WORKING PAPER[1]

"INFLATION FORECASTS WITH ARIMA AND VECTOR AUTOREGRESSIVE MODELS IN GUATEMALA"

by

Héctor A. Valle S.

May 2002

ECONOMIC RESEARCH DEPARTMENT
BANCO DE GUATEMALA

[1] The views expressed in this paper are those of the author and do not necessarily represent the views of Banco de Guatemala.

"INFLATION FORECASTS WITH ARIMA AND VECTOR AUTOREGRESSIVE MODELS IN GUATEMALA"

by

Héctor A. Valle S.

May 2002

ABSTRACT

This paper has two objectives, one is to forecast inflation and the other to identify a set of variables that Banco de Guatemala should monitor to achieve its inflation target. VAR models are employed to identify this set of variables. To do so, this kind of models are used just as forecasting instruments, ignoring any structural or theoretical interpretation of them. The variables included in the VAR models are those that intuitively contain some information about inflation and therefore could be good forecasters of it. Each model is estimated using a specific set of variables and a particular definition for each them. The set of variables in the model that proves to be the best forecaster of inflation is the set of variables that Banco de Guatemala should follow closely. ARIMA models are constructed as benchmarks. The idea is to determine whether the variables included in the VAR models have more information to forecast inflation than the information contained in the past behavior of the inflation series.

As a result of the adoption of a new CPI a structural change was produced in the inflation series. This fact provides the opportunity to test the forecasting capabilities of VAR and ARIMA models in the framework of intervention analysis. Finally, the year 2002 is forecasted with the best models and the results are judged under the plausibility criterion, considering the recent development of the series.

JEL: C53

The author wishes to thank Juan Carlos Castañeda, Victor Manuel Guerrero, Erick Roberto Vaides and Lorena Ramírez Orellana for their comments, suggestions and valuable help.

Contents

Introduction

It is now generally accepted that keeping low and stable rates of inflation is the primary objective of central banks. Economic agents, private and public alike, monitor closely the evolution of prices in the economy, in order to make decisions that allow them to optimize the use of their resources. In this context, it is very important to forecast inflation. A number of central banks have adopted an inflation targeting scheme to pursue their monetary policy. In this framework, the inflation target is the nominal anchor and therefore forecasting inflation plays a key role. Even though the inflation target usually originates from a political consensus among different government institutions, it also should be statistically plausible. Therefore, central banks must have good models to forecast inflation. There is the possibility that Banco de Guatemala will adopt an inflation target regime in the near future and the existence of inflation forecasts will be a must for the Bank in order to succeed in this policy implementation. This paper has two objectives, the first one is to forecast inflation in Guatemala, while the second one is to select a group of variables that Banco de Guatemala should follow closely to achieve its inflation target. For this purpose, vector autoregressive models are employed. The variables included in the models are variables that intuitively contain information about inflation, such as interest rates, money aggregates, exchange rates, output, and so forth.

Many possible combinations of these variables are tested in the VAR models. The group of variables that help to predict inflation best should be followed closely by Banco de Guatemala as it conducts its monetary policy.

In addition, several ARIMA models for inflation are built. These models are well known for being simple, robust, and parsimonious, and for providing good results. The idea is to determine whether the variables included in the VAR models have more information to forecast inflation than the information contained in the past behavior of the inflation series. In order to evaluate forecasts and rank them, out-of-sample forecast evaluation is performed. The criterion to assess the goodness of the inflation forecast is the minimization of the root mean square error of the forecast errors.

The base of the Consumer Price Index was March – April 1983 until the year 2001, when a new CPI was adopted. The base of this new CPI is December 2000. The new index includes more items and new weights. It was found that the new CPI produced a structural change in the series. Therefore, intervention analysis was performed to model the structural change, which gives an opportunity to test the forecasting capabilities of the models in this context.

In the first part of the paper, the inflation series is examined graphically to develop some general intuition about the structure of the series. This structure determines the strategy to model the series. In the second part, the methodology employed in the paper is exposed. Then, the models and their results for the period previous to the structural change (1993 – 2000) are presented in the third part. The intervention analysis is performed in the fourth part. Finally, the conclusions are presented.

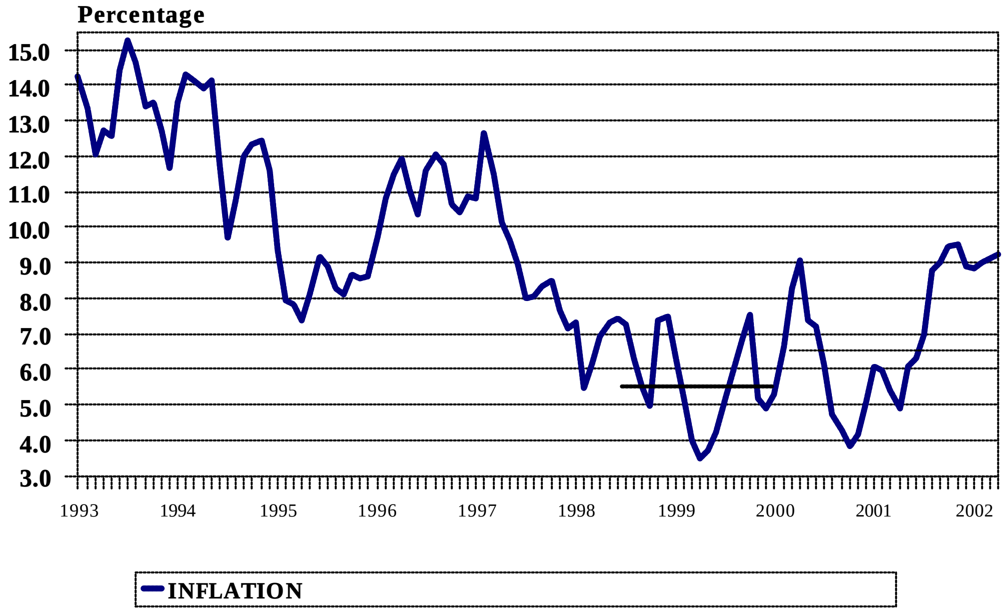
1. Analysis of the inflation series

The first step in forecasting a series is to check the structure of the data to be forecasted. In this section, an informal analysis of the inflation series is developed, in order to obtain some preliminary knowledge about the stationarity of the series (existence of a trend or a seasonal pattern.) To accomplish these first steps, both the graph and the autocorrelation function of the series are going to be examined.

It is important to notice that from 1983 to 2000 the base of the CPI was March-April 1983. The base for the new index ¡s December 2000. The new index was built according to the income and expenditure survey of 1999 – 2000, and it includes 422 items (the old index included 212). Considering the big time gap between both base-years and the fact that the new CPI has more items and also different weights for each article, there might be a structural change in the series.

The measure of inflation used in this work is the annual rate of change in the monthly consumer price index (Figure 1.)

