Papers ESADE núm. 155,
septiembre, 1997

BAYESIAN VECTOR AUTOREGRESSIONS

Fernando C. Ballabriga

Depósito Legal: B-4.761-1992
ISSN: 1132-7278

Index

Abstract

This paper containts both an heuristic and a formal description of the Bayesian Vector Autoregression methodology, a development within the Vector Autoregression literature that provides a solution to the problem of degrees of freedom inherent in empirical macroeconomics, refraining from the use of exclusion restrictions. The paper also presents an application of the methodology to the Spanish macroeconomy, exploring some probability aspects of the future path of the economy.

Key Words: Empirical Macroeconomics, Bayesian VARs.

1. Introduction

Vector Autoregressions (VAR) were introduced by C. A. Sims in two seminal papers. In Sims (1972) a VAR model was used to analyze the long standing issue of the existence of causality between aggregate money and income. In Sims (1980) VAR models were explicitly proposed as an alternative to standard simultaneous equations macroeconometric models, with the basic argument that these models were specified and identified using incredible economic restrictions. A reasonable alternative, Sims argued, should refrain from using controversial economic restrictions, relying on the data as the most important source of information to identify key macroeconomic interactions. In these terms, VAR models appeared as suitable alternatives.

Soon, the proposal found obstacles that have been the source of discussion and research in the VAR literature during the 80´s and the 90´s. The first was the issue of identification: as raw VAR models are reduced forms they do not go beyond correlations and so cannot be used for macroeconomic analysis. The second obstacle was actually a paradox: in the spirit of the VAR methodology is to avoid a priori exclusions, but because of its generous parameterization nature this means that VAR models are not actually operational alternatives to standard macroeconometric models, as the degrees of freedom are usually too few if the analyst includes more than a relatively reduced number of variables in the model. It is uncommon, in fact, to find VAR models including more than five or six variables.

Both of these problems have been reasonably well solved; to the extend that nowadays it is probably widely accepted that VAR methods are among the most successful innovations in empirical macroeconomics of the last two decades.

In particular, the search for a solution to the degrees of freedom paradox gave rise to the Bayesian dimension of the VAR methodology: the so called Bayesian Vector Autoregression (BVAR) models. These models are the focus of this article, which has three core sections. Section 2 provides a detailed intuitive description of the BVAR methodology. Section 3 gives the formal description. Section 4 fits a BVAR model to a set of Spanish macroeconomic indicators and explores some aspects concerning the evolution of inflation which may be relevant in the context of Maastricht convergence criteria.

2. An Heuristic Description of the BVAR Methodology

2.1. VAR models as a general frame of reference

As Todd (1984) points out, it is useful to think of the construction of an econometric model as a process that combines, in accordance with certain criteria, the historical information contained in the sample data with a priori statistical and economic information provided by the econometrician. The different modelling techniques can then be compared in terms of the type of a priori information used and the weight attributed to it.

All modelling techniques naturally require a minimum of prior information in order to be operational. At the very least they need the information for selecting a group of variables which are relevant for the purposes of the analysis and for establishing a type of algebraic relation among them. In fact, the selection of a vector *Y* of *n* components, together with the assumption that each of these components linearly depends on its own past, the past of the remaining components and a vector *Z* with *d* variables of a deterministic character (for example, a constant term or seasonal dummy variables) leads to a model which in recent years has become part of the tool kit commonly used by empirical economists:

$$Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + B_m Y_{t-m} + DZ_t + \varepsilon_t \qquad (1)$$

where t is a time index, B_i represents $n \times n$ matrices, D is $n \times d$ and ε is an n-dimensional vector of random disturbances. Because model (1) relates a vector of variables with its own past it is called a VAR, or Vector Autoregression, model. Furthermore, because a minimum of a priori restrictions are involved, the model is also referred to as UVAR (Unrestricted Vector Autoregression).

As a theoretical framework, the UVAR model is very general. Granger and Newbold (1986) indicate that if the number of lags is not restricted and the coefficient matrices are allowed to be time-dependent, the UVAR model can be used to represent any stochastic process. This generality makes the UVAR model an attractive starting point for econometric modelling and a frame of reference that reveals the type of restrictions actually incorporated in alternative models, inasmuch any simultaneous equations or time series econometric model can be nested in the representation of type (1).

2.2. The spirit of the BVAR methodology

The generality of the UVAR representation is theoretically attractive but, in practice, its generous parameterization is the source of its principal weakness. Indeed, the number of coefficients to be estimated in a model such as (1) is $n(nm+d)$, a number which increases exponentially as the dimension of the Y vector increases and proportionately with the number of lags included. So, for example, a model with five endogenous variables, four lags per variable and a constant term for each equation will contain a total of 105 coefficients to be estimated.

This is a serious problem in terms of empirical economic research, which is characterized by the existence of sample information which tends to be at once sparse and highly contaminated by random variability. In fact, it means that econometricians cannot estimate UVAR models that include more than a relatively reduced number of variables without running a serious risk of overfitting, i.e. without running the risk that their estimates will be overly influenced by noise as opposed to signal. Overfitting is most likely to occur when three circumstances coincide in empirical analysis: there are a large number of parameters to be estimated; sample information is relatively sparse, and the method of estimation is designed to explain (fit) the sample data as closely as possible (for example, the least squares method). These three circumstances certainly coincide in UVAR models when the objectives of the analysis require that the number of variables to be included is relatively high, as is usually the case, for example, when the aim is to model the macroeconomic environment of an economy. Generally speaking, UVAR models are not recommendable in this type of context.

The BVAR methodology was originally developed by Litterman (1980) and Doan, Litterman and Sims (1984) in an attempt to find an alternative to the usual solution to the problem of overfitting in UVAR models. This solution involves strict adherence to economic theory as a source of exclusion restrictions and is used in simultaneous equation structural models. In other words, the authors attempted to find a way of avoiding the influence of noise on estimates without being forced to confront the include/exclude dichotomy with the lags of the different variables, a procedure that normally does not allow the analyst to realistically express the available a priori information because it is not absolutely certain that the value of any coefficient is zero nor is the analyst absolutely ignorant of the value of the coefficients included in the model.

When looking at it from this angle, adopting a Bayesian perspective would appear to be the natural solution to the problem. That is, we could start out with a probability distribution for the model's coefficients. Without placing all the weight on a single value and without being absolutely devoid of information, this distribution would give a reasonable range of uncertainty and could therefore be altered by sample information if both sources of information were substantially different. So long as the a priori information was not too loose (uninformative) it would be altered only by the signal and not by noise, reducing the risk of overfitting.

Putting this idea into practice involves combining model (1) with an a priori probability distribution for its coefficients. This combination results in what is known as a Bayesian Vector Autoregression (BVAR) model.

2.3. Specification of a BVAR model

Without a doubt, with BVAR models the distinctive and most important characteristic of the specification process is the choice of the prior information. This information can, of course, have many sources and many different forms. The information described in this section was conceived as part of an empirical analysis of macroeconomic data and is known as "the Minnesota prior" because it was first used by econometricians at the Federal Reserve Bank in Minneapolis. The prior information used in more recent applications tends to be more elaborate, though its objective and its basic traits remain the same.

As mentioned in Section 2.2 the purpose of the a priori information is to help reduce the risk of overfitting. And this is the first aspect that must be stressed: this information is purely instrumental and as such does not pretend to be necessarily true on average. It does, however, aim to contain a realistic range of possible data-generating mechanisms, from which the analyst can select the most appropriate for explaining the variability of sample data.

The second noteworthy characteristic of the a priori information included in the BVAR model specification process is its empirical-statistical origin and its consequent lack of economic content. The information used includes three empirical regularities resulting from statistical time series analysis:

1) The hypothesis that the best forecast of the future value of a series is its current value (the random walk theory) is a good approximation to the behavior of numerous economic series.

2) The most recent lagged values of a particular variable usually contain more information about the variable's current value than do earlier lagged values.

3) The lagged values of a particular variable usually contain more information about the current value of a variable than do the lagged values of other variables.

The easiest way to formulate regularities 1-3 above is by defining independent normal distributions for each and every one of the coefficients of model (1). Aiming for an individualized specification of each and every one of the distributions would, however, cause overfitting, which is exactly what we are trying to avoid. Therefore a way to make the idea operational is to introduce a functional dependency among all the distributions and a reduced set of parameters (hyperparameters in the BVAR jargon) which make it possible to control their basic dimensions in line with regularities 1-3.

Figure 1 illustrates the density of the prior distribution we are describing here as applied to the coefficients of a representative equation of system (1) and shows how empirical regularities 1-3 above are introduced:

1) is represented by specifying a mean equal to one for the distribution of the first own lag coefficient and a mean equal to zero for the distribution of the remaining coefficients.

2) is represented by reducing the variance of the distribution of the coefficients as the lag increases. Thus, the more distant the lag the greater the certainty that its coefficient is zero.

