FORECASTING MILK PRICES USING MILK SUPPLY AND INFLATION

Aysegül Tercan

105622006

İSTANBUL BİLGİ UNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

EKONOMİ YÜKSEK LİSANS PROGRAMI

TEZ DANIŞMANI: DOÇ. DR. M. EGE YAZGAN

2009

FORECASTING MILK PRICES USING MILK SUPPLY AND INFLATION

SÜT MİKTARI VE ENFLASYON İLE SÜT FİYAT TAHMİNİ

Aysegül Tercan

105622006

İMZA

EGE YAZGAN ....
KORAY AKAY ....
ORHAN ERDEM ....

Tezin Onaylandığı Tarih : Şubat 2009

Toplam Sayfa Sayısı: 39

Anahtar Kelimeler (Türkçe)
1) Tahmin
2) Süt fiyatı
3) Enflasyon
4) süt miktarı
5) karşılaştırma

Anahtar Kelimeler (İngilizce)
1) Forecast
2) Milk price
3) Inflation
4) Milk supply
5) Compared

ABSTRACT

We study the forecasting of milk prices in Turkey using milk supply and inflation. Three different models are proposed: Vector Autoregressive Model, Vector Error Correction Model and an Autoregressive Model. The standard approach to evaluate performance of forecast models are comparison of root mean square error (RMSE) and mean absolute percentage error (MAPE) values. We conclude that an autoregressive model has a better forecasting performance compared to the other two models in terms of both MAPE and RMSE criterion.

ÖZET

Bu çalışmada, Türkiye'deki süt miktarını ve enflasyonu kullanarak süt fiyatlarını inceledik ve gelecek için tahmin yaptık. Tahminler Vektör Otoregresif Model(VAR), Vektör Hata Düzeltme Modeli (VECM) ve Otoregresif Model (AR) olmak üzere üç farklı model ile yapıldı. Modeller arasında kıyaslama genel olarak Karesel Ortalama Hata (RMSE) ve Ortalama Mutlak Yüzde Sapma (MAPE) kriterleri üzerinden yapılmaktadır. Bu kıyaslamalar sonucunda Otoregresif modelinin diğer modellere göre hem MAPE hem de RMSE kriterlerine göre daha iyi performans gösterdiği sonucuna vardık.

Appendix

1. INTRODUCTION

Forecasting prices of raw materials is an important and challenging area of research for both academia and business world. Obtaining accurate forecasts of prices will enable companies to have better forecasts of their profits, hence to take precautionary action if a negative financial situation is expected. This concept becomes more crucial as the raw material becomes more expensive since in this case, the profits are more sensitive to variance in prices. Having observed a high fluctuation in historical monthly milk prices, it is clear that a more accurate forecast of milk prices will highly contribute to the financial decision process of companies producing dairy products.

Estimating the price relationships and analyzing the nature of price changes from supplier to producer is an important tool to gain insight along the competition in food markets. Bakucs and Fertö (2006) state that two forces determine the access and price of the raw materials in food market: supply and demand. The growth in supply of raw materials depends on the weather conditions, which can be explained by the concept of seasonality. If due to seasonality, the supply of raw materials is low, resulting scarcity will drive the price of raw materials upward. Hence, the supply of materials constitutes a crucial variable in the estimation of material prices.

The purpose of this study is to estimate the dynamics of the milk prices in a given time period using milk supply and inflation and to choose the best performing model among Vector Autoregressive Model, Vector Error Correction Model and an Autoregressive Model in terms of MAPE and RMSE criteria.

The data used in this study is from January 1979 to September 2007. The regressions are performed using monthly data from Turkish milk market. We obtain results in favor of

Autoregressive Model compared to the other two models in terms of forecasting performance.

The remainder of paper is organized as follows. In Section 2, we present a brief review on literature on models forecasting the raw material prices. In Section 3, we describe forecasting models in detail: Vector autoregressive model, vector error correction model and basic autoregressive model. In Section 4, we explain the data on Turkish milk market and provide a careful analysis of the stationarity properties of the data. Empirical results of the forecasting models are compared in Section 5. We conclude in Section 6.

2. LITERATURE REVIEW

There is a growing literature on the dynamics of raw material prices. Bakucs and Fertö (2006) examine the dynamics of the marketing margin on the Hungarian beef market. Using monthly data from January 1992 to March 2000 for exogeneity tests, they conclude that causality runs from producer to retail prices, despite the common belief that price transmission on the Hungarian beef market is symmetric on both short and long run. They basically employ vector auto regression model for price forecasts.

Ayadi (2005) focuses on the relationship between oil prices and levels of economic activity. With a data period of 1980-2004, he concludes that oil producing countries experience an increase in wealth when oil prices rise due to the effect of oil price change on industrial production index. He also employs a unique model, although different to Bakucs and Fertö (2006), namely the vector auto regression model.

Shin & Sohn (2006) propose a new approach that combines Generalized Auto Regressive Conditional Heteroscedasticity (GARCH), neural network and random walk models on a weight that reflects the inverse of the exponentially weighted moving average of the Mean Absolute Percentage Error (MAPE) of each model. They focus on predicting currency exchange rates using daily exchange rates between January 1999 and February 2003. They apply their model to the prediction of the exchange rates between the Korean won and Japanese yen. Consistent with our study, they compare results using MAPE and conclude that the proposed model performs better than the GARCH, neural networks and random walk.

In their study, Gronewold, Guoping and Anping (2006) examine the regional spillover of production in China using a Vector Autoregressive Model. They try to find out the flow direction and the amount of a shock to GDP in different regions. Working on a set of data from 1953 to 2003, they find evidence on strong spillovers from the coast region to both regions but shocks to the western region prove to have no flow-on effect on the other two regions. As in the study of Ayadi (2005), they employ a VAR model for forecasting prices.

Despite the abundant literature on the forecast of material prices, literature on forecasting milk prices is not yet developed. Moreover, the comparison of price forecast models is not a well-explored subject in the literature. This paper constitutes a unique study by offering a comparison of various models of forecasting milk prices.

3. FORECASTING MODELS

In this study, we analyze three different forecasting models: Vector autoregressive model, vector error correction model and an autoregressive model. Our aim is choosing the best performing model among these three models by comparing their forecast accuracies. Two commonly used criteria are employed to evaluate and compare the performance of the models. The first criterion is root mean square error (RMSE).

Assuming $\hat{y}_k$ is the prediction variable of y_k :

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{k=1}^{N}(y_k-\hat{y}_k)^2}$$

The second performance indicator is mean absolute percentage error (MAPE). Using the same notation, MAPE is defined as:

$$\text{MAPE}=\frac{1}{N}\sum_{k=1}^{N}\left|\frac{y_k-\hat{y}_k}{y_k}\right|x100\%$$

3.1. Vector Autoregressive Model

The first model is Vector Autoregressive Model, adopted from the multivariate modeling of Sims (1980). VAR model has been advocated by Sims (1980) as a way to estimate dynamic relationship among jointly endogenous variables without imposing strong a priori restrictions such as exogeneity of the variables (Harris, 1995). As Enders (1996) states, VAR analysis serves for determining the interrelationships among variables.

Let x_t be an AR (p) process:

$$x_t = m + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ... + \beta_p x_{t-p} + \varepsilon_t$$

Consider a column vector with k different variables, $x_t = [x_{1t} x_{2t} .... x_{kt}]'$ where model is expressed in its past time values. This is called vector autoregression, namely; VAR. For a positive integer p, VAR (p) is denoted as:

$$x_t = m + B_1 x_{t-1} + B_2 x_{t-2} + ... + B_p x_{t-p} + \varepsilon_t$$

where B_i are k x k matrices of coefficients, m is a k x 1 vector of constants and ε_t is a vector white noise process with the following properties.

$$\text{E}(\varepsilon_t) = 0 \quad \text{for all t}$$

$$\text{E}(\varepsilon_t \varepsilon_s') = \begin{cases} \Omega \_\_\_ s = t \\ 0 \_\_\_ s \neq t \end{cases}$$

where Ω is the covariance matrix and is positive definite, implying that ε 's are not correlated (Johnson and Dinardo, 1997).

