A HIGH-FREQUENCY MODEL FOR THE MEXICAN ECONOMY

Lawrence R. Klein[*/]
University of Pennsylvania, U.S.A.

Alfredo Coutiño[*/]
Temple University, U.S.A.
CIEMEX-WEFA
July 1999

CIEMEX WEFA

[*/] Lawrence R. Klein is Emeritus Professor at the University of Pennsylvania and a Nobel Laureate in Economics 1980, and Alfredo Coutiño is Director of Macroeconomic Analysis at Ciemex-Wefa.

A HIGH-FREQUENCY MODEL FOR THE MEXICAN ECONOMY

Lawrence R. Klein [*/]
University of Pennsylvania, U.S.A.
Alfredo Coutiño [*/]
Temple University, U.S.A.
CIEMEX-WEFA
July 1999

I.- INTRODUCTION

During the last decade the development of information systems and telecommunications has made possible the immediate availability of economic information. This flow of economic indicators, on a daily, weekly or monthly basis, permits to monitor the economy more frequently.

The use of high-periodicity economic information is essential not only for forecasting purposes but also for improving the predictive accuracy of structural econometric models. This is possible through the use of a high-frequency model, which is a system of pure econometric relationships between economic and financial indicators. This model permits to forecast the quarterly GDP and its main components using high-frequency indicators, and without any subjective personal judgement.

Using monthly economic indicators, the high-frequency model is intended to forecast the quarterly GDP from three different approaches: the expenditure side, the income side, and principal components of leading economic indicators.

The first two approaches are based on the same monthly indicators that the statisticians in charge of the national accounts use to estimate the components of GDP. Thus, with these major monthly indicators we can establish the corresponding entries in the NIPA. Future values of monthly indicators are gotten from ARIMA equations.

The third approach uses a set of strategic monthly indicators which are closely related to real GDP, and another set for GDP price deflator. Then, we can extract principal components for each set of variables. Having these two sets of principal components we estimate regressions of GDP on its corresponding principal components, and the same for the GDP deflator.

[*/] Lawrence R. Klein is Emeritus Professor at the University of Pennsylvania and a Nobel Laureate in Economics 1980, and Alfredo Coutiño is Director of Macroeconomic Analysis at Ciemex-Wefa and Professor at Temple University.

At the end, we have three different estimates of quarterly GDP (expenditure, income and principal components) which are estimated independently of each other to some extent. This procedure permits us to reduce error variances in forecasting the real GDP through three independent ways, and averaging the final estimate.

For the purpose of this paper we simply present the third approach: GDP by principal components. The general idea of the principal components refers to the fact that, having a set of major important indicators, we can extract the main independent sources of variation from this set of variables. In other words, the principal components are the variables that explain the most variation of the total set of main indicators selected. Statistically, the principal components theory deals with the problem of defining a number of mutually uncorrelated (independent) variables exhibiting maximal variance. Note that all variables are standardized so that there is no change in the units of measurement.

For the case of the real GDP equation, we have selected a set of monthly indicators from industrial production, personal income, money stock, interest rates, wages, employment, etc, which are a sort of strategic indicators. Then we extract the principal components from this set of variables (can be done automatically by computer software), which represent the variables that account for the total variation of the entire set. Having the monthly principal components, we compute the 3-month average in order to get quarterly series, then we regress the quarterly GDP on its associated quarterly components. This will generate an equation to estimate the quarterly GDP as a function of quarterly indicators.

Given that the principal components represent the independent sources of variation of the major indicators, then the GDP estimate will be a function of the variation of the main economic indicators.

II.- STRATEGIC INDICATORS FOR PRINCIPAL COMPONENTS

A.- Selecting the Strategic Indicators

The principal components theory refers to the idea of extracting the main independent sources of variation from a set of major indicators. We can use these principal components as independent variables in the regression of a particular dependent variable. For example, in the case of GDP we can have a set of monthly leading indicators closely related to the behavior of real GDP, then we can extract the principal components from this set of variables and regress the GDP on them.

Given that the principal components methodology is based on pure statistical correlations among variables, it could be said that this model is an exercise in "measurement without theory"[1/]. This model, however, is just one part of three different ways to compute the GDP. The remaining two approaches of GDP (Income and Expenditure) will use one branch of economic theory which is named social accounting.