FIGURE 1
RATE OF INFLATION
ANNUAL CHANGE IN MONTHLY CPI
1993.01 - 2002.04



It can be seen on Figure 1 that the inflation series presents a trend, and consequently the series is not stationary. At first sight, there seems to be a change in the mean of the series for the period 1998 – 2000, and also for the period following 2000 (which might suggest a structural change in the series that coincides with introduction of the new CPI). However, these previous conjectures will have to be tested through formal statistical procedures.

The autocorrelation function, shown in Figure 2, also provides very useful information.

FIGURE 2
INFLATION
Autocorrelation and Partial Correlation Functions

Sample: 1993:01 2002:04
Included observations: 112

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |*******|	. |*******|	1	0.934	0.934	100.31	0.000
. |*******|	**|. |	2	0.846	-0.207	183.30	0.000
. |****** |	. |* |	3	0.778	0.150	254.22	0.000
. |****** |	. |. |	4	0.726	0.024	316.57	0.000
. |***** |	. |** |	5	0.708	0.248	376.45	0.000
. |***** |	. |. |	6	0.706	0.041	436.40	0.000
. |***** |	.*|. |	7	0.684	-0.097	493.27	0.000
. |***** |	.*|. |	8	0.632	-0.168	542.29	0.000
. |**** |	. |. |	9	0.566	-0.054	581.98	0.000
. |**** |	.*|. |	10	0.491	-0.131	612.14	0.000
. |*** |	.*|. |	11	0.414	-0.135	633.80	0.000
. |*** |	. |* |	12	0.365	0.067	650.79	0.000
. |*** |	. |** |	13	0.360	0.248	667.54	0.000
. |*** |	. |. |	14	0.362	-0.001	684.57	0.000
. |*** |	. |. |	15	0.356	0.062	701.22	0.000
. |*** |	. |. |	16	0.337	0.010	716.35	0.000
. |** |	.*|. |	17	0.288	-0.092	727.46	0.000
. |** |	. |. |	18	0.237	0.028	735.08	0.000
. |** |	. |* |	19	0.217	0.073	741.54	0.000
. |** |	.*|. |	20	0.210	-0.090	747.65	0.000
. |** |	.*|. |	21	0.206	-0.061	753.62	0.000
. |* |	.*|. |	22	0.196	-0.141	759.05	0.000
. |* |	. |. |	23	0.179	0.047	763.65	0.000
. |* |	. |. |	24	0.151	-0.016	766.95	0.000
. |* |	. |* |	25	0.128	0.131	769.36	0.000
. |* |	. |* |	26	0.127	0.135	771.74	0.000
. |* |	. |* |	27	0.132	0.084	774.36	0.000
. |* |	. |. |	28	0.135	-0.037	777.15	0.000
. |* |	.*|. |	29	0.138	-0.058	780.07	0.000
. |* |	. |. |	30	0.135	0.026	782.91	0.000
. |* |	. |. |	31	0.126	0.019	785.41	0.000
. |* |	. |. |	32	0.125	0.013	787.92	0.000
. |* |	.*|. |	33	0.128	-0.079	790.55	0.000
. |* |	.*|. |	34	0.122	-0.157	793.00	0.000
. |* |	.*|. |	35	0.115	-0.104	795.18	0.000
. |* |	. |* |	36	0.116	0.095	797.44	0.000

The inflation's autocorrelation function is typical of a non-stationary process, where the autocorrelation declines slowly as the number of lags increases. This behavior is what we expected from the graphic analysis of the data.

In order to get more information about the inflation series, the autocorrelation function in first differences was calculated (Figure 3).

FIGURE 3
INFLATION
FIRST DIFFERENCES AUTOCORRELATION AND PARTIAL CORRELATION FUNCTIONS

Sample: 1993:01 2002:04
Included observations: 112

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |** |	. |** |	1	0.264	0.264	8.0219	0.005
.*|. |	**|. |	2	-0.137	-0.222	10.197	0.006
.*|. |	.*|. |	3	-0.188	-0.096	14.342	0.002
***|. |	**|. |	4	-0.325	-0.308	26.864	0.000
**|. |	.*|. |	5	-0.248	-0.158	34.186	0.000
. |* |	. |* |	6	0.135	0.138	36.397	0.000
. |** |	. |* |	7	0.326	0.161	49.349	0.000
. |** |	. |. |	8	0.204	0.034	54.451	0.000
. |* |	. |. |	9	0.082	0.035	55.288	0.000
. |. |	. |* |	10	0.061	0.179	55.748	0.000
**|. |	.*|. |	11	-0.215	-0.090	61.617	0.000
****|. |	***|. |	12	-0.534	-0.439	98.077	0.000
.*|. |	. |* |	13	-0.103	0.078	99.453	0.000
. |* |	.*|. |	14	0.073	-0.108	100.15	0.000
. |* |	. |. |	15	0.129	-0.003	102.35	0.000
. |** |	. |. |	16	0.293	0.006	113.78	0.000
. |* |	.*|. |	17	0.132	-0.091	116.13	0.000
**|. |	. |. |	18	-0.190	-0.005	121.04	0.000
.*|. |	. |* |	19	-0.188	0.130	125.90	0.000
.*|. |	. |. |	20	-0.103	0.006	127.38	0.000
. |. |	. |* |	21	0.045	0.174	127.67	0.000
. |. |	. |. |	22	0.043	-0.017	127.93	0.000
. |* |	. |. |	23	0.155	0.026	131.36	0.000
. |. |	**|. |	24	0.062	-0.295	131.91	0.000
.*|. |	.*|. |	25	-0.138	-0.098	134.68	0.000
. |. |	. |. |	26	-0.052	-0.052	135.09	0.000
. |. |	. |* |	27	0.045	0.105	135.40	0.000
. |. |	. |* |	28	-0.029	0.068	135.52	0.000
. |. |	.*|. |	29	-0.008	-0.084	135.53	0.000
. |* |	. |. |	30	0.067	-0.012	136.24	0.000
. |. |	. |. |	31	-0.051	0.000	136.64	0.000
. |. |	. |. |	32	-0.017	0.065	136.69	0.000
. |. |	. |** |	33	0.042	0.202	136.98	0.000
. |. |	. |. |	34	0.033	0.011	137.16	0.000
.*|. |	.*|. |	35	-0.091	-0.092	138.54	0.000
. |. |	.*|. |	36	0.010	-0.164	138.56	0.000

The variable seems to be stationary in first differences, with a few significant autocorrelations and then an exponential drop. Since there is significant structure in 4^{th} and 12^{th} lags, we can infer that the series presents seasonality. Interestingly, according to the partial autocorrelation function, there is significant structure in the 24^{th} lag.

At this point, we can infer that the inflation series is non-stationary and that it might be integrated of order one (this statement needs to be confirmed through the Dickey Fuller test.) Moreover, the series might present a seasonal pattern and it shows the existence of structure up to lag 24. Altogether, this means that it is not an easy series to be modeled.