The introduction of 3) can be observed by noting that the first own lag coefficients (row 1) have a greater variance than the lags of other variables (row 2), which makes it more certain that the value of these latter is zero.

This representation also gives an idea of the nature of the reduced set of controlling hyperparameters. Thus, one hyperparameter usually controls the value of the mean of the distribution of the first own lag coefficient; a second controls the variance of the distribution of the own lags; and a third controls the variance of the distribution of the lags of other variables. In order to avoid specifying different hyperparameters to control the variance of each lag, one usually selects a functional form by which the variance is inversely related to the size of the lag, introducing a fourth hyperparameter in order to control the speed at which the variance shrinks as the lag increases. It's normally assumed, on the other hand, that the analyst does not have any specific knowledge about the deterministic component of the model and, accordingly, an uninformative prior is used for that component (row 3 of Figure 1).

An additional hyperparameter is usually specified in order to control the degree of global uncertainty of all the coefficients. This is crucial in determining the weight to be attributed to the sample information when estimating the model. In terms of Figure 1, an increase in this last hyperparameter will cause a general increase in the variance of all the distributions. This means that the weight of prior information to sample information is reduced.

Certainly, in specific applications one might wish to control other dimensions of the prior distribution which are considered relevant for the analysis (e.g. seasonal or long-run dimensions), but the dimensions just described are common to all applications of the methodology.

The utter lack of economic content in the prior information used in BVAR analysis may seem surprising. This becomes more understandable when one recalls the instrumental nature of this information. Although instrumentality and economic content are not incompatible, the instrumental nature of the information is purer if it is not contaminated by dubious economic assumptions. It therefore seems advisable to opt for a priori economic neutrality so that a single specification can be accepted by economists whose visions of the true structure of the economy may differ greatly.

The specification of a BVAR model is completed by combining the prior information with the sample information. This is done by applying the Bayesian rule for computing the posterior distribution of the model's coefficients. The mean and the variance of this posterior distribution provide the point estimates and the variances of the different coefficients of (1), respectively.

2.4. The BVAR methodology: advantages and drawbacks

The main ideas described thus far in this section can be summed up by saying that the BVAR methodology is an option that makes the process of specifying econometric models more flexible and objective. This is most likely the principal virtue of the methodology.

BVAR models are in fact flexible enough to include a variety of information that realistically and systematically expresses the uncertainty that exists with respect to the relevant interactions in the economy under analysis. One intuitively assumes that a flexible and systematic method of imposing restrictions will lead to a more accurate extraction of the empirical regularities that underlie the variability of the sample. Should this be the case, the estimates obtained will be more accurate, although they could be biased, than the estimates obtained when applying the least square method, and consequently the final estimated model would in principle be of better quality. And all obtained by means of an objective process which uses explicit statistical mechanisms that can be perfectly reproduced and are able to characterize the probable future evolution of the modelled variables.

The available evidence[1] actually confirms that BVAR models successfully compete as forecasting tools. However, in contrast with most models built along the lines of the Box and Jenkins tradition, they do it from a multivariate perspective. This is an important aspect that must be emphasized. It means that BVAR applications have a primary interest in extracting stable interactions among the variables analyzed, a necessary condition to attempt economic interpretation. In fact, a BVAR model may aim to go beyond the "black box" stage by attaching an economic interpretation to a set of stable interactions and thus be of some use in, for example, projecting the potential impact of certain economic policy measures. In such applications the models aim to compete with multivariate methods of a structural nature. We can therefore view BVAR methods as a bridge between pure forecasting time series models and simultaneous equations interpretable models.

The question that immediately comes to mind is: where do you get a possible economic interpretation when the model has been described as utterly lacking in economic content? And the reply is that, if you want to extract an economic interpretation of the analysis, the BVAR model must in fact be complemented by an additional set

---

1 Among the most noted works are Litterman (1986), McNess (1986), Runkle (1990) and Artis and Zhang (1990). Moreover, Canova (1995) includes a quite exhaustive review of the literature on the forecasting accuracy of VAR methods.

of economic restrictions. This brings us to the point where the model is identified, a point which, in VAR methodology in general and BVAR methodology in particular, occurs after the model is specified and involves applying a minimum set of economic restrictions which intends to be as uncontroversial as possible[2].

Leaving the identification of the model until last enables us to clearly separate the restrictions used when specifying the probabilistic mechanism of the model from those aimed at giving it economic content. However, the fact that identification is based on a minimum set of restrictions means that the economic interpretation is much less clear than the interpretation of an empirical model derived from a set of explicit theoretical hypotheses. This lack of clear interpretation is probably the biggest drawback to the VAR (UVAR or BVAR) methodology; it is the opportunity cost of a model that is closer to the stable regularities of the sample data.

---

2 Since it is not a BVAR-specific issue we will not deal with the identification problem in this article. Pioneer work in the literature on identification in VAR analysis can be found in Bernanke (1986), Blanchard and Watson (1986), Sims (1986) and Blanchard and Quah (1989). In the bayesian framework, Ballabriga, Sebastián and Vallés (1995) and Leeper, Sims and Zha (1996) estimate identified BVAR models.

3. A Formal Description of the Methodology

3.1. The framework

For exposition purposes we find convenient to use the conventional notation in regression analysis. Thus, we write model (1) as follows

$$\underset{n \times 1}{Y_t} = X_{t-1}\,\beta + \underset{n \times 1}{\varepsilon_t} \tag{2}$$

where

$$\underset{n \times nk}{X_{t-1}} = \begin{pmatrix} X'_{1t-1} & 0' & \cdots & 0' \\ 0' & X'_{2t-1} & \cdots & 0' \\ \cdots & \cdots & \cdots & \cdots \\ 0' & 0' & \cdots & X'_{nt-1} \end{pmatrix}$$

$$\underset{k \times 1}{X_{it-1}} = \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \cdot \\ \cdot \\ \cdot \\ Y_{t-m} \\ Z_t \end{pmatrix} \qquad i = 1, \ldots, n$$

$$\underset{nk \times 1}{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{pmatrix}$$

Each 0-block in X_{t-1} is a *k*-dimensional vector of zeros, $k = nm + d$, with *n* being the number of endogenous variables, *m* the number of lags, and *d* the number of deterministic variables. The subvectors β_i, $i = 1, ..., n$, are also $k \times 1$ and contain stacked the *ith* rows of the coefficient matrices B_l, $l = 1, ..., m$, and D. The " ' " sign here and throughout indicates transposition.

From a Bayesian perspective β is a random vector. Therefore, a complete description of the stochastic behavior of Y_t conditional on X_{t-1} requires explicit assumptions about β and ε_t. The following will be assumed for every t

$$\begin{aligned} &\beta \mid X_{t-1} \sim N(\bar{\beta}_{t-1}, \Omega_{t-1}) \\ &\varepsilon_t \mid X_{t-1} \sim N(0, \Sigma) \qquad (2a) \\ &\beta \text{ and } \varepsilon_t \text{ independent} \end{aligned}$$

As it will be seen in what follows, the set of assumptions (2a) will allow us to exploit the convenient gaussian framework to come out with a flexible formal model for incorporating prior information into the analysis.

To be more concrete, let us start by noting that under a Bayesian perspective the problem of specifying our econometric model boils down to the problem of obtaining a posterior distribution $\beta \mid X_{t-1}, Y_t$ from a prior distribution $\beta \mid X_{t-1}$. We will first focus on the process of obtaining the posterior distribution, leaving for latter discussion the choice of the prior information[3].

A convenient two-steps strategy to obtain the posterior distribution is to characterize first the joint distribution of Y_t and β and proceed then to condition on Y_t. Specifically, from (2) and the last two assumptions in set (2a) we can write

$$Y_t \mid X_{t-1}, \beta \sim N(X_{t-1}\beta, \Sigma) \qquad (3)$$

On the other hand, according to standard probability rules the joint density of β and Y_t conditional on X_{t-1} is given by the product of the densities in (3) and in the first line of set (2a). This joint density turns out to be multivariate normal. To see this, observe that for any real vector *c*, $c \in IR^{n+nk}$, we can write[4]

---

3 From a strict bayesian standpoint Σ should be also part of the specification problem; i.e. the problem should be to obtain a posterior $\beta, \Sigma \mid X_{t-1}, Y_t$ from a prior $\beta, \Sigma \mid X_{t-1}$. For the most part, however, the bayesian VAR literature has proceeded conditioning on Σ and focusing the attention on the coefficient vector β. We will stick to this framework.