Based in a purely empirical approach, however, we can say that real GDP is closely correlated with the most important principal components derived from the total production indicators. In this way, we can choose a set of variables closely related to GDP from different sectors such as industrial production, employment, wages, sales, money, and trade. These major economic indicators must represent the monthly signals which permit us to monitor the short-run behavior of GDP.

For the case of the Mexican economy, we selected a preliminary set of fifteen variables as the most representative monthly indicators for computation of the principal components for real GDP.

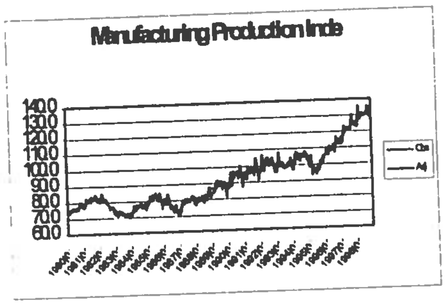
1) Manufacturing Production Index (IMI)
2) Construction Industry Index (ICI)
3) Industrial Production Index (IPI)
4) Gross Fixed Investment Index (GFII)
5) Wholesale Index (WSI)
6) Retail Sales Index (RSI)
7) Man-Hours Worked in Manufacturing Index (HOUR)
8) Average Real Wages in Manufacturing Index (WAR)
9) Employment Rate (EMR%)
10) Maquiladora Exports (MAQ)
11) Crude Oil Exports (EXOV)
12) Real Money Supply: M1 (MSR)
13) Real Interest Rate (IRR)
14) Real Exchange Rate (ERR)
15) Tourism Balance (TOUN)

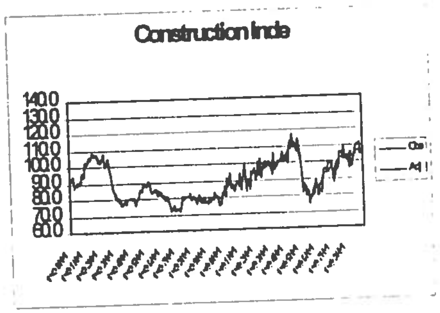
[1/] Klein (1993)

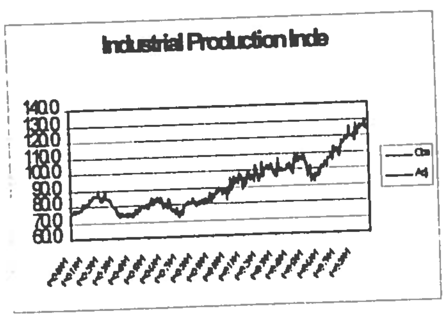
These series are released monthly by the National Institute of Statistics (INEGI) and by the Central Bank of Mexico (BANXICO) in the form of unadjusted data, that is, not seasonally adjusted. A few of them, however, are officially adjusted for seasonal.

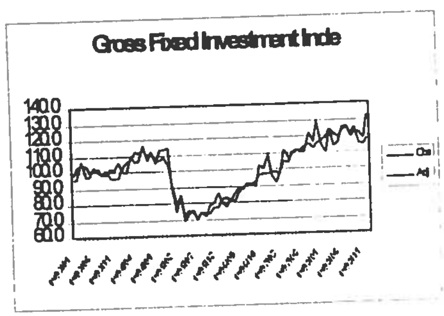
It is helpful to see the graphical behavior of each of these main indicators in order to determine whether there exists a seasonal component or a trend. The graphs show that, in most of the monthly indicators, there exist both components: seasonality and trend, in particular during the last four years. Given that we need to work with highly serially correlated indicators, seasonality could be a problem at the high-frequency level of correlated indicators, seasonality could be a problem at the high-frequency level of information. In order to deal with the seasonal problem we can work with adjusted series.

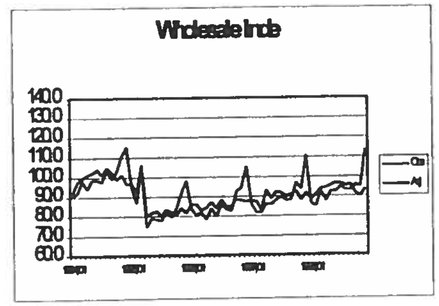
The following graphs show each series against time in both formats, observed data (obs) and seasonally adjusted (adj).