For a simpler case, let $\{r_t\}$ be affected by current and past realizations of $\{z_t\}$ sequence and visa versa. Consider a bivariate system:

$$r_t = a_{10} - a_{12} z_t + \gamma_{11} r_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{rt}$$

$$z_t = a_{20} - a_{21} r_t + \gamma_{21} r_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt}$$

With the underlying assumptions:

(1) r_t and z_t are stationary

(2) ε_{rt} and ε_{zt} are white-noise disturbances with deviations of σ_r and σ_z

(3) $\{\varepsilon_{rt}\}$ and $\{\varepsilon_{zt}\}$ are uncorrelated white noise disturbances. (Enders, 2006)

Using matrix algebra, we can denote these equations in one as follows.

$$\begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix} \begin{bmatrix} r_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} r_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{rt} \\ \varepsilon_{zt} \end{bmatrix}$$

Let B= $\begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix}$, $y_t = \begin{bmatrix} r_t \\ z_t \end{bmatrix}$, $\Gamma_0 = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix}$, $\Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$, $\varepsilon_t = \begin{bmatrix} \varepsilon_{rt} \\ \varepsilon_{zt} \end{bmatrix}$.

It is clear that

$$\mathrm{B}\, y_t = \Gamma_0 + \Gamma_1 y_{t-1} + e_t$$

Multiplying both sides by B^{-1} , we obtain the VAR model in the standard form:

$$y_t = A_0 + A_1 y_{t-1} + e_t$$

where $A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1$ and $e_t = B^{-1}\varepsilon_t$.

3.2. Vector Error Correction Model

The underlying concept in a Vector Error Correction Model is cointegration of regression variables. Let ρ denote the vector $(\rho_1, \rho_2, \ldots\ldots, \rho_n)$ and x_t denote the vector $(x_{1t}, x_{2t}, \ldots\ldots, x_{nt})'$. Assume that these variables are in long-run equilibrium when

$$\rho_1 x_{1t} + \rho_2 x_{2t} + \ldots\ldots + \rho_n x_{nt} = 0$$

Or as in matrix notation, when $\rho\, x_t = 0$. The deviation from long-run equation is denoted as e_t, where

$$e_t = \rho x_t$$

The components of the vector $x_t = (x_{1t}, x_{2t}, \ldots, x_{nt})'$ are cointegrated of order d, b, denoted as $x_t \sim CI(d,b)$, if the following properties are satisfied.

i. All components of x_t are integrated of order d

ii. There is a vector $\rho = (\rho_1, \rho_2, \ldots\ldots, \rho_n)$ such that the linear combination $\rho x_t = \rho_1 x_{1t} + \rho_2 x_{2t} + \ldots\ldots + \rho_n x_{nt}$ is integrated of order (d-b) where b>0.

The vector ρ is called the cointegration vector (Enders, 2006).

For vector auto regression model, let z_t be a vector of n endogenous variables. Define an n x 1 vector auto regression model for z_t as follows.

$$z_t = B_1 z_{t-1} + \ldots + B_k z_{t-k} + u_t \qquad\qquad u_t \sim \text{IN}\ (0, \Sigma\ )$$

where each B_i is an (n x n) matrix. The vector auto regression equation can be reformulated into a vector error-correction form as below.

$$\Delta z_t = \Gamma_1 \Delta z_{t-1} + \ldots + \Gamma_{k-1} \Delta z_{t-k+1} + \Pi z_{t-k} + u_t$$

where $\Gamma_i = -(I - B_1 - \ldots - B_i)$, (i=1,...,k-1) , and $\Pi = -(I - B_1 - \ldots - B_k)$.

Both long-run and short-run information are included in this equation by the estimates of $\overline{\Gamma}_i$ and $\overline{\Pi}$. $\Pi = \alpha\beta'$ where α represents the speed of adjustment and β is a matrix of long-run coefficients (Harris, 1995).

3.3. Autoregressive Model

Autoregressive model is a time series model which is based on using own historical time series values to forecast the future time series values of a variable. Let $Y_1, Y_2, \ldots\ldots, Y_n$ be a time series. For a dependent variable Y_k let an estimated regression equation relating Y_k to the three most recent time series values $Y_{k-1}, Y_{k-2}, Y_{k-3}, \ldots$ be:

$$\hat{Y}_k = a_0 + a_1 Y_{k-1} + a_2 Y_{k-2} + a_3 Y_{k-3} + \ldots$$

where $a_0, a_1, a_3, \ldots$ are coefficients and $\hat{Y}$ is the estimated value. (Williams, 1996) Regression models, in which the independent variables are past realizations of the dependent variable, are called autoregressive models.

First-order autoregression is denoted by AR (1) and stated as follows.

$$Y_t = c + \varphi Y_{t-1} + \varepsilon_t \qquad\qquad \forall \ |\varphi| \geq 1$$

where $\{\varepsilon_t\}$ is a sequence of white noise. Note that E (ε_t) =0 and E ($\varepsilon_t{}^2$) = σ^2 , i.e. $\{\varepsilon_t\}$ has zero mean and a variance equal to σ^2 (Hamilton, 1994).

4. THE DATA

4.1. Data Analysis

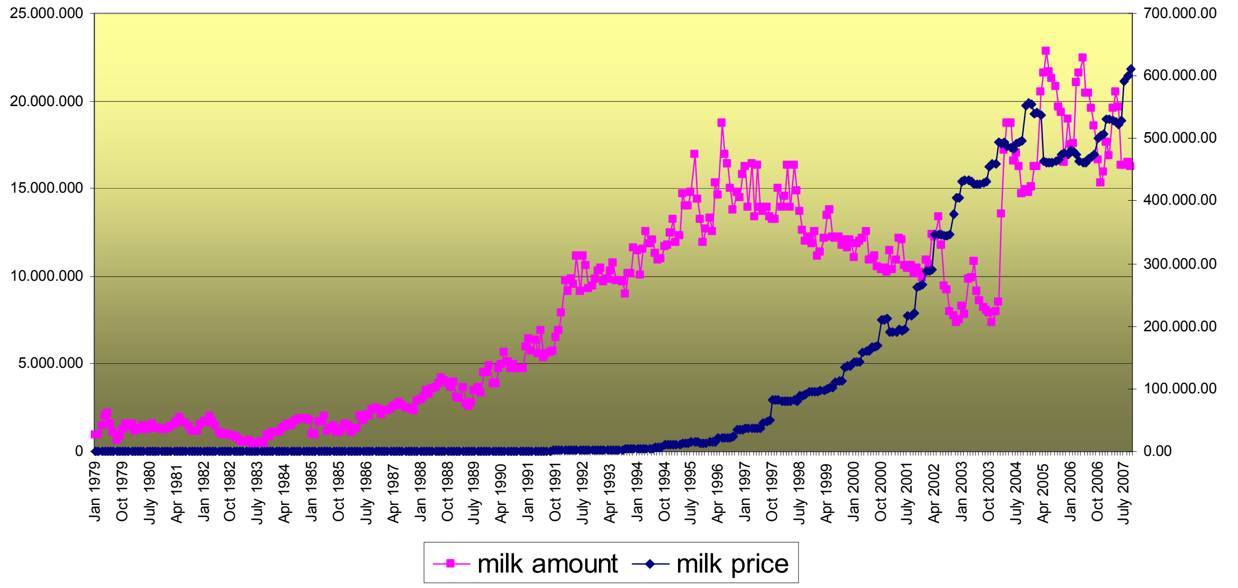
The data for this study is provided by the database of Ac-Nielsen. Ac-Nielsen is a company that collects data from the market including the variables such as volume, market share, prices and numeric distribution of products, in order to inform the market players about the market conditions, dynamics and trends. Sampling is designed to yield data that is representative of the population and the market including the data all over the country.

The country of study is Turkey, a developing country where dairy products are consumed in large amounts. The yearly raw milk supply in Turkey reached 227.071.636 tons by the end of 2007. The monthly average of raw milk supply is around 18.922.636 tons. Turkey is a growing market in dairy products; the total market value of the raw milk increased by an amount of % 69 during the period 2000-2007. Ac-Nielsen collects data over seven regions of Turkey on the volume and price of milk on a monthly basis.

