You can use VARs for structural analyses. A comment to VARs and the Great Moderation

Fabio Canova
ICREA, Universitat Pompeu Fabra, CREI and CEPR *

November 2006

1 Introduction

A number of papers have recently criticized VAR methods. In particular, their potential inability to recover the structural relationship produced by a DSGE model has been highlighted in Giordani (2004), Chari et. al. (2005) and Benati and Surico (2006). Critiques to VAR approaches have a long history, but in its current articulation they have acquired a new dimension. In the past Cooley and Leroy (1985), Faust and Leeper (1997), Cooley and Dweyer (1998) and Canova and Pina (2005) among others, have raised doubts about traditional ways of identifying structural disturbances from VAR shocks. Often in fact, conventional identification restrictions, lack connection with the theoretical models that are then used to interpret the results. The more recent critiques appear to make a more pervasive point: in relevant situations VARs are incapable of recovering the true dynamics and the true shocks, even when appropriate identification restrictions are used. The reason is that dynamic stochastic general equilibrium models, when log-linearized around the deterministic (or stochastic) steady state, produce a VARMA DGP, where the MA component may have, at least, one large root. When a sample is short, a finite order VAR will not be able to capture the true dynamics of the structural model. In some extreme cases, and even when data is abundant, a VAR representation may not exist. The argument is theoretical sound and conditions to check for the invertibility of VARMA models as well as ways to examine for the presence of misspecification have been proposed (see e.g. Christiano et. al. (2005), Fernandez-Villaverde et. al. (2005), Canova (2005)). However, its empirical relevance still needs to be proved. Given that different authors provide apparently contradictory prescriptions to avoid the problem, a substantial confusion on the usefulness of VAR methods has emerged. This confusion has been translated in an increasing perception

*I thank Luca Benati and Paolo Surico for kindly providing their codes which made it considerably easier to effectively criticize their work. It would have taken me at least a year to write this note without their help. I also thank Luca Gambetti and Evi Pappa for useful comments.

that VAR methods are at least debatable, if not flawed, and should not be used to conduct meaningful economic and policy exercises.

The purpose of this note is to show that this perception is incorrect: VAR methods when properly used trace out the true dynamics of the endogenous variables in response to structural shocks. Clearly, as any other method, both of theoretical or empirical orientation, they give silly answers when they are improperly used. Therefore, as often happens, the warning bells that recent papers have rang are, to a large extent, cautionary footnotes that should not deter conscientious researchers from employing VARs in their empirical investigations.

To give credence to this argument, we examine one of the papers that have spurred the recent debate and show that the problems it highlights have nothing to do with the VAR methodology. Instead, they are specific to the exercise it conducts. In particular, the misspecifications encountered is the result of the choice of DGP and the lack of proper connection between the underlying theory and the estimated VAR and it is, by no means, an intrinsic failure of VAR approaches. Since most of the arguments are applicable to several of the papers which have articulated such a critique, one can conclude that structural VARs are valid methods to conduct empirical analyses, and that it is their careless use that causes interpretation problems. Nevertheless, the literature criticizing VAR methods teach us an important lesson: the link between the theory one has in mind and the VAR one estimates is often weak and must be made more explicit before empirical investigations become solid and credible.

2 A case study: VAR analysis and the Great Moderation

Benati and Surico (2006) have written a provocative paper which has stirred considerable discussion in academic and policy circles. Their starting point is that VAR analyses seems to contradict the majority of the narrative accounts of the monetary history of a number of developed countries. In particular, VAR analyses favor the so-called "bad-luck" hypothesis - the high level of inflation of the 1970s and, to some extent, the high volatility of output and inflation in the same period were the result of a series of bad shocks drawn from distributions with large variances - while the common wisdom points to the "bad-policy" hypothesis, as the main explanation for both of these phenomena. According to the folklore, central banks were loose in fighting inflation in the 1970s and this led to undesired consequences. When they chose to make nominal interest rates react to inflation more than proportionally, the problems of the 1970s disappeared. While this common wisdom is not unanimously shared, it is often used as benchmark to compare alternative explanations of the facts. Also, while the results present in the literature are far from being as polarized as the authors depict them - we will came back on this issue in section 3 - one can take their claim as a useful starting point to examine whether the methodology is flawed and responsible for the somewhat unconventional outcome.

The exercise of Benati and Surico is simple: they take the three-equations New-Keynesian model and use a number of parameterizations to design an experiment where, under the null, there is a structural break in the coefficients of the policy rule. They then ask: what would structural VAR analyses tell us if this DGP was the truth? Would they be able to recognize that there was a break in the structural coefficients or would they mistakenly tell us that the variance of the shocks has changed? With the aid of a Monte Carlo exercise, they show, to the surprise of many, that reduced form VAR analyses detect a change in variances but not a change in the dynamic coefficients. Furthermore, they show that a researcher employing the correct identification scheme would mistakenly confuse changes in the structural impact coefficients for changes in the variances of the structural shocks.