2. Method

2.1 ARIMA Models

ARIMA models are univariate models that consist of an autoregressive polynomial, an order of integration (d), and a moving average polynomial. ARIMA models are well known for being simple, robust, and parsimonious, and for providing good results. ARIMA models for the inflation series are constructed with two objectives in mind: to forecast inflation and to provide a benchmark for other forecasts. The main idea is to determine whether or not the variables included in the VAR models have more information to forecast inflation than the information contained in the past behavior of the inflation series, which is modeled with the ARIMA models.

In order to create ARIMA models, the steps of the Box and Jenkins' method are followed; that is:

a) Identify.
b) Estimate.
c) Verify.

2.2 VAR Models

As it was pointed out earlier, one of the objectives of this paper is to forecast inflation using multivariate models to identify the variables that Banco de Guatemala should monitor to achieve its proposed inflation target. The univariate models will be used as benchmarks to compare the performance of the multivariate ones.

Vector autoregressive models are used to build the multivariate models. These VAR models are used only as forecasting instruments, which means that neither theoretical considerations nor structural analysis are carried out. In this sense, VAR models offer a lot of flexibility in their construction and they were originally developed with that purpose.

In this context, the variables in the VAR models, which were included without regards to any particular economic theory, are those that intuitively contain information about inflation and their role in the model will be just as predictors of such variable. Those that prove to be good forecasters of inflation will remain in the model and the rest will be excluded.

The variables taken into account in the VAR models are, as it was already said, those that might contain information about inflation and consequently might also be good forecasters of it: output, interest rate, exchange rate, money and oil prices. For some of these variables there are more than one definitions, like M1 and M2 for money; therefore the different definitions are alternatively included in the models.

In other words, at this point, after having chosen the variables and their definitions, the VAR models are constructed with different variables. Hence, the variables that Banco de Guatemala should follow closely to conduct its monetary policy ought to be those variables (and their definitions) included in the model that prove empirically to give the best forecast of inflation.

The variables and their different definitions are presented in Table 1.

TABLE 1
MODEL VARIABLES AND DEFINITIONS

VARIABLES / DEFINITIONS	*OIL PRICES*	*MONEY*	*INTEREST RATE*	*EXCHANGE RATE*	*OUTPUT GROWTH*	*PRICES*
DEFINITIONS	INTERNATL. OIL PRICES (NEW YORK MERCANTILE EXCHANGE)	M1	DEPOSIT SHORT TERM	EXCHANGE RATE (BUY)	MIEA[1/]	INFLATION
		M2	DEPOSIT LONG TERM			
		EMISION[2/]				
		MONETARY BASE				

1/ Due to lack of GDP's monthly data, the Monthly Index of Economic Activity is used as a proxy for output growth.
2/ *EMISION.* Currency issue: currency outside deposit money banks plus cash in vaults of the banks.

2.3 Criterion to establish the ranking of the different forecasts

The criterion to establish the ranking of the different forecasts is the minimization of the Root Mean Square Error -RMSE -.

$$RMSE = \sqrt{\frac{\sum (Y_i - -\hat{Y})^2}{n}}$$

Where:

Y = Observed

$\hat{Y}$ = Forecasted

n = Number of observations

The RMSE tells us how close the forecasted series is to the original series. The closer the RMSE is to zero, the better the forecast is.

As it was said before, there is the possibility of structural change in the inflation series from January 2001 on. This gives us the opportunity to test the different models and its forecasting capabilities in two different frameworks, that is, in periods with and without structural change in the series.

For the period without structural change, the models will be estimated with information from January 1993 to December 2000, then the forecasted series will be computed from January to December 2000. The RMSE will then be obtained with this information (the observed and forecasted series from January to December 2000).

2.4 Intervention analysis

An intervention can be interpreted as the occurrence of an exogenous event, which exerts its influence on the historical behavior of a variable. Such intervention could be a change in the economic policy, like the implementation of a new set of fiscal policies, or the occurrance of natural phenomena, political events, and so forth.

This methodology provides the framework to model the structural change in the series and it will be performed as follow:

a) First, it has to be determined whether or not there is statistical evidence of structural change. This will be done through the test of Box and Tiao, which consists on the statistic *C*, as follows:

$$C = \left(\sum_{h=1}^{H} e^2_{I-2+h} \right) / \hat{s}^2_a$$

Where:

$\hat{\sigma}^2_{\alpha}$ = residual variance of the model built for the previous period (1993 –2000 in this case)

H= Each forecast
I= Number of forecasted periods
t=*I* and $e_I(1),\ldots,e_{I+H-2}(1)$ forecast errors one period ahead.

The test has a Chi square distribution.

If the model does not represent the series up to the observation $t=I+H-1$, the value of *C* will be big (according to the Chi square distribution). If this happens, it means that there is a structural change in the series and an intervention model will have to be built.

a) If there is statistical evidence of structural change, according to Box and Tiao test, an intervention variable has to be added to the model. On the other hand, if there is no evidence of structural change, then the model is enough good to represent and forecast the series after the occurrence of the exogenous event.

b) In the case when there is a structural change, the intervention variable has to be significant in the model and, therefore, it has to help to produce better forecasts.

3. Results for the period 1993 - 2000

3.1 ARIMA model

ARIMA models were built following Box and Jenkins methodology. According to the results, the model that best fits the inflation series is shown in Figure 4.

FIGURE 4

Dependent Variable: DLRITP
Method: Least Squares
Sample(adjusted): 1995:02 2000:12
Included observations: 71 after adjusting endpoints
Convergence achieved after 12 iterations
Backcast: 1994:09 1995:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.011024	0.002677	-4.118698	0.0001
AR(12)	-0.787864	0.113930	-6.915330	0.0000
AR(24)	-0.424795	0.128205	-3.313414	0.0015
MA(1)	0.371872	0.093367	3.982920	0.0002
MA(4)	-0.351205	0.090873	-3.864786	0.0003
MA(5)	-0.565482	0.074379	-7.602682	0.0000

R-squared	0.570046	Mean dependent var	-0.008547
Adjusted R-squared	0.536973	S.D. dependent var	0.135128
S.E. of regression	0.091949	Akaike info criterion	-1.854440
Sum squared resid	0.549552	Schwarz criterion	-1.663228
Log likelihood	71.83264	F-statistic	17.23583
Durbin-Watson stat	2.143281	Prob(F-statistic)	0.000000

The seasonal regressors and the long lags reflect the complexity of the series. The correlogram of the residuals is also shown in Figure 5:

FIGURE 5

Date: 05/30/02 Time: 15:06
Sample: 1995:02 2000:12
Included observations: 71
Q-statistic probabilities adjusted for 5 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.*| . |	.*| . |	1	-0.087	-0.087	0.5612	
.*| . |	.*| . |	2	-0.097	-0.105	1.2629	
**| . |	**| . |	3	-0.204	-0.226	4.4286	
. |*. |	. | . |	4	0.082	0.028	4.9524	
. |*. |	. |*. |	5	0.161	0.135	6.9804	
. | . |	. | . |	6	0.010	0.011	6.9880	0.008
.*| . |	. | . |	7	-0.071	-0.016	7.3918	0.025
. | . |	. |*. |	8	0.036	0.093	7.4966	0.058
. | . |	. | . |	9	0.055	0.053	7.7513	0.101
. | . |	. | . |	10	0.037	0.020	7.8652	0.164
. |*. |	. |*. |	11	0.068	0.120	8.2634	0.219
.*| . |	.*| . |	12	-0.173	-0.137	10.887	0.144
. | . |	.*| . |	13	-0.042	-0.085	11.041	0.199
. | . |	. | . |	14	-0.001	-0.030	11.041	0.273
. | . |	. | . |	15	0.047	-0.051	11.249	0.338
. |*. |	. | . |	16	0.073	0.040	11.748	0.383
. | . |	. | . |	17	-0.042	0.017	11.914	0.453
. | . |	. | . |	18	-0.003	0.034	11.915	0.535
. | . |	. | . |	19	0.017	0.033	11.945	0.611
. | . |	. | . |	20	-0.055	-0.050	12.247	0.660
. | . |	. | . |	21	0.028	0.031	12.330	0.721
.*| . |	.*| . |	22	-0.095	-0.098	13.292	0.716
. | . |	. | . |	23	-0.015	-0.040	13.317	0.772
.*| . |	.*| . |	24	-0.089	-0.149	14.200	0.772
. | . |	.*| . |	25	0.003	-0.088	14.201	0.820
. | . |	.*| . |	26	-0.050	-0.114	14.489	0.848
. | . |	. | . |	27	0.028	-0.034	14.581	0.880
.*| . |	.*| . |	28	-0.092	-0.070	15.590	0.872
. | . |	.*| . |	29	-0.044	-0.060	15.834	0.894
. | . |	. | . |	30	0.024	0.054	15.905	0.918
.*| . |	.*| . |	31	-0.113	-0.119	17.572	0.891
. |*. |	. | . |	32	0.066	0.043	18.156	0.899

The 12 months of year 2000 were forecasted with this model and the RMSE obtained was 0.09039. This value is higher than the results obtained through the best VAR models, which means that the ARIMA model cannot predict inflation better than the best VAR model does.

In the following part, the forecasting capabilities of these models will be used to model structural change.

3.2 Results from the VAR models

As it was pointed out previously, the VAR models are used only as forecasting instruments and they include variables that might contain information about inflation and therefore could contribute to its forecast. Each one at a time, VAR models including different number of variables and different definitions of them, were solved. The most general form of the VAR models is presented here:

$$\Pi_t = a_1 + \sum_{j=1}^{k} b_{1j}\Pi_{t-j} + \sum_{j=1}^{k} g_{1j}TC_{t-j} + \sum_{j=1}^{k} q_{1j}M_{t-j} + \sum_{j=1}^{k} l_{1j}i_{t-j} + \sum_{j=1}^{k} d_{1j}Y_{t-j} + \sum_{j=1}^{k} f_{1j}Pp_{t-j} + u_{1t}$$

$$TC_t = a_2 + \sum_{j=1}^{k} b_{2j}\Pi_{t-j} + \sum_{j=1}^{k} g_{2j}TC_{t-j} + \sum_{j=1}^{k} q_{2j}M_{t-j} + \sum_{j=1}^{k} l_{2j}i_{t-j} + \sum_{j=1}^{k} d_{2j}Y_{t-j} + \sum_{j=1}^{k} f_{2j}Pp_{t-j} + u_{2t}$$

$$M_t = a_3 + \sum_{j=1}^{k} b_{3j}\Pi_{t-j} + \sum_{j=1}^{k} g_{3j}TC_{t-j} + \sum_{j=1}^{k} q_{3j}M_{t-j} + \sum_{j=1}^{k} l_{3j}i_{t-j} + \sum_{j=1}^{k} d_{3j}Y_{t-j} + \sum_{j=1}^{k} f_{3j}Pp_{t-j} + u_{3t}$$

$$i_t = a_4 + \sum_{j=1}^{k} b_{4j}\Pi_{t-j} + \sum_{j=1}^{k} g_{4j}TC_{t-j} + \sum_{j=1}^{k} q_{4j}M_{t-j} + \sum_{j=1}^{k} l_{4j}i_{t-j} + \sum_{j=1}^{k} d_{4j}Y_{t-j} + \sum_{j=1}^{k} f_{4j}Pp_{t-j} + u_{4t}$$

$$Y_t = a_5 + \sum_{j=1}^{k} b_{5j}\Pi_{t-j} + \sum_{j=1}^{k} g_{5j}TC_{t-j} + \sum_{j=1}^{k} q_{5j}M_{t-j} + \sum_{j=1}^{k} l_{5j}i_{t-j} + \sum_{j=1}^{k} d_{5j}Y_{t-j} + \sum_{j=1}^{k} f_{5j}Pp_{t-j} + u_{1t}$$

$$Pp_t = a_6 + \sum_{j=1}^{k} b_{6j}\Pi_{t-j} + \sum_{j=1}^{k} g_{6j}TC_{t-j} + \sum_{j=1}^{k} q_{6j}M_{t-j} + \sum_{j=1}^{k} l_{6j}i_{t-j} + \sum_{j=1}^{k} d_{6j}Y_{t-j} + \sum_{j=1}^{k} f_{6j}Pp_{t-j} + u_{1t}$$

Where:

P = inflation

TC = exchange rate

M = money

i = interest rate

Y = output

Pp = oil prices

All the variables included in the VAR models must be stationary. So, before their inclusion in the models, the variables were checked for stationarity. The Dickey – Fuller test was employed to test for unit roots and it was found that all the variables were integrated of order one, *I*(1). This means that they are stationary after only one differencing operation.

All the possible VAR models were estimated considering different number of variables through some of their different definitions. A total of 114 models and their correspondent forecasts were obtained at the first stage. To determine the value of *k*, which represents the number of lags of the variables, the Schwarz Criterion was adopted. It was chosen on the basis of its asymptotic properties with large samples.

According to the RMSE criterion, not even one of the VAR models estimated proved to forecast better than the ARIMA model. At first it was thought that the moving average component of the ARIMA model might be an advantage for this kind of models, in this particular case, in the forecasting of inflation. However, after a closer look it was found that most of the VAR models were specified with two and four lags while the autocorrelation function of inflation showed to have a longer structure (as it was pointed out in part number two of this work). As it will be shown below, the ARIMA model of the series has significant coefficients in lags 12 and 24 in the autoregressive part. As a result, another approach to select the lag length was adopted.