Loy and Weaver (2006) mention that the only way of working with products that have seasonality, such as milk, is using monthly data. Following their advice, we do analysis on a monthly basis for the period 1979-2007. Below are the tables analyzing the time series of the variables. The variable volume is the total amount of milk supply and the variable price is the average price of one kilogram of milk in that specific month. This average price includes all the relevant sale prices, on a wide range from the largest multinationals to the smallest regional shops. The variable inflation data is the logarithm of the monthly CPI (Consumer Price Index) data.

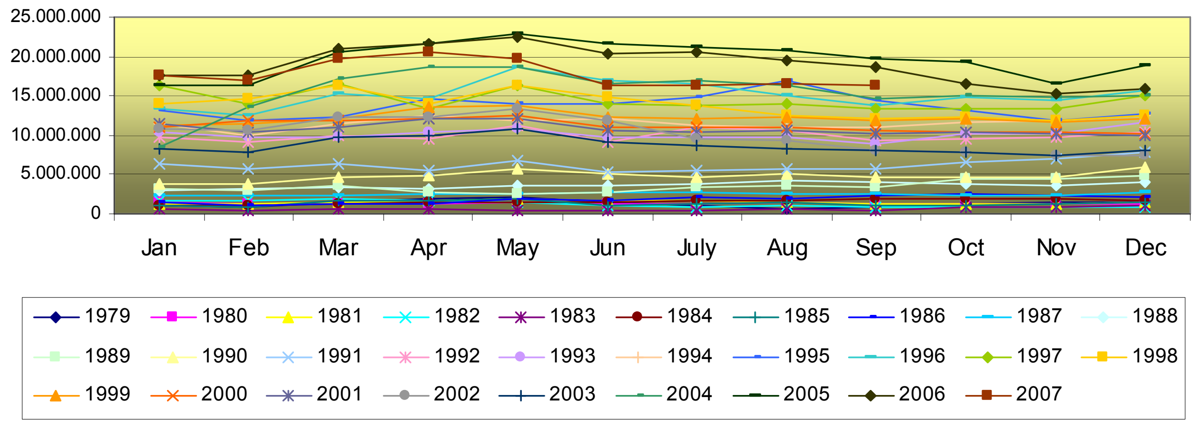
Following the literature, we analyze milk amounts and milk prices in logarithms. Hence, our final variables are denoted as log_amount, log_price and inflation (as inflation is already denoted in logarithm). The differences of these logarithmic variables are denoted as; diff_lgamount, difflgprice and diff_inf, respectively.

Table-4.1: The milk price&amount combinations over time



Seasonality is clearly observed in the milk volume data. The low season for milk supply starts from June and continues until March. The high season for milk supply is the month of May. In the range of our data, the minimum ratio of the price of the highest season to that of the lowest season is 0.8 (observed in 1989) whereas the highest ratio is 2.3 (observed in 1979).

Table-4.2: Milk amounts over time



As supply of milk fluctuates, so does the price of milk. The milk price starts to increase every year in June as the milk supply in the market starts to decrease. From 1979 to 2007, the ratio of milk price at the beginning of the year to that at the end of the year is at its minimum at 0.9 in 2005. The maximum of the ratio is at 2.6 in 1994.

Table-4.3: Milk prices over time

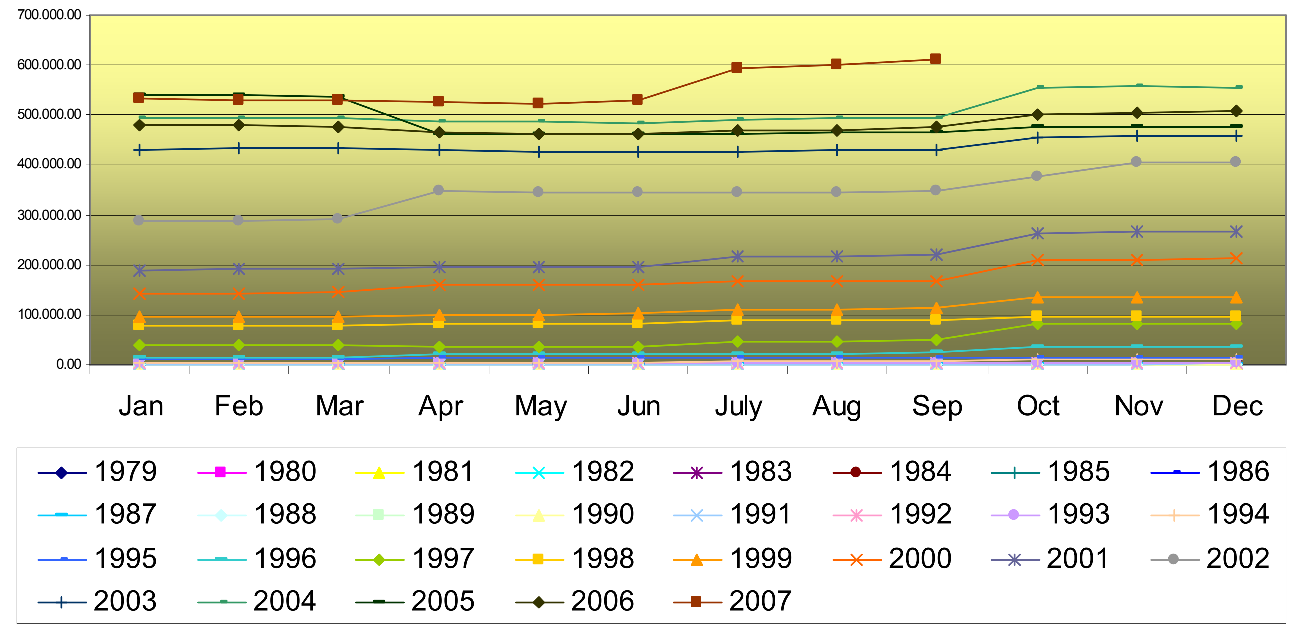


Table-4.4: Descriptive statistics of the data

Summary descriptive statistics			
Variable name	**log_amount**	**log_price**	**inflation**
Mean	6.737403	3.584142	0.28517
Median	6.962369	3.403498	0.18315
Maximum	7.359164	5.785885	2.613645
Minimum	5.554368	0.943	-2.539
Std. Dev.	0.459733	1.603213	1.681454

As shown in Table 4.4, milk amount has a mean of 6.74 and a median of 6.96 and the standard deviation is 0.46 (all numbers are rounded). Milk price has a mean of 3.58 and a median of 3.40 and the standard deviation is 1.60. The relevant statistics for inflation are 0.29, 0.18 and 1.68, respectively. The histograms of the variables are demonstrated in the Appendix 8.1.2.

4.2. Stationarity Analysis

The concept of stationarity is crucial in model building. If a model is based on past realizations of a time series, then a researcher assumes implicitly that there is some regularity in the process that generates the time series. This concept of regularity in the time series is what is referred to as stationary. In a more technical form, a stationary time series is time series that has constant mean and variance across time. If $\{y_t\}$ is a stationary series, the mean, variance, and autocorrelations can usually be well approximated by sufficiently long time averages based on the single set of realizations (Enders, 2005). If the variables are not stationary series, they will need to be differenced once. Hence, the first step of analyzing time series models is performing stationary tests.

The models containing non-stationary variables will often lead to a problem of spurious regression, whereby the results suggest that there are statistically significant relationships between the variables in the regression model when in fact all this is obtained is evidence of contemporaneous correlations rather than meaningful causal relations (Harris, 1995).

Park (1990) states that starting point of the specification of time series models is obtaining stationary series. The presence of unit root implies non-stationarity. Hence, we begin the empirical analysis with testing of unit root properties of our three basic variables: log_amount, log_price and inflation. Depending on the outcome of these tests, different model structures will be proposed.

Dickey and Fuller (1979) consider three different regression equations that can be used to test for the presence of a unit root:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$$

The first is a pure random walk model[1], the second additionally includes an intercept and the third includes both a drift and a linear trend. The parameter of interest in all equations is γ, if γ=0, the $\{y_t\}$ sequence contains a unit root. Comparing the resulting t-statistic with the appropriate value reported in the Dickey-Fuller tables allows us to determine whether to accept or reject the null hypothesis γ=0 (Enders, 2005).