4 The argument applies the following characterization of the multivariate normal distribution: a p-dimensional random vector V is multivariate normal if and only if c'V is normal for any $c \in IR^p$.

$$
\begin{aligned}
c' \begin{pmatrix} Y_t \\ \beta \end{pmatrix} &= c_1' \, Y_t + c_2' \, \beta \\
&= c_1' \, (X_{t-1} \, \beta + \varepsilon_t) + c_2' \, \beta \\
&= (c_1' \, X_{t-1} + c_2') \, \beta + c_1' \, \varepsilon_t
\end{aligned} \tag{4}
$$

where (2) and the appropriate partition of c in c_1 , $n \times 1$, and c_2 , $nk \times 1$, have been used. The set of assumptions (2a) guarantees that, conditioning on X_{t-1} , (4) is a linear combination of independent normals, and therefore normal; which in turn proves the joint normality of Y_t and β. The characterization of their density function is straightforward from (2) and the first assumption in (2a). Specifically, we have

$$
\left(\begin{matrix} Y_t \\ \beta \end{matrix} \,\middle|\, X_{t-1} \right) \sim N \left(\begin{pmatrix} X_{t-1} \bar{\beta}_{t-1} \\ \bar{\beta}_{t-1} \end{pmatrix}, \begin{pmatrix} X_{t-1} \Omega_{t-1} \, X'_{t-1} + \Sigma & X_{t-1} \Omega_{t-1} \\ \Omega_{t-1} \, X'_{t-1} & \Omega_{t-1} \end{pmatrix} \right) \tag{5}
$$

From (5) we can then obtain the posterior distribution for β using the well-known properties of the multivariate normal to condition on Y_t . We specifically get that

$$
\beta \mid X_{t-1} \, , \, Y_t \sim N \, (\, \bar{\beta}_t \, , \, \Omega_t) \tag{6}
$$

where

$$
\bar{\beta}_t = \bar{\beta}_{t-1} + \Omega_{t-1} \, X'_{t-1} \, M \, (Y_t - X_{t-1} \, \bar{\beta}_{t-1})
$$

$$
\Omega_t = \Omega_{t-1} - \Omega_{t-1} \, X'_{t-1} \, M \, X_{t-1} \, \Omega_{t-1}
$$

$$
M = (X_{t-1} \, \Omega_{t-1} \, X'_{t-1} + \Sigma)^{-1}
$$

Thus, starting from a set of sample observations Y_t , $t = 1, \ldots, T$, a prior information for β expressed in the form of a normal density for $\beta \mid X_0$ and interpreted as conditional on presample information, and a covariance matrix Σ, we can effectively use the updating scheme specified in (6) to come out with a posterior distribution for $\beta \mid X_T$ whose mean and variance provide, respectively, the point estimates and the standard errors of the coefficients of our model.

3.2. Time-variation

The analysis has proceeded so far under the assumption that the coefficient vector β has a time-invariant distribution which each new sample observation helps to estimate with a higher precision. But quite often the analyst may believe that the sample contains structural shifts. This belief can be incorporated into the prior information set by allowing a time-varying distribution for β.

The prior possibility for time-variation is a standard feature of BVAR models which increases the flexibility of the specification process and also provides a convenient mechanism to account for potential in-sample structural shifts without having to model explicitly the source of the shift (e.g. a policy regime change).

The most common implementation of time-variation specifies a first order autoregressive law of motion for β. This law of motion seems a priori sufficient to capture possible linear shifts of β. At the same time it keeps the analysis within the convenient gaussian framework. In fact, the framework described in 3.1 can be extended quite easily to accomodate this type of time-variation. Specifically, model (2) will now read

$$Y_t = X_{t-1}\,\beta_t + \varepsilon_t \tag{7}$$

where

$$\beta_t = \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{nt} \end{pmatrix}$$

The attached time index indicates that the stochastic properties of the coefficient vector are now time-dependent. As a consequence, the characterization of the stochastic behavior of Y_t conditional on X_{t-1} requires to extend the set of assumptions (2a) in order to make explicit this time-dependence. To be concrete, we use the following extended version of (2a)

$$\beta_t = A\beta_{t-1} + u_t$$
$$\beta_{t-1} \mid X_{t-1} \sim N(\bar{\beta}_{t-1}, \Omega_{t-1})$$
$$u_t \mid X_{t-1} \sim N(0, \varphi) \qquad (7a)$$
$$\varepsilon_t \mid X_{t-1} \sim N(0, \Sigma)$$
$$\beta_{t-1},\ u_t \text{ and } \varepsilon_t \text{ independent}$$

where A and φ are *nk x nk* matrices.

The prior distribution for β_t conditional on X_{t-1} (the analogous to the first line in (2a)) is now obtained by combining the first three and last lines of set (7a) to get

$$\beta_t \mid X_{t-1} \sim N(\beta^*_{t-1}, \Omega^*_{t-1}) \qquad (8)$$

where

$$\beta^*_{t-1} = A\,\bar{\beta}_{t-1}$$
$$\Omega^*_{t-1} = A\,\Omega_{t-1}\,A' + \varphi$$

On the other hand, the analogous to (3) is obtained from (7) and the last two assumptions in (7a). That is

$$Y_t \mid X_{t-1}, \beta_t \sim N(X_{t-1}\beta_t, \Sigma) \qquad (9)$$

Finally, using (7), (7a), (8) and (9) and following exactly the same line of reasoning that led us to (6) we come out with the corresponding posterior distribution for β_t. Specifically, we can write

$$\beta_t \mid X_{t-1}, Y_t \sim N(\bar{\beta}_t, \Omega_t) \qquad (10)$$

where

$$\bar{\beta}_t = \beta^*_{t-1} + \Omega^*_{t-1}\,X'_{t-1}\,M\,(Y_t - X_{t-1}\,\beta^*_{t-1})$$
$$\Omega_t = \Omega^*_{t-1} - \Omega^*_{t-1}\,X'_{t-1}\,M\,X_{t-1}\,\Omega^*_{t-1}$$
$$M = (X_{t-1}\,\Omega^*_{t-1}\,X'_{t-1} + \Sigma)^{-1}$$

It should be noted that this more general time-varying framework delivers as a particular case the time-invariant setting described in 3.1 when A = I and $\varphi = 0$, in which case (7a) reduces to (2a) and (10) to (6).

3.3. The prior information

As it was already mentioned in reference to (6), for the updating scheme described in (10) to be operational we need in $t = 1$, the first sample period, an initial specification of Σ and of the distribution in (8), which in turn requires to specify matrices A, φ and Ω_0, and a vector β_0. This initial specification is what defines the prior information set of the model. In section 2 we carefully described the basic characteristic of this information. Our objective now is just to give a formal transcription of those basic traits.

Starting with vector $\bar{\beta}_0$, we take the following specification

$$\bar{\beta}_{i0} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \Upsilon_1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \qquad i = 1, \dots, n \tag{11}$$

where the hyperparameter Υ_1 is in the *ith* position and represents the prior mean of the first own lag coefficient of the dependent variable in equation i. Coefficients for lags other than first-own have prior mean equal to zero.

The prior information usually makes the individual components of β_0 independent of each other; i.e. it makes the covariance matrix Ω_0 diagonal. The diagonal elements are defined as follows

$$\sigma_{ij}^2(l) = \left(\frac{\Upsilon_2}{l^{\Upsilon_4}}\right)\sigma_{\varepsilon_i}^2 \qquad i = 1, \dots, n \; ; \; i = j \; ; \; l = 1, \dots, m$$

$$\sigma_{ij}^2(l) = \left(\frac{\Upsilon_2\,\Upsilon_3}{l^{\Upsilon_4}}\right)\left(\frac{\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_j}^2}\right) \qquad i = 1, \dots, n \; ; \; i \neq j \; ; \; l = 1, \dots, m \tag{12}$$

$$\sigma_{ih}^2 = \Upsilon_2\,\Upsilon_5\,\sigma_{\varepsilon_i}^2 \qquad i = 1, \dots, n \; ; \; h = 1, \dots, d$$

where i represents equation, j endogenous variable, l lag, and h deterministic variable. Υ_2 controls the prior overall tightness, determining the global degree of uncertainty with which the prior information is incorporated in the specification process; by allowing Υ_2 to approach infinity the prior is made diffuse. Υ_3 controls the tightness of own lags relative to the tightness of lags of the other variables in the equation; in the limiting case, when Υ_3 equals zero, the prior defines a set of n univariate AR (m) processes. Υ_4 controls the lag-decay in the prior variance. Υ_5 controls the degree of uncertainty with respect to the coefficients of the deterministic variables. Finally, $\sigma^2_{\varepsilon_i}$ and $\sigma^2_{\varepsilon_j}$ are the diagonal elements of Σ, taken to be a measure of the scale of fluctuations in variables *i* and *j*. Their role in the prior is twofold: on the one hand, they allow to compare the degree of prior uncertainty relative to the scale of fluctuations; on the other hand, they introduce a correction for potential differences in the units of measurement of variables.

The hyperparameterization of Σ itself is of course a possibility. However, as we have already mentioned in 3.1, the Bayesian VAR literature has concentrated[5] on applications that condition on Σ. It has been common in practice to estimate Σ from a set of univariate AR (m) models.