This unexpected outcome led the authors to question interpretations of the Great Moderation episode obtained with reduced form or structural VARs analyses and to call, as an alternative, for direct structural estimation, presumably by maximum likelihood or by Bayesian methods, as these approaches are supposed to provide a more solid evidence about the nature of the structural changes modern economies display. As an aside, one should note that direct structural estimation gives far from univocal answers to the question of what drives the "Great Moderation". For example, while Lubik and Schorfheide (2003) find evidence consistent with the bad-policy paradigm, Canova (2004) find evidence of structural changes in the parameters describing the private sector behavior but no statistically change in the parameters regulating the policy rule or in the variance of the policy shock (see, for a complementary evidence, Cogley and Sbordone (2005)), while Collard and Dellas (2006) find that once learning is allowed, the evidence for breaks is inexistent.

Benati and Surico attribute the problems to the inability of VAR models to approximate the DGP they consider, and question the idea that detected changes in the variance of reduced form and of structural shocks give reliable information about the nature of the changes experienced in the actual economies. But why are VAR bad? Is it because they are unable to capture the truth or because the design of the experiment is poor? Are any of the auxiliary assumptions employed crucial to determine the conclusions the authors reach? We break our argument in two parts. First, since the experiment involves several potential sources of misspecification, we isolate those which we believe are critical to deliver the main results. Second, we explain why VARs detect a change in the variance of the shocks and why they fail to detect changes in the coefficients of the model.

To summarize our points. There are two reasons why VAR methods find changes in the variances across regimes. First, the population variance of the reduced form shocks does changes across regimes, because the matrix of impact coefficients in the structural representation is different and because expectational shocks play a role in one regime but not in the other. Second, in one regime the estimated VAR omits a variable (expected inflation) which is jointly generated with the three (the nominal interest rate, inflation and the output gap) which are used in the VAR exercise. Omitted variables induce, among other things, biases in the coefficients if omitted variables are correlated with the included ones, which is true by construction here, and overestimate the variance of the shocks. Since in the determinate regime, there are no omitted variables, expectational errors are absent and the population variance of the reduced form shocks falls, it is far from surprising that reduced form VAR methods will detect a fall in the variances when moving from one regime to the other. Structural analyses, on the other hand, detect with high probability that the change in reduced form variances is due to changes the structural impact coefficients, if estimation biases are taken care of.

As far as the failure to detect difference in the dynamics, it turns out that the population dynamics in response to shocks are qualitatively similar in the two regimes and the quantitative differences are small. Since relatively small samples are used, small sample biases dominated these differences and bias the tests employed toward uniformity across regimes. Clearly, the small sample bias is independent of the empirical methodology used to recover the structural relationships and is present with any classical estimation method, both of structural or reduced form nature.

We conclude by showing that direct estimation of the structural relationships is not necessarily the answer to examine the Great Moderation question. Direct estimation faces important identification problems and, for example, it is possible to have two equally reasonable structures, featuring different characteristics, which produce similar dynamics in response to shocks. Under these conditions, information external to the models and the data under consideration must be used to select the appropriate DGP. Given that this information is scarce, structural methods leave us in a limb. Since no method is uniformly superior and since misspecification is likely to induce problems no matter which approach one uses, progress on the issues of interest can come only from integrating, rather than counterpoising, VARs and direct structural analyses, as suggested, for example, in Canova (2002), Del Negro and Schorfheide (2004) or Del Negro et. al. (2005).

2.1 The design of the experiment

The model considered is a standard three equation New-Keynesian model which includes a log-linearized Euler condition, a log-linearized Phillips curve and a policy rule describing how nominal interest rates react to the output gap and inflation. In log deviation from a non-stochastic steady state, the equations are:

$$x_t = x_{t+1|t} + \tau(R_t - \pi_{t+1|t}) + g_t \tag{1}$$

$$\pi_t = \beta\pi_{t+1|t} + \kappa x_t - z_t \tag{2}$$

$$R_t = \phi_r R_{t-1} + (1-\phi_r)(\phi_\pi \pi_t + \phi_x x_t) + e_{R,t} \tag{3}$$

where $g_t = \rho_g g_{t-1} + e_{g,t}$, $z_t = \rho_z z_{t-1} + e_{z,t}$, x_t is the output gap, π_t the inflation rate, R_t the nominal rate and the notation $t+1|t$ denotes conditional expectations. Here, g_t is a generic demand shifter, z_t exogenously shifts the marginal cost of production while $\beta, \kappa, \tau, \phi_r, \phi_\pi, \phi_x, \rho_g, \rho_x, \sigma_e, \sigma_g, \sigma_z$ and ρ_{gz}, the contemporaneous correlation between g_t and z_t, are parameters.