For higher-order equations consider p-th order autoregressive process:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots\dots\dots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_{py} t - p + \varepsilon_t$$

For Augmented Dickey Fuller (ADF) test, add and subtract the term $a_p y_{t-p+1}$ to obtain:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots\dots\dots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} + a_p \Delta y_{t-p+1} + \varepsilon_t$$

Continue by adding and subtracting the term $(a_{p-1} + a_p) y_{t-p+2}$ and more terms to obtain:

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_T$$

where: γ=-(1-$\sum_{i=1}^{p} a_i$) and $\beta_i = -\sum_{j=i}^{p} a_j$ (Enders, 2004).

Augmented Dickey- Fuller (ADF) test is performed on all the variables. Intercept and trend terms are introduced to the ADF tests. The results indicate that null hypothesis of stationarity is rejected at 90% confidence level for all the variables. Following this, first differences of these variables are created further to be tested by the ADF test. The intercept is included whereas the trend is not included as this test is performed for the first differences. Maximum lag length is set at 16.

The unit root null hypothesis is rejected for the first differences of all three series. This implies that the first differences of the variables are stationary. We conclude that our three

[1] Random walk is demonstrated as $\varepsilon_t = \varepsilon_{t-1} + z_t$ where $z_t \sim IN(0,\sigma^2)$)

variables, log_amount, log_price and inflation, are not stationary at level however their first differences are stationary. Table 4.5 contains the results of the stationary tests.

Table-4.5: ADF test results

Augmented Dickey Fuller Test (ADF Test) Results

Variable	**prob.**	**max lag**	**lag length**	**Remarks**
log_price	0.9764	16	12	not stationary
log_amount	0.5726	16	12	not stationary
inflation	0.8643	16	13	not stationary
first differences	**prob.**	**max lag**	**lag length**	**Remarks**
diff_lgprice	0.0003	16	11	stationary
diff_lgamount	0.0000	16	12	stationary
diff_inf	0.0569	16	12	stationary

5. EMPIRICAL RESULTS

5.1. Forecast of Milk Prices

VAR :

The first multivariate model for forecasting milk prices is Vector Auto Regression model. The model is performed using 12 lags of the stationary variables diff_lgprice, diff_lgamount and diff_inf to select the lag-length. The choice of 12 lags is relies on the aim of guaranteeing the inclusion of any possible year-based trend.

Regarding selection of the lag-length of a model, there are three criteria: Akaike Information Criteria (AIC), Schwarz Criteria (SC) and Hannan Quinn Criteria (HQ). Below is the explicit formula of each criterion, where *n* is the number of parameters estimated, T is the number of usable observations (Enders, 2005) and RSS[2] is the residual sum of squares.

$$AIC = T\,ln(RSS) + 2n$$

$$SC = T\,ln(RSS) + n\,ln(T)$$

$$HQC = T\,ln\left(\frac{RSS}{T}\right) + 2n\,ln(T)$$

All these three criteria results in favor of the model which returns the lowest level of test value. Hence, lag 8, lag 1 and again lag 1 are chosen regarding AIC, SC and HQ methodologies. These selections are presented in Table-5.1, where the selected lag of each criterion is denoted by "<". We'll be using the AIC selection which is lag 8 despite the fact that SC selection and HQ selection is the first lag.

---

[2] Let $y_i = a + bx_i + \varepsilon_i$ where a and b are coefficients, y and x are the regressand and the regressor and ε is the error term. The sum of squares of residuals is the sum of squares of estimates of ε_i, that is

$$RSS = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

Table-5.1: Lag-lenght selection

Order	AIC	SC	HQ
lag 12	-7.421	-6.151	-6.914
lag 11	-7.415	-6.248	-6.950
lag 10	-7.452	-6.388	-7.028
lag 9	-7.485	-6.525	-7.102
lag 8	**-7.514<**	-6.650	-7.166
lag 7	-7.508	-6.759	-7.213
lag 6	-7.504	-6.852	-7.244
lag 5	-7.464	-6.915	-7.245
lag 4	-7.454	-7.008	-7.276
lag 3	-7.444	-7.101	-7.308
lag 2	-7.477	-7.237	-7.381
lag 1	-7.509	**-7.372<**	**-7.454<**

As lag 8 is chosen, VAR (8) will be evaluated. The estimation will be performed using data for the period January 1979 - August 2006. Forecasting will be done for the period September 2006 - September 2007(data points from 334 to 345), which is the last one year period of the data. We'll be comparing the estimated values for this period with the actual data in order to figure out the accuracy of the model. In Table-5.2 the forecasts and the quality of the forecasts of VAR from data points 334 to 345 are shown.

Table-5.2: Forecast evaluation of VAR

Horizon	Forecast	SE	Actual	Error	t-value
334	0.012165	0.03367	0.003149	-0.009016	-0.268
335	0.011594	0.03402	0.002451	-0.009143	-0.269
336	0.006676	0.0341	0.019626	0.012951	0.38
337	0.001784	0.03438	-0.000897	-0.002681	-0.078
338	0.001924	0.03475	-0.001252	-0.003175	-0.091
339	0.003779	0.03508	-0.001264	-0.005043	-0.144
340	0.004287	0.03576	-0.003209	-0.007497	-0.21
341	0.010774	0.0361	0.003891	-0.006883	-0.191
342	0.013211	0.03614	0.050041	0.03683	1.019
343	0.013076	0.03628	0.006409	-0.006666	-0.184
344	0.013195	0.03633	0.006743	-0.006452	-0.178
345	0.013756	0.0364	-0.005786	-0.005799	-0.159

RMSE =	1.674	**MAPE =**	205.07

As can be seen in table 5.2, the absolute t-values are below the critical value of 1.96, hence each forecast is significant at 95% confidence level. RMSE of this forecast is 1.674 and MAPE of this forecast is 205.07. These two criteria will be taken into consideration while comparing the models with each other.

VECM :

As mentioned in Section 3.2, the underlying concept in a Vector Error Correction Model is cointegration of regression variables. In order to have a long-run equation we apply cointegration tests to the lags of non- stationary variables log_amount, log_price and inflation, also accounting for a trend variable. Johansen trace test results are given in Table-3.1. Trace test results provide strong evidence of cointegration of rank one.

Table-3.1: Trace test results

Null hypothesis H0	Alternative hypothesis H1	Trace test result	Prob.
r=0	r>1	54.160	0.002
r≤1	r>2	22.677	0.119
r≤2	r>3	3.1542	0.848

The presence of cointegration suggests that there is a long-run equilibrium relationship between the series. So, cointegrated series have an error correction representation.

The long-run equation is defined as:

$$Log\_price=0.41171 \text{ X } log\_amount + 1.1728 \text{ X } inflation - 0.0056699 \text{ X } t + Error$$

Details about the cointegration giving out the long run equation are displayed the table in Appendix 8.2.1.

Long-run equation implies that as a result of a %1 increase in milk amount, milk price increases by 0.41%. This may be explained by the complex relations of demand and supply in the market. On the other hand, %1 increase in inflation results in a 1.17% increase in milk price. This is consistent with the existing literature on raw material prices.

From the above equation, error term is derived as:

$$Error=log\_price-0.41171*log\_amount - 1.1728*inflation + 0.0056699*t$$

Regarding vector error correction model, diff_lgprice, diff_lgamount and diff_inf of lag 7 and one period lagged errors are used as variables. The forecast is done for the data points from 334 to 345. In Table-5.3 the forecasts and the quality of the forecasts of VECM are presented.

Table-5.3: Forecast evaluation of VECM

Horizon	Forecast	SE	Actual	Error	t-value
334	0.010281	0.03313	0.003149	-0.007132	-0.215
335	0.008927	0.03334	0.002451	-0.006477	-0.194
336	0.002687	0.03342	0.019626	0.002598	0.507
337	-0.003496	0.03361	-0.000897	0.002598	0.077
338	-0.004572	0.03407	-0.001252	0.00332	0.097
339	-0.004566	0.03433	-0.001264	0.003302	0.096
340	-0.003483	0.03468	-0.003209	0.000274	0.008
341	0.00287	0.0349	0.003891	0.001021	0.029
342	0.00331	0.0349	0.050041	0.046731	1,339
343	0.000275	0.03494	0.006409	0.006134	0.176
344	-0.004001	0.03496	0.006743	0.010744	0.307
345	-0.005066	0.03498	0.578588	-0.005781	-0.165

RMSE =	1.669	**MAPE =**	156.35

As can be seen in the table, forecast value is significant at each time point. RMSE of this forecast model is 1.669 and MAPE of this forecast is 156.35. These two criteria will be taken into consideration while comparing the models with each other.