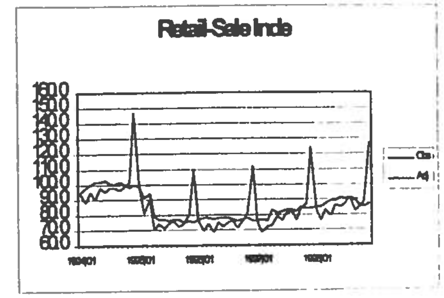


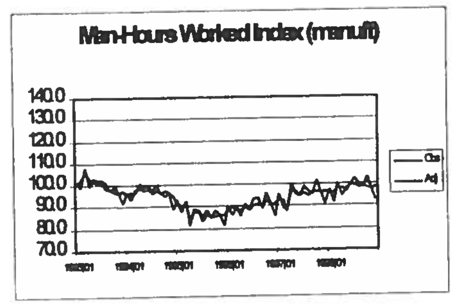


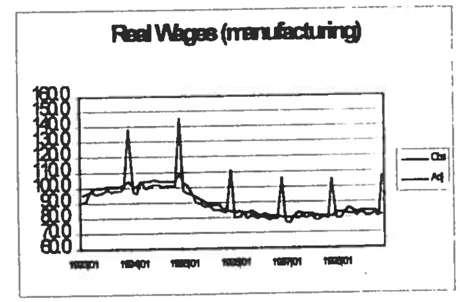


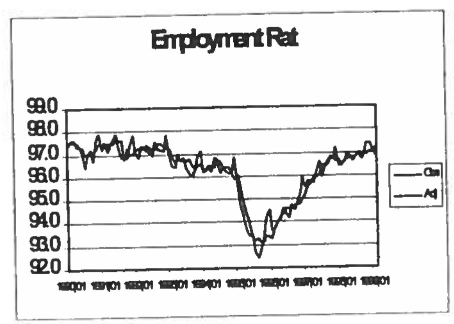


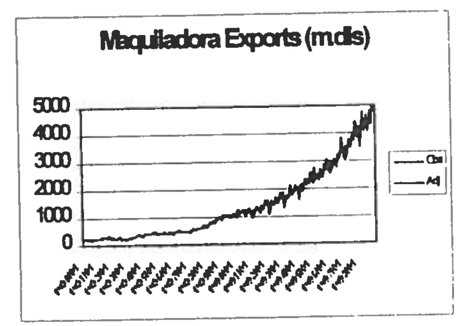


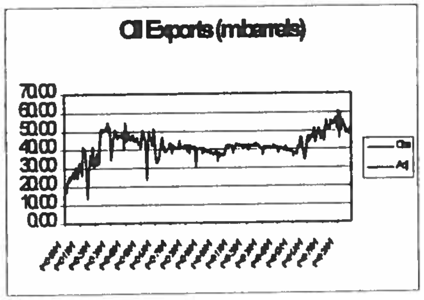


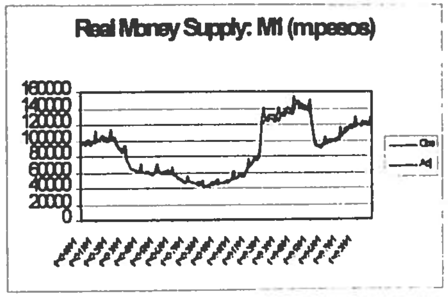


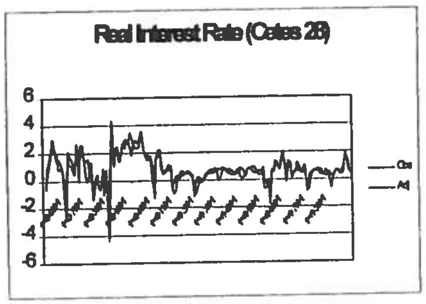


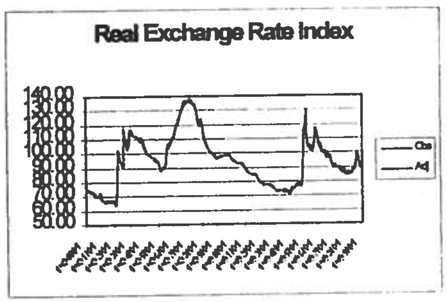


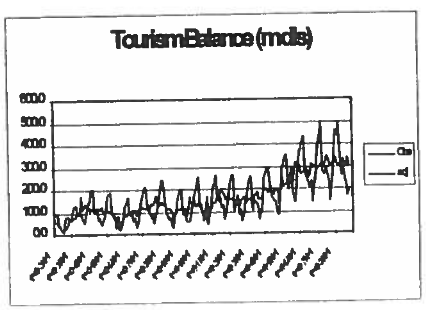


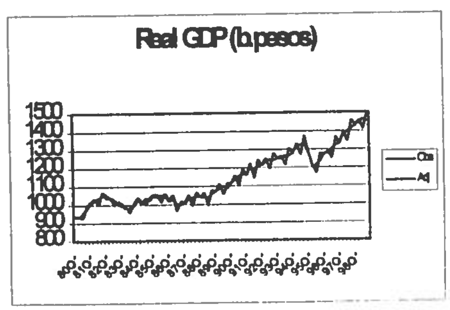












B.- Filtering the Strategic Indicators

Commonly, economic data at high-frequency intervals are highly correlated. In particular, if we select a set of major indicators, each of which has to do with a specific variable (as in the case of real GDP), we can see the intercorrelation. In this case, we selected a set of fifteen indicators which are correlated with GDP.

The presence of high serial correlation in the set of main indicators represents both a benefit and a cost. It is a benefit because it improves the forecast accuracy, but it is a cost because introduces inefficiency in parameter estimates. High correlation in economic data can be explained by two factors: seasonality and trend. As we saw in the previous graphs, almost all the indicators chosen show both seasonal movement and trend, with the exception of real interest rate and real exchange rate.

In order to deal with these two serial problems in our set of major indicators, and clean them up to get better parameter estimators, we need to filter these series. With respect to seasonality, we can use seasonally adjusted series either officially or personally adjusted. Given that most of the Mexican economic data are not seasonally adjusted, then we use the X11 methodology to adjust our leading indicators.

With respect to the second serial problem, we need to "detrend" our series, i.e. take the trend component out. This can be done by regressing each indicator on a smooth fraction of time, and then computing the residuals. These residuals represent the variation of each indicator which is not explained by the trend. Of course, for this detrending process we use the series seasonally adjusted.

