An Introductory Review of a Structural VAR-X Estimation and Applications

Por:
Sergio Ocampo
Norberto Rodríguez

Borradores de ECONOMÍA

Núm. 686
2011

tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

BANCO DE LA REPÚBLICA COLOMBIA

An Introductory Review of a Structural VAR-X Estimation and Applications*

Sergio Ocampo
socampdi@banrep.gov.co

Norberto Rodríguez
nrodrini@banrep.gov.co

Macroeconomic Modeling Department, Banco de la República.

Abstract

This document presents how to estimate and implement a structural VAR-X model under long run and impact identification restrictions. Estimation by bayesian and maximum likelihood methods is presented. Applications of the structural VAR-X for impulse response functions to structural shocks, multiplier analysis of the exogenous variables, forecast error variance decomposition and historical decomposition of the endogenous variables are also described, as well as a method for computing HPD regions in a bayesian context. Some of the concepts are exemplified with an application to US data.

Keywords: S-VAR, B-VAR, VAR-X, IRF, FEVD, historical decomposition.
JEL Classification: C11, C18, C32.

The use of VAR-X and structural VAR-X models in econometrics is not new, yet textbooks and articles that use them often fail to provide the reader a concise (and moreover useful) description of how to implement these models (Lütkepohl (2005) constitutes an exception of this statement). The use of bayesian techniques in the estimation of VAR-X models is also largely neglected from the literature, as is the construction of the historical decomposition of the endogenous variables. This document builds upon the S-VAR and B-VAR literature and its purpose is to present a review of some of the basic features that accompany the implementation of a structural VAR-X model.

Section 1 presents the notation and general setup to be followed throughout the document. Section 2 discuses the identification of structural shocks in a VAR-X, with both long run restrictions, as in Blanchard and Quah (1989), and impact restrictions, as in Sims (1980, 1986). Section 3 considers the estimation of the parameters by maximum likelihood and bayesian methods. In Section 4 it is shown how to use the marginal density of the model for choosing the lag structure. Finally, in Section 5, four of the possible applications of the model are presented, namely the construction of impulse response functions to structural shocks, multiplier analysis of the exogenous variables, forecast error variance decomposition and historical decomposition of the endogenous variables. Appendix A exemplifies some of the concepts developed in the document using Galí (1999)'s structural VAR augmented with oil prices as an exogenous variable.

1 General setup

In all sections the case of a structural VAR-X whose reduced form is a VAR-X(p, q) will be considered. It is assumed that the system has n endogenous variables (y_t) and m exogenous variables (x_t). The

*The results and opinions expressed in this document do not compromise in any way Banco de la República or its board of directors. We wish to thank Eliana González, Martha Misas, Andrés González, Luis Fernando Melo, and Christian Bustamente for useful comments on earlier drafts of this document, of course, all remaining errors are our own.

variables in y_t and x_t may be in levels or in first differences, this depends on the characteristics of the data, the purpose of the study, and the identification strategy, in all cases no co-integration is assumed. The reduced form of the structural model includes the first p lags of the endogenous variables, the contemporaneous values and first q lags of the exogenous variables and a constant vector.[1] Under this specification it is assumed that the model is stable and presents white-noise Gaussian residuals (e_t), i.e. $e_t \overset{iid}{\sim} N(0,\Sigma)$.

The reduced form VAR-X can be represented as in equation (1) or equation (2), where v is a n-vector, B_i are $(n \times n)$ matrices and Θ_i are $(n \times m)$ matrices. In equation (2) one has $B(L) = B_1 L + \ldots + B_p L^p$ and $\Theta(L) = \Theta_0 + \ldots + \Theta_q L^q$, both matrices of polynomials in the lag operator L.

$$y_t = v + B_1 y_{t-1} + \ldots + B_p y_{t-p} + \Theta_0 x_t + \ldots + \Theta_q x_{t-q} + e_t \tag{1}$$
$$y_t = v + B(L) y_t + \Theta(L) x_t + e_t \tag{2}$$

Defining $\Psi(L) = \Psi_0 + \Psi_1 L + \ldots = [I - B(L)]^{-1}$ with $\Psi_0 = I$ as an infinite polynomial on the lag operator L, one has the VMA-X representation of the model, equation (3).[2]

$$y_t = \Psi(1) v + \Psi(L) \Theta(L) x_t + \Psi(L) e_t \tag{3}$$

Finally, there is a structural VAR-X model associated with the equations above, most of the applications are obtained from it, for example those covered in Section 5. Instead of the residuals (e), which can be correlated among them, the structural model contains structural disturbances with economic interpretation (ϵ), this is what makes it useful for policy analysis. It will be convenient to represent the model by its VMA-X form, equation (4),

$$y_t = \mu + C(L)\epsilon_t + \Lambda(L) x_t \tag{4}$$

where the endogenous variables are expressed as a function of a constant n-vector (μ), and the current and past values of the structural shocks (ϵ) and the exogenous variables. It is assumed that ϵ is a vector of white noise Gaussian disturbances with identity covariance matrix, i.e. $\epsilon_t \overset{iid}{\sim} N(0,I)$. Both $C(L)$ and $\Lambda(L)$ are infinite polynomials in the lag operator L, each matrix of $C(L)$ $(C_0, C_1, \ldots)$ is of size $(n \times n)$, and each matrix of $\Lambda(L)$ $(\Lambda_0, \Lambda_1, \ldots)$ is of size $(n \times m)$.

2 Identification of structural shocks in a VAR-X

The identification of structural shocks is understood here as a procedure which enables the econometrician to obtain the parameters of a structural VAR-X from the estimated parameters of the reduced form of the model. As will be clear from the exposition below, the identification in presence of exogenous variables is no different from what is usually done in the S-VAR literature.

Equating (3) and (4) one has:

$$\mu + \Lambda(L) x_t + C(L)\epsilon_t = \Psi(1) v + \Psi(L) \Theta(L) x_t + \Psi(L) e_t$$

then the following equalities can be inferred:

[1] The lag structure of the exogenous variables may be relaxed allowing different lags for each variable. This complicates the estimation and is not done here for simplicity. Also, the constant vector or intercept may be omitted according to the characteristics of the series used.

[2] The models stability condition implies that $\Psi(1) = \left[I - \sum_{i=1}^{p} B_i\right]^{-1}$ exist and is finite.

$$\begin{aligned}\mu &= \Psi(1)\,v & (5)\\ \Lambda(L) &= \Psi(L)\,\Theta(L) & (6)\\ C(L)\,\epsilon_t &= \Psi(L)\,e_t & (7)\end{aligned}$$

Since the parameters in v, $B(L)$ and $\Theta(L)$ can be estimated from the reduced form VAR-X representation, the values of μ and $\Lambda(L)$ are also known.[3] Only the parameters in $C(L)$ are left to be identified, the identification depends on the type of restrictions to be imposed. From equations (5), (6) and (7) is clear that the inclusion of exogenous variables in the model has no effect in the identification of the structural shocks. Equation (7) also holds for a structural VAR model.

The identification restrictions to be imposed over $C(L)$ may take several forms. Since there is nothing different in the identification between the case presented here and the S-VAR literature, we cover only two types of identification procedures, namely: impact and long run restrictions that allow the use of the Cholesky decomposition. It is also possible that the economic theory points at restrictions that make impossible a representation in which the Cholesky decomposition can be used, or that the number of restrictions exceeds whats needed for exact identification. Both cases complicate the estimation of the model, and the second one (over-identification) makes possible to carry out test over the restrictions imposed. For a more comprehensive treatment of this problems we refer to Amisano and Giannini (1997).

There is another identification strategy that won't be covered in this document, identification by sign restrictions over some of the impulse response functions. This kind of identification allows to avoid some puzzles that commonly arise in the VAR literature. References to this can be found in Uhlig (2005), Mountford and Uhlig (2009), Canova and De Nicolo (2002), Canova and Pappa (2007) and preceding working papers of those articles originally presented in the late 1990's. More recently, the work of Moon et al. (2011) presents how to conduct inference over impulse response functions with sign restrictions, both by classical and bayesian methods.

2.1 Identification by impact restrictions

In Sims (1980, 1986) the identification by impact restrictions is proposed, the idea behind it is that equation (7) is equating two polynomial in the lag operator L, for them to be equal it must be the case that:

$$\begin{aligned}C_i L^i \epsilon_t &= \Psi_i L^i e_t\\ C_i \epsilon_t &= \Psi_i e_t & (8)\end{aligned}$$

Equation (8) holds for all i, in particular it holds for $i = 0$. Knowing that $\Psi_0 = I$, the following result is obtained:

$$C_0 \epsilon_t = e_t \tag{9}$$

then, by taking the variance on both sides one gets:

$$C_0 C_0^{'} = \Sigma \tag{10}$$

Since Σ is a symmetric, positive definite matrix it is not possible to infer in an unique form the parameters of C_0 from equation (10), restrictions over the parameters of C_0 have to be imposed. Because C_0 measures the impact effect of the structural shocks over the endogenous variables, those

[3] Lütkepohl (2005) presents methods for obtaining the matrices in $\Psi(L)$ and the product $\Psi(L)\,\Theta(L)$ recursively in Sections 2.1.2 and 10.6, respectively. $\Psi(1)$ is easily computed by taking the inverse on $I - B_1 - \ldots - B_p$.

restrictions are called here impact restrictions. Following Sims (1980), the restrictions to be imposed ensure that C_0 is a triangular matrix, this allows to use the Cholesky decomposition of Σ to obtain the non-zero elements of C_0. This amount of restrictions account $n \times (n-1)/2$ and make the model just identifiable.