AR:

An autoregressive model AR will be used as a univariate model; AR (1) is applied to the data.

Test is issued to select the lag-length to log_price from lag 1 to lag 12. From the least squares model the lag-length selections of the three criteria's among 12 lags are shown in Table 5.4.

Table-5.4: Lag-lenght selection

Order	AIC	SC	HQ
lag 12	**-3.893<**	-3.745	-3.834
lag 11	-3.893	-3.756	-3.838
lag 10	-3.891	-3.766	-3.841
lag 9	-3.875	-3.760	-3.829
lag 8	-3.877	-3.774	-3.836
lag 7	-3.877	-3.786	-3.841
lag 6	-3.883	-3.803	-3.851
lag 5	-3.881	-3.812	-3.853
lag 4	-3.861	-3.804	-3.838
lag 3	-3.866	-3.820	-3.848
lag 2	-3.871	**-3.837<**	**-3.857<**
lag 1	-3.857	-3.834	-3.848

Among the AIC results -3.893 is the smallest value so lag 12 is chosen by AIC. Among the SC results -3.837 is the smallest value so lag 2 is chosen by SC. Among the HQ results -3.857 is the smallest value so lag 2 is chosen by HQ. We'll be using the AIC selection which is lag 12 where SC selection and HQ selection is the second lag.

Forecasting will be done for the data between 333 and 345 to be able to compare AR forecast quality with other models. In Table-5.5 the forecasts and the quality of the forecasts of AR from 334 to 345 are shown.

Table-5.5: Forecast evaluation of AR

Horizon	Forecast	SE	Actual	Error	t-value
334	5.691	0.03443	5.700	0.009055	0.263
335	5.705	0.05269	5.703	-0.00126	-0.024
336	5.719	0.06549	5.706	-0.013149	-0.201
337	5.731	0.07652	5.725	-0.005898	-0.077
338	5.743	0.08424	5.725	-0.01871	-0.222
339	5.753	0.08986	5.723	-0.029705	-0.331
340	5.764	0.09472	5.722	-0.042205	-0.446
341	5.774	0.0988	5.719	-0.054842	-0.555
342	5.784	0.1019	5.723	-0.060874	-0.597
343	5.795	0.1064	5.773	-0.022456	-0.211
344	5.806	0.1104	5.779	-0.026707	-0.242
345	5.818	0.1148	5.786	-0.031896	-0.278

RMSE =	0.031954	**MAPE =**	0.46025

For all the forecast values, the t-values are far below the critical value, hence each value is significant at 95% confidence level. RMSE of this forecast is 0.031954 and MAPE of this forecast is 0.46025. These two values will be taken into consideration while comparing the models with each other.

5.2 Comparison of Forecasting Performance

Two commonly used criteria are employed to evaluate and compare these models. The first criterion is root mean square error (RMSE). The second performance indicator is mean absolute percentage error (MAPE). Since these indicators measure the degree of deviation of forecast values from the actual time series values, the smaller the indicator values, the higher is the prediction ability of the model.

For forecasting, one year period is eliminated from the actual data and performances after the 333rd month are forecasted. Table 6.1 shows the forecasting performances of the models in terms of RMSE and MAPE.

Table-6.1: Forecasting evaluation of the models

	RMSE	MAPE
VAR	1.674	205.07
VECM	1.669	156.35
AR	0.031954	0.46025

As shown in the table 6.1, AR has the smallest RMSE and MAPE values so AR is chosen as the best performing model among three proposed forecasting models.

6. CONCLUSION

Using monthly milk price, milk supply and inflation data from Turkey for the period January 1979 – September 2006, we forecast milk prices with three different methods: Vector Autoregressive Model, Vector Error Correction Model and an Autoregressive Model. Regarding the comparison of root mean square error and mean absolute percentage error values, we conclude that Autoregressive Model achieves the highest forecasting performance.

7. REFERENCES

Ayadi, O.F. (2005), "Oil price fluctuations and Nigerian economy" Organization of Petroleum Exporting Countries, pp.199-215.

Aydın, D.(2007), "A comparison of the nonparametric regression models using smoothing spline and Kernel Regression" Internal Journal of Mathematical, Physical and Engineering Sciences, 2(2) , pp 75-79.

Bakucs, L.Z. and Fertö I. (2006), "Marketing margins and price transmissions on the Hungarian beef market" Food economics-Acta Agricult Scand C, 3, pp.151-160.

Bose, S. and Bodman, P.M. and Campbell H.F. (2006), "Estimation and significance of long-run relationships among fish prices" International Journal of Oceans and Oceanography, 1(2), pp. 167-182.

Enders, W. (2004), "Applied Econometric Time Series", Hamilton Printing, NJ

Groewold, N. and Guoping, L. and Anping, C. (2007), "Regional output spillovers in China: Estimates from a VAR model" Papers in Regional science, 86(1), pp.102-122.

Hamilton, J.D. (1994), "Time Series Analysis" Princeton University Press, Princeton

Harris, R. (1995), "Using Cointegration Analysis in Econometric Modeling", T.J. Press, Great Britain.

Johnson, J. and Dinardo, J. (1997), "Econometric Models", The McGraw-Hill Companies, Inc., New York.

Loy, J. and Weaver, R. (2006), "How do price based sales affect market basket prices? Evidence for fresh food products in German grocery stores" Food Economics-Acta Agricult Scand C, 3, pp 64-80.

Park,T. (1990), "Forecast Evaluation for Multivariate Time-Series Models: The U.S. Cattle Market" Western Journal of Agricultural Economics,15(1), pp133-143.

Shın, H. W. and Sohn, S. Y. (2007), “Application of an EWMA combining technique to the prediction of currency exchange rates” IIE Transactions, 39 , pp. 639-644.

Sims, C. (1980), “Macroeconomics and reality” Econometrica, 48, pp 1-48.

William, A.S. (1996), “Statistics for Business and Economics”, West Publishing Company Printing, St. Paul.

8. APPENDIX

8.1. Data

8.1.1. Stationary Analysis

Null Hypothesis: LOG_PRICE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 12 (Automatic based on AIC, MAXLAG=16)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.627270	0.9764
Test critical values:	1% level	-3.985941	
	5% level	-3.423418	
	10% level	-3.134664	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOG_PRICE)
Method: Least Squares
Date: 10/26/08 Time: 10:45
Sample (adjusted): 14 345
Included observations: 332 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG_PRICE(-1)	-0.006017	0.009592	-0.627270	0.5309
D(LOG_PRICE(-1))	0.149596	0.056028	2.670026	0.0080
D(LOG_PRICE(-2))	-0.030779	0.056620	-0.543609	0.5871
D(LOG_PRICE(-3))	0.011063	0.056462	0.195938	0.8448
D(LOG_PRICE(-4))	-0.113925	0.055980	-2.035112	0.0427
D(LOG_PRICE(-5))	-0.057115	0.056080	-1.018457	0.3092
D(LOG_PRICE(-6))	-0.009448	0.056031	-0.168629	0.8662
D(LOG_PRICE(-7))	-0.028216	0.055979	-0.504045	0.6146
D(LOG_PRICE(-8))	-0.059053	0.055773	-1.058795	0.2905
D(LOG_PRICE(-9))	0.163780	0.055315	2.960848	0.0033
D(LOG_PRICE(-10))	-0.087009	0.056055	-1.552188	0.1216
D(LOG_PRICE(-11))	0.053927	0.056149	0.960427	0.3376
D(LOG_PRICE(-12))	0.176375	0.055687	3.167278	0.0017
C	0.019390	0.007904	2.453263	0.0147
@TREND(1)	7.91E-05	0.000157	0.502932	0.6154

