Building CPI Forecasting Model by Combining the Smooth Transition Regression Model and Mining Association Rules

Do Van Thanh[1], Cu Thu Thuy[2], Pham Thi Thu Trang[1]
[1] National Center for Socio-Economic Information and Forecast,
Ministry of Plan and Investment
Email: hieuthanhdo@yahoo.com, trang_p3t@yahoo.com
[2] Faculty of Economic Information System,
Academy of Finance, Ha Noi, Viet Nam
Email: cuthuthuy@hvtc.edu.vn

***Abstract*: Inflation forecast plays a very important role for stabilizing the economy. In Vietnam, inflation is measured via consumer price index (CPI). CPI's changes depend on many factors in which the merchandises' price changes are direct factors and those changes are not difficult to observe.**

The aim of our research is to propose a CPI forecasting model based on the change of merchandise pricing since such a model has not been built so far. A comprehensive study has been carried out to understand the effects of price changes of merchandises on CPI. After that Nonlinear Smooth Transition Regression Model and Mining Association Rules are applied to build the model. The model parameters were configured and justified using actual data collected in two years 2008-2009. The results showed the accuracy of the model for CPI forecast in Vietnam and the model can also be used to predict the price changes of merchandises.

Keywords: CPI Forecasting Model, Association Rules, Nonlinear Smooth Transition Regression.

I. INTRODUCTION

In 2008, the inflation rate in Vietnam was very high, merchandise prices changed irregularly. The Government had to introduce many economic and monetary policies to stabilize merchandise prices and to restrain the inflation. Although the inflation rate was restrained in 2009, it is possible to increase highly in 2010. Hence it is essential and urgent to build inflation forecasting models for the economy.

In general, the GDP Price Index (IGDP) is used to measure the inflation status of the economy. However, the Consumer Price Index (CPI), the Producer Price Index (PPI) or the WholeSale Price Index (WPI),... can also be used as well. Forecasting models for these indicators in different countries are very different even though they were built using the same method. Nowadays there are many methods to build inflation forecasting models such as using leading indicators [2,14], the time series model [3,9,14-15], or the structural econometric model [6,11,14],…

The use of smooth transition models, as means of representing deterministic structural change in a time series model, has been considered in [12,13]. These models allow the possibility of a smooth transition between two different trend paths over time.

The OECD (Organization for Economic Co-operation Development) countries use the smooth transition models to build inflation forecasting models for CPI, where CPI is considered in economic relations with some other socio-economic indicators such as GDP growth rate, unemployment rate, exchanges rate, import and export price indexes,…[6,10]. Smooth transition analysis was used to endogenously determine the transition path in the trend of price series. This specifies a speed of transition and the midpoint of the dynamic process between two monetary policy regimes [10,11].

In Vietnam, inflation is evaluated via CPI and the CPI forecasting models are in fact the inflation forecasting models. So far, main social-economic factors effecting the formations and changes of CPI are determined under economic theories. In year 2008, an assumption has been raised up by some famous economic researchers that there should be an existance of many hidden economic relations. These relations can be mined in a real dataset by using techiques of data mining, however they cannot be explained by the current economic theories. For CPI, a question has risen: which merchandises' price changes affect the most the CPI and how exactly are these effects? Until now, this question has not been answered in the economic theories.

The purpose of our research is to provide the answer for this question. We will propose an approach of applying the mining association rules on the real datasets of merchandise prices and CPI to find out the hidden relations between CPI and merchandise prices. The nonlinear smooth transition regression model is then used to analyze quantitatively the correlations between CPI and merchandise prices, and forecast the CPI.

The approach of building CPI forecasting model in this paper is very different from the previous inflation forecasting models for CPI. It is a combination of the mining association rules in Information Technology and the nonlinear smooth transition regression (STR) model in economics.

The mining association rules were cited for the first time in 1993 [1,16]. It was applied very succesfully in many fields such as commerce, finance, monetary, security, science research, medicine, bioinformatics.... In this paper, mined association rules provide new relations which have not yet been known between CPI and merchandise prices.

The STR model can be considered as a hybrid one of nonlinear econometric and time series models. Its goal is to analyze and forecast nonlinear economic phenomena. It has been showed that the forecast accuracy of the nonlinear smooth transition models is higher than the other models such as the Autoregressive Moving Average Integrated model (ARIMA) or the Autoregressive Conditional Heteroscedastic model (ARCH),…[14,15]. Building forecast models based on the STR model could be implemented by using the tool JMULTI [9, 18]. It can be said that JMULTI is the first Open – Source Software supporting for building forecast models based on the STR model.

Dataset for building CPI forecasting models includes CPI, the pricing of some main export and import merchandises, and some major essential merchandises for living.

