Thailand rice production analysis: alternative approaches without P-value

Thunyawadee Sucharidtham and Satawat Wannapan

1 Faculty of Administration and Liberal of Art, Rajamangala University of Technology, Chiang Mai, Thailand
2 Faculty of Economics, Chiang Mai University, Thailand

thunyawadee@gmail.com; lionz1988@gmail.com

Abstract. This paper was contributed to seek a supporting evidence for the use of spatial analysis in panel data estimations. Yearly panel data regarding rice production and crucial factors, as well as quantities of rice planting areas, fertilizer usages, rice farmer families, and precipitations were observed as a time-series trend during 2009 to 2017. Methodologically, to avoid traditionally statistical assumptions like probabilistic significant test and to improve sensibility of model estimations, Bayesian statistics and inferences were applied to the entire methods of this paper. Stationary testing was accomplished by employing the Bayesian-ADF unit method. Additionally, the findings estimated via Bayesian panel regression models that indicated the spatial dummy added into the models also influenced both rice products and related factors. Thus, rice production needed to be regionally investigated and proposed critical issues should be practically considered.

1. Introduction

Historically, the conceptual framework of homogeneity has been politically and widely developed in Thailand. This concept is instilled in farmers’ perception that they have to grow a certain kind of economic crop throughout the country. A critical issue of the paper is to provide implementation question of this idea. Therefore the question should be ‘Is it the really best solution for managing quantities and quality of agricultural products by producing them in everywhere without zoning considerations?’ Econometrics and alternative statistics might be a potential tool to find the answer.

However, the problems of classical statistics are the points of estimation consistency and reliability along with econometrical predictions. Previously, probability and significant testing is highly trusted, but this is intensively criticized at present. On the other hand, dealing with probabilistic systems where statistical inference plays a major role seems to be useful when Fisher’s contributions in statistical science influenced a vast number of noble econometric papers. Somehow, the problem is referred as Null Hypothesis Significant Testing (NHST). However, this paper is not a critical academic topic. The alternative statistical inference called “Bayesian statistics” is applied to find a better path for the econometric interpretation of farming forecasting, especially in terms of ‘rice products’ in Thailand.

Literarily, Bayesian statistical inference is a branch of statistics which primarily focuses on an initial, called “prior”. This term seems to be close to common aspects of human behaviors. According to the literature, Bayesian approaches have been continuously applied in economics, econometrics, and other academic fields. In tourism economics, [1] the panel data of tourism sectors in Thailand was explored throughout Malaysia, and Singapore. The main statistical approach was Bayesian inference. In
macroeconomics, [2] Bayesian approaches were analyzed and employed in the extreme events that could be described as crises or unusual times-series trends among the macroeconomic variables. Moreover, [3] Bayesian inference would be part of stationary testing in the section of data classification. In agricultural economics, [4] Bayesian statistics would be employed as essential component of the non-stationary model to estimate rubber prices in further exchange markets. Furthermore, in other fields, [5] researches were conducted to provide theoretical framework of sustainable economics that could gather any provided information from panel data and extraneous sources by applying Bayesian statistics. [6] Consideration of linear panel data model with time varying heterogeneity in financial economics is based on Bayesian inference approaches with MCMC simulations. [7] Bayesian Model Averaging (BMA) was applied to hamper consensus on the key factors of economic growth. As a result of panel data model’s efficiency and flexibility of Bayesian statistics, reliability of estimated results and guarantee could be improved, along with more accuracy in the prediction of Thai rice productivity estimation.

2. Methodology

Informatively, it is undeniable that the trend of rice production has been continuously mentioned from academic researchers and political authorities in Thailand. Due to its deep harmonious blending with traditional culture crucially centered throughout the country, rice is Thai’s identity and Thailand is referred as “rice bowl of Asia”. In terms of technical process, time-series panel data of rice production and factors potentially affecting rice production were observed and collected into spatial information. Seventy-six provinces were divided into six regions; northern, northeastern, central, eastern, western, and southern regions. Yearly data was collected during 2009 to 2017, for 10-year data consideration. According to the panel observations, a vast amount of dimensions could be powerful tool to reliably explain estimated results based on real situations. Details of descriptive information are presented in Table.1.