R-squared	0.127740	Mean dependent var	0.013827
Adjusted R-squared	0.089217	S.D. dependent var	0.035124
S.E. of regression	0.033520	Akaike info criterion	-3.909192
Sum squared resid	0.356188	Schwarz criterion	-3.737274
Log likelihood	663.9260	F-statistic	3.315981
Durbin-Watson stat	2.023474	Prob(F-statistic)	0.000056

Null Hypothesis: DIFF_LGPRICE has a unit root
Exogenous: Constant
Lag Length: 11 (Automatic based on AIC, MAXLAG=16)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.440494	0.0003
Test critical values:	1% level	-3.449857	
	5% level	-2.870031	
	10% level	-2.571363	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DIFF_LGPRICE)
Method: Least Squares
Date: 10/26/08 Time: 10:45
Sample (adjusted): 14 345
Included observations: 332 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DIFF_LGPRICE(-1)	-0.853691	0.192251	-4.440494	0.0000
D(DIFF_LGPRICE(-1))	0.000960	0.183592	0.005229	0.9958
D(DIFF_LGPRICE(-2))	-0.032915	0.173223	-0.190017	0.8494
D(DIFF_LGPRICE(-3))	-0.024528	0.164114	-0.149459	0.8813
D(DIFF_LGPRICE(-4))	-0.141410	0.150942	-0.936848	0.3495
D(DIFF_LGPRICE(-5))	-0.200685	0.138288	-1.451207	0.1477
D(DIFF_LGPRICE(-6))	-0.211788	0.125351	-1.689559	0.0921
D(DIFF_LGPRICE(-7))	-0.241597	0.111407	-2.168589	0.0309
D(DIFF_LGPRICE(-8))	-0.302026	0.097595	-3.094696	0.0021
D(DIFF_LGPRICE(-9))	-0.139097	0.086249	-1.612744	0.1078
D(DIFF_LGPRICE(-10))	-0.227661	0.071381	-3.189369	0.0016
D(DIFF_LGPRICE(-11))	-0.174679	0.054912	-3.181058	0.0016
C	0.011754	0.003278	3.586027	0.0004

R-squared	0.488706	Mean dependent var	8.58E-06
Adjusted R-squared	0.469472	S.D. dependent var	0.045971
S.E. of regression	0.033484	Akaike info criterion	-3.917142
Sum squared resid	0.357651	Schwarz criterion	-3.768145
Log likelihood	663.2455	F-statistic	25.40894
Durbin-Watson stat	2.022340	Prob(F-statistic)	0.000000

Null Hypothesis: LOG_AMOUNT has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 12 (Automatic based on AIC, MAXLAG=16)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.047630	0.5726
Test critical values:	1% level	-3.985941	
	5% level	-3.423418	
	10% level	-3.134664	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOG_AMOUNT)
Method: Least Squares
Date: 10/26/08 Time: 10:42
Sample (adjusted): 14 345
Included observations: 332 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG_AMOUNT(-1)	-0.033120	0.016175	-2.047630	0.0414
D(LOG_AMOUNT(-1))	-0.036191	0.055660	-0.650223	0.5160
D(LOG_AMOUNT(-2))	0.038473	0.055097	0.698281	0.4855
D(LOG_AMOUNT(-3))	-0.073635	0.055057	-1.337426	0.1820
D(LOG_AMOUNT(-4))	-0.011395	0.054628	-0.208592	0.8349
D(LOG_AMOUNT(-5))	0.018178	0.054212	0.335308	0.7376
D(LOG_AMOUNT(-6))	-0.154066	0.052975	-2.908286	0.0039
D(LOG_AMOUNT(-7))	0.032646	0.052809	0.618185	0.5369
D(LOG_AMOUNT(-8))	0.048751	0.052494	0.928706	0.3537
D(LOG_AMOUNT(-9))	0.030743	0.052313	0.587666	0.5572
D(LOG_AMOUNT(-10))	0.026532	0.050837	0.521892	0.6021
D(LOG_AMOUNT(-11))	-0.100025	0.050352	-1.986507	0.0478
D(LOG_AMOUNT(-12))	0.208730	0.050573	4.127293	0.0000
C	0.203620	0.096966	2.099910	0.0365
@TREND(1)	0.000131	7.53E-05	1.732942	0.0841

R-squared	0.140026	Mean dependent var	0.003089
Adjusted R-squared	0.102046	S.D. dependent var	0.057563
S.E. of regression	0.054547	Akaike info criterion	-2.935366
Sum squared resid	0.943207	Schwarz criterion	-2.763448

Log likelihood	502.2708	F-statistic	3.686846
Durbin-Watson stat	1.954772	Prob(F-statistic)	0.000010

Null Hypothesis: DIFF_LGAMOUNT has a unit root
Exogenous: Constant
Lag Length: 12 (Automatic based on AIC, MAXLAG=16)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.257865	0.0000
Test critical values:	1% level	-3.449917	
	5% level	-2.870057	
	10% level	-2.571377	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DIFF_LGAMOUNT)
Method: Least Squares
Date: 10/26/08 Time: 10:44
Sample (adjusted): 15 345
Included observations: 331 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DIFF_LGAMOUNT(-1)	-1.256242	0.238926	-5.257865	0.0000
D(DIFF_LGAMOUNT(-1))	0.215700	0.230043	0.937651	0.3491
D(DIFF_LGAMOUNT(-2))	0.218910	0.216785	1.009799	0.3134
D(DIFF_LGAMOUNT(-3))	0.131852	0.204868	0.643598	0.5203
D(DIFF_LGAMOUNT(-4))	0.119832	0.191976	0.624204	0.5329
D(DIFF_LGAMOUNT(-5))	0.136701	0.178286	0.766753	0.4438
D(DIFF_LGAMOUNT(-6))	-0.058246	0.162478	-0.358486	0.7202
D(DIFF_LGAMOUNT(-7))	-0.063810	0.144706	-0.440968	0.6595
D(DIFF_LGAMOUNT(-8))	-0.044001	0.127956	-0.343878	0.7312
D(DIFF_LGAMOUNT(-9))	-0.034372	0.108826	-0.315843	0.7523
D(DIFF_LGAMOUNT(-10))	-0.020816	0.088852	-0.234282	0.8149
D(DIFF_LGAMOUNT(-11))	-0.118640	0.071845	-1.651329	0.0997
D(DIFF_LGAMOUNT(-12))	0.072730	0.051292	1.417948	0.1572
C	0.004417	0.003100	1.425068	0.1551

R-squared	0.611231	Mean dependent var	0.000310
Adjusted R-squared	0.595288	S.D. dependent var	0.085414
S.E. of regression	0.054338	Akaike info criterion	-2.945813
Sum squared resid	0.935977	Schwarz criterion	-2.784998
Log likelihood	501.5321	F-statistic	38.33807
Durbin-Watson stat	1.984993	Prob(F-statistic)	0.000000

Null Hypothesis: INFLATION has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 13 (Automatic based on AIC, MAXLAG=16)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.383412	0.8643
Test critical values:	1% level	-3.986026	
	5% level	-3.423459	
	10% level	-3.134688	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(INFLATION)
Method: Least Squares
Date: 10/25/08 Time: 14:55
Sample (adjusted): 15 345
Included observations: 331 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INFLATION(-1)	-0.003863	0.002793	-1.383412	0.1675
D(INFLATION(-1))	0.376448	0.051967	7.243961	0.0000
D(INFLATION(-2))	0.014704	0.055141	0.266663	0.7899
D(INFLATION(-3))	0.026794	0.054921	0.487872	0.6260
D(INFLATION(-4))	-0.059053	0.054644	-1.080676	0.2807
D(INFLATION(-5))	0.077900	0.054504	1.429248	0.1539
D(INFLATION(-6))	0.095661	0.054680	1.749461	0.0812
D(INFLATION(-7))	0.013551	0.054811	0.247238	0.8049
D(INFLATION(-8))	0.018750	0.054601	0.343401	0.7315
D(INFLATION(-9))	0.089469	0.054468	1.642595	0.1015
D(INFLATION(-10))	-0.075683	0.054214	-1.396001	0.1637
D(INFLATION(-11))	0.134349	0.053630	2.505103	0.0127
D(INFLATION(-12))	0.224249	0.053875	4.162376	0.0000
D(INFLATION(-13))	-0.084094	0.051154	-1.643940	0.1012
C	-0.008420	0.008315	-1.012580	0.3120
@TREND(1)	6.45E-05	4.82E-05	1.340205	0.1811