The rest of the paper is structured as follows: Section 2 presents briefly the theoretical bachground of Minning Association Rules and STR. Section 3 described the datasets used in this study and the methods to deal with missing data and transform the dataset into a binary dataset. In Section 4, we present mining association rules concerning CPI. CPI forecasting model based on the minning association rules and the smooth transition regression model is shown in Section 5. Conclusion is given in the last Section.

II. MINING ASSOCIATION RULES AND THE SMOOTH TRANSITION REGRESSION MODEL

A. Association Rules

An important task in data mining is the discovery of association rules. The aim of association rule mining is to identify the relationships between items in very large datasets [1,16]. Let $\mathbf{I} = \{i_1, i_2,..., i_m\}$ be the universe of items, and **D** be the set of transactions where each transaction T is a set of items such that $T \subseteq \mathbf{I}$. Let A be a set of items. Transaction T is said to contain A if and only if $A \subseteq T$. The number (or percentage) of transactions in **D** containing A is said to be the support of A, supp(A). An association rule is an implication of the form $A \rightarrow B$, where $A \subseteq \mathbf{I}$, $B \subseteq$

I, and A ∩ B = ϕ. A is referred to as an antecedent of the rule and B as a consequent.

Support and confidence are two terms associated with association rules. The support of the rule is given as supp(AB) (meaning the probability of transaction containing both A and B). The confidence of the rule is given as conf(A →B) = supp(AB)/supp(A) (it means the conditional probability that a transaction contains B, given that it contains A).

An association rule mining problem is broken into two sub-problems: (1) Find all the item sets whose support is greater or equal to a user-determined minimum support. Such item sets are called frequent item sets, and (2) For each frequent item set found, generate all association rules that satisfy a user-determined minimum confidence. The second sub-problem can be solved in a straightforward manner when all frequent item sets and their support are known. In the problem of mining association rules, the first sub-problem is most complicated and difficult.

B. Tool for Mining Association Rules

We applied the CBA software [17] to mine association rules in binary datasets. CBA is a data mining tool built at School of Computing, National University of Singapore. An association rule mined in the CBA software is in a format:

A_1 =Y, ..., A_n = Y → B_1 =Y, ..., B_m = Y (Cover%, Conf%, CoverCount, SupCount, Sup%)

where A_i, B_j are merchandise codes, A_i = Y means A_i's price was changed. The meaning of 5 parameters of the association rule Cover%, Conf%, CoverCount, SupCount, Sup% is as follows: The first value Cover% is a percentage of the weeks that satisfy the conditions A_1 =Y, ..., A_n = Y in the dataset. The third number CoverCount shows the number of weeks in the dataset can satisfy the conditions A_1 =Y, ..., A_n = Y. Hence, Cover% = CoverCount/Total weeks in the dataset (or total transactions in the dataset). The fourth number, SupCount, shows the number of weeks satisfying both conditions A_1 =Y, ..., A_n = Y and B_1 =Y, ..., B_m = Y. The second value is the confidence (Conf%) of this rule. The confidence is calculated by (SupCount/Cover Count)*100. The last value, Sup%, shows the percentage of the total transactions that satisfy both conditions and conclusions. It can be calculated by (SupCount/Total transactions)* 100.

C. The Smooth Transition Regression Model

In our approach, the smooth transition regression model is used to build CPI forecasting models. It is a nonlinear regression model. The standard STR model is defined as follows [13,15]:

$$\begin{aligned} y_t &= \phi' Z_t + \theta' Z_t G(\gamma, c, s_t) + u_t \\ &= \{\phi + \theta G(\gamma, c, s_t)\}' Z_t + u_t, \quad t = 1,..., T. \end{aligned} \tag{1}$$

where $Z_t = (W_t' X_t')'$ is a vector of explanatory variables, $W_t' = (1, y_{t-1}, ..., y_{t-p})'$, and $X_t = (x_{1t}, ..., x_{kt})'$ is a vector of exogenous variables. Furthermore, $\phi = (\phi_0, \phi_1, ..., \phi_m)'$ and $\theta = (\theta_0, \theta_1, ..., \theta_m)'$ are parameter vectors and $u_i \sim$ iid $(0, \sigma^2)$. Transition function $G(\gamma, c, s_t)$ is a bounded function of the continuous transition variable s_t, continuous everywhere in the parameter space for any value of s_t, γ is the slope parameter, and $c = (c_1, ..., c_K)'$ is a vector of location parameters, $c_1 \le ... \le c_K$.

The STR is called Logistic Smooth Transition Regression Model (LSTR) if the transition function *G()* is given of a form:

$$G(\gamma, c, s_t) = \left[1 + \exp\left\{-\gamma \prod_{k=1}^{K} (s_t - c_k)\right\}\right]^{-1}, \gamma > 0 \tag{2}$$

The most common choices for K are K=1 and K= 2.