| Variable (type) | Definition | Unit  | Obs  | Mean   | Std.Dev. | Min   | Max  |
|-----------------|------------|-------|------|--------|----------|-------|------|
| Rice product    | The panel samples of Thai rice products were collected from 76 provinces during 2009 to 2017. Data source was collected from the Office of Agricultural Economics, Bangkok, Thailand. | Tons  | 684  | 388250.1| 492639.3 | 31.0  | 4314831 |
| (Dependent factor) | | | |
| Crop area       | The panel samples of quantities of rice growing areas were collected from 76 provinces during 2009 to 2017. Data source was collected from the Office of Agricultural Economics, Bangkok, Thailand. | Rai  | 684  | 800546.9| 951749.7 | 63.0  | 4314831 |
| (Independent factor) | | | |
| Fertilizer usage| The panel samples of fertilizer usages were collected from 76 provinces during 2009 to 2017. Data source was from the Office of Agricultural Economics, Bangkok, Thailand. | Tons  | 684  | 23047.4 | 26686.3 | 1.0   | 112045 |
| (Independent factor) | | | |
| Rice farmer household | The panel samples of numbers of rice farming families were collected from 76 provinces during 2009 to 2017. Data source was from the Office of Agricultural Economics, Bangkok, Thailand. | Units | 684  | 49020.6 | 55342.1 | 7.0   | 243328 |
| (Independent factor) | | | |
| Precipitation   | The panel samples of quantities of average rainfalls per year were collected from 76 provinces during 2009 to 2017. Data source was from GISTHAI, the Department of Geology, Faculty of Science, Chulalongkorn University. | mm.   | 684  | 1666.0  | 733.4    | 654.2 | 5559 |
| (Independent factor) | | | |

Generally, collected data was spatially observed throughout Thailand’s seventy-six provinces during 2009 to 2017 as yearly details. Descriptive information evidently stated that most areas could be potential for rice production. This is graphically shown in Fig.1 to Fig.6 in the appendix.

1 Rai is a unit of area which equals to 1,600 square meters (16 acres, 0.16 hectares, 0.3954 acres)
Considering Fig.1 and Fig.2 (in the appendix), rice producing activities were mostly observed in northern and northeastern provinces. The areas of 46 provinces contain mountain curves inside the ten-year trends of rice production. Especially, the northeastern area presented the highest frequency of production activities. In terms of Fig.3 and Fig.4 (in the appendix), there are varied kinds of economic plants in eastern and central regions, along with seashore commercial activities. Several curves appeared in 9 of 11 central province areas. However, there are some places within the eastern region that could be chosen as rice production areas. After considering Fig. 5 and Fig. 6, it was found that rice was not able to grow well, with close to a half amount ratio from 23 western and southern provinces where rice production activities could not be found. Lastly, the descriptive details presented in Table 1 reveal some highlighted points. For example, the fluctuations between maximum and minimum values imply that rice is not planted everywhere and it can be influenced by many seasonal factors. Additionally, a huge variation of standard deviations indicates that the trends of rice production can be roughly estimated or predicted by traditional econometric coefficient analyses.

Fundamentally, the major framework in this paper was divided into two parts: panel data classification and panel data estimation. The former was conducted to use data stationary test and ADF unit testing based on Bayesian inference, and finally, all of selected variables were checked. From Bayes’ theorem, we obtained the posterior density \( \pi(\theta|y) \) which was referred to as probability distribution of parameters \( \theta \) as in [8],

\[
\pi(\theta|y) \propto L(y|\theta)\pi(\theta),
\]

As showed in Equation (1), \( \pi(\theta|y) \) is a prior density of \( \theta \). Bayesian modeling, as mentioned, requires a joint distribution function expressed in Equation (2).

\[
L(y|\theta)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{y_i^2}{2\sigma^2}\right),
\]

where \( y \) refers to a time-series data of \( n \) observations, \( y = (y_1, y_2, \ldots, y_n) \) and \( \theta \) stands for the estimated parameters. Bayesian statistics is applied to consider hypotheses regarding multiple parameters by adapting Bayes factor comparisons [9]. The ADF test analyzes the null hypothesis that a time-series data is I(1) against the alternative I(0) [10]. This is presented in Equation (3) and the result is described in Table 2.

\[
\Delta y_t = c + \alpha D_t + \phi y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + \epsilon_t,
\]