The regressions were obtained using monthly data for the last four years (1995-98), after the 1994 peso crisis, when the data show more definite seasonal and trend components. Almost all the regressions show that the trend component is present and it has a high correlation with its corresponding indicator, as shown by the R squared statistic in Table 1. With the exception of the real interest rate (irr) and retail sales (rsi), the R Squared shows a correlation higher than 25%.

This filtering process of our main economic indicators, for seasonality and trend, enable us to be ready for the computation of the principal components for real GDP.

TABLE 1

Estimated Regressions with respect to Time

Equation	Estimated Coefficients Constant	Trend	t-Statistic for Trend	R-Squared
IMI	-174.0	0.08952	30.3	0.95
ICI	-149.2	0.07699	12.8	0.78
IPI	-160.8	0.08287	26.2	0.94
GFII	-296.2	0.15060	18.1	0.88
WSI	-49.3	0.02691	3.9	0.26
RSI	-47.6	0.02604	3.1	0.18
HOUR	-73.0	0.03885	11.6	0.75
WAR	65.9	-0.03077	5.3	0.38
EMR%	16.0	0.01029	10.3	0.70
MAQ	-341.0	0.17481	26.5	0.94
EXOV	-176.7	0.09042	8.7	0.63
MSR	-121.9	0.06685	7.8	0.58
IRR	-214.9	0.10787	1.2	0.03
ERR	15584.4	-7.75547	9.7	0.67
TOUN	-146.4	0.07615	6.4	0.49
GDP	-100.6	0.05400	13.0	0.92

Note: All series are seasonally adjusted and the equations were estimated in logs using OLS. All regressions use monthly data with the exception of GDP which is quarterly.

III.- THE MODEL OF PRINCIPAL COMPONENTS FOR REAL GDP

A.- Estimation of Principal Components

Having a set of strategic economic indicators, the principal component method gives an answer to the question: "How many independent sources of variation are there?"[2]. For example, in the regression of Y on a number of independent variables X_1, X_2, ..., X_m, we wish to find some linear functions of the X's which in some sense capture most of their variablility[3].