In the case of the Exponential Smooth Transition Regression Model (ESTR) the transition function is given as follows:

$$G_E(\gamma, c, s_t) = 1 - \exp\{-\gamma (s_t - c_1^*)^2\}, \gamma > 0 \tag{3}$$

This function is symmetric around $s_t = c_1^*$.

In practice, in general the transition variable s_t is a stochastic variable and belongs to Z_t. It can also be a linear combination of several variables. In some cases, it can be a difference of an element of Z_t. A special case, $s_t = t$, yields a linear model with deterministically changing parameters.

When X_t is absent from (1) and $s_t = y_{t-d}$ or $s_t = \Delta y_{t-d}$, d>0 (Δ is the difference of y_{t-d}), the STR model becomes a univariable smooth transition autoregressive model.

D. The Modeling Cycle

A modeling cycle for the STR model consists of three stages: specification, estimation, and evaluation.

- *Model specification*

The specification stage includes two phases. First, the starting point is subjected to linearity tests, and then the type of STR model (ESTR or LSTR, LSTR1 or LSTR2) is selected. Economic theory may give an idea of which variables should be included in the linear model. However, this may not be helpful in specifying the dynamic structure of the model. The linear specification, including the dynamics, in that case may be obtained by various model selection techniques.

The purpose of linearity tests is twofold. First, they are used to test linearity against different directions in the parameter space. If no rejections to the null hypothesis occur, we accept the linear model and do not proceed with the STR model. Second, the test results are used for model selection. If the null hypothesis is rejected for at least one of the variables, the variable with the strongest rejection of linearity (measured in the p-value) is chosen as the transition variable. The next step is to choose the transition function and to estimate the STR model. The available choices are K= 1 and K= 2 in (2). In practice the chosen STR models are LSTR1 or LSTR2.

- *Estimation of Parameters*

The parameters of the STR model are estimated using conditional maximum likelihood. Finding good starting values for the algorithm is very important. One way of obtaining them is the following: When γ and c in the transition function (2) are fixed, the STR model is linear in parameters. This suggestion will help construct a grid. Then estimate the remaining parameters ϕ and θ conditionally on (γ, c_1) for K =1 or (γ, c_1, c_2) for K= 2. Compute the sum of squared residuals and repeat this process for N combinations of these parameters. Select the parameter values that minimize the sum of squared residuals.

Once good starting values have been found, the unknown parameters c, γ, θ, ϕ can be estimated by using a form of the Newton-Raphson algorithm to maximize the conditional maximum likelihood function [9,15].

- *Model Evaluation*

The procedure to evaluate and test the STR model is as follows:

Test of no error autocorrelation: The test consists of regressing the residual $\tilde{u}_t$ of the estimated STR model on the lagged residuals $\tilde{u}_{t-1},...,\tilde{u}_{t-q}$ and the partial derivatives of the log-likelihood function with respect to the parameters of the model evaluated at the maximizing value.

Test of no additive nonlinearity: After a STR model has been fitted to the data, it is important to ask whether there are some nonlinearities remaining un-modeled by applying testing of no additive nonlinearity. In the STR framework, a natural alternative to consider in this context is an additive STR model. It can be defined as follows:

$$y_t = \phi' z_t + \theta' z_t G(\gamma_1, c_1, s_{1t}) + \Psi' z_t H(\gamma_2, c_2, s_{2t}) + u_t \quad (4)$$

where $H(\gamma_2, c_2, s_{2t})$ is another transition function

of the equation type (2) and $\varepsilon_t \sim$ iid $N(0,\sigma^2)$. Then the null hypothesis with no additive nonlinearity can be defined as $\gamma_2 = 0$ in (4).

Test of parameter constancy: In the economic relation described by the model, parameter non-constancy may indicate misspecification of the model or change over the time. So parameter constancy is one of the hypotheses that have to be tested before the estimated model can be used for forecasting. The parameter constancy allows smooth continuous change in parameters.

Other tests: Although the tests discussed above may be the most obvious ones to use when an estimated STR model is evaluated, other tests may also be useful, e.g. to test the null hypothesis of no Autoregressive Conditional Heteroscedastic Model (ARCH). Applied to macroeconomic equations, most of these tests may be conveniently regarded as general misspecification tests. However, such tests cannot be expected to be very powerful against misspecification in the conditional means. The Lomnicki-Jarque-Bera normality test is also available here. It is sensitive to outliers, and the result should be considered jointly with a visual examination of the residuals.

E. Tool for Building Price Forecasting Models Based on the STR

The software used in this study for building the STR model is JMULTI [18]. It is an interactive software for economic analysis. JMULTI can be used for building multiple time series, analyzing and forecasting models such as the Autoregressive Conditional Heteroscedastic Model (ARCH), the Autoregressive Integrated Moving Average Model (ARIMA), the Nonlinear Smooth Transition Regression Model (STR), the Vector Autoregressive Model (VAR), or the Vector Error Correction Model (VECM), etc.