The Schwarz and Akaike criteria have been criticized for penalizing the number of coefficients in the model, without taking into account how the model looks like, Wickens (2002). Therefore, it is recommended to start with a general model, and then find the maximum lag length by testing the significance of the coefficients using standard *t*-tests.

The outcome is usually a model with too many lags, while it has been shown that for most purposes it is better to have a model with as few lags as possible. This can be accomplished by having a very strict significance criterion. After this new approach was defined, the models were estimated again, using twelve and eighteen lags to start the regressions. This time, some of the VAR models proved to be better forecasters than ARIMA models, according to the RMSE criterion. The models with best results were chosen and are shown in table 2.

TABLE 2
BEST VAR MODELS TO FORECAST RESULTS
PERIOD 1993 – 2000

Models of 18 Lags:	**Variables:**	**RMSE[a/]:**
(1)	Inflation, MIEA[b/], IOILP[c/]	0.02290978
(2)	Inflation, MIEA, LTDIR[d/]	0.03550484
(3)	Inflation, MIEA, ER[e/]	0.03750642
Models of 12 Lags:		
(1)	Inflation, MIEA, ER, *EMI* [f/], LTDIR, IOILP	0.03137564
(2)	Inflation, MIEA, ER, BASE[g/], IOILP	0.03248904
(3)	Inflation, MIEA, *EMI*, LTDIR	0.05481890

a/ Root Mean Square Error.
b/ MIEA = Monthly Index of Economic Activity.
c/ IOILP = International Oil Prices.
d/ LTDIR = Long Term Deposit Interest Rate.
e/ ER = Exchange Rate
f/ *EMI = Emision* (Currency Issue) = *Currency outside deposit money banks, plus cash in vaults of the banks.*
g/ Monetary base

According to these results, Banco de Guatemala should monitor the Monthly Index of Economic Activity (MIEA) and international oil prices (IOILP). However, it should be noted that the central bank exerts no influence on these variables, and being so, they do not contribute with relevant information for policy purposes. The second model in the ranking includes MIEA, exchange rate (ER), *Emision* (EMI), Long Term Interest Rate on Deposits (LTDIR) and IOILP. This information is more useful for monetary policy because Banco de Guatemala can exert influence over some of these variables, like the ER, *EMI* and LTDIR.

4. Intervention analysis

As it was explained before, the Consumer Price Index was re-defined in December 2000, and this event could have caused a structural change in the inflation series. In order to test this hypothesis, the Box and Tiao Test for structural change was performed. The null hypothesis of non-existence of structural change was rejected. In other words, the results obtained support the existence of a structural change in January 2001.

Now, the problem is to establish the form of the intervention variable. In fact, there are two different ways to conceive the structural change: it can be viewed either as a permanent change or as a temporary one. A permanent change would mean that since the new CPI has more items and more up to date sample and weights, it would always show higher rates of inflation. On the other hand, a temporary change means that during the year 2001 data of the one year change in the CPI is distorted for being the change between two different baskets of products and services. But in the year 2002 that effect would disappear because the change corresponds to the same basket.

In order to model the permanent change scenario, the intervention variable has the following form: zeros for the period previous to the intervention (January 1993 to December 2000) and ones after it (January 2001 to April 2002).

In the case of the temporary change scenario, the intervention model is composed of zeros before and after the intervention (January 1993 to December 2000 and January 2002 to date). For the intervention itself, ones are used (January to December 2001).

4.1 ARIMA model with intervention

4.1.1 Permanent change

This is the same model built for the period 1993 – 2000 but it now includes an intervention variable to account for the structural change (dum1). This time, the sample has been expanded to include April 2002. The estimated model and the forecasted series are presented in Figures 6 and 7. It is important to point out that the intervention variable is statistically significant.

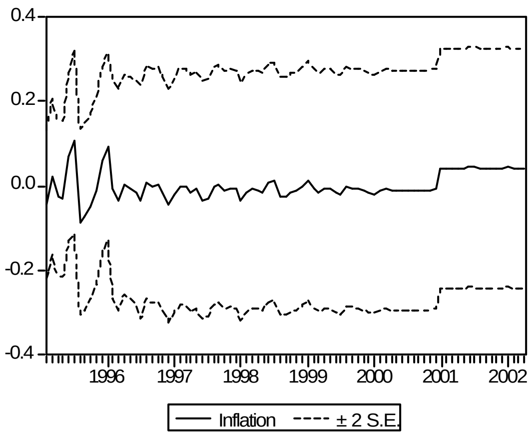
FIGURE 6

Dependent Variable: DLRITP
Method: Least Squares
Sample(adjusted): 1995:02 2002:04
Included observations: 87 after adjusting endpoints
Convergence achieved after 16 iterations
Backcast: 1994:09 1995:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.010879	0.002222	-4.895685	0.0000
DUM1	0.051023	0.014816	3.443860	0.0009
AR(12)	-0.751912	0.100354	-7.492573	0.0000
AR(24)	-0.325762	0.102665	-3.173047	0.0021
MA(1)	0.329322	0.086927	3.788475	0.0003
MA(4)	-0.406278	0.082110	-4.947979	0.0000
MA(5)	-0.565953	0.069440	-8.150200	0.0000

R-squared	0.577130	Mean dependent var	-8.67E-05
Adjusted R-squared	0.545415	S.D. dependent var	0.130133
S.E. of regression	0.087740	Akaike info criterion	-1.951851
Sum squared resid	0.615858	Schwarz criterion	-1.753445
Log likelihood	91.90552	F-statistic	18.19726
Durbin-Watson stat	2.080945	Prob(F-statistic)	0.000000

FIGURE 7



Forecast: DLRITPF	
Actual: DLRITP	
Sample: 1995:02 2002:04	
Include observations: 87	
Root Mean Squared Error	0.128085
Mean Absolute Error	0.099753
Mean Abs. Percent Error	144.7339
Theil Inequality Coefficient	0.793420
Bias Proportion	0.000000
Variance Proportion	0.577467
Covariance Proportion	0.422533

Also, it is important to check the autoregression function of the residuals. This information can help to find out whether or not the permanent-change model is a better model than the temporary-change one. The residuals of the model without intervention are better than those of the permanent-change and the temporary-change models. However, it is important to keep in mind that the residuals of these two models correspond to a period of time, which is characterized for having more difficulties to be dealt with in the forecasting.

