARIMA. and BATS techniques (Erdoğdu & Yorulmaz. 2019)then forecasting tax revenues using ARMA and GARCH (Cyril. 2017)Next. forecast types of VAT taxes in Ghana using the ARIMA method with intervention and the Holt linear trend method (Ofori et al.. 2020)and forecast taxes for the types of VAT taxes using the ARIMA Box – Jenkins Method (Fathoni & Saputra. 2023)

Forecasting tax revenues is an important thing because it can be an effective and targeted benchmark and can be used as a baseline for calculating tax revenue targets in the following year. Therefore. the tax revenue target in the State Budget will be a reference and performance for the government in generating the amount of tax revenue for one year so that it can be used in making policies for preparing the State Budget. which influence the design of government activity programs. LGM with a Bayesian approach is one of the forecasting methods used in time series data analysis by offering a more accurate representation without the assumptions of stationarity and cointegration as it is accommodated in the model. Correct model specification is necessary to produce accurate forecasts.

The data used in this research involved 14 temporal data. and in obtaining MAPE values. only the training data is used. Bayesian approaches are generally carried out on small samples with limited information. The use of the Bayesian method through the prior distribution can be a solution to incomplete information obtained from the data and provide good results for small sample sizes compared to the Maximum Likelihood method (Jaya & Andriyana. 2020)As a result. one of the advantages of using a Bayesian approach is that it does not have to use large sample data (Van De Schoot et al.. 2014).

Conclusion

The Latent Gaussian Model using the Bayesian Approach in parameter estimation can be utilized in forecasting Personal Income Tax Article 25/29. This is supported by the highly accurate MAPE value of 0.01%. The implementation of the developed model is not limited to forecasting Personal Income Tax Article 25/29. but can also be used in various other fields. With its hierarchical structure. the Bayesian approach proves to be an effective method for addressing complex modeling challenges.

Acknowledgments

The authors would like to thanks to Department of Statistics. Faculty of Mathematics and Natural Sciences. Padjadjaran University. Directorate General of Taxes (DJP) and Indonesia Endowment Fund for Education
(LPDP) for give ocasion for this research.

Author Contributions

Investigation. E.P. Y. S and G.D; formal analysis. E.P. Y. S and G.D ; investigation. E.P. Y. S and G.D; resources. E.P. Y. S and G.D; data curation. E.P. Y. S and G.D: writing—original draft preparation. E.P. Y. S and G.D ; writing—review and editing. E.P. Y. S and G.D: visualization. E.P. Y. S and G.D ; supervision. E.P. Y. S and G.D; project administration. E.P. Y. S and G.D; funding acquisition. E.P. Y. S and G.D. All authors have read and agreed to the published version of the manuscript.

Funding

This research is fully supported by the author's funds without any external funding sources

Conflicts of Interest

We certify that there is no conflict of interest with any financial. personal and other relationships with other peoples or organisation related to the material discussed in the manuscript.

References

Aguerrebere. C.. Almansa. A.. Gousseau. Y.. Delon. J.. & Musé. P. (2014). A Hyperprior Bayesian Approach for Solving Image Inverse Problems. *International Conference on Computational Photography*. Retrieved from https://shorturl.asia/d98P2

Badan Kebijakan Fiskal. (2016). *Mengelola Tantangan untuk Meningkatkan Kesejahteraan*. Kementerian Keuangan. Retrieved from https://fiskal.kemenkeu.go.id/files/tekf/file/tinjauanekonomi_edisi5-2016.pdf

Bawazier, F. (2018). Reformasi Pajak di Indonesia Tax Reform In Indonesia. *Jurnal Legislasi Indonesia*, *8*(1), 1-28. Retrieved from https://shorturl.asia/5dG4J

Blangiardo. M.. & Cameletti. M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley and Sons. Inc.

Chiuchiolo. C. (2022). *Joint Posterior Inference for Latent Gaussian Models and extended strategies using*

INLA. (Doctoral dissertation). Retrieved from https://repository.kaust.edu.sa/handle/10754/679224

Cyril. C. (2017). Forecasting Tax Revenue and its Volatility in Tanzania. *African Journal of Economic Review*. *V*(I). Retrieved from https://www.ajol.info/index.php/ajer/article/view/149255

Erdoğdu. H.. & Yorulmaz. R. (2019). Comparison of Tax Revenue Forecasting Models for Turkey. In S. Oktar & Y. Taşkın. *34. International Public Finance Conference* (pp. 482–492). Istanbul University Press. https://doi.org/10.26650/PB/SS10.2019.001.075

Fathoni. M. I.. & Saputra. A. (2023). Forecasting Value-Added Tax (VAT) revenue using Autoregressive Integrated Moving Average (ARIMA) Box-Jenkins method. *Scientax*. *4*(2). 205–218. https://doi.org/10.52869/st.v4i2.568

Gelman. A.. Carlin. J. B.. Stern. H. S.. Dunson. D. B.. Vehtari. A.. & Rubin. D. B. (2013). *Bayesian Data Analysis Third edition*. Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Giacomini. R.. & Kitagawa. T. (2021). Robust Bayesian Inference for Set-Identified Models. *Econometrica*. *89*(4). 1519–1556. https://doi.org/10.3982/ECTA16773

Gómez-Rubio. V.. Palmí-Perales. F.. López-Abente. G.. Ramis-Prieto. R.. & Fernández-Navarro. P. (2019). Bayesian joint spatio-temporal analysis of multiple diseases. *SORT. Statistics and Operations Research Transactions*. *43*. 51–74. https://doi.org/10.2436/20.8080.02.79

Jaya. I. G. N. M.. & Andriyana. Y. (2020). *Analisis Data Spasial Perspektif Bayesian*. Alqaprint.