The prior time-varying features of the model are determined by the matrices A and φ, which we specify as follows

$$\begin{aligned} A &= \text{diag}\,(A_1, \ldots, A_n) \\ \underset{K\times K}{A_i} &= \text{diag}\,(\Upsilon_6) \qquad i = 1, \ldots, n \\ \varphi &= \text{diag}\,(J_1, \ldots, J_n)\,\Omega_0 \\ J_i &= \text{diag}\,(\Upsilon_7) \qquad i = 1, \ldots, n \end{aligned} \tag{13}$$

where Υ_6 controls the coefficients of the first-order autoregressive law of motion for β and Υ_7 the amount of time variation actually introduced in the model. Observe that if $\Upsilon_6 = 1$ and $\Upsilon_7 = 0$ we get the time-invariant model. Observe also that for convenience time variation is proportional to the prior covariance matrix of the coefficient vector β_0 , which allows a relative assessment of the amount of time variation.

Obviously, as it stands, the specification of the prior information is incomplete, as it contains an unknown hyperparameter vector $\Upsilon = (\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_7)'$. From a strict Bayesian standpoint our prior information should not contain unknown (hyper) parameters; we should also attach a prior distribution to them. A full Bayesian implementation would actually require to specify that distribution and then go through the appropriate integration process to obtain the posterior distribution. However, two shortcuts to this computationally demanding procedure are usual practice in the BVAR literature.

The first is to use the posterior distribution associated with a setting of Υ which directly reflects the empirical rules of thumb concerning time series behavior that were described in section 2. For instance,

---

5 Sims (1989) is probably the only exception.

$$\Upsilon = (1, 0.2, 0.5, 1, 10^6, 1, 0.001) \tag{14}$$

This approach was a characteristic of the early BVAR applications, and formally amounts to assuming that Υ is a degenerate random vector with probability weight one on the specific choice.

The second approach calls for a "fine tuning" of the prior information. What this means is to use the posterior distribution associated with a setting of Υ which has been selected according to some goodness of fit criteria. Two commonly used criteria are to minimize a loss function defined in terms of forecasting performance statistics and to maximize the sample likelihood of the model.

Focusing on the latter, observe that the sample likelihood of our model can be written as follows

$$\prod_{t=1}^{T} L\,(Y_t \mid X_{t-1}, \Sigma, \gamma) = \tag{15}$$

$$= (2\pi)^{-T/2} \prod_{t=1}^{T} |\sigma_{t-1}|^{-1/2} exp\left(-1/2\,(Y_t - X_{t-1}\beta^*_{t-1})'\,\sigma^{-1}_{t-1}(Y_t - X_{t-1}\beta^*_{t-1})\right)$$

where

$$\sigma_{t-1} = X_{t-1}\,\Omega^*_{t-1}\,X'_{t-1} + \Sigma$$

So under the likelihood criteria we would choose Υ to maximize (15) and then compute the posterior associated with that hyperparameter setting. The Bayesian rational for this criteria is that it provides a sensible approximation to the full Bayesian integration process. Specifically, with a diffuse prior for Υ, the posterior would be a weighted average of the posterior associated with each setting for Υ, with weights given by the corresponding likelihood value of that setting. Thus, by choosing the posterior associated with the maximum likelihood value of Υ we actually focus on the posterior with the highest weight in the integration process. In the case where the Υ settings with high likelihood values have similar posteriors the approach will deliver a posterior closer to the true full Bayesian posterior.

4. Application: A Small Macroeconometric Model for the Spanish Economy

As an illustration of the use of the Bayesian VAR methodology we will now fit a model to a small set of quarterly key indicators of the Spanish macroeconomy during the period 1970:1-1995:4. The selected variables measure the money stock, the price and wage levels, and real output and employment levels. Thus, we include three nominal and two real indicators *(n=5)*. The appendix gives more detailed information about the data.

The presentation below will be divided in two parts. First, we will assess the goodness of fit properties of the model. With that purpose some within sample statistics will be computed, and the unrestricted (UVAR) version of the model will be also estimated as a benchmark for comparison. Second, we will perform some stochastic simulations of the model under alternative scenarios for the period 1996-97, focusing our attention on the likely evolution of the inflation rate.

4.1. Goodness of Fit

Preliminary unrestricted regressions showed that the inclusion of four lags of each of the five endogenous variables was sufficient to deliver a stochastic structure for the error term compatible with the white noise hypothesis. Based on this, we set the prior number of lags of the model equal to four *(m=4)*. Also, alongside with the endogenous component, the model includes a deterministic component with a constant term and a set of seasonal dummies *(d=4)*. All the endogenous variables are logged.

The prior distribution for the model's 120 coefficients was defined in terms of a 7-dimensional hyperparameter vector. The components of this vector are described in Table 1. The Table also contains the maximum likelihood setting of the vector (i.e. the value that maximizes (15)) for the period 1970-90. Noteworthy is the combination of a small overall tightness (Υ_2) with a quite large value for the component that controls the tightness of the coefficients of other variables' lags (Υ_3). Given the zero prior mean for coefficients other than those of the first own lags (whose prior mean is .96), this combination actually implies that cross-effects have more prior weight than higher than one own lags, which is an a priori signal of the existence of useful interactions to explain our data set sample variabiiity. We should also mention that the prior variance shrinks at lag speed (Υ_4), and that there is a moderate amount of prior time variation (Υ_7). Observe finally that the deterministic component has the largest prior variance of the model (Υ_5).

It will have been noticed that the maximization period for the likelihood has been 1970-90 rather than the entire sample period. The reason is that, besides the likelihood, we also want to assess the fit of the model in terms of forecasting performance, and selecting Υ

as a function of the whole sample could give rise to the criticism that a better forecasting performance would just be reflecting the optimal prior selection of the model[6].

Table 2 reports the likelihood of the model estimated with the prior information described in Table 1 along with the values of two forecasting performance statistics: MSE4 and MSE8. These statistics are define, respectively, as the weighted average of each variable mean square forecasting error at forecasting steps (quarters) 4 and 8, with the weights given by the inverse standard deviation of the corresponding error term. Thus, they define an average forecasting error for the model. Although computed with sample information, these statistics actually reflect out-of-sample-type of performance, as they are computed through an iterative process that updates the coefficient estimates and then makes forecasts for the range of sample information not incorporated in the updated coefficients.

Observe that the statistics have been caculated for the period 1980-95. We think of this choice as appropriate and "fair" for several reasons. First, it contains an acceptable number of error observations to compute the corresponding averages. Second, it includes the subperiod 91-95 that has not been used in the selection of the prior information. And, finally, it generates the first forecasting errors once ten years of data have been incorporated into the coefficient estimates, a fair amount of time for the unrestricted version of the model (UVAR) to get a good feeling of data patterns. Table 2 also reports the goodness of fit measures for the UVAR model.

As can be seen in the Table, the BVAR performs clearly better than the UVAR in both the likelihood and the forecasting dimensions. The forecasting improvement is specially interesting as bad forecasting performance often signals overfitting. Thus, we may claim some success of our BVAR specification process in dealing with the probable overparameterization problem of the UVAR version. The merit is higher if we consider the fact that the estimated model is a small one; actually a size that UVAR models may handle relatively well.

The statistics reported in Table 2 do not give information about the relative forecasting performance of each equation. Table 3 complements Table 2 with the single equation Theil-U forecasting performance statistics for steps 1, 4 and 8. This statistic is define as the ratio between the mean square forecasting error for the estimated equation and the mean square forecasting error for the random walk model at the corresponding step. A value less than one means, therefore, that the estimated model outperforms the random walk forecast for the variable. As can be seen in the Table, both the BVAR and the UVAR models outperform the random walk hypothesis (except for employment at step 8), and the BVAR actually provides a

---

6 Notice, however, that, in purity, this criticism would be fully legitimate only if the prior were selected to optimize a function of the forecasting statistics that are later used to assess the forecasting performance of the model, which is not the case here.

clear improvement over the UVAR version in basically all the equations and steps.

As an additional and final piece of goodness of fit evidence we present in Figure 2 the answer of the BVAR and UVAR models to a question cocerning a specific sample episode: the recession of the first half of the 90s.

As it is well known, the real output growth was negative for the Spanish economy in 1992:4 and during the four quarters of 1993. Say we define a recession as two periods in a row of negative growth. It is then interesting to ask what was[7] in 1992:4 the projected probability of recession for 1993. Figure 2 plots the BVAR and UVAR answers obtained from a stochastic simulation[8] of each model. As it is evident from the Figure, the projected probability is similar in 1993:1 (0.91 the UVAR and 0.88 the BVAR), but while it decreases substantially in the UVAR case (0.52 in 1993:4), the BVAR probability is always greater or equal than 0.75.