F. Process for Building CPI Forecasting Models

The process is implemented in two stages**.** The first stage involves mine association rules that present price changing correlations of merchandises and CPI. These correlations, in general, are not introduced in current economic theories. In this paper they are discovered by mining association rules in a real dataset.

The real dataset includes the price of merchandises, collected weekly, and CPI, collected monthly, from 3 Jan 2008 to 31 Dec 2009. In order to mine the association rules, we have to deal with some missing and error data on the real dataset first. The data set was transformed into a transactional dataset with negation. Association rules mined from such transactional datasets are also called association rules with negation [7]. These rules were introduced as follows: Assume $\bar{I} = \{\bar{i}_1, \bar{i}_2, ..., \bar{i}_j, ..., \bar{i}_n\}$ is a set of negational items in the set of items **I** above, where $\bar{i}_j$ is defined as a negational item of i_j. $\bar{i}_j$ implies that the item i_j must be absent in the transactional database **D**. Then associaton rules with negation are in the form A $\rightarrow$ B, where $A = A_1 \cup A_2$ and $B = B_1 \cup B_2$; A_1, B_1 $\subset$ **I** and A_2, $B_2 \subset \bar{I}$ [7]. Although there are some important researching results related to mining association rules with negation, there is no available algorithm for mining them completely and effectively. Association rules mined in this paper are ones with negation. It implies that in this case, we used a technique to transform the problem of mining association rules with negation to one of mining association rules from transactional datasets.

The second stage is to build CPI forecasting models based on the smooth transition regression model and the mined relations from the first stage. A support tool for implementing the modeling cycle is the softwate JMULTI mentioned before. Many hypothesis and statistical tests have been applied in the second stage, their details can be found in [9,13-15].

For every association rule, where its consequent includes only one item CPI, we can build a forecasting model for CPI from the price of merchandises belonging to the rule's antecedent. Since many

Table 1. Absolute error of forecasted CPI compared to the statistical CPI

Month	Week	Weekly CPI			Monthly CPI		
		Forecasted CPI	*Statistical CPI*	*% of absolute error*	*Forecasted CPI*	*Statistical CPI*	*% of absolute error*
Nov. 2009	95	100.47	100.48	0.0112%	100.51	100.55	0.04 %
	96	100.62	100.68	0.0640%			
	97	100.50	100.57	0.0678%			
	98	100.45	100.47	0.0196%			
Dec.2009	99	100.50	100.62	0.1221%	101.342	101.380	0.039 %
	100	100.88	100.98	0.1011%			
	101	101.60	101.46	0.1370%			
	102	101.80	101.87	0.0645%			
	103	101.93	101.97	0.0405%			

association rules have been found in which their consequent includes only the item CPI , thus many CPI forecasting models can be built. However, these models are built by the same method. We will present briefly the process of building one of these models and implementing test forecast for that model.

III. DATASET FOR BUILDING CPI FORECASTING MODELS

A. Dataset for Merchandise Prices

Merchandise prices were collected weekly in two years, 2008 and 2009. Prices of main export and import merchandises were collected from the Customs Office and they are the weekly average values. Prices of essential merchandises for living were collected in Hanoi from 3 Jan 2008 to 31 Dec 2009 on Monday, Wednesday and Friday. The average value of these three days' prices is considered the weekly price.

By analyzing the collected dataset, we find that the price fluctuation of some merchandises is very small or their prices change only once every several months (includes 14 merchandises that their price are stabilized by the Government). We deleted these merchandises from the studying scope. The prices of all merchandises in the studying scope were collected in the duration of 103 weeks from 3 Jan 2008 to 31 Dec 2009.

The CPI is used to evaluate the inflation levels of the Vietnamese economy. In our data, the CPI is collected monthly, while the prices of other merchandises are collected weekly. To overcome the differences in the granularities of these 2 datasets we have to estimate the CPI values for the missing weeks. The following method was applied:

- If the CPI of a current month is higher (lower) than the previous month and lower (higher) than the next month, then the CPI-s of 4 weeks in that month are estimated using linear trend (decreasing or increasing).

- If the CPI of a current month is higher (lower) than both of the adjacent months, then the CPI-s of 4 weeks in that month are estimated using increasing (decreasing) trend for the first 2 weeks and in decreasing (increasing) trend for the remaining 2 weeks.

In fact, the estimates for weekly CPI-s presented above are very close to the real situation of CPI fluctuation in Viet Nam.