### Table 2: The Bayesian ADF unit-root testing for data stationary.

| Index               | Hypothesis          | Posterior odds ratio (MCMC/Burn-in) | Interpretation                  | Result                        |
|---------------------|---------------------|-------------------------------------|---------------------------------|-------------------------------|
| Rice products       | \( H_0(M_i) \): Stationary data \( H_1(M_i) \): Non-stationary data | 133,101 (55,000/5,000) | Strong evidence of \( M_i \) | Strong evidence for Stationary data |
| Crop areas          | \( H_0(M_i) \): Stationary data \( H_1(M_i) \): Non-stationary data | 0.472 (55,000/5,000) | Weak evidence of \( M_i \) | Weak evidence for non - stationary data |
| Fertilizer usages   | \( H_0(M_i) \): Stationary data \( H_1(M_i) \): Non-stationary data | 0.866 (55,000/5,000) | Weak evidence of \( M_i \) | Weak evidence for non - stationary data |
| Rice farmer households | \( H_0(M_i) \): Stationary data \( H_1(M_i) \): Non-stationary data | 0.312 (55,000/5,000) | Weak evidence of \( M_i \) | Weak evidence for non - stationary data |
| Precipitation       | \( H_0(M_i) \): Stationary data \( H_1(M_i) \): Non-stationary data | 648,948 (55,000/5,000) | Strong evidence of \( M_i \) | Weak evidence for non - stationary data |

The ADF unit-root testing based on Bayes statistics was employed to guarantee that suspicious information would not occur during the long-run period. Without restrictions of traditional statistics and the ability to efficiently simulate observations by the MCMC approach (50,000 MCMC simulations and 5,000 burn-in samples, outliners and data trails could provide sensible results of stationary testing. Empirically, based on Jeffrey’s scales, all variables were evidences for stationary action.

The latter model is the Bayesian panel regression model. The fixed-effect model can be understood as an OLS model with a matrix of explanatory variables, \( X \), being augmented by vectors of dummy variables. The model can be considered as
\[ Y_{it} = \eta + \sum_{j=1}^{K} \beta_j X_{jt} + \epsilon_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T, \]  

(4)

where \( Y_{it} \) is a value of response variable for \( i^{th} \) unit at time \( t \). \( X_{jt} \) is an explaining variables and \( \eta, \beta_j, j = 1, \ldots, K \) are fixed parameters. \( \epsilon_{it} \) is a residual term with IID assuming and \( \epsilon_{it} \approx N(0, \sigma_\epsilon^2) \). The parameter \( \eta \) could be specifically expressed as \( \eta = \beta_0 + u_i \), where \( u_i \approx N(0, \sigma_u^2) \). Then, the model in Equation (4) can be re-written as [12]:

\[ Y_{it} = \beta_0 + \sum_{j=1}^{K} \beta_j X_{jt} + u_i + \epsilon_{it}, \]  

(5)

where \( u_i + \epsilon_{it} = \omega_{it}, \omega_{it} \approx N(0, \sigma_\omega^2) \), \( \sigma_\omega^2 = \sigma_u^2 + \sigma_\epsilon^2 \). Hence, by using matrix notation, Equation (4) is \( Y = F\theta + \omega \). From Equation (5), \( Y \approx N(F\theta, \lambda) \). Such likelihood function is a joint density of the \( Y \)'s, that is,

\[
L(Y; \theta, \lambda) = (2\pi)^{-NT/2} |\lambda|^{N/2} \exp \left[ -\frac{1}{2} (Y - F\theta)^T \lambda^{-1} (Y - F\theta) \right] = (2\pi)^{-NT/2} \left( \sigma_\epsilon^2 \right)^{-N/2} \left( \sigma_u^2 \right)^{-N/2} \exp \left[ -\frac{1}{2} (Y - F\theta)^T \frac{1}{\sigma_u^2 + \sigma_\epsilon^2} (Y - F\theta) \right] \]  

(6)

To specify a complete Bayesian model, the uniform prior distribution \( U(0,1) \) of the vector parameters \( \theta \) is required. \( \sigma_\epsilon^2, \sigma_u^2 \) are the inverse gammas of parameters \( \alpha_\epsilon, \beta_\epsilon, \alpha_1, \beta_1 \) which are hyperparameters. In terms of the posterior distribution, from Equation (5), \( Y|\theta, \sigma_\epsilon^2, \sigma_u^2 \approx N_{NT}(F\theta, \lambda) \). Then, the likelihood function is expressed as

\[
L(Y|\theta, \sigma_\epsilon^2, \sigma_u^2) = \prod_{i=1}^{N} \left( 2\pi \right)^{-T/2} \exp \left[ -\frac{1}{2} \left( Y - F\hat{\theta} \right)^T \lambda^{-1} \left( Y - F\hat{\theta} \right) \right]. \]  

(7)

Thus, the deduction form of the conditional and marginal posterior distributions is

\[
\pi_1 \left( \sigma_\epsilon^2 | \theta, \sigma_u^2, Y \right) \propto \left( \frac{1}{\sigma_\epsilon^2} \right)^{\frac{N}{2}} \exp \left[ -\frac{1}{2} \left( Y - F\hat{\theta} \right)^T \frac{1}{\sigma_u^2 + \sigma_\epsilon^2} \left( Y - F\hat{\theta} \right) + \sigma_\epsilon^2 \right]. \]  