I.2. Projections

The type of exercise reported in Figure 2 illustrates an important feature of VAR methods: as VAR models always incorporate a complete stochastic description of the included variables, they invite the analyst to characterize aspect of the future stochastic evolution of the economy, as opposed to just performing deterministic simulations and focusing the discussion on statistically meaningless point forecasts. In fact, the interesting questions about the future generally involve probability statements.

In this spirit, we present in Figures 3 and 4 the 96% levels and growth rates projected bands of our BVAR for 1996 and 1997, with information up to 1995:4. Tables 4 and 5 contain the corresponding numerical values for a set of selected periods. Overall, these unconditional[9] projections look quite reasonable. Noteworthy is the rather large range of uncertainty they imply about the future, a standard feature of economic projections that highlights the covenience of thinking and talking in probability terms.

One aspect of the projections that has drawn the attention of economic agents, in general, and of policymakers in particular is the likely evolution of the inflation rate of the economy; an indicator explicitly included in the Maastricht convergence criteria to qualify as a participant in the European Monetary Union process. We now focus on this indicator.

The models for this exercise are estimated with data up to 1992:4.

These and the projections presented below are all obtained from stochastic simulations involving a thousand draws from the error term component of the corresponding model.

In the VAR jargon, the term "unconditional" refers to projections conditional on just the sample information, whereas the term "conditional" is used when the projection is made subject to constraints on the future path of the economy.

The end of 1995 official target for the end of 1996 inflation rate was 3.5%. A first interesting question then is how likely was to meet that target according to our model. Based on the unconditional projection presented in Figure 4 we actually get a probability of 0.17 of ending the year with an inflation rate greater than 3.5%. Thus, the target was highly likely[10]. This probability is reported in Table 6 along with other probabilities to which we will refer next.

How about the projected probability for the end of 1997? Currently (end of 1996) it is widely considered as reasonable to target a rate of 2.5%. Again based on the projection in Figure 4 (and so with information up to 1995:4) our model´s unconditional probability for the event that the end of 1997 inflation rate will be above 2.5% is 0.44; rather large actually.

Can something be done to reduce this probability? Well, let us compute some alternative conditional projections of the model (see Table 6).

Consider first a projection which constraint the wage rate to grow at a rate of 3.5% during 1996:1,2 (which roughly coincides with published provisional data) and at a rate of 4.5% from 1996:3 to 1997:4. Under this scenario the end of 1997 inflation rate will be above 2.5% with probability 0.70. However, if we keep the rate of 3.5% for 1996:1,2 and constraint the wage to grow at 3% during 1996:3,4 and 2% during 1997 that probability is reduced to 0.15.

We can also consider a scenario characterized by a deceleration of money growth. For instance, suppose we impose the observed growth rates for money during 1996:1,2,3 (see Table 6) and a rate of 6% from 1996:4 to 1997:4. The result is that the end of 1997 probability range for inflation is basically unaffected by the condition, and so we estimate it will be above 2.5% with probability 0.46.

What do these projections suggest? Policymakers have been asking lately for moderate wage increases. In their mind they may still consider that wage indiscipline would have to be offset by a tighter monetary policy. According to our model wage moderation certainly is a crucial determinant factor of the future path of inflation. But a deceleration of money growth does not seem to be an effective strategy.

Some clues for a possible interpretations of these results can be found in Figures 5 and 6. They contain the 96% projected effects on the complete system of the two "anti-inflation" alternatives considered: wage and money growth deceleration. The wage strategy shifts down the projected probability structure for wage an price rates, implying a very slight shift downward of the real wage projection and upward of the output and employment projections. The money growth strategy, however, leaves basically unaffected the wage and price projections, but clearly shifts downward the probability projections for output and employment growth rates.

---

10 At the time we are writing this (December 96) the inflation rate is actually running below 3.5%.

Thus, the money strategy not only appears as ineffective in terms of price effects, but also costly in terms of real growth.

All this points to the conclusion that the average behavior of the modeled economy during the period 1970-1995 has been characterized by a rather autonomous wage evolution with clear impact on the price evolution; an autonomous behavior that clearly limits the impact of monetary policy on prices.

Of course, this is tentative, as we have not explored the identification of the model. But this empirical evidence is consistent with the view that wage growth moderation would be an effective anti-inflation strategy that could also incentivate aggregate economic activity at, if any, a very slight cost in terms of average purchasing power loss.

5. Concluding Remark

Objectivity, reproducibility, systematization and statistical completeness. These words summarize the most relevant characteristics of the Bayesian methodology described in this paper; a methodology that above all views empirical macroeconomic analysis as an activity closer to science than to art.

References

ARTIS, M.; ZHANG, W.: "BVAR Forecasts for the G-7". *International Journal of Forecasting*, 1990, 6, pp. 349-362.

BALLABRIGA, F.C.; SEBASTIÁN, M.; VALLÉS, J.: *European Asymmetries*. Barcelona: ESADE, 1995. (Papers ESADE, 141).

BERNANKE, B.: "Alternative Explanations of the Money Income Correlation". *Carnegie-Rochester Conferences Series in Public Policy*, 1986, 25, pp. 49-100.

BLANCHARD, O.; WATSON, M.: "Are Business Cycles All Alike?". In GORDON, R. (ed.). *The American Business Cycle*. Chicago: University of Chicago Press, 1986.

BLANCHARD, O.; QUAH, D.: "The Dynamic Effects of Aggregate Demand and Supply Disturbances". *American Economic Review*, 1989, 79, pp. 655-673.

CANOVA, F.: "Vector Autoregressive Models: Specification, Estimation, Inference and Forecasting". In PESARAN; HASHEM; WICKENS (eds.). *Handbook of Applied Econometrics*. Blackwell Publishers Ltd., 1995.

DOAN, T.; LITTERMAN, R.; SIMS, C.: "Forecasting and Conditional Projections Using Realist Prior Distributions". *Econometric Review*, 1984, 3 (1), pp. 1-100.

GARCÍA PEREA, P.; GÓMEZ, R.: *Elaboración de series históricas de empleo a partir de la Encuesta de Población Activa (1964-1992).* Madrid: Bank of Spain, 1994. (Working Paper, 9409).

GRANGER, C.W.J.; NEWBOLD, P.: *Forecasting Economic Time Series*. Academic Press, 1986.

LEEPER, E.; SIMS, C.; ZHA, T.: "What Does Monetary Policy Do?". *Brooking Papers of Economic Activity*. [Forthcoming].

LITTERMAN, R.: "Techniques for Forecasting with Vector Autoregressions". University of Minnesota. Ph. D. Thesis, 1980.

LITTERMAN, R.: "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience". *Journal of Business and Economic Statistics*, 1986, 4(1), pp. 25-38.

MATEA, Mª LL.; REGIL, A.V.: *El cambio de base del IPC. Construcción de series históricas homogéneas con base 1992.* Madrid: Bank of Spain, 1993. (Document C, 73).

MCNEES, S. K.: "Forecasting Accuracy of Alternatives Techniques: A Comparison of US Macroeconomic Forecasts". *Journal of Business and Economic Statistics*, 1986, 4(1), pp. 5-15.

RUNKLE, D.: "Bad News from a Forecasting Model of the US Economy". *Quarterly Review*, Fall, pp. 2-10. Minneapolis: Federal Reserve Bank of Minneapolis, 1990.

SIMS, C.: "Money, Income and Causality". *American Economic Review*, 1972, 62(1), pp. 540-553.

SIMS, C.: "Macroeconomics and Reality". *Econometrica*, 1980, 48(1), pp. 1-48.

SIMS, C.: "Are Forecasting Models Usable for Policy Analysis?". *Quarterly Review,* Winter, pp. 2-16. Minneapolis: Federal Reserve Bank of Minneapolis, 1986.

SIMS, C.: *A Nine Variable Probabilistic Model of the US Economy.* Minneapolis: Federal Reserve Bank of Minneapolis. Institute for Empirical Macroeconomics, 1989. (Discussion Paper, 14).

TODD, R.: "Improving Economic Forecasting with Bayesian Vector Autoregression". *Quarterly Review*, Fall, pp. 18-29. Minneapolis: Federal Reserve Bank of Minneapolis, 1984.

Appendix: Data

The following time series for the period 1970:1-1995:4 have been used in this paper.

M	→	Broad Money (ALP). Average of monthly data. Millions of pesetas. Bank of Spain.
W	→	Wage per Worker. Thousands of pesetas. Bank of Spain.
P	→	Consumer Price Index. 1992 = 100. Average of monthly data. INE and Matea & Regil (1994).
Y	→	GDP at 1986 prices. INE.
EMPL	→	Employment. Thousands of workers. INE and García Perea & Gómez (1994).