Jaya. I. G. N. M.. Chadidjah. A.. Darmawan. G.. Princidy. J. C.. & Kristiani. A. F. (2022). *Does the Restriction of Human Mobility Significantly Control COVID-19 Transmission in Jakarta. Indonesia? Global Versus Local Regression Models* [Preprint]. Mathematics & Computer Science. https://doi.org/10.20944/preprints202209.0271.v1

Jaya. I. G. N. M.. & Folmer. H. (2020). Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung. Indonesia. *Journal of Geographical Systems*. *22*(1). 105–142. https://doi.org/10.1007/s10109-019-00311-4

Jenkins. G. P.. Kuo. C. Y.. & Shukla. G. P. (n.d.). *Tax Analysis and Revenue Forecasting: Issues and Techniques*. Harvard Institute for International Development.

Kemenkeu. (2023. May 5). *Reformasi Perpajakan untuk Penciptaan Keadilan. Peningkatan Kepatuhan. dan Penguatan Fiskal*. Retrieved from https://fiskal.kemenkeu.go.id/publikasi/siaran-pers-detil/326

Kim. S.. & Kim. H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*. *32*(3). 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Lawson. A. B. (2013). *Bayesian Disease Mapping Hierarchical Modeling in Spatial Epidemiology*. CRC Press.

Lee. D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-Temporal Epidemiology*. *2*(2). 79–89. https://doi.org/10.1016/j.sste.2011.03.001

Martino. S.. & Riebler. A. (2020). Integrated Nested Laplace Approximations (INLA). *Wiley StatsRef: Statistics Reference Online*. 1–19. https://doi.org/10.1002/9781118445112.stat08212

Morrison. K. T.. Shaddick. G.. Henderson. S. B.. & Buckeridge. D. L. (2016). A latent process model for forecasting multiple time series in environmental public health surveillance: A latent process model for forecasting multiple time series in environmental public health surveillance. *Statistics in Medicine*. *35*(18). 3085–3100. https://doi.org/10.1002/sim.6904

Ofori. M. S.. Fumey. A.. & Nketiah-Amponsah. E. (2020). *Forecasting Value Added Tax Revenue in Ghana*. *4*(2). Retrieved from https://ojs.tripaledu.com/jefa/article/view/58

Oleson. J. J.. Smith. B. J.. & Kim. H. (2021). Joint Spatio-Temporal Modeling of Low Incidence Cancers Sharing Common Risk Factors. *Journal of Data Science*. *6*(1). 105–123. https://doi.org/10.6339/JDS.2008.06(1).382

Opitz, T. (2017). Latent Gaussian modeling and INLA: A review with focus on space-time applications. *Journal de la société française de statistique*, *158*(3), 62-85. Retrieved from http://www.numdam.org/item/JSFS_2017__158_3_62_0/

Richardson, S., Abellan, J. J., & Best, N. (2006). Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). *Statistical methods in medical research*, *15*(4), 385-407. https://doi.org/10.1191/0962280206sm458oa

Robert. C.. & Casella. G. (2011). A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical*

Science. *26*(1). https://doi.org/10.1214/10-STS351

Rue. H.. & Held. L. (2005). *Gaussian Markov Random Fields*. Chapman and Hall/CRC.

Rue. H.. Martino. S.. & Chopin. N. (2009). Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. *71*(2). 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Van De Schoot. R.. Depaoli. S.. King. R.. Kramer. B.. Märtens. K.. Tadesse. M. G.. Vannucci. M.. Gelman. A.. Veen. D.. Willemsen. J.. & Yau. C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*. *1*(1). 1. https://doi.org/10.1038/s43586-020-00001-2

Van De Schoot. R.. Kaplan. D.. Denissen. J.. Asendorpf. J. B.. Neyer. F. J.. & Van Aken. M. A. G. (2014). A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development*. *85*(3). 842–860. https://doi.org/10.1111/cdev.12169

Wagenmakers. E.-J.. Lee. M.. Lodewyckx. T.. & Iverson. G. J. (2008). Bayesian Versus Frequentist Inference. In H. Hoijtink. I. Klugkist. & P. A. Boelen (Eds.). *Bayesian Evaluation of Informative Hypotheses* (pp. 181–207). Springer New York. https://doi.org/10.1007/978-0-387-09612-4_9

Wasserman, L. A. (1989). A robust Bayesian interpretation of likelihood regions. *The Annals of Statistics*, *17*(3), 1387-1393. Retrieved from https://projecteuclid.org/journals/annals-of-statistics/volume-17/issue-3/A-Robust-Bayesian-Interpretation-of-Likelihood-Regions/10.1214/aos/1176347277.short

Wei. W. W. S. (2019). *Multivariate Time Series Analysis and Applications*. John Wiley & Sons Ltd.