To simulate data from this model in two different regimes Benati and Surico use a variety of parameterization, including the one of Lubik and Schorfheide (2004) (LS) obtained estimating the model on US data with Bayesian methods, and the one of Clarida et. al. (2000) (CGG), obtained via calibration of the relevant relationships. In both cases, the parameters imply that the first is a regime where indeterminacies are present - due to the weak response of the nominal interest rate to inflation- and the second, a regime where a determinate equilibrium exists.

Table 1: Parameterization and population statistics

	CGG		LS	
Parameter	Regime 1	Regime 2	Regime 1	Regime 2
β	0.99	0.99	0.99	0.99
τ	1.0	1.0	1.45^{-1}	1.45^{-1}
κ	0.3	0.3	0.77	0.77
ρ_g	0.9	0.9	0.68	0.68
ρ_z	0.9	0.9	0.82	0.82
σ_g	1.0	1.0	0.27	0.27
σ_z	1.0	1.0	1.13	1.13
σ_R	0.25	0.25	0.23	0.23
ϕ_π	0.83	2.15	0.77	2.19
ϕ_x	0.1	0.93	0.17	0.30
ϕ_R	0.68	0.79	0.60	0.84
ρ_{gz}	0	0	0.14	0.14
var(R)	284.45	404.62	17.22	11.31
var(π)	389.34	344.8	30.36	9.56
var(x)	774.8	601.8	31.01	43.11
$AR(1)_\pi$	0.61	0.65	0.46	0.41
$AR(2)_\pi$	0.41	0.49	0.25	0.18
$AR(3)_\pi$	0.29	0.41	0.16	0.09
$AR(4)_\pi$	0.22	0.35	0.11	0.05
$AR(5)_\pi$	0.18	0.31	0.08	0.05

While the major points we will make are independent of the parameterization, it worth stressing that both parameterizations fail to capture important features of the experience they are supposed to mimic. Table 1 shows the values of the parameters together with the unconditional population variance of the nominal rate, the inflation and the output gap and the autocovariance function of inflation at lags from 1 to 5 in the two regimes. Two features are evident. While with the CGG parameterization the variances of inflation and the output gap fall across regimes, the variability of the nominal interest rate counterfactually increases. In addition, the persistence of inflation increases in the determinate regime, exactly the opposite of what we observe in US data. Second, with the LS parameterization the population variance of the inflation and interest rates fall considerably, while the population variance of the output gap almost doubles when we move from indeterminate to determinate regime. The persistence of inflation marginally falls but this comes together with a substantial increase in all the entries of the autocovariance function of the nominal interest rate. Hence, regardless of the problems we discuss below, the choice of parameterization (and probably of the model) is unfortunate as several important features of the

Great Moderation episode are absent DGPs used [1].

There are many ways to understand the nature of the results Benati and Surico present. We find it most informative to look at the law of motion of the controls (the nominal rate, the inflation rate and the output gap) as a function of lagged values of the states and of reduced form shocks. To avoid lengthy repetitions, we just present the arguments for the LS parameterization but, it should be clear, that both choices face exactly the same problems.

The log-linearized solution in the indeterminate regime is:

$$\begin{bmatrix} \widehat{R}_t \\ \widehat{\pi}_t \\ \widehat{x}_t \end{bmatrix} = \begin{bmatrix} -0.24 & -0.41 & -0.33 & -0.28 \\ 0.23 & -0.19 & -0.59 & -0.15 \\ 0.19 & -0.45 & 0.07 & 0.19 \end{bmatrix} \begin{bmatrix} \widehat{R}_{t-1} \\ \widehat{\pi}_{t-1} \\ \widehat{y}_{t-1} \\ \widehat{\pi}^e_{t-1} \end{bmatrix} + \begin{bmatrix} \widehat{u}_{1t} \\ \widehat{u}_{2t} \\ \widehat{u}_{3t} \end{bmatrix}, \Sigma_u = \begin{bmatrix} 0.31 & & \\ 0.96 & 3.39 & \\ -0.15 & -0.42 & 0.40 \end{bmatrix}$$

where π^e_t is expected inflation, while the log-linearized solution in the determinate regime is:

$$\begin{bmatrix} \widehat{R}_t \\ \widehat{\pi}_t \\ \widehat{x}_t \end{bmatrix} = \begin{bmatrix} -0.39 & -0.31 & 0.11 \\ -0.15 & 0.30 & -0.12 \\ -0.23 & -0.16 & 0.44 \end{bmatrix} \begin{bmatrix} \widehat{R}_{t-1} \\ \widehat{\pi}_{t-1} \\ \widehat{y}_{t-1} \end{bmatrix} + \begin{bmatrix} \widehat{u}_{1t} \\ \widehat{u}_{2t} \\ \widehat{u}_{3t} \end{bmatrix}, \Sigma_u = \begin{bmatrix} 0.09 & & \\ 0.24 & 0.98 & \\ -0.21 & -0.55 & 0.89 \end{bmatrix}$$

For later use, we rewrite these solutions as $y^i_t = A^i_1 y^i_t + A^i_2 \pi^e_t + u^i_t$ in the case of indeterminacy and $y^d_t = A^d_1 y^d_t + u^d_t$ in the case of determinacy where $u^j_t = B^j e_t + F^j \epsilon_t$ are reduced form shocks $j = i, d$, e_t are the structural disturbances and ϵ_t are expectational errors.