Once C_0 is known, equations (8) and (9) can be used to calculate C_i for all i as:

$$C_i \quad = \quad \Psi_i C_0 \tag{11}$$

The steps for the identification by impact restrictions are summarized in Algorithm 1.

Algorithm 1 Identification by impact restrictions

1. Estimate the reduced form of the VAR-X.
2. Calculate the VMA-X representation of the model (matrices Ψ_i) and the covariance matrix of the reduced form disturbances e (matrix Σ).
3. From the Cholesky decomposition of Σ calculate matrix C_0 (equation 10).
$$C_0 = \text{chol}\,(\Sigma)$$
4. For $i = 1, \ldots, R$, with R given, use equation 11 to calculate the matrices C_i.
$$C_i = \Psi_i C_0$$

Step 4 completes the identification in the sense that all matrices of the structural VMA-X are known.

2.2 Identification by long run restrictions

Another way to identify the matrices of the structural VMA-X is to impose restrictions on the long run impact of the shocks over the endogenous variables. This method is proposed in Blanchard and Quah (1989). For the model under consideration, if the variables in y_t are in differences, the matrix $C\,(1) = \sum_{i=0}^{\infty} C_i$ measures the long run impact of the structural shocks over the levels of the variables.[4] Matrix $C\,(1)$ is obtained by evaluating equation (7) in $L = 1$. As in the case of impact restrictions, the variance of each side of the equation is taken, the result is:

$$C\,(1)\,C^{'}\,(1) \quad = \quad \Psi\,(1)\,\Sigma\Psi^{'}\,(1) \tag{12}$$

Again, since $\Psi\,(1)\,\Sigma\Psi^{'}\,(1)$ is a symmetric, positive definite matrix it is not possible to infer the parameters of $C\,(1)$ from equation (12), restrictions over the parameters of $C\,(1)$ have to be imposed. It is conveniently assumed that those restrictions make $C\,(1)$ a triangular matrix, as before, this allows to use the Cholesky decomposition to calculate the non-zero elements of $C\,(1)$. Again, this amount of restrictions account $n \times (n-1)/2$ and make the model just identifiable.

Finally, it is possible to use $C\,(1)$ to calculate the parameters in the C_0 matrix, with it, the matrices C_i for $i > 0$ are obtained as in the identification by impact restrictions. Combining (10) with (7) evaluated in $L = 1$ the following expression for C_0 is derived:

[4] Of course, not all the variables of y_t must be in differences, but the only meaningful restrictions are those imposed over variables that enter the model in that way. We restrict our attention to a case in which there are no variables in levels in y_t.

$$C_0 = [\Psi(1)]^{-1} C(1) \tag{13}$$

The steps for the identification by long run restrictions are summarized in Algorithm 2.

Algorithm 2 Identification by long run restrictions

1. Estimate the reduced form of the VAR-X.
2. Calculate the VMA-X representation of the model (matrices Ψ_i) and the covariance matrix of the reduced form disturbances e (matrix Σ).
3. From the Cholesky decomposition of $\Psi(1)\Sigma\Psi'(1)$ calculate matrix $C(1)$ (equation 12).
$$C(1) = \text{chol}\left(\Psi(1)\Sigma\Psi'(1)\right)$$
4. With the matrices of long run effects of the reduced form, $\Psi(1)$, and structural shocks, $C(1)$, calculate the matrix of contemporaneous effects of the structural shocks, C_0 (equation 13).
$$C_0 = [\Psi(1)]^{-1} C(1)$$
5. For $i = 1, \ldots, R$, with R sufficiently large, use equation (11) to calculate the matrices C_i.
$$C_i = \Psi_i C_0$$

Step 5 completes the identification in the sense that all matrices of the structural VMA-X are known.

3 Estimation

The estimation of the parameters of the VAR-X can be carried out by maximum likelihood or bayesian methods, as will become clear it is convenient to write the model in a more compact form. Following Zellner (1996) and Bauwens et al. (2000), equation (1), for a sample of T observations, plus a fixed presample, can be written as:

$$Y = Z\Gamma + E \tag{14}$$

where $Y = \begin{bmatrix} y_1' \\ \vdots \\ y_t' \\ \vdots \\ y_T' \end{bmatrix}$, $Z = \begin{bmatrix} 1 & y_0' & \ldots & y_{-(p-1)}' & x_1' & \ldots & x_{1-q}' \\ \vdots & & & & & & \\ 1 & y_{t-1}' & \ldots & y_{t-p}' & x_t' & \ldots & x_{t-q}' \\ \vdots & & & & & & \\ 1 & y_{T-1}' & \ldots & y_{T-p}' & x_T' & \ldots & x_{T-q}' \end{bmatrix}$, $E = \begin{bmatrix} e_1' \\ \vdots \\ e_t' \\ \vdots \\ e_T' \end{bmatrix}$ and $\Gamma = \begin{bmatrix} v' \\ B_1' \\ \vdots \\ B_p' \\ \Theta_o' \\ \vdots \\ \Theta_q' \end{bmatrix}$.

For convenience we define the auxiliary variable $k = (1 + np + m(q+1))$ as the total number of regressors. The matrices sizes are as follow: Y is a $(T \times n)$ matrix, Z a $(T \times k)$ matrix, E a $(T \times n)$ matrix and Γ a $(k \times n)$ matrix.

Equation (14) is useful because it allows to represent the VAR-X model as a multivariate linear regression model, with it the likelihood function is derived. The parameters can be obtained by maximizing that function or by means of Bayes theorem.

3.1 The likelihood function

From equation (14) one derives the likelihood function for the error terms. Since $e_t \sim N(0, \Sigma)$, one has: $E \sim \mathrm{MN}(0, \Sigma \otimes I)$, a matricvariate normal distribution with I the identity matrix with dimension $(T \times T)$. The following box defines the probability density function for the matricvariate normal distribution.

The matricvariate normal distribution

The probability density function of a $(p \times q)$ matrix X that follows a matricvariate normal distribution with mean $M_{p\times q}$ and covariance matrix $Q_{q\times q} \otimes P_{p\times p}$ $(X \sim \mathrm{MN}(M, Q \otimes P))$ is:

$$\mathrm{MN}_{\mathrm{pdf}}(M, Q \otimes P) \quad \propto \quad |Q \otimes P|^{\frac{-1}{2}} \exp\left(-\frac{1}{2}\left[\mathrm{vec}(X - M)\right]^{'} (Q \otimes P)^{-1} \left[\mathrm{vec}(X - M)\right]\right) \tag{15}$$

Following Bauwens et al. (2000), the vec operator can be replaced by a trace operator (tr):

$$\mathrm{MN}_{\mathrm{pdf}}(M, Q \otimes P) \quad \propto \quad |Q|^{\frac{-p}{2}} |P|^{\frac{-q}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(Q^{-1}(X - M)^{'} P^{-1}(X - M)\right)\right) \tag{16}$$

Both representations of the matricvariate normal pdf are useful when dealing with the compact representation of the VAR-X model. Note that the equations above are only proportional to the actual probability density function. The missing constant term has no effects in the estimation procedure.

Using the definition in the preceding box and applying it to $E \sim \mathrm{MN}(0, \Sigma \otimes I)$ one gets the likelihood function of the VAR-X model, conditioned to the path of the exogenous variables:

$$\mathcal{L} \quad \propto \quad |\Sigma|^{\frac{-T}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1} E^{'} E\right)\right)$$

From (14) one has $E = Y - Z\Gamma$, replacing:

$$\mathcal{L} \quad \propto \quad |\Sigma|^{\frac{-T}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}(Y - Z\Gamma)^{'}(Y - Z\Gamma)\right)\right)$$

Finally, after tedious algebraic manipulation, one gets to the following expression:

$$\mathcal{L} \quad \propto \quad \left[|\Sigma|^{\frac{-(T-k)}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1} S\right)\right)\right] \left[|\Sigma|^{\frac{-k}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\left(\Gamma - \hat{\Gamma}\right)^{'} Z^{'} Z \left(\Gamma - \hat{\Gamma}\right)\right)\right)\right]$$

where $\hat{\Gamma} = \left(Z^{'} Z\right)^{-1} Z^{'} Y$ and $S = \left(Y - Z\hat{\Gamma}\right)^{'}\left(Y - Z\hat{\Gamma}\right)$.