FIGURE 8

Date: 05/30/02 Time: 20:29
Included observations: 87
Q-statistic probabilities adjusted for 5 ARMA term(s)

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>. | .</td><td>. | .</td><td>1</td><td>-0.053</td><td>-0.053</td><td>0.2546</td><td></td></tr>
<tr><td>.*| .</td><td>.*| .</td><td>2</td><td>-0.097</td><td>-0.100</td><td>1.1044</td><td></td></tr>
<tr><td>**| .</td><td>**| .</td><td>3</td><td>-0.203</td><td>-0.216</td><td>4.8915</td><td></td></tr>
<tr><td>. |*.</td><td>. | .</td><td>4</td><td>0.095</td><td>0.060</td><td>5.7276</td><td></td></tr>
<tr><td>. |*.</td><td>. |*.</td><td>5</td><td>0.181</td><td>0.158</td><td>8.8118</td><td></td></tr>
<tr><td>. | .</td><td>. | .</td><td>6</td><td>-0.002</td><td>-0.006</td><td>8.8121</td><td>0.003</td></tr>
<tr><td>. | .</td><td>. | .</td><td>7</td><td>-0.026</td><td>0.037</td><td>8.8793</td><td>0.012</td></tr>
<tr><td>. |*.</td><td>. |*.</td><td>8</td><td>0.080</td><td>0.155</td><td>9.5137</td><td>0.023</td></tr>
<tr><td>. | .</td><td>. | .</td><td>9</td><td>-0.024</td><td>-0.040</td><td>9.5727</td><td>0.048</td></tr>
<tr><td>. | .</td><td>. | .</td><td>10</td><td>0.018</td><td>0.004</td><td>9.6068</td><td>0.087</td></tr>
<tr><td>. | .</td><td>. |*.</td><td>11</td><td>0.064</td><td>0.119</td><td>10.020</td><td>0.124</td></tr>
<tr><td>.*| .</td><td>**| .</td><td>12</td><td>-0.146</td><td>-0.192</td><td>12.227</td><td>0.093</td></tr>
<tr><td>. | .</td><td>.*| .</td><td>13</td><td>-0.021</td><td>-0.073</td><td>12.275</td><td>0.139</td></tr>
<tr><td>. | .</td><td>. | .</td><td>14</td><td>-0.021</td><td>0.004</td><td>12.322</td><td>0.196</td></tr>
<tr><td>. | .</td><td>. | .</td><td>15</td><td>0.063</td><td>-0.047</td><td>12.747</td><td>0.238</td></tr>
<tr><td>. | .</td><td>. | .</td><td>16</td><td>0.035</td><td>0.000</td><td>12.881</td><td>0.301</td></tr>
<tr><td>.*| .</td><td>. | .</td><td>17</td><td>-0.063</td><td>0.027</td><td>13.325</td><td>0.346</td></tr>
<tr><td>. | .</td><td>. | .</td><td>18</td><td>-0.018</td><td>-0.003</td><td>13.360</td><td>0.420</td></tr>
<tr><td>. | .</td><td>. | .</td><td>19</td><td>0.016</td><td>0.010</td><td>13.389</td><td>0.496</td></tr>
<tr><td>. | .</td><td>. | .</td><td>20</td><td>-0.048</td><td>-0.016</td><td>13.660</td><td>0.551</td></tr>
<tr><td>. | .</td><td>. | .</td><td>21</td><td>0.055</td><td>0.048</td><td>14.012</td><td>0.598</td></tr>
<tr><td>.*| .</td><td>.*| .</td><td>22</td><td>-0.081</td><td>-0.084</td><td>14.786</td><td>0.611</td></tr>
<tr><td>. | .</td><td>. | .</td><td>23</td><td>-0.029</td><td>-0.039</td><td>14.890</td><td>0.670</td></tr>
<tr><td>.*| .</td><td>.*| .</td><td>24</td><td>-0.135</td><td>-0.176</td><td>17.133</td><td>0.581</td></tr>
<tr><td>. | .</td><td>. | .</td><td>25</td><td>0.022</td><td>-0.041</td><td>17.194</td><td>0.640</td></tr>
<tr><td>. | .</td><td>.*| .</td><td>26</td><td>-0.031</td><td>-0.102</td><td>17.317</td><td>0.692</td></tr>
<tr><td>. | .</td><td>.*| .</td><td>27</td><td>-0.049</td><td>-0.111</td><td>17.629</td><td>0.728</td></tr>
<tr><td>.*| .</td><td>.*| .</td><td>28</td><td>-0.135</td><td>-0.115</td><td>20.023</td><td>0.641</td></tr>
<tr><td>. | .</td><td>.*| .</td><td>29</td><td>-0.051</td><td>-0.087</td><td>20.374</td><td>0.675</td></tr>
<tr><td>. |*.</td><td>. |*.</td><td>30</td><td>0.084</td><td>0.066</td><td>21.336</td><td>0.674</td></tr>
<tr><td>.*| .</td><td>.*| .</td><td>31</td><td>-0.133</td><td>-0.163</td><td>23.792</td><td>0.588</td></tr>
<tr><td>. | .</td><td>. | .</td><td>32</td><td>-0.018</td><td>0.008</td><td>23.836</td><td>0.639</td></tr>
</table>

4.1.2 Temporary change

This time the intervention variable (dum2) included in the original ARIMA model accounts for the temporary change as it is shown below.

FIGURE 9

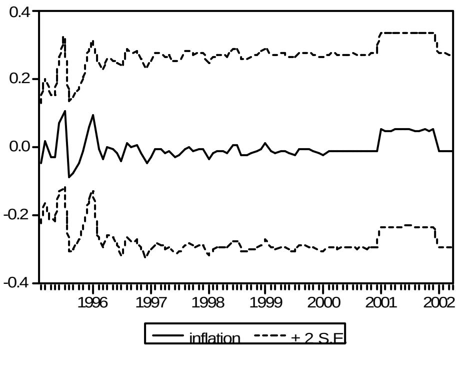
Dependent Variable: DLRITP
Method: Least Squares
Date: 05/30/02 Time: 20:41
Sample(adjusted): 1995:02 2002:04
Included observations: 87 after adjusting endpoints
Convergence achieved after 16 iterations
Backcast: 1994:09 1995:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.011150	0.002215	-5.033616	0.0000
DUM2	0.062372	0.017573	3.549386	0.0006
AR(12)	-0.754012	0.100085	-7.533680	0.0000
AR(24)	-0.314817	0.102193	-3.080598	0.0028
MA(1)	0.328679	0.086282	3.809338	0.0003
MA(4)	-0.407721	0.081117	-5.026336	0.0000
MA(5)	-0.571949	0.068465	-8.353912	0.0000

R-squared	0.579246	Mean dependent var	-8.67E-05
Adjusted R-squared	0.547690	S.D. dependent var	0.130133
S.E. of regression	0.087520	Akaike info criterion	-1.956868
Sum squared resid	0.612776	Schwarz criterion	-1.758461
Log likelihood	92.12374	F-statistic	18.35584
Durbin-Watson stat	2.092016	Prob(F-statistic)	0.000000

The intervention variable is significant, as it is in the permanent-change model.

FIGURE 10



Forecast: DLRITPF
Actual: DLRITP
Sample: 1995:02 2002:04
Include observations: 87

Root Mean Squared Error	0.128047
Mean Absolute Error	0.099608
Mean Abs. Percent Error	142.7025
Theil Inequality Coefficient	0.787223
Bias Proportion	0.000083
Variance Proportion	0.563644
Covariance Proportion	0.436273

As it can be seen in the forecast graph, the RMSE for the whole period is not really different to that of the previous model.