As these expressions show, the model has important implications for expected inflation under indeterminacy - it becomes a state variable - while this is not the case when determinacies are present. Also, while the structural model differs across regimes only in the coefficients of the policy equation, the solution is such that both lag dynamics as well as the variance of the reduced form shocks change across regimes. Interestingly, the variance-covariance matrix of the reduced form shocks display the same features as the unconditional variances we have reported in table 1 (with the variance of the reduced form shocks hitting the output gap almost doubling in size) and relatively speaking, the regime change is more pronounced in the variance of the reduced form shocks than in the lagged dynamics of model. In these conditions, any unbiased empirical methods is bound to find larger changes in the estimated variances than estimated lagged coefficients.

[1] The numbers that Benati and Surico present in table 2 are different from those appearing in this table because they present conditional moments (in particular, conditional on the path for expected inflation and expected output) while here we present marginal moments, once the influence of expected inflation and expected output movements is intergrated out. Clearly, since neither expected inflation nor expected output are used in the exercises, our numbers are more informative about the features of the DGP than those reported by the authors.

2.2 Step 1: Should we expect estimated variances to stay constant across regimes?

To understand why the population variances of reduced form shocks change across regimes and vary in the way they do is worth comparing the solutions obtained in the two regimes. First, the matrices B^d and B^i differ reflecting changes in the coefficients of the policy rule - this change would be present even when changes in the coefficients of the policy rule would occur within a regime. Second, the matrix F^d is identically zero while F^i has non-negligible entries - this change is present because, in the indeterminate regime, expectations matter. Third, omitted variables play a role in a regime but not the other. In fact, under indeterminacy the model has implications for the output gap, inflation, the nominal interest rate and inflation expectations but only the first three variables are used in the VAR. Since the loadings of the three endogenous variables on lagged expected inflation (the matrix A_2^i) are non-negligible relative to the size of the other coefficients (the matrix A_1^i), expected inflation plays the role of an additional regressor omitted in all three equations. These three observations imply that in the determinate regime, the reduced form error is $u_t = B^d e_t$ while in the indeterminate regime the composite reduced form error is $A_2^i \pi_t^e + B^i e_t + F \hat{\imath} \epsilon_t$.

Given this structure, it is not particularly surprising to find that VAR methods detect a change in the variance of estimated reduced form shocks - it would be surprising if they do not. The evidence that Benati and Surico present is overwhelming because in addition to population differences, small sample estimation biases play a role (in both regimes only 100 data points are used) and small sample lag selection distortions are also likely to be present (standard lag selection criteria like AIC and BIC are biased in small samples). However, it should be clear that, even when the sample is of infinite length and there are no biases in the criteria to choose the lag length, one *must* find differences in the estimated variances of VAR shocks.

If one is interested in measuring the ability of VAR methods to answer the questions of interest, one should also design an experiment where only the lagged coefficients change. Benati and Surico try to do this when they consider variations within a determinacy regime - this knocks out the effects due to misspecification and expectational errors - but even this design implies that both the contemporaneous and the lagged coefficients change. To have a design with the required features, one should study within regime changes when the nominal interest rate reacts to *lagged* output gap and *lagged* inflation since, by construction, the matrix of impact coefficients is fixed across regimes while the matrix of lagged coefficients changes. We have performed the same simulations of Benati and Surico on this alternative setup and found the VAR methods correctly recognize the true DGP when data is abundant (see table 2). The conclusion is therefore that VARs have good size and power properties, when small sample biases are taken care of.

Since the presence of omitted variable biases is partially responsible for the change in the variance of reduced form shocks and unnecessarily clouds the message of the exercise, it is natural to ask whether such misspecification could be detectable and whether it could be removed from the analysis. Any econometric textbook tells us that, theoretically, omitting a variable from a regression model causes a) biased estimates of the coefficients on the included variables, whenever omitted and included ones are correlated (which is likely to be the case here because of the general equilibrium nature of the model and the choice of the "continuity" solution (see Lubik and Schorfheide (2003)); b) biased estimates of the variance of the residuals, which now include the true sources of disturbances plus the variability of the omitted variable. In practice, it is easy to design diagnostics to detect this type of misspecification in the course of the analysis. For example, if any measure of inflation expectations is available and it reasonably proxies for true inflation expectations, then correlating such a measure with reduced form (and/or structural) residuals would indicate the presence of an omitted variable. This type of specification checks are indeed conducted in the literature on the Great Moderation, see e.g. Canova, Gambetti and Pappa (2006), and are informative about the nature of the true DGP in non-experimental settings.