One last thing is noted, the second factor of the right hand side of the last expression is proportional to the pdf of a matricvariate normal distribution for Γ, and the first factor to the pdf of an inverse Wishart distribution for Σ (see the box below). This allows an exact characterization of the likelihood function as in equation (17).

$$\mathcal{L} \quad = \quad \mathrm{iW}_{\mathrm{pdf}}(S, T - k - n - 1)\, \mathrm{MN}_{\mathrm{pdf}}\left(\hat{\Gamma}, \Sigma \otimes \left(Z^{'} Z\right)^{-1}\right) \tag{17}$$

The parameters of the VAR-X, Γ and Σ, can be estimated by maximizing equation (17). It can be shown that the result of the likelihood maximization gives:

$$\Gamma_{ml} = \hat{\Gamma} \qquad\qquad \Sigma_{ml} = S$$

The inverse Wishart distribution

If the variable X (a square, positive definite matrix of size q) is distributed $\mathrm{iW}(S, v)$, with parameter S (also a square, positive definite matrix of size q), and v degrees of freedom, then its probability density function $\left(\mathrm{iW}_{\mathrm{pdf}}\right)$ is given by:

$$\mathrm{iW}_{\mathrm{pdf}}(S, v) = \frac{|S|^{\frac{v}{2}}}{2^{\frac{vq}{2}}\Gamma_q\left(\frac{v}{2}\right)}|X|^{\frac{-(v+q+1)}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(X^{-1}S\right)\right) \tag{18}$$

where $\Gamma_q(x) = \pi^{\frac{q(q-1)}{4}} \prod_{j=1}^{q} \Gamma\left(x + \frac{1-j}{2}\right)$ is the multivariate Gamma function. It is useful to have an expression for the mean and mode of the inverse Wishart distribution, these are given by:

$$\mathrm{Mean}(X) = \frac{S}{v-q-1} \qquad \mathrm{Mode}(X) = \frac{S}{v+q+1}$$

3.2 Bayesian estimation

If the estimation is carried out by bayesian methods the problem is to elect an adequate prior distribution and, by means of Bayes theorem, obtain the posterior density function of the parameters. The use of bayesian methods is encouraged because they allow inference to be done conditional to the sample, and in particular the sample size, giving a better sense of the uncertainty associated with the parameters values; it also facilitate to compute moments not only for the parameters but for their functions as is the case of the impulse responses, forecast error variance decomposition and others; it is also particularly useful to obtain a measure of skewness in this functions, specially for the policy implications of the results. As mentioned in Koop (1992), the use of bayesian methods gives an exact finite sample density for both the parameters and their functions.

The election of the prior is a sensitive issue and wont be discussed in this document, we shall restrict our attention to the case of the Jeffreys non-informative prior (Jeffreys, 1961) which is widely used in bayesian studies of vector auto-regressors. There are usually two reasons for its use. The first one is that information about the reduced form parameters of the VAR-X model is scarce and difficult to translate into an adequate prior distribution. The second is that it might be the case that the econometrician doesn't want to include new information to the estimation but only wishes to use bayesian methods for inference purposes. Besides the two reasons already mentioned, the use of the Jeffreys non-informative prior constitute a computational advantage because it allows a closed form representation of the posterior density function, thus allowing to make draws for the parameters by direct methods or by the Gibbs sampling algorithm (Geman and Geman, 1984).[5]

For a discussion of other usual prior distributions for VAR models we refer to Kadiyala and Karlsson (1997) and, more recently, to KociĘcki (2010) for the construction of feasible prior distributions over impulse response in a structural VAR context. When the model is used for forecast purposes the so called Minnesota prior is of particular interest, this prior is due to Litterman (1986), and is generalized in Kadiyala and Karlsson (1997) for allowing symmetry of the prior across equations. This generalization is recommended and is of easy implementation in the bayesian estimation of the model. It should me mentioned that the Minnesota prior is of little interest in the structural VAR-X context, principally because the model is conditioned to the path of the exogenous variables, adding difficulties to the forecasting process.

In general the Jeffreys Prior for the linear regression parameters correspond to a constant for the parameters in Γ and for the covariance matrix a function of the form: $|\Sigma|^{\frac{-(n+1)}{2}}$, where n represents the size of the covariance matrix. The prior distribution to be used is then:

[5] For an introduction to the use of the Gibbs sampling algorithm we refer to Casella and George (1992).

$$P(\Gamma,\Sigma) \quad = \quad C\,|\Sigma|^{\frac{-(n+1)}{2}} \tag{19}$$

where C is the integrating constant of the distribution. Its actual value will be of no interest.

The posterior is obtained from Bayes theorem as:

$$\pi(\Gamma,\Sigma|Y,Z) \quad = \quad \frac{\mathcal{L}(Y,Z|\Gamma,\Sigma)\,P(\Gamma,\Sigma)}{m(Y)} \tag{20}$$

where $\pi(\Gamma,\Sigma|Y,Z)$ is the posterior distribution of the parameters given the data, $\mathcal{L}(Y,Z|\Gamma,\Sigma)$ is the likelihood function, $P(\Gamma,\Sigma)$ is the prior distribution of the parameters and $m(Y)$ the marginal density of the model. The value and use of the marginal density is discussed in Section 4 and will be omitted in the current Section.

Combining equations (17), (19) and (20) one gets an exact representation of the posterior function as the product of the pdf of an inverse Wishart distribution and the pdf of a matricvariate normal distribution:

$$\pi(\Gamma,\Sigma|Y,Z) \quad = \quad \mathrm{iW}_{\mathrm{pdf}}(S,T-k)\,\mathrm{MN}_{\mathrm{pdf}}\left(\hat{\Gamma},\Sigma\otimes\left(Z^{'}Z\right)^{-1}\right) \tag{21}$$

Equation (21) implies that Σ follows an inverse Wishart distribution with parameters S and $T-k$, and that the distribution of Γ given Σ is matricvariate normal with mean $\hat{\Gamma}$ and covariance matrix $\Sigma\otimes\left(Z^{'}Z\right)^{-1}$. The following two equations formalize the former statement:

$$\Sigma|Y,Z \sim \mathrm{iW}_{\mathrm{pdf}}(S,T-k) \qquad\qquad \Gamma|\Sigma,Y,Z \sim \mathrm{MN}_{\mathrm{pdf}}\left(\hat{\Gamma},\Sigma\otimes\left(Z^{'}Z\right)^{-1}\right)$$

Although further work can be done to obtain the unconditional distribution of Γ it is not necessary to do so. Because equation (21) is an exact representation of the parameters distribution function, it can be used to generate draws of them, moreover it can be used to compute any moment or statistic of interest, this can be done by means of the Gibbs sampling algorithm.

4 Marginal densities and lag structure

The marginal density ($m(Y)$) can be easily obtained under the Jeffreys prior and can be used afterward for purposes of model comparison. The marginal density gives the probability that the data is generated by a particular model, eliminating the uncertainty due to the parameters values. Because of this $m(Y)$ is often used for model comparison by means of the Bayes factor: the ratio between the marginal densities of two different models that explain the same set of data ($\mathrm{BF}_{12} = {}^{m(Y|M_1)}/_{m(Y|M_2)}$). If the Bayes factor is bigger than one then the first model (M_1) would be preferred.

From Bayes theorem (equation 20) the marginal density of the data, given the model, is:

$$m(Y) \quad = \quad \frac{\mathcal{L}(Y,Z|\Gamma,\Sigma)\,P(\Gamma,\Sigma)}{\pi(\Gamma,\Sigma|Y,Z)} \tag{22}$$

its value is obtained by replacing for the actual forms of the likelihood, prior and posterior functions (equations 17, 19 and 21 respectively):

$$m(Y) \quad = \quad \frac{\Gamma_n\left(\frac{T-k}{2}\right)}{\Gamma_n\left(\frac{T-k-n-1}{2}\right)}\,|S|^{\frac{-n-1}{2}}\,2^{\frac{n(n+1)}{2}}C \tag{23}$$

Algorithm 3 Bayesian estimation

1. Select the specification for the reduced form VAR-X, that is to chose values of p (endogenous variables lags) and q (exogenous variables lags) such that the residuals of the VAR-X (e) have withe noise properties. With this the following variables are obtained: T, p, q, k, where:

$$k = 1 + np + m(q+1)$$

2. Calculate the values of $\hat{\Gamma}$, S with the data (Y, Z) as:

$$\hat{\Gamma} = \left(Z'Z\right)^{-1} Z'Y \qquad\qquad S = \left(Y - Z\hat{\Gamma}\right)'\left(Y - Z\hat{\Gamma}\right)$$

3. Generate a draw for the covariance matrix of the reduced form VAR-X (Σ) from an inverse Wishart distribution with parameter S and $T-k$ degrees of freedom.

$$\Sigma \sim \mathrm{iW}_{\mathrm{pdf}}\left(S, T-k\right)$$

4. Generate a draw for the parameters of the reduced form VAR-X (Γ) from a matricvariate normal distribution with mean $\hat{\Gamma}$ and covariance matrix $\Sigma \otimes \left(Z'Z\right)^{-1}$.

$$\Gamma|\Sigma \sim \mathrm{MN}_{\mathrm{pdf}}\left(\hat{\Gamma}, \Sigma \otimes \left(Z'Z\right)^{-1}\right)$$

5. Repeat steps 2-3 as many times as desired, save the values of each draw.

The draws generated (step 4) can be used to compute moments of the parameters.
For every draw the corresponding structural parameters, impulse responses functions, etc. can be computed, then, their moments and statistics can also be computed.
The algorithms for generating draws for the inverse Wishart and matricvariate normal distributions are presented in Bauwens et al. (2000), Appendix B.