For each merchandise we attached a code to make our study and analysis more simple. As the result, we have a data set of 121 merchandises (CPI is also considered as a merchandise). In the dataset, there are 13 export merchandises (coded from XA1 to XA9 and from XB1 to XB4), 16 import merchandises (coded

```
, XA52, XA61, XA71, XA82, XA91, XB21, XB31, XB42, CPI2
, WA32, WA41, WA61, WA72, WA82, WA91, WB12, WB21, CPI1
, NB62, XA12, XA21, XA42, XA52, XA61, XA71, XA81, XA92, XB21, XB31, XB41, CPI1
, NB32, NB42, NB51, NB71, XA32, XA41, XA52, XA61, XA71, XA81, XA91, XB11, XB22, XB32,
, NB51, XA12, XA21, XA32, XA42, XA51, XA62, XA71, XA81, XA92, XB22, XB31, XB41, CPI1
, XA42, XB32, XB41, CPI1
, XA12, XA22, XA32, XA41, XA51, XA61, XA71, XA82, XA92, XB22, XB32, XB42, CPI2
, NB52, NB62, NB71, XA32, XA42, XA52, XA61, XA71, XA81, XA91, XB11, XB21, XB31, XB41,
, NB52, XA41, XA51, XA62, XA72, XA81, XA91, XB22, XB31, CPI2

, XA72, XA82, XA92, XB22, XB32, XB42, CPI2
, NB72, XA11, XA21, CPI2
, NB12, NB22, NB42, NB51, XA11, XA21, XA32, XA41, XA51, XA61, XA71, XA81, XA92, XB21,
, XB32, XB41, CPI2
, XB42, CPI1
, NB51, NB72, XA11, XA21, XA41, XA51, XA61, XA71, XA81, XA92, XB21, XB31, XB42, CPI1
, NB62, XA32, XA41, XA52, XA61, XA72, XA82, XA91, XB11, XB21, XB31, XB41, CPI1
, XA41, XA51, XA61, XA71, XA81, XA91, XB22, XB32, XB41, CPI1
, NB22, NB52, XA11, XA21, XA32, XA42, XB32, CPI1
, NB72, XA11, XA22, XA42, XA52, XA61, XA71, XA81, XA92, XB21, XB31, XB42, CPI2
, NB62, XA32, XA41, XA52, XA62, XA72, XA81, XA91, XB12, XB21, XB31, XB41, CPI2
, NB11, NB42, NB52, NB62, XA11, XA21, XA41, XA51, XA61, XA71, XA81, XA91, XB22, XB41,
, XA32, XA42, XB32, XB41, CPI2
```

Figure 1. Samples of the dataset used in the study

from NA1 to NA9 and from NB1 to NB7), 80 essential merchandises for living (coded from DA1 to DA9, from DB1 to DB9, ..., from DK1 to DK9) and CPI.

B. Transform the Dataset to the Binary Dataset

Association rules mined in our research are binary. They illustrate the correlations between price changes of merchandises and CPI's change. To mine such rules, the dataset needs to be formatted in the binary form. This new dataset is created from the original dataset as followings: If a merchandise's price in a current week is higher than one in the previous week (price increased), value "1" is added in the right of its code; value "2" is added if the price is lower (price decreased). For example, DA2 is the code for Rice then DA21 indicates that in current week the price of Rice is higher than the previous week. A part of the binary dataset is presented in Figure 1.

IV. CORRELATIONS BETWEEN PRICE CHANGES OF MERCHANDISES AND CPI CHANGE

Using the CBA Software for the binary dataset with minSupp = 10%, minConf = 90% , 214 associations rules were mined. Among them there are 12 rules whose consequent includes only CPI. These rules are the following:

Rule 92:

XB41 = Y, XA81 = Y, NA31 = Y, NB12 = Y

→ CPI1 = Y (11.765% 91.67% 12 11 10.784%)

Rule 93:

XB41 = Y, XA81 = Y, NB12 = Y

→ CPI1 = Y (13.725% 92.86% 14 13 12.745%)

Rule 102:

XA92 = Y, XA71 = Y, NB62 = Y

→ CPI1 = Y (11.765% 91.67% 12 11 10.784%)

Rule 118:

DB12 = Y, XA21 = Y, XA32 = Y

→ CPI2 = Y (11.765% 91.67% 12 11 10.784%)

Rule 124:

XA62 = Y, XA82 = Y, XA52 = Y

→ CPI2 = Y (11.765% 91.67% 12 11 10.784%)

Rule 165:

XA92 = Y, XA81 = Y, XA21 = Y, XA71 = Y

→ CPI1 = Y (12.745% 92.31% 13 12 11.765%)

Rule 169:

NB31 = Y, XA21 = Y, XA71 = Y,

→ CPI1 = Y (13.725% 92.86% 14 13 12.745%)

Rule 174:

XA62 = Y, XA91 = Y

→ CPI2 = Y (11.765% 91.67% 12 11 10.784%)

Rule 181:

XA92 = Y, XA81 = Y, XA21 = Y, XB21 = Y

→ CPI1 = Y (11.765% 91.67% 12 11 10.784%)

Rule 195:

NB31 = Y, XA51 = Y, XA11 = Y

→ CPI1 = Y (11.765% 91.67% 12 11 10.784%)

Rule 203:

DK61 = Y, XA41 = Y, NB21 = Y

→ CPI1 = Y (11.765% 91.67% 12 11 10.784%)

Rule 205:

XB41 = Y, XA81 = Y, XA21 = Y

→ CPI1 = Y (12.745% 92.31% 13 12 11.765%)

There are 9 rules where CPI increases and 3 remaining rules where CPI decreases. Here, most mined association rules are the ones with negations. It is still unclear what the real meaning of the relations presented in the mined is.