Table 2: Wald tests: Backward looking policy, within regime changes

		Coefficients			Variances		
Sample size	Percentile	Interest rate	Inflation	Output	Interest rate	Inflation	Output
T = 100	5	0.00	0.00	0.00	0.24	0.00	0.00
	50	0.00	0.00	0.00	0.94	0.00	0.26
	95	0.00	0.96	0.85	1.00	0.94	1.00
T = 200	5	0.00	0.00	0.00	0.50	0.03	0.04
	50	0.00	0.20	0.00	0.96	0.32	0.66
	95	0.00	0.84	0.76	1.00	0.99	1.00

In the context of the exercise they run, the authors should have noticed that under indeterminacy, adding lags to model did not reduce the serial correlation present in the residuals and that lagged endogenous variables are correlated with the VAR residuals. Therefore, failure to link the specification of the VAR to the features of the theoretical model is responsible for the extreme nature of results the authors present.

There is another, and probably more informative way to show what happens when one omits a variable which is jointly generated with the included ones in a DSGE model. Consider for this purpose a VAR(q) of the form

$$\begin{bmatrix} A_{11}(\ell) & A_{12}(\ell) \\ A_{21}(\ell) & A_{22}(\ell) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} e_t \end{bmatrix}$$

.

where both y_{1t} and y_{2t} are vectors of variables and say y_{1t} are the observable variables (say, output gap, inflation and the nominal rate) and y_{2t} are the nonobservable variables (say, inflation and output expectations). Then, the marginal representation for y_{1t} is

$$[A_{11}(\ell) - A_{12}(\ell)A_{22}(\ell)^{-1}A_{21}(\ell)]y_{1t} = (I - A_{12}(\ell)A_{22}(\ell)^{-1})e_t \equiv v_t$$

which is a $VARMA(\infty, \infty)$, Four important conclusions can be drawn from this representation. First, if the true model has a finite order VAR representation when all the variables it generates are included in the VAR, it has a VARMA(∞, ∞) representation when only observable variables are used. While under determinacy the matrix A_{12} is close to zero, this is not the case under indeterminacy - so that misspecification is likely to be more important in the latter regime. Second, the variance of v_t is always greater or equal to the variance of e_t, except in the trivial case in which $A_{12}(\ell) = 0$ and will decrease as y_{2t} looses importance in the determination of y_{1t} - which is exactly what happens when we move from the indeterminacy to the determinacy regime. Third, by construction, v_t will be correlated with the omitted variable y_{2t} and, as a consequence, with the lags of y_{1t}, and this will occur even when the correct identification scheme is used. Finally, v_t fails to recover the timing of the information in the true residuals (both current and lagged values of e_t are present). That is, estimated VAR innovations will have little to do with the true innovations of the model.

What does structural analysis add to this picture? Very little. When the composite error term is of the form $A_2^i \pi_t^e + B^i e_t + F^i \epsilon_t$ and no a-prior information is available on the relative importance of the three components, it is difficult to attribute changes in the reduced form variances to its various effects and hard to guess exactly what structural decompositions would do. To gain some intuition on how sign based structural decompositions would interpret the change in the variance of reduced form shocks, we have run the following Monte Carlo exercise. We computed the population variance of u_t in the two regimes, performed an eigenvalue eigenvector decomposition and set $A = PD^{0.5}$, where P is the matrix of eigenvectors and D the matrix of eigenvalues. Then, we drew a matrix of N(0,1) random variables, applied the QR decomposition to this matrix and check whether $A_0 = A * Q'$ satisfies the sign restrictions necessary for the identification of the three shocks used by Benati and Surico - that monetary shocks imply a positive contemporaneous response of nominal rates and a negative response of the output gap and inflation; that demand shocks imply positive contemporaneous effects on all three variables; and that a cost push shock has a contemporaneously positive impact on inflation and the nominal rate and a negative impact on the output gap. We found that the probability that the square difference between the impact matrix of the two regimes is larger than 0.5 is 0.35, 0.48, 0.19 for the three equations and that the square difference between the median impact coefficients in the two regimes in the three equations are 0.32, 0.33, 0.17, respectively. Given the complexity of the structure of the error term, this is a quite good outcome. For comparison, when the change is within a regime and the policy rule is specified so that interest rates react to lagged output gap and lagged inflation, the square difference between the median impact coefficients in the two regimes in the three equations are 0.14, 0.13, 0.09. These results could be improved if a stronger set of sign restrictions is imposed, for example, if we add some restrictions at horizons one or two - what may look like change in the variance in the impact matrix may in fact be inconsistent with the longer run dynamics the shocks induce. But even with this minimalist set of restrictions, it is clear that sign restrictions give a reasonable account of the structural relationships when applied to population quantities. Two qualifications are important to put these results into perspective. First, sign restrictions are weak and not designed to recover the exact DGP of the data - this is why, even when there are no changes, some differences exist across regimes. Second, these conclusions are obtained using population quantities since the addition of estimation errors strongly bias the results, in particularly, when structural relationships are examined.