Although the exact value of the marginal density for a given model cannot be known without the constant C, this is no crucial for model comparison if the only difference between the models is in their lag structure. In that case the constant C is the same for both models, and the difference between the marginal density of one specification or another arises only in the first two factors of the right hand side of equation (23) $\left[\frac{\Gamma_n\left(\frac{T-k}{2}\right)}{\Gamma_n\left(\frac{T-k-n-1}{2}\right)}|S|^{\frac{-n-1}{2}}\right]$. When computing the Bayes factor for any pair of models the result will be given by those factors alone.

The Bayes factor between a model, M_1, with k_1 regressors and residual covariance matrix S_1, and another model, M_2, with k_2 regressors and residual covariance matrix S_2, can be reduced to:

$$\begin{aligned} \mathrm{BF}_{12} &= \frac{m\left(Y|M_1\right)}{m\left(Y|M_2\right)} = \frac{\Gamma_n\left(\frac{T-k_1}{2}\right)}{\Gamma_n\left(\frac{T-k_1-n-1}{2}\right)}|S_1|^{\frac{-n-1}{2}} 2^{\frac{n(n+1)}{2}} C \\ \mathrm{BF}_{12} &= \frac{\frac{\Gamma_n\left(\frac{T-k_1}{2}\right)}{\Gamma_n\left(\frac{T-k_1-n-1}{2}\right)}|S_1|^{\frac{-n-1}{2}}}{\frac{\Gamma_n\left(\frac{T-k_2}{2}\right)}{\Gamma_n\left(\frac{T-k_2-n-1}{2}\right)}|S_2|^{\frac{-n-1}{2}}} \end{aligned} \tag{24}$$

5 Applications

There are several applications for the structural VAR-X, all of them useful for policy analysis. In this Section four of those applications are covered, they all use the structural VMA-X representation of the model (equation 4).

5.1 Impulse response functions (IRF), Multiplier analysis (MA), and Forecast error variance decomposition (FEVD)

Impulse response functions (IRF) and multiplier analysis (MA) can be constructed from the matrices in $C(L)$ and $\Lambda(L)$. The IRF shows the endogenous variables response to a unitary change in a structural shock, in an analogous way the MA shows the response to a change in an exogenous variable. The construction is simple and is based on the interpretations of the elements of the matrices in $C(L)$ and $\Lambda(L)$.

For the construction of the IRF consider matrix C_h. The elements of this matrix measure the effect of the structural shocks over the endogenous variables h periods ahead, thus c_h^{ij} (i-th row, j-th column) measures the response of the i-th variable to a unitary change in the j-th shock h periods ahead. The IRF for the i-th variable to a change in j-th shock is constructed by collecting elements c_h^{ij} for $h = 0, 1, \ldots, H$, with H the IRF horizon.

Matrices C_h are obtained from the reduced form parameters according to the type of identification restrictions (see Section 2). For a more detailed discussion on the construction and properties of the IRF we refer to Lütkepohl (2005), Section 2.3.2.

The MA is obtained in a similar fashion from the matrices Λ_h, these are also a function of the reduced form parameters.[6] The interpretation is the same as before.

A number of methods for inference over the IRF and MA are available. If the estimation is carried out by classical methods intervals for the IRF and MA can be computed by means of their asymptotic distributions or by bootstrapping methods.[7] Nevertheless, because the OLS estimators are biased, as proved in Nicholls and Pope (1988), the intervals that arise from both asymptotic theory and usual bootstrapping methods are also biased. As pointed out by Kilian (1998) this makes necessary to conduct the inference over IRF, and in this case over MA, correcting the bias and allowing for skewness in the intervals. Skewness is common in the small sample distributions of the IRF and MA and arises from the non-linearity of the function that maps the reduced form parameters to the IRF or MA. A double bootstrapping method that effectively corrects the bias and accounts for the skewness in the intervals is proposed in Kilian (1998).

In the context of bayesian estimation, it is noted that, applying Algorithm 1 or 2 for each draw of the reduced form parameters (Algorithm 3), the distribution for each c_h^{ij} and λ_h^{ij} is obtained. With the distribution function inference can be done over the point estimate of the IRF and MA. For instance, standard deviations in each horizon can be computed, as well as asymmetry measures and *credible sets* (or intervals), the bayesian analogue to a classical confidence interval.

In the following we shall restrict our attention to *credible sets* with minimum size (length), these are named Highest Posterior Density regions (HDP from now on). An $(1-\alpha)\,\%$ HPD for the parameter θ is defined as the set $C = \{\theta \in \Theta : \pi\left(\theta/Y\right) \geq k(\alpha)\}$, where $k(\alpha)$ is the largest constant satisfying $P(C|y) = \int_\theta \pi\left(\theta/Y\right) d\theta \geq 1-\alpha$.[8] From the definition just given is clear that HPD regions are of minimum size and that each value of $\theta \epsilon C$ has a higher density (probability) than any value of θ outside the HPD. The second property makes possible direct probability statements about the likelihood of θ falling in C, i.e., "The probability that θ lies in C given the observed data Y is at least $(1-\alpha)\%$", this contrast with the interpretation of the classical confidence intervals. An HPD region can be disjoint if

[6] See Lütkepohl (2005), Section 10.6

[7] The asymptotic distribution of the IRF and FEVD for a VAR is presented in Lütkepohl (1990). A widely used non-parametric bootstrapping method is developed in Runkle (1987).

[8] Integration can be replaced by summation if θ is discrete.

Algorithm 4 Highest Posterior Density Regions

As in Chen and Shao (1998), let $\left\{\theta^{(i)},\, i = 1\,,\,\ldots\,,\,N\right\}$ be an ergodic sample of $\pi\left(\theta/Y\right)$, the posterior density function of parameter θ. $\pi\left(\theta/Y\right)$ is assumed to be unimodal. The $(1-\alpha)\,\%$ HPD is computed as follows:

1. Sort the values of $\theta^{(i)}$. Define $\theta_{(j)}$ as the $j-th$ larger draw of the sample, so that:

$$\theta_{(1)} = \min_{i\epsilon\{1,\ldots,N\}} \left\{\theta^{(i)}\right\} \qquad\qquad \theta_{(N)} = \max_{i\epsilon\{1,\ldots,N\}} \left\{\theta^{(i)}\right\}$$

2. Define $\overline{N} = \lfloor (1-\alpha)\,N \rfloor$ the integer part of $(1-\alpha)\,N$. The HPD will contain $\overline{N}$ values of θ.

3. Define $C_{(j)} = \left(\theta_{(j)}\,,\,\theta_{\left(j+\overline{N}\right)}\right)$ an interval in the domain of the parameter θ, for $j\epsilon\left\{1,\ldots,N-\overline{N}\right\}$. Note that although $C_{(j)}$ contains always $\overline{N}$ draws of θ, its size may vary.

4. The HPD is obtained as the interval $C_{(j)}$ with minimum size. $\text{HPD}\,(\alpha) = C_{(j^\star)}$, with $j^\star$ such that:

$$\theta_{\left(j^\star+\overline{N}\right)} - \theta_{(j^\star)} = \min_{j\epsilon\left\{1,\ldots,N-\overline{N}\right\}} \left(\theta_{\left(j+\overline{N}\right)} - \theta_{(j)}\right)$$

the posterior density function $\left(\pi\left(\theta/Y\right)\right)$ is multimodal. If the posterior is symmetric, all HPD regions will be symmetric about posterior mode (mean).

Koop (1992) presents a detailed revision of how to apply bayesian inference to the IRF in a structural VAR context, his results can be easily adapted to the structural VAR-X model. Another reference on the inference over IRF is Sims and Zha (1999). Here we present, in Algorithm 4, the method of Chen and Shao (1998) for computing HPD regions from the output of the Gibbs sampler.[9]

It is important to note that bayesian methods are by nature conditioned to the sample size and, because of that, avoid the problems of asymptotic theory in explaining the finite sample properties of the parameters functions, this includes the skewness of the IRF and MA distribution functions. Then, if the intervals are computed with the HPD, as in Chen and Shao (1998), they would be taking into account the asymmetry in the same way as Kilians method. This is not the case for intervals computed using only standard deviations although, with them, skewness can be addressed as in Koop (1992), although bootstrap methods can be used to calculate approximate measures of this and others moments, for instance, skewness and kurtosis, Bayesian methods are preferable since exact measures can be calculated.

Another application of the structural VAR-X model is the forecast error variance decomposition (FEVD), this is no different to the one usually presented in the structural VAR model. FEVD consists in decomposing the variance of the forecast error of each endogenous variable h periods ahead, as with the IRF, the matrices of $C\,(L)$ are used for its construction. Note that, since the model is conditioned to the path of the exogenous variables, all of the forecast error variance is explained by the structural shocks. Is because of this that the FEVD has no changes when applied in the structural VAR-X model. We refer to Lütkepohl (2005), Section 2.3.3, for the details of the construction of the FEVD. Again, if bayesian methods are used for the estimation of the VAR-X parameters, the density function of the FEVD can be obtained and several features of it can be explored, Koop (1992) also presents how to apply bayesian inference in this respect.

[9] The method presented is only valid if the distribution of the parameters of interest is unimodal. For a more general treatment of the highest posterior density regions, including multimodal distributions, we refer to the work of Hyndman (1996).