R-squared	0.399277	Mean dependent var	0.014469
Adjusted R-squared	0.370671	S.D. dependent var	0.010876
S.E. of regression	0.008628	Akaike info criterion	-6.620552
Sum squared resid	0.023448	Schwarz criterion	-6.436764
Log likelihood	1111.701	F-statistic	13.95788

Durbin-Watson stat	1.918343	Prob(F-statistic)	0.000000

Null Hypothesis: DIFF_INF has a unit root
Exogenous: Constant
Lag Length: 12 (Automatic based on AIC, MAXLAG=16)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.817568	0.0569
Test critical values:	1% level	-3.449917	
	5% level	-2.870057	
	10% level	-2.571377	

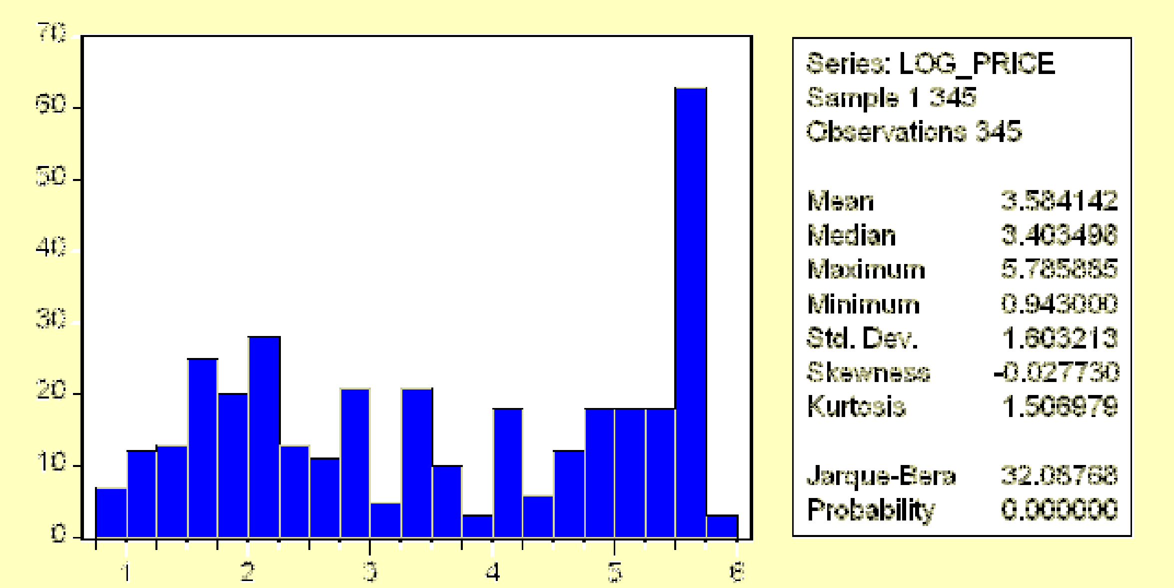
*MacKinnon (1996) one-sided p-values.

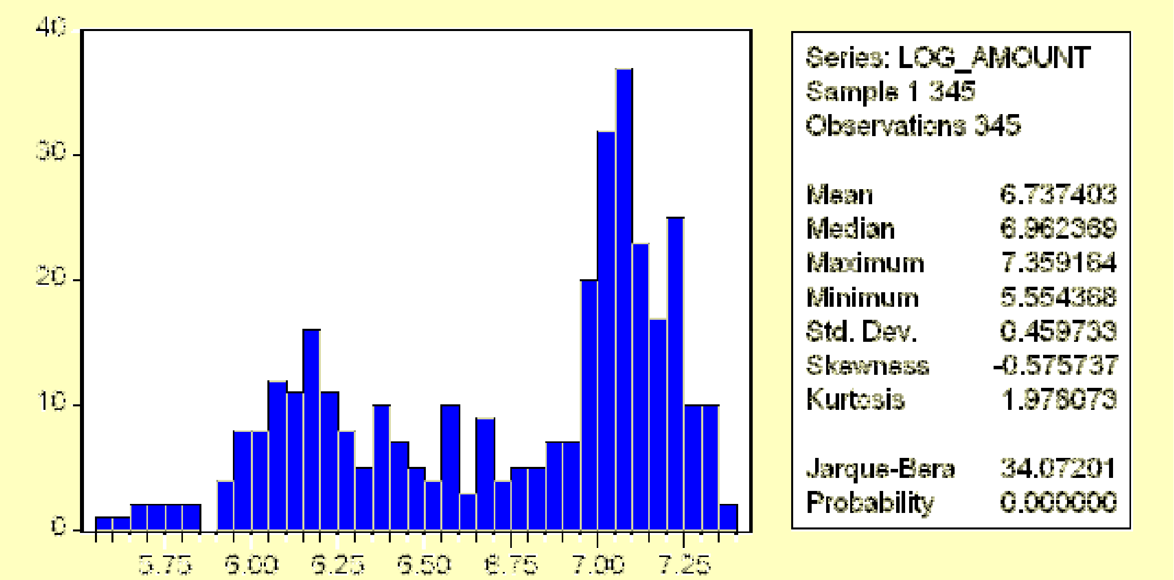
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DIFF_INF)
Method: Least Squares
Date: 10/25/08 Time: 15:04
Sample (adjusted): 15 345
Included observations: 331 after adjustments

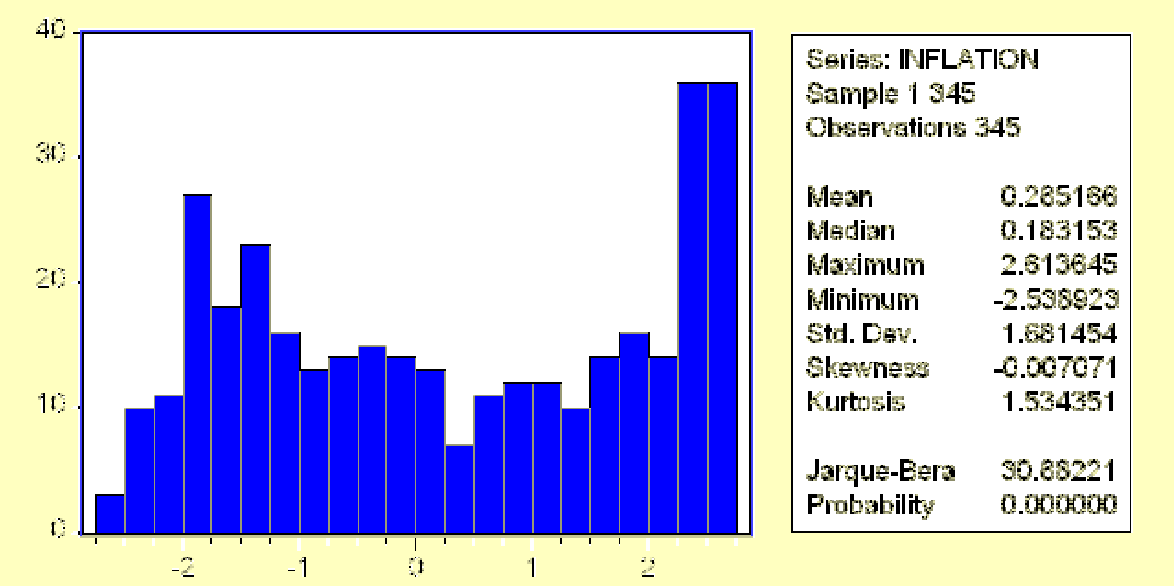
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DIFF_INF(-1)	-0.197892	0.070235	-2.817568	0.0051
D(DIFF_INF(-1))	-0.423588	0.082657	-5.124633	0.0000
D(DIFF_INF(-2))	-0.411002	0.084165	-4.883304	0.0000
D(DIFF_INF(-3))	-0.387124	0.084253	-4.594752	0.0000
D(DIFF_INF(-4))	-0.448325	0.084874	-5.282214	0.0000
D(DIFF_INF(-5))	-0.373145	0.085092	-4.385209	0.0000
D(DIFF_INF(-6))	-0.280829	0.083929	-3.346031	0.0009
D(DIFF_INF(-7))	-0.270866	0.081965	-3.304651	0.0011
D(DIFF_INF(-8))	-0.256695	0.077948	-3.293174	0.0011
D(DIFF_INF(-9))	-0.172864	0.071611	-2.413939	0.0163
D(DIFF_INF(-10))	-0.253397	0.065803	-3.850835	0.0001
D(DIFF_INF(-11))	-0.124281	0.058823	-2.112784	0.0354
D(DIFF_INF(-12))	0.094444	0.050552	1.868281	0.0626
C	0.002428	0.001165	2.083797	0.0380