TABLE 1

PRIOR INFORMATION

Υ_1	First Own Lag Prior Mean	0.96
Υ_2	Overall Tightness	0.399E-3
Υ_3	Relative Tightness of Other Lags	8.90
Υ_4	Lag Decay Tightness	1.0
Υ_5	Deterministic Component Tightness	0.9E+3
Υ_6	Time Variation	0.18E-5
Υ_7	Law of Motion for β	1.0

TABLE 2

OVERALL GOODNESS OF FIT MEASURES

	BVAR	UVAR
Likelihood (1970-90)	3460.5	1723.2
MSE4 (1980-95)	51.9	69.5
MSE8 (1980-95)	190.2	260.7

TABLE 3

SINGLE EQUATION U-THEIL STATISTIC

(1980-1995)

	Forecasting Step	BVAR	UVAR
M	1	0.264	0.253
	4	0.162	0.210
	8	0.161	0.229
W	1	0.318	0.332
	4	0.262	0.394
	8	0.307	0.442
P	1	0.346	0.670
	4	0.199	0.424
	8	0.225	0.447
Y	1	0.405	0.508
	4	0.461	0.716
	8	0.567	0.811
EMPL	1	0.860	0.865
	4	0.833	0.859
	8	1.135	1.123

TABLE 4

BVAR MODEL

UNCONDITIONAL PROJECTIONS
(Levels. 96% bands in parenthesis)

	1996: 4	1997: 4
M	79308951 (± 1632868)	85949092 (± 2413854)
W	849.4 (± 13.2)	880.0 (± 19.0)
P	118.5 (± 2.2)	121.3 (± 3.4)
Y	10821 (± 108)	11165 (± 167)
EMPL	12463 (± 211)	12816 (± 294)

TABLE 5

BVAR MODEL

UNCONDITIONAL PROJECTIONS
(Growth Rates. 96% bands in parenthesis)

	1996: 4	1997: 4
M	8.51 (± 2.2)	8.37 (± 2.2)
W	3.64 (± 1.6)	3.54 (± 1.7)
P	2.52 (± 2.0)	2.35 (± 2.2)
Y	2.97 (± 1.0)	3.18 (± 1.1)
EMPL	2.64 (± 1.7)	2.82 (± 1.6)

TABLE 6

PROBABILITIES OF FUTURE EVENTS
(estimation period: 1970-1995)

Unconditional Probabilities

$\Pr(\Pi_{96:4} > 3.5) = 0.1700$

$\Pr(\Pi_{97:4} > 2.5) = 0.4400$

Conditional Probabilities

Condition on W Growth	3.5	1996:1,2
	4.5	1996:3 - 1997:4

$\Pr(\Pi_{97:4} > 2.5) = 0.7010$

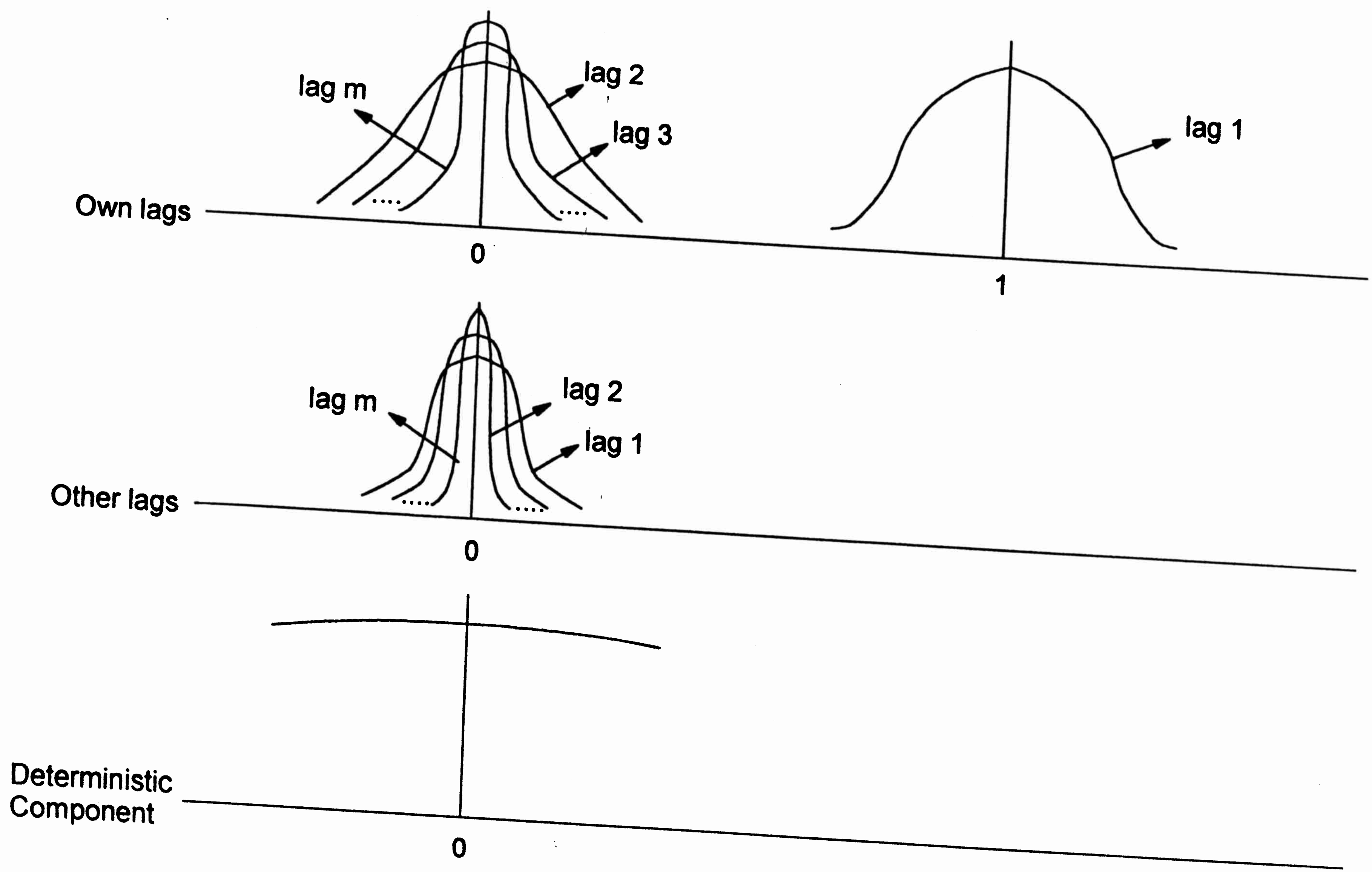
Condition on W Growth	3.5	1996:1,2
	3.0	1996:3,4
	2.0	1997:1,2,3,4