We can also discover CPI changing signs from the price changing signs of some merchandises in a few mixed groups. This includes import, export, and essential merchandises. These groups contain merchandises with increasing prices while others have decreasing prices.

V. BUILDING CPI FORECASTING MODELS

A. Building CPI forecasting models

The abovementioned mined rules indicate the correlations of some merchandises price and the CPI. In fact, these correlations mainly show the qualitative relations. We can not see how much the price changes of these merchandises effect the change of CPI. Our goal, however, is not only to forecast the CPI changing behaviors, but also to analyze the affects of changes of merchandises prices on the CPI.

Here after we briefly present the process to build a CPI forecasting model using one of the mined association rules. Other CPI forecasting models can be implemented in the same way with the remaining mined association rules.

Suppose that we need to build a CPI forecasting model from the following association rule:

Rule 93

XB41 = Y, XA81 = Y, NB12 = Y

→ CPI1 = Y (13.725% 92.86% 14 13 12.745%)

This rule presents the relation between CPI and the import price of American cotton type 1 (NB1), the export prices of SVR rubber type 1 (XA8) and of Shrimp type 20-30 shrimps per kilo (XB4). It also shows that there are 14 of 103 weeks (13.725% of the total weeks of year 2008 and 2009), in which the import price of NB1 decreases while the export prices of XA8 and of XB4 increase. There are only 13 in the 14 weeks (12.7455% of the total weeks) where the import price of NB1 decreases while the export prices of XA8 and of XB4 and CPI increase together. In other words, the support of this Rule is 12.745%. Rule 93 has the confidence value of 92.86%, i.e. when the import price of American cotton type 1 decreases, the export prices of SVR rubber type 1 and of Shrimp type 20-30 shrimps per kilo increase then CPI will increase with a confidence at least 92.86%.

In order to build the forecasting model for CPI from the import price of American cotton type 1 (NB1), the export prices of SVR rubber type 1 (XA8) and of Shrimp type 20-30 shrimps per kilo (XB4), the original dataset of CPI and prices of NB1, XA8 and XB4 are divided into two sub-datasets. The first

dataset, containing first 94 weeks of year 2008 and 2009, is used to build a forecasting model for CPI. The second dataset of 9 remaining weeks, which are the weeks of November and December 2009, will be used later for the verification of the model.

In the first stage of the modeling cycle, by applying the unit root test provided by the JMULTI software on the time series of CPI, XA8, XB4 and NB1, we found that the time series CPI, XA8 and NB1 are not stationary while XB4 is. However, the differences order 1 of these time series are all tested to be stationary. Hence, we choose to build the forecast model for the difference order 1 of CPI (noted as CPI_d1) from the differences order 1 of the time series XA8, XB4 and NB1 (noted as XA8_d1, XB4_d1, and NB1_d1, respectively). The linearity test results indicates that the type of the model for CPI_dl in this case is LSTR1, the selected smooth transition variable is CPI_d1(t-3) and the maximum lag number of the dependent variable CPI_d1 and the independent variables XA8_d1, XB4_d1, NB1_d1 are the same and equal to 4.

In the second stage of the modeling cycle, we estimated the parameters of the model and the results are presented in Figure 2. It shows:

p-values of the t-statistic for all independent variables are smaller than 0.1. This implies that all the variables in both linear and nonlinear parts of the model have the significance level being more than 90%.

The variables XA8_d1(t), XB4_d1(t) as well as their lags such as XA8_d1(t-1), XA8_d1(t-2), XA8_d1(t-3), XA8_d1(t-4),… do not effect the change of CPI_d1(t).

The variable NB1_d1(t-4) and lagged variables of CPI_d1 such as CPI_d1(t-1), CPI_d1(t-2), CPI_d1(t-3) effect strongly and directly the change of CPI_d1(t).