FIGURE 11

Date: 05/30/02 Time: 21:33
Sample: 1995:02 2002:04
Included observations: 87
Q-statistic probabilities adjusted for 5 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. | . |	. | . |	1	-0.056	-0.056	0.2850	
.*| . |	.*| . |	2	-0.108	-0.112	1.3511	
**| . |	**| . |	3	-0.209	-0.226	5.3916	
. |*. |	. | . |	4	0.090	0.048	6.1403	
. |*. |	. |*. |	5	0.181	0.152	9.2298	
. | . |	. | . |	6	-0.003	-0.008	9.2305	0.002
. | . |	. | . |	7	-0.028	0.037	9.3061	0.010
. |*. |	. |*. |	8	0.088	0.166	10.065	0.018
. | . |	. | . |	9	-0.023	-0.032	10.117	0.039
. | . |	. | . |	10	0.019	0.011	10.152	0.071
. |*. |	. |*. |	11	0.066	0.130	10.592	0.102
.*| . |	**| . |	12	-0.147	-0.189	12.812	0.077
. | . |	.*| . |	13	-0.016	-0.070	12.838	0.118
. | . |	. | . |	14	-0.012	0.013	12.853	0.169
. |*. |	. | . |	15	0.070	-0.043	13.379	0.203
. | . |	. | . |	16	0.039	0.001	13.547	0.259
.*| . |	. | . |	17	-0.078	0.013	14.226	0.286
. | . |	. | . |	18	-0.029	-0.019	14.319	0.352
. | . |	. | . |	19	0.014	-0.003	14.342	0.425
.*| . |	. | . |	20	-0.058	-0.039	14.724	0.471
. | . |	. | . |	21	0.056	0.040	15.095	0.518
.*| . |	.*| . |	22	-0.081	-0.086	15.886	0.532
. | . |	. | . |	23	-0.024	-0.035	15.956	0.596
.*| . |	.*| . |	24	-0.138	-0.177	18.299	0.503
. | . |	. | . |	25	0.039	-0.007	18.493	0.555
. | . |	.*| . |	26	-0.038	-0.103	18.677	0.606
. | . |	.*| . |	27	-0.046	-0.105	18.947	0.649
.*| . |	.*| . |	28	-0.125	-0.086	21.013	0.580
. | . |	.*| . |	29	-0.038	-0.082	21.202	0.627
. |*. |	. |*. |	30	0.103	0.075	22.633	0.599
.*| . |	.*| . |	31	-0.122	-0.140	24.705	0.536
. | . |	. | . |	32	-0.005	0.028	24.708	0.591

Checking the residuals of the model does not help to find a significant difference between both models either.

4.2 VAR models with intervention

The best VAR models to forecast inflation for year 2000 are used to model the structural change of January 2001. To do so, an intervention variable is introduced in the models. The intervention variables are the same employed in the ARIMA model. Alternatively, one of them accounts for a permanent change and the other for a temporary one. The VAR models are the six ones shown in Table 2.

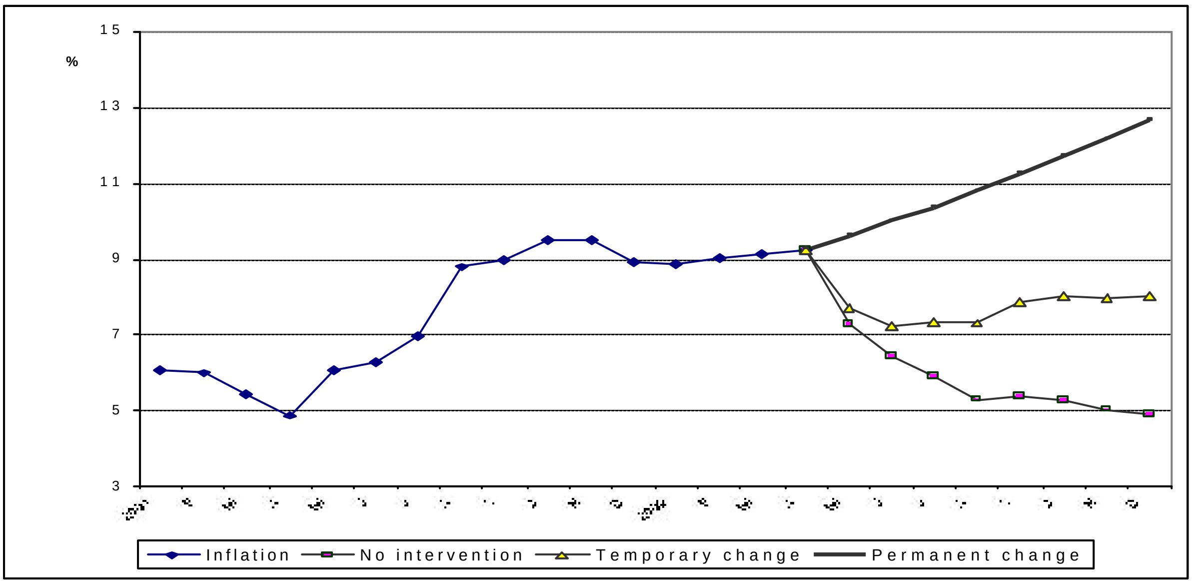
5. Forecasts for the year 2002

The main interest for the decision-makers of Banco de Guatemala is not in the historic behavior of the inflation series but in the forecasts for the year 2002. At this stage, it is already possible to follow, for forecasting purposes, the different models developed in this work. To this date, the last CPI data published is April 2002. Therefore, inflation is forecasted from May to December 2002.

5.1 Forecasts with ARIMA models

ARIMA models specified with intervention, both temporary and permanent, and also those without intervention, were run with data up to April 2002. An estimate for the rest of the year was produced, and these forecasted results are shown in the figure below.