Consider the linear functions Z's of the independent variables X's[4]:

$$Z_1 = a_1 X_1 + a_2 X_2 + \dots\dots\dots + a_m X_m$$
$$Z_2 = b_1 X_1 + b_2 X_2 + \dots\dots\dots + b_m X_m$$
$$\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots$$
$$Z_m = \dots \quad \dots \quad \dots \quad \dots$$

The principal component method extracts these linear functions Z's choosing the a's and b's so that the variances of Z's are maximized. Therefore, the principal components Z's, which are the linear combinations of the variables X's, have the highest variance. Z_1 is called the first principal component with the highest variance, Z_2 is the second principal component with the next highest variance, and so on. At the end, we will have a set of m principal components that accounts for the total variation of the X's.

In addition to the fact that the principal components have the highest variance, they are also uncorrelated or orthogonal. This orthogonality is an important characteristic because it allows us to avoid the multicollinearity problem in the regression of Y on Z's, instead of Y on X's.

We need to remember that the original set of X's variables, which are the main economic indicators, was chosen as the set of most correlated variables to Y, then multicollinearity is present among the X's. The presence of multicollinearity implies that we cannot estimate individual coefficients with good precision in the regression of Y on X's. We can get, however, uncorrelated linear functions (Z's) from the X's and regress Y on Z's, and then get estimates of the coefficients that are not subject to multicollinearity.

The extraction of principal components from a set of strategic indicators involves a detailed derivation. The use of a computer program, however, facilitates the computational problem. The software used is Time Series Processor (TSP).

---

[2] Maddala (1992).
[3] Dhrymes (1970).
[4] A similar theoretical example is in Maddala (1992), p.284.

The attached tables show the TSP computational output for extracting the principal components from the set of our fifteen major economic indicators. The period used is from January 1995 to December 1998. The series are adjusted for seasonality and trend, so that they are the residuals obtained from the regressions of each indicator against time.

There are some key points to remark in the tables attached. In Table 2, we have the correlation matrix among all the fifteen indicators, in which we can see the degree of multicollinearity among the original variables. In order to avoid the multicollinearity problem in our regression we estimate linear combinations of the original variables, which are the principal components. In Table 3, we have the fifteen principal components extracted, their corresponding eigenvalues, and the cumulative R-Squared. The eigenvalue is a characteristic root of the correlation matrix. If we take an eigenvalue divided by the number of variables (15), this is exactly equal to the fraction of the variance of the original variables explained by that principal component. The R-Squared column means that if we use all the fifteen principal components, we can explain the total variation of the main indicators, it says 100%, as indicated by the 1.00 Cumulative R-Squared in the last row of Table 3. If we take component 1, however, we can explain only 84% of the total variation of the fifteen leading indicators. If we take the first two components, then we can explain 91% of the total variation, and so on. So that the first seven components account for 99.5% of the total variation, then we can omit the remaining eight components.

In addition, we present the graphs for the first seven principal components over time. In these graphs we can see that each of them shows a different pattern, which confirms the fact that the principal components are the "uncorrelated" linear combinations of the original strategic indicators.

TABLE 2
Correlation Matrix

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
V1	1.00														
V2	0.98	1.00													
V3	0.99	0.99	1.00												
V4	0.95	0.98	0.96	1.00											
V5	0.96	0.95	0.96	0.96	1.00										
V6	0.93	0.93	0.93	0.95	0.98	1.00									
V7	0.99	0.97	0.99	0.96	0.97	0.95	1.00								
V8	0.95	0.93	0.95	0.92	0.95	0.98	0.96	1.00							
V9	0.99	0.96	0.99	0.93	0.95	0.92	0.99	0.96	1.00						
V10	0.95	0.94	0.95	0.93	0.92	0.90	0.96	0.91	0.95	1.00					
V11	0.89	0.85	0.89	0.85	0.80	0.76	0.88	0.78	0.88	0.87	1.00				
V12	0.94	0.96	0.94	0.97	0.95	0.97	0.94	0.95	0.92	0.90	0.79	1.00			
V13	0.52	0.50	0.52	0.47	0.48	0.47	0.52	0.50	0.53	0.50	0.47	0.46	1.00		
V14	0.84	0.81	0.84	0.78	0.80	0.78	0.84	0.81	0.85	0.80	0.75	0.77	0.00	1.00	
V15	0.64	0.60	0.64	0.59	0.56	0.50	0.62	0.54	0.63	0.56	0.63	0.52	0.19	0.64	1.00