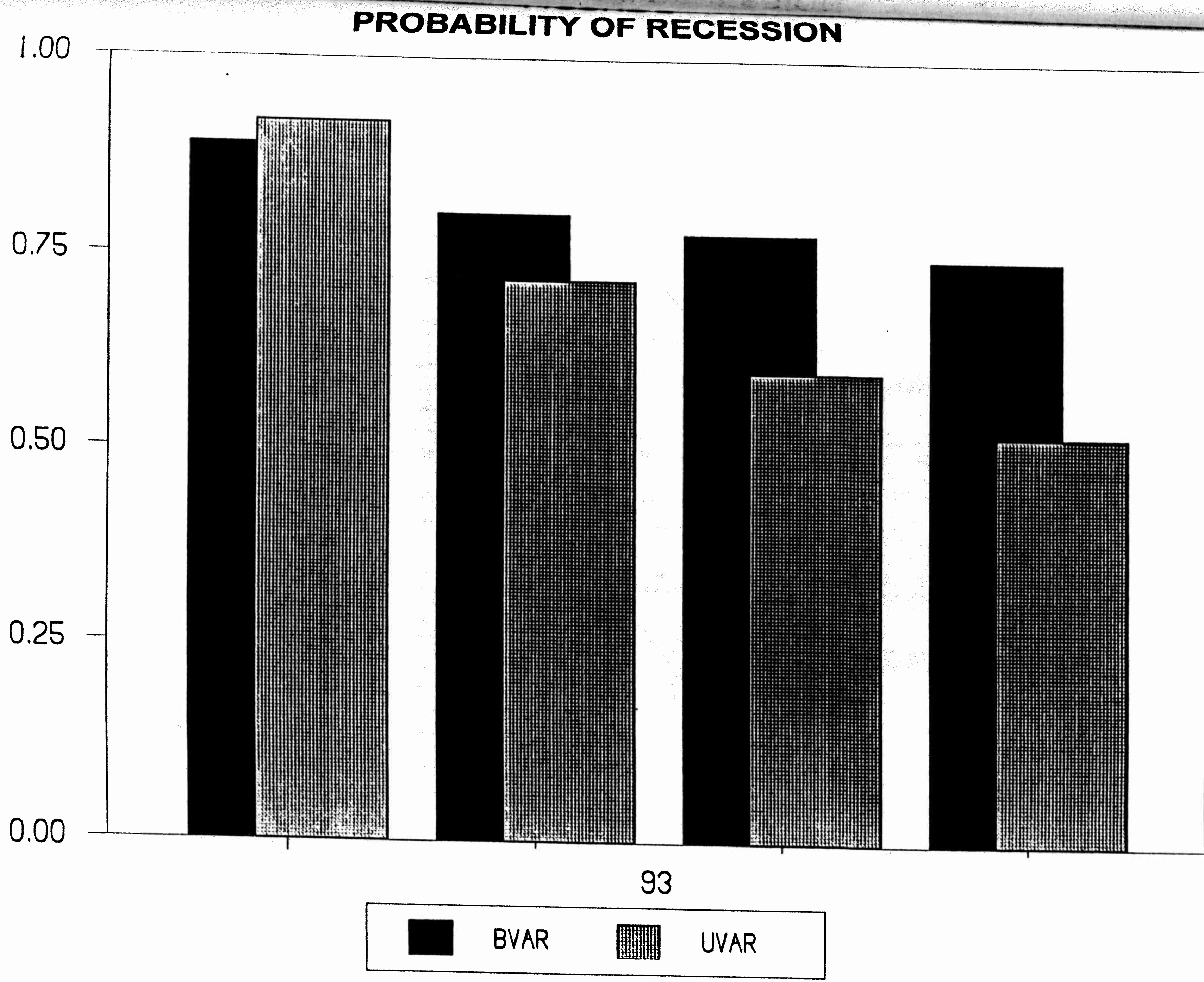
$\Pr(\Pi_{97:4} > 2.5) = 0.1500$

Condition on M Growth	8.5, 7.7, 6.06	1996:1,2,3
	6.0	1996:4 - 1997:4

$\Pr(\Pi_{97:4} > 2.5) = 0.4600$

FIGURE 1
BASIC TRAITS OF THE PRIOR INFORMATION




PROBABILITY OF RECESSION
1.00
0.75
0.50
0.25
0.00
93
BVAR
UVAR

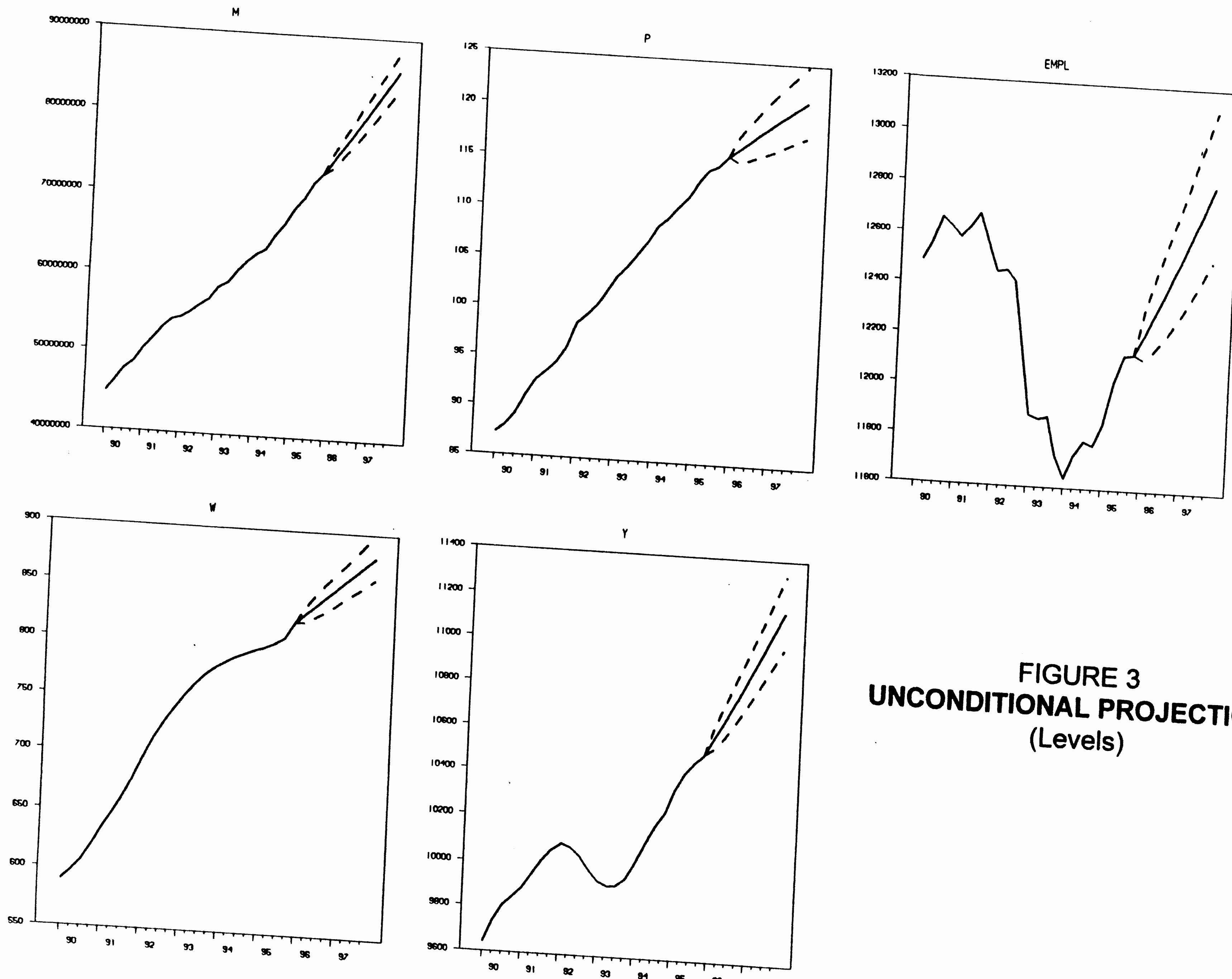


FIGURE 3
UNCONDITIONAL PROJECTIONS
(Levels)

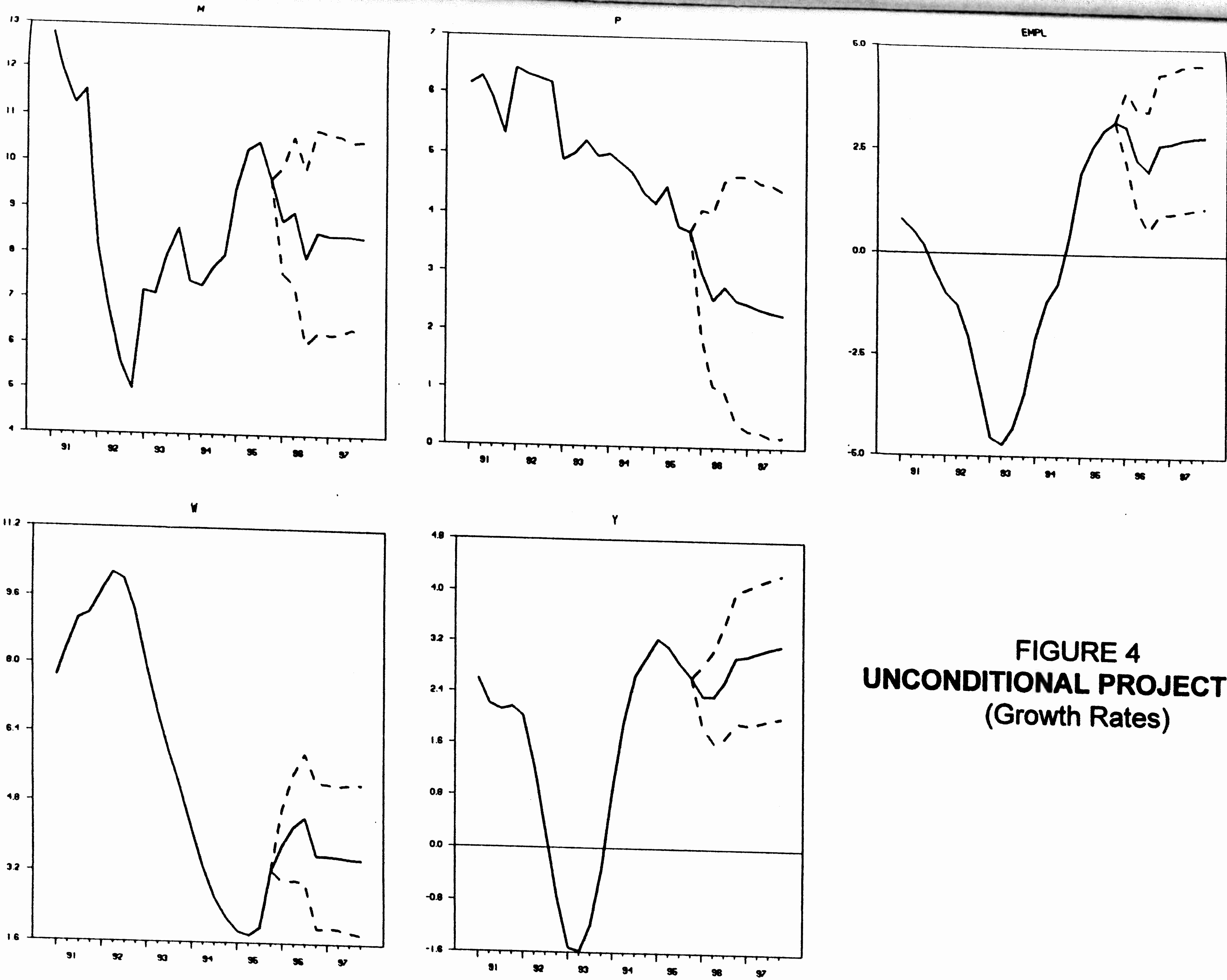


FIGURE 4
UNCONDITIONAL PROJECTIONS
(Growth Rates)