5.2 Historical decomposition of the endogenous variables (HD)

The historical decomposition (HD) consists in explaining the observed values of the endogenous variables in terms of the structural shocks and the path of the exogenous variables. This kind of exercise is present in the DSGE literature (for example, in Smets and Wouters (2007)) but mostly absent in the structural VAR literature, being Canova (2007) an exception.[10] Unlike the applications already presented, the historical decomposition allows to make an statement over what has actually happened to the series in the sample period, in terms of the recovered values for the structural shocks and the observed paths of the exogenous variables. It allows to have all shocks and exogenous variables acting simultaneously, thus making possible the comparison over the relative effects of them over the endogenous variables, this means that the HD is particularly useful when addressing the relative importance of the shocks over some set of variables. The possibility of explaining the history of the endogenous variables instead of what would happen if some hypothetical shock arrives in the absence of any other disturbance is at least appealing.

Here we describe a method for computing the HD in a structural VAR and structural VAR-X context. The first case is covered in more detail and the second presented as an extension of the basic ideas.

5.2.1 Historical decomposition for a structural VAR model

In a structural VAR context is clear, from the structural VMA representation of the model, that variations of the endogenous variables can only be explained by variations in the structural shocks. The HD uses the structural VMA representation in order to compute what the path of each endogenous variable would have been conditioned to the presence of only one of the structural shocks. It is important to note that the interpretation of the HD in a stable VAR model is simpler than the interpretation in a VAR-X. This is because in the former there is no need for a reference value that indicates when a shock is influencing the path of the variables. In that case, the reference value is naturally zero, and it is understood that deviations of the shocks below that value are interpreted as negative shocks and deviations above as positive shocks. As we shall see, when dealing with exogenous variables a reference value must be set, and its election is not necessarily "natural".

Before the HD is computed it is necessary to recover the structural shocks from the estimation of the reduced form VAR. Define $\hat{E} = [\hat{e}_1 \dots \hat{e}_t \dots \hat{e}_T]^{'}$ as the matrix of all fitted residuals from the VAR model (equation (14) in the absence of exogenous variables). Recalling equation (9), the matrix C_0 can be used to recover the structural shocks from matrix $\hat{E}$ as in the following expression:

$$\hat{\mathcal{E}} = \hat{E}\left(C_0^{'}\right)^{-1} \tag{25}$$

Because zero is the reference value for the structural shocks the matrix $\hat{\mathcal{E}} = [\hat{\epsilon}_1 \dots \hat{\epsilon}_t \dots \hat{\epsilon}_T]^{'}$ can be used directly for the HD.

The HD is an in-sample exercise, thus is conditioned to the initial values of the series. It will be useful to define the structural infinite VMA representation of the VAR model, as well as the structural VMA representation conditional on the initial values of the endogenous variables, equations (26) and (27) respectively.

$$y_t = \mu + C(L)\epsilon_t \tag{26}$$

$$y_t = \sum_{i=0}^{t-1} C_i \epsilon_{t-i} + K_t \tag{27}$$

[10] Another exception is found in King and Morley (2007) where the historical decomposition of a structural VAR is used for computing a measure of the natural rate of unemployment for the US.

Note that in equation (26) the endogenous variables depend on an infinite number of past structural shocks. In equation (27) the effect of all shocks that are realized previous to the sample is captured by the initial values of the endogenous variables. The variable K_t is a function of those initial values and of the parameters of the reduced form model, $K_t = f_t\left(y_0\,,\,\ldots\,,\,y_{-(p-1)}\right)$. It measures the effect of the initial values over the period t realization of the endogenous variables, thus the effect of all shocks that occurred before the sample. It is clear that if the VAR is stable $K_t \longrightarrow \mu$ for t sufficiently large, this is because the shocks that are too far in the past have no effect in the current value of the variables. K_t will be refer to as the reference value of the historical decomposition.

Starting from the structural VMA representation, the objective is now to decompose the deviations of y_t from K_t into the effects of the current and past values of the structural shocks (ϵ_i for i from 1 to t). The decomposition is made over the auxiliary variable $\tilde{y}_t = y_t - K_t = \sum_{i=0}^{t-1} C_i \epsilon_{t-i}$. The information needed to compute $\tilde{y}_t$ is contained in the first t matrices C_i and the first t rows of matrix $\hat{\mathcal{E}}$.

The historical decomposition of the i-th variable of $\tilde{y}_t$ into the j-th shock is given by:

$$\tilde{y}_t^{(i,j)} \quad = \quad \sum_{i=0}^{t-1} c_i^{ij}\, \hat{\epsilon}_{t-i}^{j} \tag{28}$$

Note that it must hold that the sum over j is equal to the actual value of the i-th element of $\tilde{y}_t$, $\tilde{y}_t^i = \sum_{j=1}^{n} \tilde{y}_t^{(i,j)}$. For t sufficiently large, when K_t is close to μ, $\tilde{y}_t^{(i,j)}$ can be interpreted as the deviation of the i-th endogenous variable from its mean caused by the recovered sequence for the j-th structural shock.

Finally, the endogenous variables can be decomposed as well. The historical decomposition for the i-th endogenous variable into the j-th shock is given by:

$$y_t^{(i,j)} \quad = K_t^i + \tilde{y}_t^{(i,j)} = \quad K_t^i + \sum_{i=0}^{t-1} c_i^{ij}\, \hat{\epsilon}_{t-i}^{j} \tag{29}$$

the new variable $y_t^{(i,j)}$ is interpreted as what the i-th endogenous variable would have been if only realizations of the j-th shock had occurred. The value of K_t can be obtained as a residual of the historical decomposition, since y_t is known and $\tilde{y}_t$ can be computed from the sum of the HD or from the definition.

The HD of the endogenous variables ($y_t^{(i,j)}$) can be also used to compute what transformations of the variables would have been conditioned to the presence of only one shock. For instance, if the i-th variable enters the model in quarterly differences, the HD for the annual differences or the level of the series can be computed by applying to $y_t^{(i,j)}$ the same transformation used over y_t^i, in this example, a cumulative sum.

Algorithm 5 summarizes the steps carried out for the historical decomposition.

5.2.2 Historical decomposition for a structural VAR-X model

The structure already described applies also for a VAR-X model. The main difference is that now its necessary to determine a reference value for the exogenous variables.[11] It shall be understand that realizations of the exogenous variables different to this value are what explain the fluctuations of the endogenous variables. We shall refer to $\overline{x}_t$ as the reference value for the exogenous variables in t.

[11] The reference value for the exogenous variables need not be a constant. It can be given by a linear trend, by the sample mean of the series,or by the initial value. When the exogenous variables enter the model in their differences, it may seem natural to think in zero as a natural reference value, identifying fluctuations of the exogenous variables in an analogous way to whats done with the structural shocks.

Algorithm 5 Historical decomposition for a structural VAR model

1. Estimate the parameters of the reduced form VAR.
   (a) Save a matrix with all fitted residuals $\left(\hat{E} = [\hat{e}_1 \dots \hat{e}_t \dots \hat{e}_T]^{'}\right)$.
   (b) Compute matrices C_i according to the identifying restrictions (Algorithm 1 or 2).

2. Compute the structural shocks $\left(\hat{\mathcal{E}} = [\hat{\epsilon}_1 \dots \hat{\epsilon}_t \dots \hat{\epsilon}_T]^{'}\right)$ with matrix C_0 and the fitted residuals of the reduced form VAR:

$$\hat{\mathcal{E}} = \hat{E}\left(C_0^{'}\right)^{-1}$$

3. Compute the historical decomposition of the endogenous variables relative to K_t:

$$\tilde{y}_t^{(i,j)} = \sum_{i=0}^{t-1} c_i^{ij} \hat{\epsilon}_{t-i}^{j}$$

4. Recover the values of K_t with the observed values of y_t and the auxiliary variable $\tilde{y}_t$:

$$K_t = y_t - \tilde{y}_t$$

5. Compute the historical decomposition of the endogenous variables:

$$y_t^{(i,j)} = K_t^i + \tilde{y}_t^{(i,j)}$$

Steps 3 and 5 are repeated for $t = 1, 2, \dots, T$, $i = 1, \dots, n$ and $j = 1, \dots, n$. Step 4 is repeated for $t = 1, 2, \dots, T$.