To conclude, the finding that variances of reduced form (and /or structural) shocks are changing across regimes is unsurprising and due to the fact that a) the design is such that expectational errors play a role in one regime but not in the other, b) changes in the coefficients in the policy rule translate in variations of the B matrix across regimes, c) an improper specification of the VAR in indeterminacy regime. Clearly neither a), b) or c) have anything to do with estimation or flaws in the VAR methodology.

2.3 Step 2: Why VAR fail to detect changes in the coefficients across regimes?

To answer this question it is useful to examine the population responses generated by the three shocks in the two regimes (see figure 1).

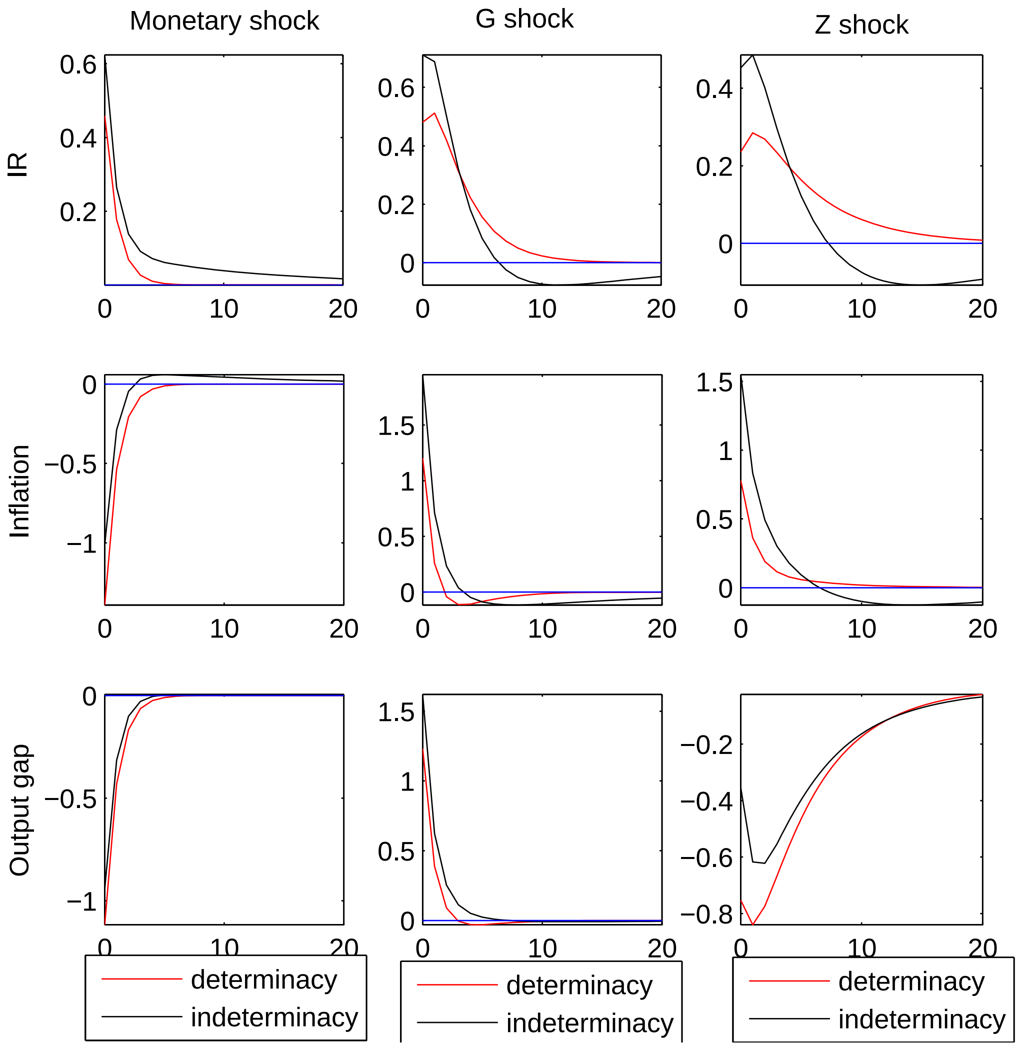

Figure 1: LS parameterization

Three features of these responses are clear: first, consistent with the discussion in the previous subsection, the impact coefficients are quantitatively different. Second, there are quantitative differences in lagged dynamics across regimes, but they tend to die out relatively quickly. Third, the qualitative features of the responses - in particular, their sign and shapes - are fairly similar under the two regimes. These features are also present in the unconditional population autocorrelation functions: since these are weighted averages of the columns of figure 1, even unconditionally, the two regimes are fairly similar.