(8)

Additionally, the model section method called Deviance Information Criterion (DIC) summarizes the fit of a model by posterior expectation of the deviance, \( D \), and the complexity of a model by its effective number of parameters, \( p_D \) [13]. The definition of DIC can be expressed as:

\[
DIC = \overline{D} + p_D, \]  

(9)

where the deviance is minus twice a logarithm of the likelihood, \( D(\theta) = -2 \log \ p(\sigma_u | \theta) \), \( \overline{D} \) is the expected deviance,

\[
\overline{D} = E_{\theta|\sigma_u}[D(\theta)], \]  

(10)

and \( p_D \) is different between \( \overline{D} \) and the deviance evaluated at a particular point estimate, \( D(\hat{\theta}) \). A complication arises when models are defined hierarchically. In hierarchical model, there is a hidden parameter \( \phi \). This is the marginal posterior distributions as in:

\[
D(\theta) = -2 \log \int \ p(\pi_1 | \phi) p(\phi | \theta) d\phi. \]  

(11)

Table 3 shows that some contributions could provide details for confirming spatially computational analyses crucially needed to estimate panel data, especially agricultural economics. The first reason is the result of model selection by using deviance information criterion (DIC). The result was obvious. The DIC value of the fixed effect model (1328.288) was lower than value of the pool model (1329.244). This
means that the regional dummy factors (d1 to d6), which refer to six regional zones in Thailand, should be included and considered as affecting factors of rice production in the fixed effect model. It also conforms to descriptive information which stated that the entire agricultural areas in Thailand were not chosen for economic rice production.

**Table 3:** Model selection between the pool and fixed effect panel regressions based on Bayesian approaches by comparing deviance information criterion (DIC)

|                  | pool panel regression | fixed effect panel regression |
|------------------|-----------------------|------------------------------|
| DIC (Deviance Information Criterion) | 1329.244              | 1328.288                     |
| Log ML            | -676.981              | -685.870                     |

In Table 4, the second reason is that spatial computation in the fixed effect panel model provides a better parametric estimation. On the other hand, a comparison of MCMC standard errors (MCSE) between the pool and fixed effect models is evident. MCSE of parametric factors included in the fixed effect panel model is interestingly lower than their counterparts. This implies that spatial variables in the fixed effect model could be efficient key to sensibly explain estimated results. The third rational is a comparison of acceptance rates. Interestingly, the rate of the fixed effect model is lower than the pool panel estimation. This means that simulation convergences estimated by the fixed panel model (13.84%) is faster than their counterparts (22.41%). The simulation is based on the same conditions, which are 12,500 MCMC iterations, 2,500 burn-in samples, and 10,000 simulated observations.

Considering coefficient outcomes, the regional panel estimation indicates that northeastern, central, western, and southern region areas are potential to economic rice growing. The parametric means are 0.8770, 0.3783, 0.5335, and 2.0969, respectively. On the other hand, northern and eastern areas are presented as negative. These regions seem to be unsuitable for rice planting due to the condition of land sizes and land conditions (alkaline soil, mountains, etc.). Moreover, the parametric details of endogenous factors are also interesting. Numbers of rice farming families and rains do not guarantee mass and efficient rice products. The parametric means are negative, which are -0.0175 and -0.1161, respectively. Conversely, planted areas (0.5061) and fertilizers (0.4908) are main positive effects for improving rice products.

**Table 4:** A comparison between the pool and fixed effect panel regressions based on Bayesian approaches

|                  | Means for pool panel regression (Std. Dev./MCSE*) | Means for fixed effect panel regression (Std. Dev./MCSE*) |
|------------------|---------------------------------------------------|---------------------------------------------------------|
| (LN function) Crop areas | 0.6132 (0.1777/0.0222)               | 0.5061 (0.1597/0.0083)                                  |
| (LN function) Fertilizer usages | 0.4204 (0.1240/0.0150)                | 0.4908 (0.1136/0.0063)                                  |
| (LN function) Farmer households | -0.0548 (0.0911/0.0079)         | -0.0175 (0.0847/0.0040)                                  |
| (LN function) Precipitation | -0.0993 (0.0793/0.0124)               | -0.1161 (0.8970/0.0067)                                  |
| Dependent variable | Rice production                           | Rice production                                        |
| Prior             | Uniform prior(normal (0,1))               | Uniform prior(normal (0,1))                             |
| Log-marginal likelihood | -676.981                                  | -685.870                                                |
| Burn-in           | 2500                                    | 2500                                                    |
| MCMC sample sizes | 10000                                   | 10000                                                   |
| Number of observations | 684                                      | 684                                                     |
| d1 (Northern region) | -0.5873 (0.4987/0.0824)                  | -0.5383                                                 |
| d2 (Northeastern region) | 0.8770                                  | 0.8770                                                   |
| d3 (Central region) | 0.5383 (0.5605/0.1476)                   | 0.3783                                                   |
| d4 (Eastern region) | -0.5558 (0.5044/0.0757)                  | -0.5558                                                 |
| d5 (Western region) | 0.5335                                  | 0.5335                                                   |
d6 (Southern region)  
(0.3993/0.0532)  
2.0969  
(0.5265/0.0529)