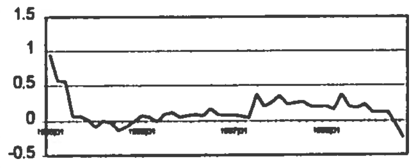
Notes: V1 = IMI, V2 = ICI, V3 = IPI, V4 = GFII, V5 = WSI, V6 = RSI, V7 = HOUR, V8 = WAR
V9 = EMR%, V10 = MAQ, V11 = EXOV, V12 = MSR, V13 = IRR, V14 = ERR, V15 =TOUN.

TABLE 3
Principal Components

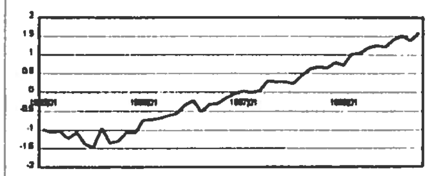
Component	Eigenvalue	Cumulative R-Squared
PC1	12.63256	0.8422
PC2	1.06762	0.9133
PC3	0.63881	0.9559
PC4	0.27925	0.9745
PC5	0.15304	0.9875
PC6	0.07904	0.9900
PC7	0.06777	0.9945
PC8	0.05148	0.9979
PC9	0.01280	0.9988
PC10	0.00796	0.9994
PC11	0.00443	0.9997
PC12	0.00304	0.9999
PC13	0.00204	0.9999
PC14	0.00009	0.9999
PC15	0.00004	1.0000

Notes: The Eigenvalue is also known as the Characteristic Root of the correlation matrix.

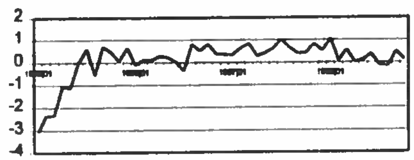
Principal Component No. 1



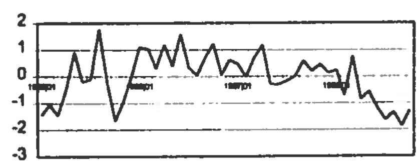
Principal Component No. 2



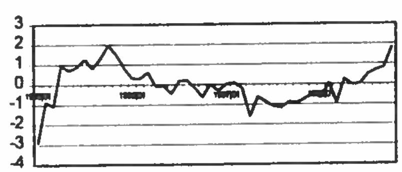
Principal Component No. 3



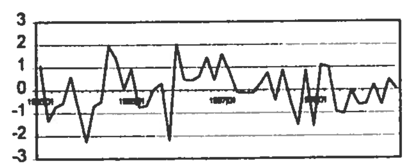
Principal Component No.4



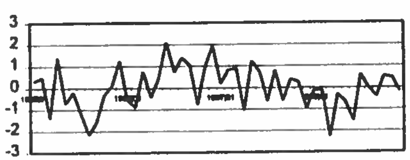
Principal Component No. 5



Principal Component No. 6



Principal Component No. 7



B.- The Equation for GDP

If we regress Y on all the principal components extracted, this is exactly equivalent to using all the original set of strategic indicators X's. We shall use just a subset of principal components Z's in the regression of Y. In other words, regress Y on the first k principal components that substantially account for almost all the variation of the X's, and omit the m-k remaining superfluous components. According to the Cumulative R-Squared we already determined that the first seven components account for almost all the variation of the original variables. Therefore, at this point we can regress the real GDP on the first seven most important principal components.

In order to estimate the regression for the quarterly GDP, we need to get quarterly data for the principal components. We do that by averaging each component per quarter. The regression covers the period from the first quarter of 1995 to the fourth quarter of 1998, and is estimated using OLS.

$$GDP = -\underset{(0.006)}{0.01603} + \underset{(4.26)}{21.9351}\,PC1 + \underset{(2.82)}{12.2111}\,PC2 - \underset{(3.48)}{14.090}\,PC3 - \underset{(0.93)}{3.7336}\,PC4$$

$$-\underset{(2.83)}{12.8152}\,PC5 + \underset{(0.35)}{1.9722}\,PC6 - \underset{(0.36)}{1.5807}\,PC7$$