R-squared	0.408424	Mean dependent var	-0.000242
Adjusted R-squared	0.384164	S.D. dependent var	0.010994
S.E. of regression	0.008628	Akaike info criterion	-6.626245
Sum squared resid	0.023598	Schwarz criterion	-6.465430
Log likelihood	1110.644	F-statistic	16.83513
Durbin-Watson stat	1.918044	Prob(F-statistic)	0.000000

8.1.2. Histograms







8.2. Empirical Results

8.2.1. Cointegration

I(1) cointegration analysis, 8 - 345

eigenvalue	loglik	for rank
	2293.172	0
0.08894	2308.914	1
0.056123	2318.675	2
0.0092885	2320.252	3

H0:rank<=	Trace test [Prob]
0	54.160 [0.002]**
1	22.677 [0.119]
2	3.1542 [0.848]

Asymptotic p-values based on: Restricted trend, unrestricted constant
Unrestricted variables:
[0] = Constant
Restricted variables:
[0] = Trend
Number of lags used in the analysis: 7

beta (scaled on diagonal; cointegrating vectors in columns)

log_price	1.000	7.255	-0.095874
log_amount	-0.41171	1,000	-0.020654
inflation	-1.1728	-7.0323	1.000
Trend	0.0056699	-0.0032313	-0.013106

alpha

log_price	-0.075996	-0.014138	-0.0065647
log_amount	0.19842	-0.0082829	-0.015353
inflation	-0.017268	0.0027519	-0.003119

long-run matrix, rank 3

	log_price	log_amount	inflation	Trend
log_price	-0.17794	0.018505	0.18199	-0.00029917
log_amount	0.13981	-0.086804	-0.18981	-0.001353
inflation	0.0029947	0.010506	-0.0022192	-6.5930

8.2.2. Forecasting Results

Output of Table-5.1:Lag-lenght selection

Progress to date

Model	T	p	log-likelihood	SC	HQ	AIC	lag
SYS(1)	333	111	1346.5053	-6.151	-6.914	-7.421	12
SYS(2)	333	102	1336.5756	-6.248	-6.950	-7.415	11
SYS(3)	333	93	1333.6839	-6.388	-7.028	-7.452	10
SYS(4)	333	84	1330.2989	-6.525	-7.102	-7.485	9
SYS(5)	333	75	1325.0942	-6.650	-7.166	-7.514<	8
SYS(6)	333	66	1316.9913	-6.759	-7.213	-7.508	7
SYS(7)	333	57	1306.3854	-6.852	-7.244	-7.504	6
SYS(8)	333	48	1290.7602	-6.915	-7.245	-7.464	5
SYS(9)	333	39	1280.0145	-7.008	-7.276	-7.454	4
SYS(10)	333	30	1269.4916	-7.101	-7.308	-7.444	3
SYS(11)	333	21	1265.8659	-7.237	-7.381	-7.477	2
SYS(12)	333	12	1262.2442	-7.372<	-7.454<	-7.509	1

Output of Table-5.2: Forecast evaluation of VAR

Dynamic (ex ante) forecasts for diff_lgprice (SE based on error variance only)

Horizon	Forecast	SE	Actual	Error	t-value
334	0.0121651	0.03367	0.00314909	-0.009016	-0.268
335	0.0115936	0.03402	0.00245061	-0.00914295	-0.269
336	0.0066756	0.0341	0.0196262	0.0129506	0.38
337	0.0017837	0.03438	-0.000897463	-0.00268115	-0.078
338	0.0019237	0.03475	-0.00125152	-0.00317518	-0.091
339	0.0037793	0.03508	-0.0012642	-0.00504348	-0.144
340	0.0042872	0.03576	-0.00320947	-0.00749667	-0.21
341	0.0107736	0.0361	0.00389097	-0.00688263	-0.191
342	0.0132105	0.03614	0.0500408	0.0368303	1,019
343	0.0130759	0.03628	0.00640943	-0.00666648	-0.184
344	0.0131952	0.03633	0.00674305	-0.00645213	-0.178
345	0.0137559	0.0364	-0.005786	-0.005799	-0.159
mean(Error)=	-0.43387		RMSE =	1674	
SD(Error) =	1,603		MAPE =	205.07	

Output of Table-5.3: Forecast evaluation of VECM

Horizon	Forecast	SE	Actual	Error	t-value
334	0.0102814	0.03313	0.00314909	-0.00713232	-0.215
335	0.00892725	0.03334	0.00245061	-0.00647664	-0.194
336	0.0026867	0.03342	0.0196262	0.00259827	0.507
337	-0.0034957	0.03361	-0.000897463	0.00259827	0.077
338	-0.0045719	0.03407	-0.00125152	0.00332041	0.097
339	-0.0045665	0.03433	-0.0012642	0.00330228	0.096
340	-0.0034832	0.03468	-0.00320947	0.00027373	0.008
341	0.00287039	0.0349	0.00389097	0.00102058	0.029
342	0.0033103	0.0349	0.0500408	0.0467305	1,339
343	0.00027501	0.03494	0.00640943	0.00613442	0.176
344	-0.004001	0.03496	0.00674305	0.010744	0.307
345	-0.0050656	0.03498	0.578588	-0.005781	-0.165
mean(Error)=	-0.47528		RMSE =	1,669	
SD(Error) =	1.6		MAPE =	156.35	

Output of Table-5.4: Lag-lenght selection of AR

Progress to date

Model	T	p	log-likelihood	SC	HQ	AIC	lag
EQ(1)	333	13	661.238960	-3.745	-3.834	**-3.893<**	12
EQ(2)	333	12	660.197430	-3.756	-3.838	-3.893	11
EQ(3)	333	11	658.917900	-3.766	-3.841	-3.891	10
EQ(4)	333	10	655.112250	-3.760	-3.829	-3.875	9
EQ(5)	333	9	654.475870	-3.774	-3.836	-3.877	8
EQ(6)	333	8	653.573560	-3.786	-3.841	-3.877	7
EQ(7)	333	7	653.539080	-3.803	-3.851	-3.883	6
EQ(8)	333	6	652.144640	-3.812	-3.853	-3.881	5
EQ(9)	333	5	647.822700	-3.804	-3.838	-3.861	4
EQ(10)	333	4	647.709130	-3.820	-3.848	-3.866	3
EQ(11)	333	3	647.565930	**-3.837<**	**-3.857<**	-3.871	2
EQ(12)	333	2	644.172110	-3.834	-3.848	-3.857	1

Output of Table-5.5: Forecast evaluation of AR

Dynamic (ex ante) forecasts for log_price (SE based on error variance only)

Horizon	Forecast	SE	Actual	Error	t-value
334	5.691	0.03443	5.700	0.00905515	0.263
335	5.705	0.05269	5.703	-0.0012602	-0.024
336	5.719	0.06549	5.706	-0.0131487	-0.201
337	5.731	0.07652	5.725	-0.0058983	-0.077
338	5.743	0.08424	5.725	-0.0187098	-0.222
339	5.753	0.08986	5.723	-0.0297048	-0.331
340	5.764	0.09472	5.722	-0.0422046	-0.446
341	5.774	0.0988	5.719	-0.0548423	-0.555
342	5.784	0.1019	5.723	-0.0608735	-0.597
343	5.795	0.1064	5.773	-0.0224558	-0.211
344	5.806	0.1104	5.779	-0.026707	-0.242
345	5.818	0.1148	5.786	-0.0318961	-0.278
mean(Error)=	-0.02489		RMSE =	0.031954	
SD(Error) =	0.020043		MAPE =	0.46025	