R2 = 4.9696e-01 and adjusted R2 = 0.5026 show that the independent variables in the linear and

variable	start	estimate	SD	t-stat	p-value
----- linear part ------					
CONST	-13.86256	-5.99704	3.2616	-1.8387	0.0698
CPI_d1(t-1)	-5.78085	-7.09577	4.1723	-1.7007	0.0930
CPI_d1(t-2)	5.48318	7.34688	4.0032	1.8353	0.0703
CPI_d1(t-3)	-10.31479	-6.26734	3.2103	-1.9522	0.0545
NB1_d1(t-4)	-0.01966	-0.01908	0.0093	-2.0548	0.0433
---- nonlinear part ----					
CONST	14.35961	6.04024	3.2552	1.8555	0.0673
CPI_d1(t-1)	6.29465	7.45941	4.1772	1.7858	0.0781
CPI_d1(t-2)	-5.39052	-7.13244	4.0042	-1.7812	0.0788
CPI_d1(t-3)	9.15263	5.58218	3.2195	1.7338	0.0869
NB1_d1(t-4)	0.01947	0.01840	0.0093	1.9862	0.0506
Gamma	0.92928	2.85916	0.0000	0.0000	0.0009
C1	-1.34000	-0.80295	0.0000	-0.0000	0.0000

AIC:	-2.5306e+00
SC:	-2.1951e+00
HQ:	-2.3954e+00
R2:	4.9696e-01
adjusted R2:	0.5026
variance of transition variable:	0.1354
SD of transition variable:	0.3680
variance of residuals:	0.0703

Figure 2. Estimated parameters of the model

nonlinear parts explained about 50% the changes of the dependent variable CPI_d1(t).

The forecasting model for CPI_d1 can be presented as follows:

$$CPI\_d1(t) = \left\{ \begin{matrix} -5.997 - 7.096\,CPI\_d1(t-1) + 7.347\,CPI\_d1(t-2) \\ -6.267\,CPI\_d1(t-3) - NB1\_d1(t-4) \end{matrix} \right\} + \frac{\left\{ \begin{matrix} 6.04 + 7.46\,CPI\_d1(t-1) - 7.132\,CPI\_d1(t-2) \\ +5.582\,CPI\_d1(t-3) + 0.018\,NB1\_d1(t-4) \end{matrix} \right\}}{1 + \exp\{-2.86\,(CPI\_d1(t-3) + 0.803)\}}$$

The linear part of this forecasting model shows that the changes of CPI_d1(t) and CPI_d1(t-2) are in the same direction but in the opposite direction with the changes of CPI_d1(t-1), CPI_d1(t-3), CPI_d1(t-4) and NB1_d1(t-4).

The nonlinear part is the product of two components. The first component is the autoregressive part. It is rather similar with the linear part but the coefficient signs of the independent variables are opposite. The second component with logistic function and smooth transition function is PCI_d1(t-3). Its location parameter is -0.803 and the slope parameter is 2.86. The nonlinear part shows two different changing regions of CPI_d1(t), before and after the value - 0.803, where the transition between two regions is very smooth.

In the third stage of the modeling cycle, several tests were applied to examine the built model. Testing results showed that the forecasting model for CPI_d1 has no error autocorrelation, no additive nonlinearity, and no parameter constancy. The next step is to evaluate how accurate the model is in the forecasting of the future CPI.

B. Testing the forecasting model

The second dataset is used for this purpose. Using the model CPI_d1 is calculated with t = 95, 96, …, 103 (the weeks of collected data in the second set), then CPI(t) is determined from CPI-d1(t). The comparison between the estimated CPI and the real CPI is shown in Table 1. As seen in the table, the absolute errors for both weekly and monthly CPI are very low. It implies that the proposed forecasting model is very accurate and can be used to forecast the CPI in Vietnam.

C. Priori Forecast

It is very interesting, and very special in the proposed model above, that all independent variables are lagged dependent variable CPI_d1 and lagged variable NB1_d1. It means that in order to forecast CPI (dependent variable) at a time t, there is no need to forecast any independent variables in this model. In other words, no other models need to forecast the independent variables. To forecast CPI(t) we only need calculate CPI_d1(t) from the defined values such as CPI_d1(t-1), CPI_d1(t-2), CPI_d1(t-3), CPI_d1(t-4) and NB1_d1(t-4).

VI. CONCLUSION

In recent years, application of the mining association rules as well as the smooth transition regression model takes much interest, especially in the filed of Information Technology and Economics. In this paper, a new approach for CPI ecasting model is proposed using mining association rules and smooth transition regression model.

The goal of mining association rules is to detect the hidden relations between the price changes of some merchandises and the CPI. These relations have not been introduced in the economic theories so far. They suggest a new approach in inflation research, though they are mainly qualitative relations. The support of mined association rules is not very high and it is natural, but its confidence is very high. This implies that the correlations of price changes, detected by association rules, are very strong and clear. The forecasting models for CPI are built by applying the smooth transition regression model on the detected relations.

The model was applied in a set of real data of CPU and merchandises prices collected in Vietnam. The results showed that it is very accurate to forecast

Vietnamese CPI. However, it is necessary to adjust the model parameters frequently. The proposed approach can also be used to forecast merchandises price changes as well.