As before, its necessary to present the structural VMA-X representation conditional to the initial values of the endogenous variables (equation 30), with K_t defined as above. It is also necessary to express the exogenous variables as deviations of the reference value, for this we define an auxiliary variable $\tilde{x}_t = x_t - \overline{x}_t$. Note that equation (30) can be written in terms of the new variable $\tilde{x}_t$ as in equation (31). In the later, the new variable $\tilde{K}_t = \sum_{i=0}^{t-1} \Lambda_i \bar{x}_{t-i} + K_t$ has a role analogous to that of K_t in the VAR context. $\tilde{K}_t$ properties depend on those of $\bar{x}_t$ and, therefore, it can't be guaranteed that it converges to any value.

$$y_t = \sum_{i=0}^{t-1} C_i \epsilon_{t-i} + \sum_{i=0}^{t-1} \Lambda_i x_{t-i} + K_t \tag{30}$$

$$y_t = \sum_{i=0}^{t-1} C_i \epsilon_{t-i} + \sum_{i=0}^{t-1} \Lambda_i \tilde{x}_{t-i} + \tilde{K}_t \tag{31}$$

The historical decomposition is now computed using matrices C_i, the recovered matrix of structural shocks $\hat{\mathcal{E}}$, matrices Λ_i and the auxiliary variables $\tilde{x}_i$, for i from 1 to T. Matrix $\hat{\mathcal{E}}$ is still computed as in equation (25). The new reference value for the historical decomposition is $\tilde{K}_t$, and the decomposition is done to explain the deviations of the endogenous variables with respect to it as a function of the structural shocks and deviations of the exogenous variables from their own reference value, $\bar{x}_t$. For

Algorithm 6 Historical decomposition for a structural VAR-X model

1. Estimate the parameters of the reduced form VAR-X.
   (a) Save a matrix with all fitted residuals $\left(\hat{E} = [\hat{e}_1 \ldots \hat{e}_t \ldots \hat{e}_T]^{'}\right)$.
   (b) Compute matrices C_i and Λ_i according to the identifying restrictions (Algorithm 1 or 2).
2. Compute the structural shocks $\left(\hat{\mathcal{E}} = [\hat{\epsilon}_1 \ldots \hat{\epsilon}_t \ldots \hat{\epsilon}_T]^{'}\right)$ with matrix C_0 and the fitted residuals of the reduced form VAR-X:
$$\hat{\mathcal{E}} = \hat{E}\left(C_0^{'}\right)^{-1}$$
3. Compute the historical decomposition of the endogenous variables relative to $\tilde{K}_t$:
$$\tilde{y}_t^{(i,j)} = \sum_{i=0}^{t-1} c_i^{ij} \hat{\epsilon}_{t-i}^{j} \qquad \tilde{y}_t^{(i,k)} = \sum_{i=0}^{t-1} \lambda_i^{ik} \tilde{x}_{t-i}^{k}$$
4. Recover the values of $\tilde{K}_t$ with the observed values of y_t and the auxiliary variable $\tilde{y}_t$:
$$\tilde{K}_t = y_t - \tilde{y}_t$$
5. Compute the historical decomposition of the endogenous variables:
$$y_t^{(i,j)} = \tilde{K}_t^i + \tilde{y}_t^{(i,j)} \qquad y_t^{(i,k)} = \tilde{K}_t^i + \tilde{y}_t^{(i,k)}$$

Steps 3 and 5 are repeated for $t = 1, 2, \ldots, T$, $i = 1, \ldots, n$, $j = 1, \ldots, n$ and $k = 1, \ldots, m$. Step 4 is repeated for $t = 1, 2, \ldots, T$.

notation, variable $\tilde{y}_t$ is redefined: $\tilde{y}_t = y_t - \tilde{K}_t = \sum_{i=0}^{t-1} C_i \epsilon_{t-i} + \sum_{i=0}^{t-1} \Lambda_i \tilde{x}_{t-i}$. The decomposition of the i-th variable of $\tilde{y}_t$ into the j-th shock is still given by equation (28), and the decomposition into the k-th exogenous variable is given by:

$$\tilde{y}_t^{(i,k)} = \sum_{i=0}^{t-1} \lambda_i^{ik} \tilde{x}_{t-i}^{k} \tag{32}$$

Variable $\tilde{y}_t^{(i,k)}$, for k from 1 to m, is interpreted as what the variable $\tilde{y}_t^i$ would have been if, in the absence of shocks, only the k-th exogenous variable is allowed to deviate from its reference value. As in the VAR model, it holds the following equation: $\tilde{y}_t^i = \sum_{j=1}^{n} \tilde{y}_t^{(i,j)} + \sum_{k=1}^{m} \tilde{y}_t^{(i,k)}$. The variable $\tilde{K}_t$ is recovered in the same way used before to recover K_t.

The historical decomposition of the endogenous variables can be computed by using the recovered values for $\tilde{K}_t$. The decomposition of the i-th variable into the effects of the j-th shock is still given by equation (29), if K_t^i is replaced by $\tilde{K}_t^i$. The decomposition of the i-th variable into the deviations of the k-th exogenous variable from its reference value is obtained from the following expression:

$$y_t^{(i,k)} = K_t^i + \tilde{y}_t^{(i,k)} \tag{33}$$

Variable $y_t^{(i,k)}$ has the same interpretation as $\tilde{y}_t^{(i,k)}$ but applied to the value of the endogenous variable, and not to the deviation from the reference value.

Although the interpretation and use of the HD in exogenous variables may seem strange and impractical, it is actually of great utility when the reference value for the exogenous variables is chosen correctly. The following example describes a case in which the interpretation of the HD in exogenous variables is more easily understood. Consider the case in which the exogenous variables are introduced in the model in their first differences. The person performing the study may be asking himself the effects of the shocks and the changes in the exogenous variables over the endogenous variables. In this context, the criteria or reference value for the exogenous variables arises naturally as a base scenario of no change in the exogenous variables and no shocks. Under the described situation one has, for all t, $\overline{x}_t = 0$ and $\tilde{K}_t = K_t$. This also allows to interpret both $y_t^{(i,k)}$ and $\tilde{y}_t^{(i,k)}$ as what would have happened to the i-th endogenous variable if it were only for the changes of the k-th exogenous variable.

Algorithm 6 summarizes the steps carried out for the historical decomposition in a structural VAR-X setup.

References

Amisano, G. and Giannini, C. (1997). *Topics in structural VAR econometrics*. Springer.

Bauwens, L., Lubrano, M., and Richard, J.-F. (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press.

Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, 79(4):655–73.

Canova, F. (2007). *Methods for applied macroeconomic research*. Number 13 in Princeton Series in Applied Mathematics. Princeton University Press.

Canova, F. and De Nicolo, G. (2002). Monetary disturbances matter for business fluctuations in the g-7. *Journal of Monetary Economics*, 49(6):1131–1159.

Canova, F. and Pappa, E. (2007). Price differentials in monetary unions: The role of fiscal shocks. *Economic Journal*, 117(520):713–737.

Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174.

Chen, M. and Shao, Q. (1998). Monte carlo estimation of bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8:69–92.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review*, 89(1):249–271.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician*, 50:120–126.

Jeffreys, H. (1961). *Theory of probability*. International Series of Monographs on Physics. Clarendon Press.

Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in bayesian VAR-models. *Journal of Applied Econometrics*, 12(2):99–132.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80(2):218–230.

King, T. B. and Morley, J. (2007). In search of the natural rate of unemployment. *Journal of Monetary Economics*, 54(2):550–564.

KociĘcki, A. (2010). A prior for impulse responses in bayesian structural VAR models. *Journal of Business & Economic Statistics*, 28(1):115–127.

Koop, G. (1992). Aggregate shocks and macroeconomic fluctuations: A bayesian approach. *Journal of Applied Econometrics*, 7(4):395–411.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions-five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.

Lütkepohl, H. (1990). Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. *The Review of Economics and Statistics*, 72(1):116–25.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Moon, H. R., Schorfheide, F., Granziera, E., and Lee, M. (2011). Inference for VARs identified with sign restrictions. NBER Working Papers 17140, National Bureau of Economic Research, Inc.

Mountford, A. and Uhlig, H. (2009). What are the effects of fiscal policy shocks? *Journal of Applied Econometrics*, 24(6):960–992.

Nicholls, D. F. and Pope, A. L. (1988). Bias in the estimation of multivariate autoregressions. *Australian Journal of Statistics*, 30A(1):296–309.

Runkle, D. E. (1987). Vector autoregressions and reality. *Journal of Business & Economic Statistics*, 5(4):437–42.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Sims, C. A. (1986). Are forecasting models usable for policy analysis? *Quarterly Review*, pages 2–16.

Sims, C. A. and Zha, T. (1999). Error bands for impulse responses. *Econometrica*, 67(5):1113–1156.

Smets, F. and Wouters, R. (2007). Shocks and frictions in US business cycles: A bayesian DSGE approach. *American Economic Review*, 97(3):586–606.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Zellner, A. (1996). *An introduction to bayesian inference in econometrics*. Wiley Classics Library. John Wiley.

A An application

In this appendix some of the concepts presented in the document are exemplified by an application of Galí (1999)'s structural VAR, augmented with oil prices as an exogenous variable. The exercise has illustrative purposes only and does not mean to make any assessment on the economics involved.

The Appendix is organized as follows: first a description of the model to be used is made, then the lag structure of the reduced form VAR-X is chosen and the estimation described. Finally, impulse response functions, multiplier analysis and the historical decomposition are presented for one of the model's endogenous variables.

A.1 The model and the data

The model used in this application is original from Galí (1999) and is a bi-variate system of labor productivity and a labor measure.[12] The labor productivity is defined as the ratio between product (GDP) and labor. The identification of the shocks is obtained by imposing long run restrictions a la Blanchard and Quah (1989). Two shocks are identified, a technology (productivity) shock and a non-technology shock, the former is assumed to be the only shock that can have long run effects on the labor productivity. As pointed out in Galí (1999) this assumption is maintain in neoclassical growth, RBC and New-Keynesian models among others.

The model is augmented with oil prices as an exogenous variable with the only purpose of turning it into a structural VAR-X model, so that it can be used to illustrate some of the concepts of the document. As mentioned in Section 2 the presence of an exogenous variable doesn't change the identification of the structural shocks.