Given these similarities, one should not be particularly surprised to find that estimated dynamics are similar across regimes. However, a closer look at details of the experimental design more clearly indicates the reasons for this outcome. Benati and Surico, in fact, simulate only 100 data points for each regime and for each simulation a trivariate VARs with up to 4 lags is used. Given this setup, small sample biases dominate the small differences existing in the dynamics across regimes. It is well known that estimates of VAR coefficients are downward biased in small samples and that this bias could reach 15-20 percent depending on the DGP.

To measure the magnitude of the problem in the present context, we have performed two simple back-of-the-envelope calculations. First, given the trivariate VAR(1) coefficients and the unconditional variance of the three variables in the two regimes we ask how large should T be when no estimation biases are present so that a $t - test$ would reject the null hypothesis of no difference in the diagonal elements of matrix of VAR(1) coefficients in the two regimes. It turns out that differences would be detectable only if at least 120 data for the interest rate equation, 410 for the inflation equation and 2300 data points for the output gap equation would be available in each regime. Second, we take the estimated VAR(1) coefficients in the two regimes across simulations and construct the distribution of sample size needed to detect differences in each simulated sample. A few percentiles of this distribution, which is highly skewed, are presented in table 3. For example, in the median simulation we would need between 124 and 199 data points to detect differences in the lagged inflation coefficient in the inflation equation and over 2200 to detect differences in the lagged output gap coefficient in the output gap equation.

Table 3: Distribution of sample size T

		1 lag VAR			4 lags VAR		
Sample size	Percentile	Interest rate	Inflation	Output	Interest rate	Inflation	Output
T = 100	5	105	41	81	33	101	97
	50	166	124	286	88	199	271
	95	516	2244	2293	746	2281	1599

In sum, differences in the lagged dynamics across regimes are fairly small and, even without estimation biases, about twice as many observations in each regime to detect differences. Since small sample biases are substantial, considerably larger samples than those used are needed to make the exercise informative.. Hence, any empirical method will have hard time to detect difference with $T=100$.

Once again, structural analysis adds little to these points. To the extent that small sample biases are present in the estimation of the covariance matrix of the reduced form shocks this will translate in biases in estimated impact coefficients. Since, as we have mentioned, sign restrictions are not designed to recover the quantitative features of the DGP, even in the ideal situation in which misspecification is absent and an infinite amount of data is available, it is very unlikely that sign restrictions can pin down exactly the true matrix of contemporaneous parameters.

What do we learn from the evidence we have provided so far? Three points can be made. First, none of the problems highlighted by Benati and Surico have to do with VAR methods, per se. It is the choice of experimental design, failure to recognize the presence of omitted variables in one regimes and the choice of relatively small sample size which drive the results they obtain. VARs deliver sensible answers when they are sensibly specified and data is abundant. Absent these two preconditions, any empirical model is bound to give dubious conclusions. Second, structural methods will not able to correctly detect the correct DGP exactly for the same reasons which generate problems for the VARs: the experimental design make the alternative regime local to the null regime; misspecification is likely to remain if one only uses the observables of the model; and the small sample problem will not go away. As a matter of fact, structural methods are more liable to problems in the presence of omitted variable misspecifications. Maximum likelihood, for example, is extremely sensitive to misspecification of the structural relationships and the omission of an important state variable is likely to produce important distortions. Third, Benati and Surico's exercise provides refreshing evidence on the fact that mechanical use of VARs, and routinary selection of the variables to be used, may be dangerous. One should always try to specify as tightly as possible the link between the empirical analysis and the theoretical model before estimation is performed and an economic interpretations of the results sought.

3 Implications for the analysis of the Great Moderation,

While the exercise Benati and Surico conduct has little to say about VAR methods, it has some implications for the literature concerned with the Great Moderation issue. The authors highlight two major conclusions of their investigation: first, variations in the variance of reduced form shocks across regimes do not necessarily indicate evidence of "bad-lack"; second, counterfactual exercises conducted switching either the parameters or the variances across regimes are uninformative about the true nature of the DGP. While, in general, one can hardly disagree with both points, it is also worth remembering that first, changes in the variances of reduced form or structural shocks is not the only thermometer used to give credit to the "back-luck" hypothesis and, second, that counterfactual exercises often violate a basic version of the Lucas critique.

First of all, most of the analyses present in the literature (including for example, Canova and Gambetti (2004), Benati and Mumtaz (2005), Primiceri (2005), Gambetti et al. (2005)), Sims and Zha (2005) detect not only changes in the variances of reduced form and structural shocks but also changes in the structural coefficients of the model. Therefore, the design the authors used (only variances change, estimated dynamics are similar across regimes) does not necessarily mimic what it is typically found in the empirical literature. Second, support for the good luck hypothesis is typically found when the timing of the changes in the structural coefficients do not correspond to the timing of the changes the narrative literature has pointed out and when changes in the structural coefficients do not translate into time variations in normalized impulse responses or normalized spectra. Third, when significant time variations in the impulse responses are detected, the timing of these changes is typically carefully analyzed and the magnitude of the variations compared with the magnitude of the variations in structural variances. Therefore, Benati and Surico's warning it is inconsequential for carefully executed VAR analyses, for the interpretation they propose and for the conclusions they reach. As far as I am aware, no investigator has claimed support for the "bad-luck" hypothesis simply because she has detected changes in the variances of reduced form shocks.