It should be noted that for each mined relation, we can build a forecasting model for CPI. But each different model will provide a different forecast for CPI. Then, a problem of combining several forecasts arises. However, we did not address this issue in current study. This issue has also attracted much attension from many economists and seems to be a challenge for future research.

REFERENCES

[1] Agrawal R., Mannila H., Srikant R., Toivonen H., "Fast Discovery of Association Rules", Advances in Knowledge discovery and DataMining, edited by U.M. fayyad, G.Platstsky-Shapiro, P.Smyth, and Uthurusamy, AAAI Press/The MIT Press, 1996, pp.306-328.

[2] Ang A., Bekaert G., Wei M., "Do macro variables, asset markets, or surveys forecast inflation better ?". Journal of Monetary Economics, Vol 54, 2007, pp. 1163- 1212

[3] Boris Kovalerchuk, and Evgenii Vityaev, "Data mining in finance – Advances in relational and hybrid methods", Kluwer Academic publisher, 2001.

[4] Cu Thu Thuy, and Do Van Thanh, "New Approach for Analysing Viet Nam Stock Market", Computer and Cybernatic, Tom 24 , N2, pp. 107-118, 2008.

[5] Eitrheim and et al., "Testing the adequacy of smooth transition autoregressive models," Journal of Econometrics, Vol. 74, 1996, pp. 59-75.

[6] Gregoriou,A., Kontonikas A. and Montagnoli, A.; " Euro area inflation differentials: Unit roots, Structural break and non-linear adjustment". Andros.gregoriou@brunel. ac.uk., 2006.

[7] Kryszkiewicz M.,and Cichon K., "Support Oriented Discovery of Generalized Disjunction-Free Representation of Frequent Patterns with Negation", PAKDD 2005, LNAI 3518, pp. 672-682, 2005.

[8] Lucas R. E., "Expectations and the Neutrality of Money", Journal of Economic Theory. Vol. 4, 1972 , pp. 103-124.

[9] Nguyen Khac Minh, "Theoretical Foundation of Nonlinearn Time Series and Application for building inflation models of Viet Nam, Time Series models and application for analyzing inflation", Lectute Document of EU Technical Assistant Program for Viet Nam, 3/ 2009.

[10] Stock J. H., and Watson M. W., "Forecasting inflation," Working Paper 7023, National Bureau of Economic Research, USA, Cambridge Press , 1999.

[11] Stock J.H., and Watson M.W., "Phillips curve inflation forecasts", Working Paper 14322, National Bureau of Economic Research, USA, Cambridge Press, 2008.

[12] Teräsvirta T., and Anderson H. M., "Characterizing Nonlinearities in Businccess Cycles Using Smooth Transition Autoregressive Models", Journal of Applied Econometrics, Vol.7, 1992, pp. 119 - 136.

[13] Teräsvirta T., "Specification, estimation, and evaluation of smooth transition autoregressive models", Journal of American Statistical Association, Vol. 89, 1994, pp. 150-189.

[14] Teräsvirta T. et al., "Linear models, smooth transition auto regression, and neural networks for forecasting macroeconomic time series: A re-examination", International Journal of Forecasting Vol. 21, 2005, pp.755-774.

[15] Teräsvirta T. "Smooth Transition Regression Modeling", Applied Time Series Econometrics, Cambridge University Press, 2007.

[16] Zaki M. J., and Ogihara M., "Theoretical Foundation of Association Rules", In 3rd ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, June 1998.

[17] CBA Software in http://www.nus.com

[18] JMULTI Open – Source Software in www.JMULTI.de

AUTHORS' BIOGRAPHIES

Do Van Thanh received BS and MS degrees in Mathematics at the National Pedagogical University at Ha Noi in 1977 and 1979 respectively and worked as a full time university lecturer and researcher in Mathematics. Since 1989 he has been worked as a Computer Science researcher. He received Doctoral degree from Institute for Information Technology, National Institute of Sciences and Technology in Vietnam. From 2004 he also has been working as an economic researcher. His research interests include: State Administration Computerization, Knowledge databases, Automated reasoning, Data mining and Socio-Economic Analysis and Forecast.

Cu Thu Thuy received BS degree in Mathematics at Hanoi National Pedagogical University in 1993. Since 1994, she has been working as a full time university lecturer at the Faculty of Economic Information System – Hanoi Academy of Finance. She received MS degree in Information Technology at Vietnam National University in Hanoi. She is now a PhD-student at College of Technology, Vietnam National University at Ha Noi.

Pham Thi Thu Trang received BS and MS degrees in Mathematical Economic at Vietnam National Econimics University at Ha Noi in 2003 and 2006 respectively. She has been working as a researcher at the Department of Economical Analysis and Forecast, National Center for Social-Economics Information and Forecasts (NCSEIF), Vietnam. Her research interests are Analysis and Forecast Economic.