Note: * implies MCMC standard errors (MCSE).

3. Conclusions
This paper presents the highlight issue of econometric panel estimation which totally employed Bayesian approaches to clarify the factors affecting rice products in Thailand. One of the interesting points is the need of spatial estimation. Applying Bayesian panel regression analysis, the pool model stood for the estimation without regional dummy considerations. Conversely, the fixed-effect panel model was mentioned as regional panel estimation. Five dummy factors which are major regions were included into the model. Empirically, the comparison presented the result which highly recommended that regional panel estimation should be necessary. The result estimated by the fixed effect model also represented efficient MCSE rather than the pool model.

With the spatial consideration by fixed-effect panel estimation, the results implied that intensive rice farming areas should be politically mentioned. If we focus on the planting area size, the northeastern region would be the highlighted point for intensive rice farming zones. After considering fertilizer usages, chemical contamination controls should be practically implemented. Additionally, water management and skillful labors for rice production should be concerned as the national long-term contribution.

References
[1] Wannapan S, Chaiboonsri C, and Sriboonchitta S. 2018 Identification of the connection between tourism demand and economic growth in ASEAN-3. *International Journal of Trade and Global Markets* 11 pp. 12-20.
[2] Wannapan S, Chaiboonsri C, and Sriboonchitta S. 2018 Macro-econometric forecasting for during periods of economic cycle using bayesian extreme value optimization algorithm. Springer, Cham pp. 706-723.
[3] Wannapan S, and Chaiboonsri C. 2018 The frontier of estimator comparison between MLE and MEboot estimation: application for optimization management of macroeconomics. *Journal of Physics: Conference Series* 1039 pp. 12-27.
[4] Romyen A, Wannapan S, and Chaiboonsri C. 2019 Bayesian extreme value optimization algorithm: application to forecast the rubber futures in futures exchange markets. International Conference of the Thailand Econometrics Society. Springer, Cham pp. 582-595.
[5] Morawetz U. 2006 Bayesian modelling of panel data with individual effects applied to simulated data. Discussion paper DP-16-2006. Institute of Sustainable Economic Development, University of Natural Resources and Applied Live Science, Vienna.
[6] Liu J, Sickles R C, and Tsionas E G. 2017 Bayesian treatments for panel data stochastic frontier models with time varying heterogeneity. *Econometrics* 5 pp. 1-21.
[7] Moral-Benito E. 2007 Determinants of economic growth: a Bayesian panel data approach. CEMFI Working Paper No. 0719. CEMFI. Casado del Alisal 5; 28014 Madrid.
[8] Takaishi T 2010 Bayesian inference with an adaptive proposal density for GARCH models *Journal of Physics: Conference Series* 221
[9] Kass RE and Raftery AE 1995 Bayes factors *Journal of the American Statistical Association* 90 pp. 773-795
[10] Chen CWS, Chen SY and Lee S 2013 Bayesian unit root test in double threshold heteroskedastic models. *Comput Econ* 42 pp. 471-490
[11] Said SE and Dickey D 1984 Testing for unit roots in autoregressive moving-average models with unknown order *Biometrika* 71 pp. 599–607
[12] Mohaisen A J, and Abdulsamad S Y. 2017 Bayesian panel data model based on Markov Chain Monte Carlo. *Mathematical Theory and Modeling* 7 pp. 12-24.
[13] Francois O, and Laval G. 2011 Deviance information criteria for model selection in approximate Bayesian computation. *Statistical Applications in Genetics and Molecular Biology* 10 pp. 1-25.
Appendix

Figure 1: rice products information from 17 provinces in the northern region of Thailand

Figure 2: rice products information from 19 provinces in the northeastern region of Thailand

Figure 3: rice products information from 11 provinces in the central region of Thailand

Figure 4: rice products information from 8 provinces in the eastern region of Thailand

Figure 5: rice products information from 7 provinces in the western region of Thailand

Figure 6: rice products information from 14 provinces in the southern region of Thailand