All variables are included in the model in their first differences, this is done partially as a condition for the long run identification (labor productivity) and partially because of the unit root behavior of the observed series. It should be clear that, in the notation of the document, $n = 2$ (the number of endogenous variables) and $m = 1$ (the number of exogenous variables).

Noting by z_t the labor productivity, l_t the labor measure and p_t^o the oil price, the reduced form representation of the model is given by equation (1) with $y_t = \begin{bmatrix} \Delta z_t & \Delta l_t \end{bmatrix}'$ and $x_t = \Delta p_t^o$:

$$y_t = v + B_1 y_{t-1} + \ldots + B_p y_{t-p} + \Theta_0 x_t + \ldots + \Theta_q x_{t-q} + e_t$$

In the last equation vector v is of size 2×1, matrices B_i are of size 2×2 for $i = 1 : p$ and all Θ_j are 2×1 vectors. The structural VMA-X form of the model is given (as in equation (4)) by:

$$y_t = \mu + C(L)\,\epsilon_t + \Lambda(L)\,x_t$$

with μ a 2×1 vector, each matrix of $C(L)$ is of size 2×2, and the "coefficients" of $\Lambda(L)$ are 2×1 vectors. $\epsilon_t = \begin{bmatrix} \epsilon_t^T & \epsilon_t^{NT} \end{bmatrix}$ is the vector of structural shocks.

The identification assumption implies that $C(1)$ is a lower triangular matrix, this allows us to use algorithm 2 for the identification of the shocks and the matrices in $C(L)$. Equations (5), (6) and (7) still hold.

The data set used to estimate the model consists in quarterly GDP, non-farm employees and oil price series for the US economy that range from 1948Q4 to 1999Q1. The quarterly GDP is obtained from the Bureau of Economic Analysis, and the non-farm employees and oil price from the FRED database of the Federal Reserve Bank of St. Louis. GDP and non-farm employees are seasonally adjusted. GDP is measured in billions of chained 2005 dollars, non-farm employees in thousands of persons and oil prices as the quarterly average of the WTI price in dollars per barrel.

[12] Galí uses total hours worked in the non-farm sector as labor measure in the main exercise but also points at the number of employees as another possible labor measure, here we take the second option and use non-farm employees.

Tab. 1: Marginal Densities

$m_0(Y)$	$m_1(Y)$	$m_2(Y)$	$m_3(Y)$	$m_4(Y)$	$m_5(Y)$	$m_6(Y)$
6.1379	6.1268	6.1664	6.1817	6.2414	6.1733	6.1115

The values presented are proportional to the marginal densities of the models by a factor of $10^{13}C$.

A.2 Lag structure and estimation

Choosing the lag structure of the model consists in finding values for p and q so that the estimated reduced form model satisfies some conditions. In this case we shall choose values for p and q so that the residuals (e_t) are not auto-correlated.[13] The tests indicate that four lags of the endogenous variables are necessary for obtaining non-auto-correlated residuals ($p = 4$), this result is independent of the lags of the exogenous variable. The change of the oil prices can be included only contemporary ($q = 0$) or with up to six lags ($q = 6$).

Since any number of lags of the exogenous variables makes the residuals satisfy the desired condition, the marginal density of the different models (under the Jeffreys prior) is used to determined the value of q. Each possible model only differs in the lags of exogenous variable, there are seven models indexed as $m_i(Y)$ with $i = 0 \dots 6$. The marginal density for each model is computed as in equation (23):

$$m_i(Y) = \frac{\Gamma_n\left(\frac{T-k_i}{2}\right)}{\Gamma_n\left(\frac{T-k_i-n-1}{2}\right)} |S_i|^{\frac{-n-1}{2}} 2^{\frac{n(n+1)}{2}} C$$

A presample is taken so that all models have the same effective T, since all have the same number of endogenous variables ($n = 2$), the only difference between the marginal density of two models is in k_i (the total number of regressors) and S_i (the estimated covariance of the residuals). Recalling from Section 3: $k_i = (1 + np + m(q_i + 1))$ and $S_i = \left(Y - Z_i\hat{\Gamma}_i\right)' \left(Y - Z_i\hat{\Gamma}_i\right)$.

Table 1 presents the results of the marginal densities, it is clear that the marginal density doesn't increase monotonically in the exogenous lag and that $m_4(Y)$ ($q = 4$) is preferred to the other models. Then, the VAR-X model is estimated with four lags in both the endogenous and the exogenous variables, and the contemporary value of the change in the oil price.

The estimation is carried out by bayesian methods under the Jeffreys prior as in Section 3.2. Algorithm 3 is applied to obtain 10,000 draws of the reduced form parameters, for every draw Algorithm 2 is applied, along with the identification restriction over the technology shock, to obtain the parameters of the structural VMA-X representation of the model.

A.3 Impulse response functions and multiplier analysis

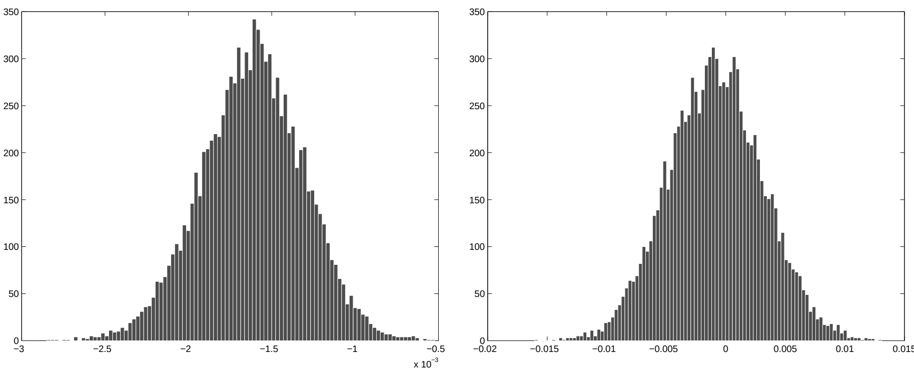
From the output of the bayesian estimation of the model the impulse response function and multipliers are computed. Note that the distributions of the IRF and the multipliers are available since the estimation allows to obtain both for each draw of the reduced form parameters. This makes possible to compute highest posterior density regions (HPD) as mentioned in Section 5.1. For doing so we presented, in Algorithm 4, the steps to be carried out in the case in which the distribution of the IRF and the multipliers in every period is unimodal. Here we present only the response of labor to a technology shock and a change in oil price as the posterior mean of the responses generated for each of the 10,000 draws of the parameters, the responses are presented along with HPD regions at 68% and 90% probability.

Before presenting the HPD for the IRF and the multipliers its necessary to check if the distribution of the responses in every period are unimodal. Although no sufficient, a preliminary test of the mentioned condition is to check the histograms of the IRF and the multipliers before computing the

[13] The auto-correlation of the residual is tested whit Portmanteau tests at a 5% significance level. See Lütkepohl (2005), Section 4.4.3.

Fig. 1: Histograms

(a) IRF: Labor to tech shock at impact (b) MA: Labor to oil price at impact



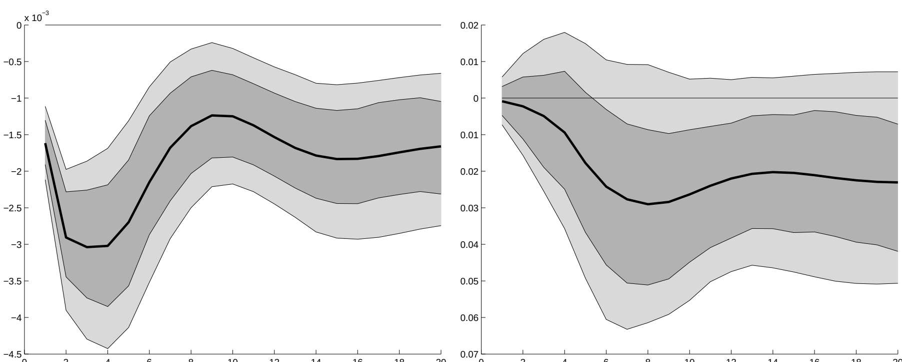
Histograms of the response of labor to a technology shock and a change in the oil price at impact. The histograms are obtained from 10000 draws of the parameters of the structural VAR-X model, and are computed with 100 bins.

HPD. Figure 1 presents the histograms for the response of labor to a technology shock (Figure 1a) and to a change in oil price (Figure 1b) at impact, the histograms for up to 20 periods ahead are also checked, but not presented. In all cases Algorithm 4 can be used.

The results are presented in Figure 2 and point to a decrease of labor in response to both a positive technology shock and an increase in oil prices, although the decrease is only significant for the response to a technology shock. The response of labor to an increase in the oil price is never significant at 90% probability and only significant at 68% probability after period 5.