FIGURE 12
INFLATION FORECASTS
ARIMA MODELS



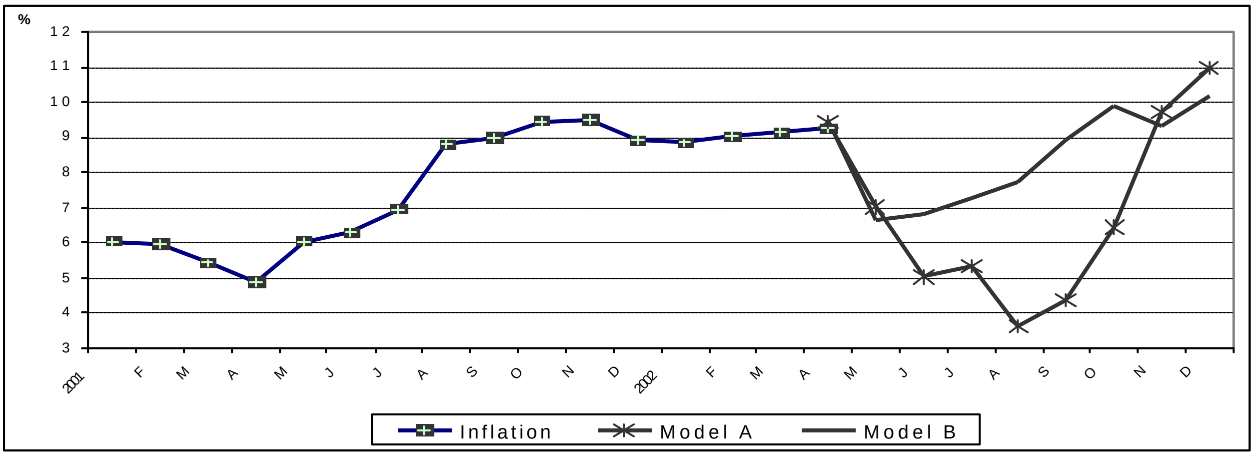
The results obtained are interesting and each one of them reflects how the model was built. The permanent-change model produces a strong trend in the forecast. The temporary-change model forecasts a strong fall in May and June because, according to this model, the effect of re-defining the CPI should have ended since January (and it should have reverted to lower first-differences in inflation rates from that month on).

The forecasts of the model without intervention are consistent with the rates of inflation registered between 1998 and 2000 (the period prior to the CPI change). However, it is not consistent with the recent history of the series (last year and the first four months of 2002). This forecast is the least probable.

5.2 Forecasts with VAR models

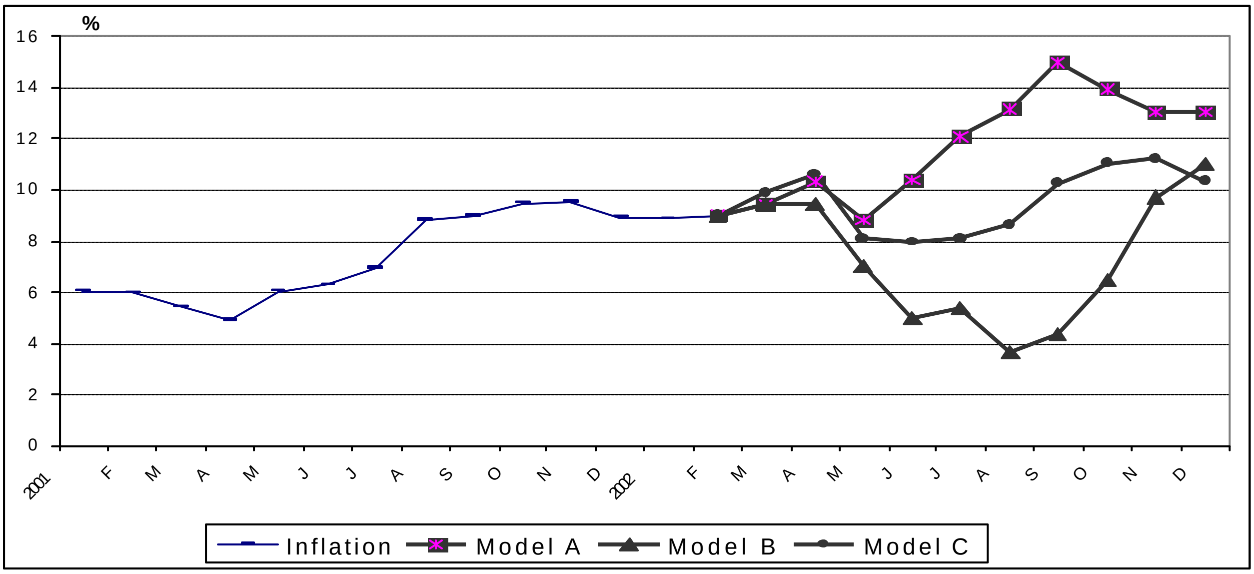
The VAR models with intervention that produced the best results were chosen to forecast the year 2000. The "best models" were chosen arbitrarily according to what seemed to be the most plausible results. Due to the lack of some information, the VAR models were estimated using data up to February 2002. Therefore, the forecasted series is from March to December 2002. These results are presented in Figures 13 and 14.

FIGURE 13
INFLATION FORECASTS
VAR MODELS WITH INTERVENTION



Where model *A* includes MIEA, exchange rate, monetary base, international oil prices and an intervention variable of temporal change. Model *B* includes MIEA, international oil prices and an intervention variable of permanent change.

FIGURE 14
INFLATION FORECASTS
VAR MODELS WITH INTERVENTION



Model *A* includes MIEA and exchange rate. Model *B* includes MIEA, EMISION, and interest rate on deposits (long term). Model *C* includes MIEA, interest rate on deposits (long term).

According to the historic development of the series, the results of model *C* could be the most plausible.

The inflation target of this year is between four and six percent; according to these results, Banco de Guatemala will not be able to reach its target.

6. Conclusions

1. There is a structural change in the inflation series in January 2001, when a new CPI with more items, updated weights and new base was adopted.

2. According to statistical evidence, VAR models can produce better forecasts of inflation in the period prior to the occurrence of the structural change.

3. Based on their ability to forecast inflation, Banco de Guatemala should monitor the following variables: the Monthly Indicator of Economic Activity, exchange rate, emision, long term deposit interest rate and the international oil prices. For policy purposes, there are some variables in the corresponding model than can be influenced by the central bank. The resultant model would be the best one. However, just for forecasting purposes, a model that includes only the Monthly Indicator of Economic Activity and the international oil prices is the best forecaster. Despite this fact, the drawback of the model would be that the central bank can do nothing to exert influence over these variables.

4. ARIMA models seem to produce better results than the VAR models for the period of structural change. The forecasts behave according to the underlying assumptions of each model. These forecasts have less variance and look more plausible than those produced with VAR models.

5. The lag length of the regressors in VAR models does not necessarily has to be determined by Schwarz or Akaike tests. Following the t-statistic significance test for each regressor proved to give better results.

6. According to the forecasting capabilities of the VAR models specified for the period prior to the structural change, these models could also be exploited with intervention analysis. It is necessary to find better ways to model the intervention variables so that the VAR models produce better results.

References

Enders, Walter. "Applied Econometric Time Series", John Wiley and Sons, Inc., New York, 1995.

Lütkepohl, Helmut. "Introduction to Multiple Time Series Analysis", 2nd. Edition, Springer-Verlag Berlin, 1993.

Hamilton, James D. "Time Series Analysis", Princeton University Press, Princeton, New Jersey. 1994.

Guerrero, Víctor M. "Análisis Estadístico de Series de Tiempo Económicas", Universidad Autónoma Metropolitana, México, 1991.

Wickens, Michael R. Course "VAR Analysis in Macroeconomics", Lecture Notes, Washington, D.C., 2002 (IMF Institute Economics Training Program).