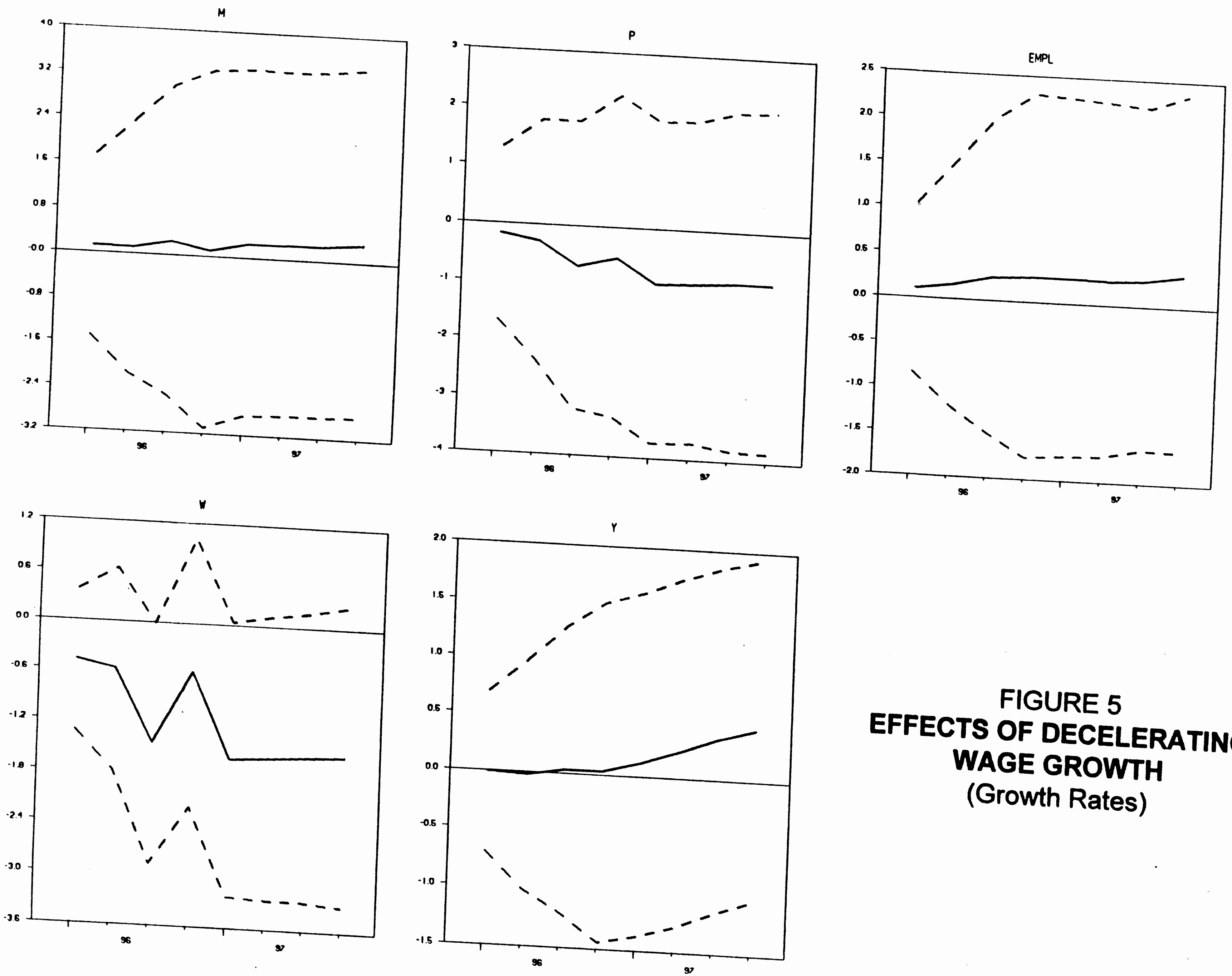


FIGURE 5
EFFECTS OF DECELERATING WAGE GROWTH
(Growth Rates)

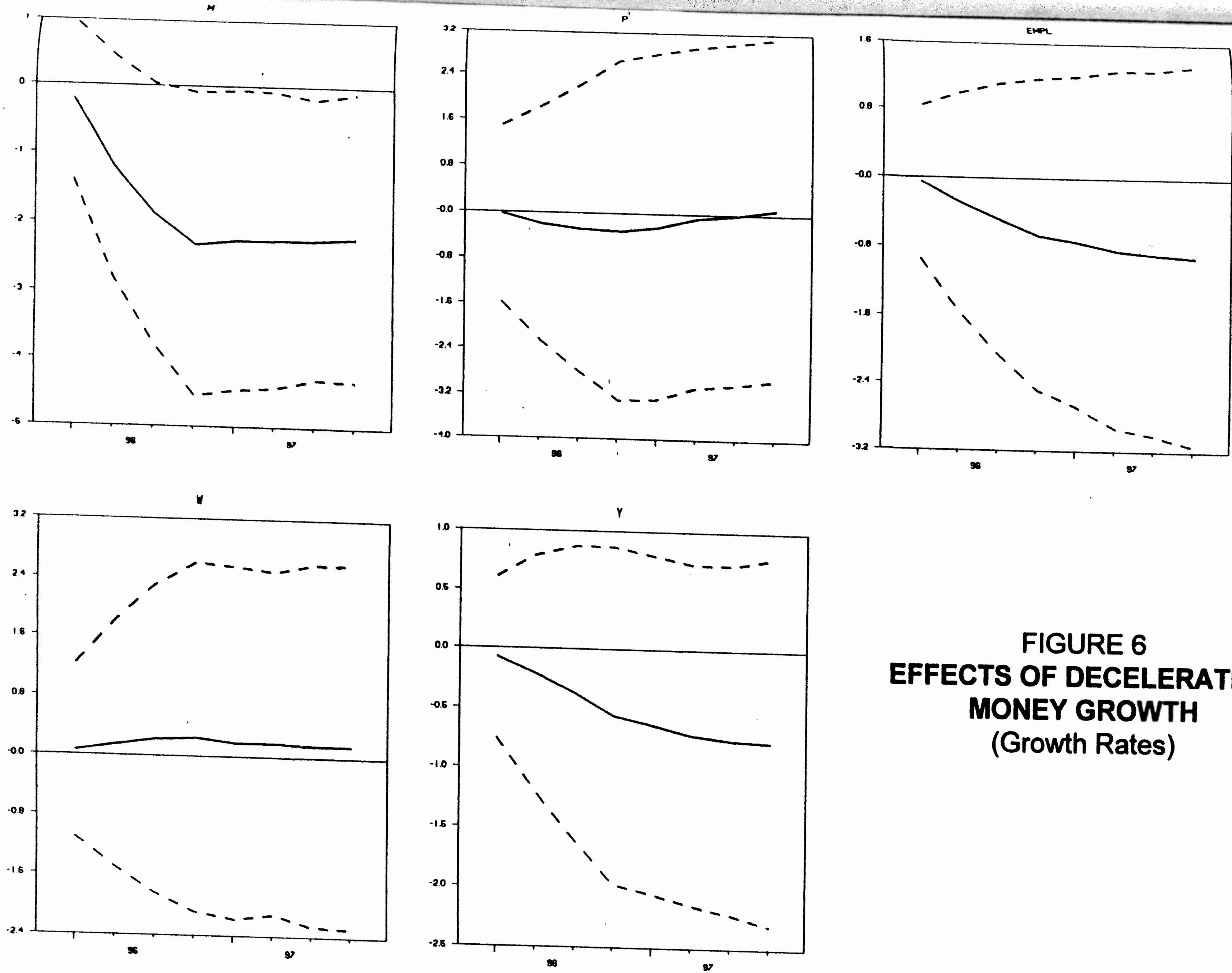


FIGURE 6
EFFECTS OF DECELERATING MONEY GROWTH
(Growth Rates)