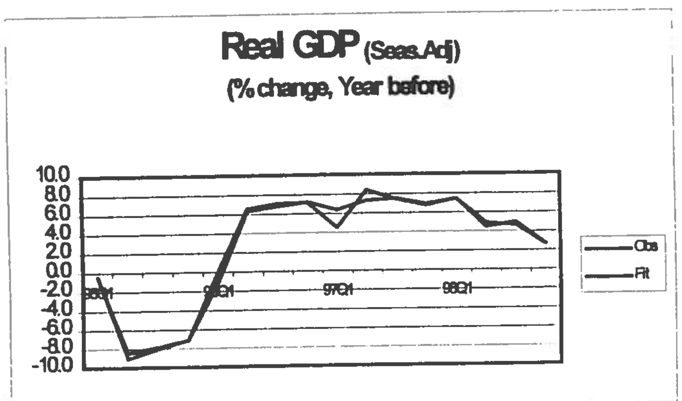
$$R^2 = 0.94 \qquad D.W. = 2.46 \qquad F_{7,8} = 17.11$$

Even though components 4, 6, and 7 are not significative in explaining the real GDP, we keep them in the equation because they all together capture most of the variation of the monthly main indicators. The fit of the regression is good, as shown by the R-Squared = 0.94, which means that all the seven principal components explain 94% of the variation of the real GDP.

Finally, in Table 4 we present the fitted values given by the regression, which show the precision of the equation in the forecast. The regression is able to reproduce the historical quarterly GDP growth rate with a high degree of accuracy, in particular during the last year. The last Graph shows the observed (obs) and fitted (fit) GDP growth rate.

TABLE 4
Real GDP
(Seasonally Adjusted)
(% change. year before)

Date	Observed	Fitted
95Q1	-0.49	-0.53
95Q2	-9.07	-8.43
95Q3	-8.05	-7.85
95Q4	-7.07	-7.18
96Q1	0.03	-1.15
96Q2	6.55	6.33
96Q3	7.08	6.78
96Q4	7.18	7.21
97Q1	4.54	6.35
97Q2	8.38	7.31
97Q3	7.38	7.48
97Q4	6.75	6.88
98Q1	7.43	7.43
98Q2	4.46	4.88
98Q3	4.94	4.63
98Q4	2.66	2.61



IV.- CONCLUSION

Using a set of fifteen strategic economic indicators, closely related to real GDP, we can compute a set of principal components which accounts for most of the variation of the original indicators, and then regress the GDP on some of them.

This methodology allows us to use high-frequency information, on a monthly basis, to estimate the quarterly real GDP for the Mexican economy. But also, it avoids the multicollinearity problem derived from the original variables, and permits to get better estimated coefficients.

With this econometric instrument we will try to monitor the Mexican economy more frequently using high-periodicity information, but also we can use it to improve the predictive accuracy of larger structural econometric models.

V.- REFERENCES

Banco de México (1999). Información Económica. México.

Dhrymes, P.J. (1970). *Econometrics, Statistical Foundations and Applications*. Harper and Row. New York.

Greene, W.H. (1997). *Econometric Analysis*. 3rd ed. Prentice Hall. New Jersey.

Hall, B. and C. Cummins (1998). *TSP Reference Manual*. Version 4.4. TSP International, California, USA.

INEGI (1999). Banco de Datos INEGI. México.

Klein, L.R. and J.Y. Park (1995). "The University of Pennsylvania Model for High-Frequency Economic Forecasting". *Economic and Financial Modeling*. Autum 1995, pp. 95-146.

Klein, L.R. and J.Y. Park (1993). "Economic Forecasting at High-Frequency Intervals". *Journal of Forecasting*, Vol. 12, pp. 301-319.

Klein, L.R. and E. Sojo (1989). "Combinations of High and Low Frequency Data in Macroeconometric Models". In Klein and Marquez (eds), *Economics in Theory and Practice: An Eclectic Approach*. Dordrecht: Kluwer, pp. 3-16.

Maddala, G.S. (1992). *Introduction to Econometrics*. 2nd ed. MacMillan, New York.

CIEMEX-WEFA, Inc.
800 Baldwin Tower
Eddystone,
Pennsylvania 19022

(610) 490-2566
Fax
(610) 490-2770