Fig. 2: IRF and MA

(a) IRF: Labor to tech shock (b) MA: Labor to oil price at impact

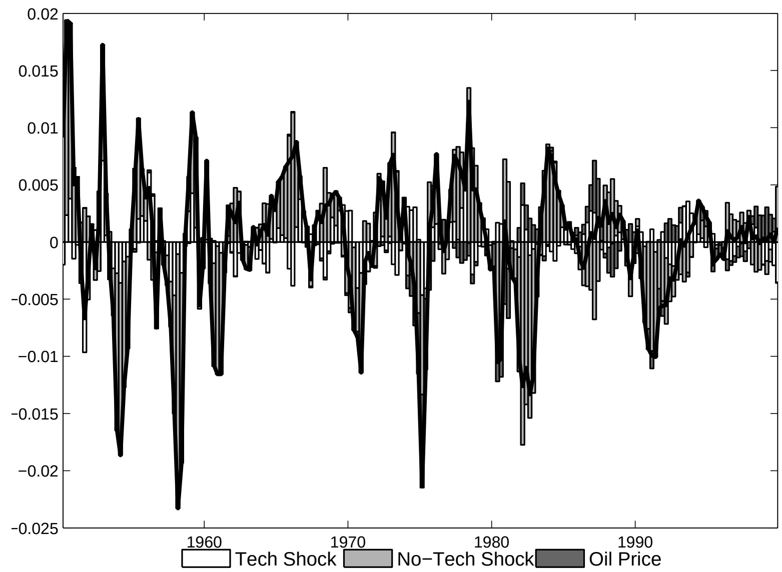


Response of labor to a unitary technology shock and a unit change in the oil price. The point estimate (dark line) corresponds to the posterior mean of the distribution of the IRF and the multipliers of labor, the distributions are obtained from 10000 draws of the parameters of the structural VAR-X model. HPD regions at 68% and 90% probability are presented as dark and light areas correspondingly.

A.4 Historical decomposition

Finally, the historical decomposition of labor into the two structural shocks and the changes in the oil price is computed. As mentioned in Section 5.2 its necessary to fix a reference value for the exogenous

Fig. 3: Historical Decomposition - Labor in first difference



variable. Since the oil price enters the model in its first difference, the reference value will be set to zero ($\forall_t\, \overline{x}_t = 0$). This means that all changes in the oil price are understood by the model as innovations to that variable.[14] In this exercise all computations are carried out with the posterior mean of the parameters. Since the Jeffreys prior was used in the estimation, the posterior mean of the parameters equals their maximum likelihood values.

Applying Algorithm 6, steps 1 to 3, the historical decomposition for the first difference of labor (relative to $\tilde{K}_t$) is obtained, this is presented in Figure 3. Yet, the results are unsatisfactory, principally because the quarterly difference of labor lacks of a clear interpretation, its scale is not the one commonly used and might be too volatile for allowing an easy understanding of the effects of the shocks.[15]

An alternative to the direct historical decomposition is to use the conditioned series (step 5 of Algorithm 6) to compute the historical decomposition of the annual differences of the series, this is done by summing up the quarterly differences conditioned to each shock and the exogenous variable. The advantage of this transformation is that it allows for an easier interpretation of the historical decomposition, since the series is now less volatile and its level is more familiar for the researcher (this is the case of the annual inflation rate or the annual GDP growth rate). The result is presented in Figure 4, it is clear that labor dynamics have been governed mostly by non-technology shocks in the period under consideration, with technology shocks and changes in the oil price having a minor effect.

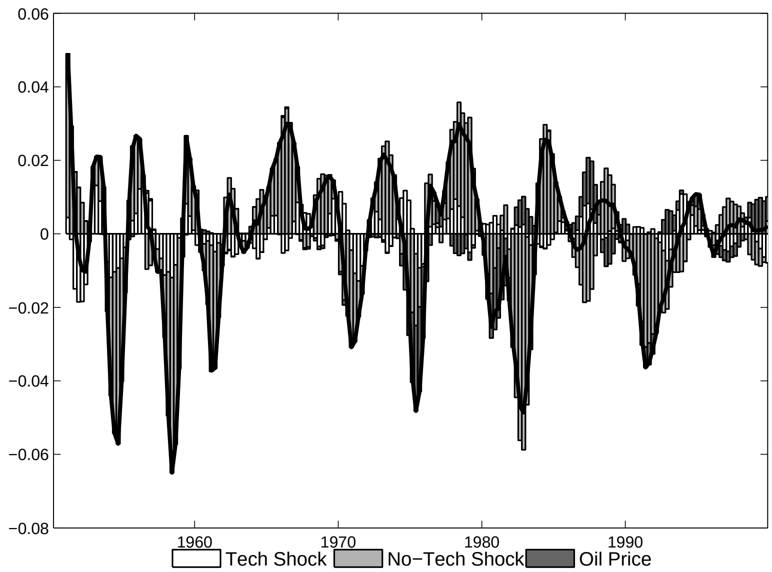
Its worth to note that decomposing the first difference of the series (as in Figures 3 and 4) has another advantage. The decomposition is made relative to $\tilde{K}_t$ with $\overline{x}_t = 0$, hence $\tilde{K}_t = K_t$ and $\tilde{K}_t \longrightarrow \mu$, this means, for Figure 3, that the decomposition is made relative to the sample average of the quarterly growth rate of the series, in that case if the black solid line is, for example, 0.1 at some point it can be read directly as the growth rate of labor being 10% above its sample average. Since Figure 4 is also presenting differences it can be shown that the new $\tilde{K}_t$ converges to the sample mean of the annual growth rate of the series, making interpretation of the decomposition easier to read.

Another alternative is to accumulate the growth rates (conditioned to each shock and the exogenous

[14] Another possibility is to use the sample mean of the change in the oil price as a reference value, in this case the innovations are changes of the oil price different to that mean.

[15] In fact the series used is not too volatile, but there are other economically relevant series whose first difference is just too volatile for allowing any assessment on the results, the monthly inflation rate is usually an example of this.

Fig. 4: Historical Decomposition - Labor in annual differences



variable) starting from the observed value of the series in the first period, this generates the historical decomposition of the level of the variable. The results of this exercise are presented in Figure 5.

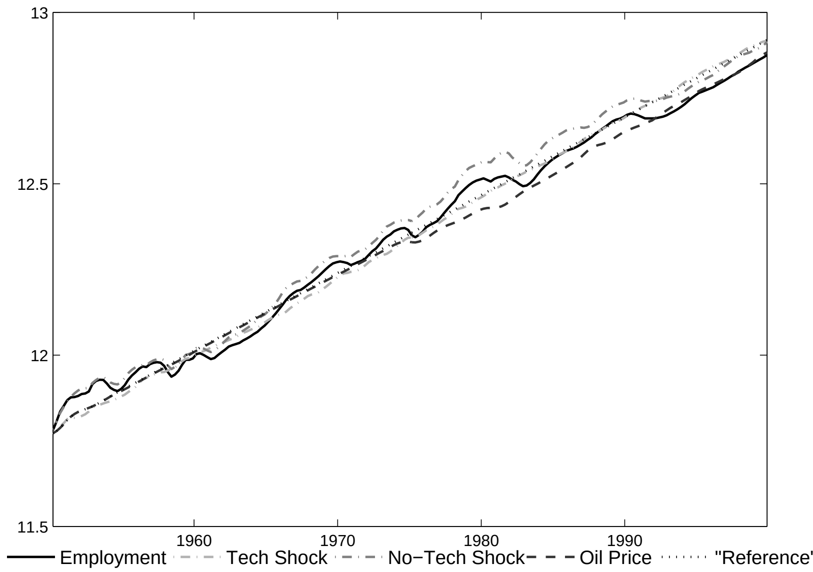
There are several points to be made about the historical decomposition in levels, the first one is that, since $\tilde{K}_t$ is also being accumulated from some initial value, the decomposition is not made relative to a constant but relative to a line, this line corresponds to the linear tendency of the series. Figure 5a plots the actual path of labor along with path conditioned to each shock and the exogenous variable and the "Reference" line, which is the accumulation of $\tilde{K}_t$. Interpretation of Figure 5a is difficult because the effect of each shock and the exogenous variable is obtained as the difference between its conditioned path and the "Reference" line, because all are moving in each period identifying that effect becomes a challenging task.

The second point arises from the interpretation of Figure 5b, which presents the decomposition of the level of labor relative to the "Reference" line, this is similar to what was presented in Figures 3 and 4. The interpretation is nevertheless more complicated. In the former Figures the decomposition was made relative to a constant, but the decomposition in levels is made relative to a line, whose value is changing in each period, this makes the reading of the level of the bars and the line more difficult. If the line is in 3 it means that the observed series is 3 units above its linear tendency.

Another characteristic of decomposition in level must be mentioned, although it is not clear from Figure 5b, the accumulated effects of the shocks over any series in the first and last period are, by construction, equal to zero. This means that the bars associated with the structural shocks are not present in both the first and last period of the sample, and that the value of the observed variable has to be explained entirely by the exogenous variables, moreover, it means that the accumulated effect of the shocks has to be dis-accumulated when the sample is getting to its end. This occurs because the accumulated effect of the shocks has to be zero at the beginning of the sample, since the effect of the shocks before that point is summarized in the initial value of the series, and because the mean of the shocks over the sample is zero (one of the properties of the estimation), this implies that $\sum_{t=1}^{T} \epsilon_t^i = 0$. When the conditioned difference series is accumulated, the effect of the shock is accumulated so that it also sums to zero. This last problem is not present in the historical decomposition in differences (or annual differences) and makes the results of the decomposition in levels to be unreliable.

Fig. 5: Historical Decomposition - Labor in level

(a) Decomposition in level



(b) Decomposition around reference value