Second, while some authors perform counterfactual exercises to make their point stronger, the conclusions they reach is never solely based on the results of counterfactuals. What is really problematic is not the conclusions researchers reach but the design of the experiments they conduct. If the regime is truly changing, one should expect estimated coefficients and estimated variances to have different distributions across regimes. If the distributions do not overlap, at least partially, it is very unwise to consider a counterfactual where a parameter value is set to, say, the 99th percentile of the estimated distribution, or worse, with probability one it was never experienced in that sample. The literature is typically silent on this issue. However, there are instances where a more careful analysis is performed. For example, Canova and Gambetti (2004) obtain the posterior distribution of the coefficients of the model and change coefficients in the policy rule in way which is consistent with this distribution. Sims and Zha (2006) conduct a similar exercise in the context of their estimated Markov switching specification. These experiments are likely to be informative about the issues of interest, they do not suffer for Lucas critique type arguments and are, by construction, well designed. Interestingly, when carefully performed these exercises suggest that policy has a role to play, even though it is not a dominant one. Hence, the conclusion that counterfactuals are uninformative can not be generally made. It is only when the estimated uncertainty is not taken into account that these exercises are uninformative about the issues of interest.

Although Benati and Surico do not stress this aspect in their paper, one of the important conclusion of their exercise is that, under indeterminacy, expected inflation plays a role. Since empirical analyses typically neglect inflation expectations when estimating VARs in samples where these are potentially important, the conclusions they reach may be flawed. But is it this really the case? Canova, Gambetti and Pappa (2006) analyze whether the structural residuals of their TVC-VAR are correlated with inflation expectations, constructed using the term structure of interest rates. They find little evidence that the structural shocks they construct are, on average, contaminated by the absence of a measure of inflation expectations from the system. They also find that the recursive contemporaneous correlations with inflation expectations are approximately of the same magnitude up to 1979 and after 1982 but switch sign (from positive to negative) roughly in the middle of the 1980s. This last implication goes against the conventional wisdom formalized in the indeterminacy-determinacy story, and cast doubts about the idea that to characterize the dynamics of US data in the 1970s, we need to resort to indeterminate equilibrium.

4 Is structural estimation is the answer?

While in the current version of the paper the author do not propose any alternative to structural VAR when analyzing the Great Moderation issue, one would presume that the only feasible alternative coming out of their exercise is some form of direct structural estimation. Such an approach is typically believed to allow solid economic inference and the advantages of structural estimations are well known. However, one should also remember that structural estimation face a number of important problems. First, estimates obtained, for example by ML, are consistent and enjoy good properties only when the model is correctly specified. A three equation model can be hardly considered a well specified economic structure. However, even larger version of the same model are subject to important specification problems. Hence, measurement errors or other arbitrary devices are used to complete its probabilistic structure, therefore making structural estimation not so structural. Second, as already mentioned in the introduction, structural estimation does not necessarily provide a unique answer to the question at stake. Depending on a set of auxiliary assumptions one may reach one conclusion or another. Third, as perhaps more importantly, structural estimation faces important identification problems and it is relatively easy to reproduce, for example, the dynamics generated by the model under indeterminacy in setup where only a determinate equilibrium exists. Following Canova and Sala (2005), we take the impulse responses produced by the indeterminate regime as if they were the truth and try to find a parameterization of the model under determinacy which goes as close as possible to match these dynamics. We plot the "true" and the "alternative" dynamics in figure 2.

While the match is not perfect, one can easily see that the serial correlation properties of interest rates, output and inflation in response to shocks are closely reproduced, at least for the first five horizons. If we had randomized the parameters of the determinate solution, using the standard errors associated with the solution, the impulse responses obtained under indeterminacy would have be contained in the 68 percent band for the determinate responses (and viceversa). Hence, the two impulse responses are for all purposes, indistinguishable. The parameters which generate the two impulse responses are in table 3. Three interesting features need to be stressed. First, we need to change all the parameters of the model to match the dynamics produced under indeterminacy with a determinate parameterization. Second, the variance of the shocks, need to be adjusted mainly downward to match the dynamics. Therefore, if both regimes were characterized by a determinate equilibria, one is supposed to find an increase in the variance of the structural shocks driving the economy, rather than a decline. Third, as noted also in Canova and Sala (2005) the two policy parameters are not separately identifiable. Since the objective function displays a positive ridge in these two dimensions, a high ϕ_y is found when ϕ_π is forced to be greater than one.

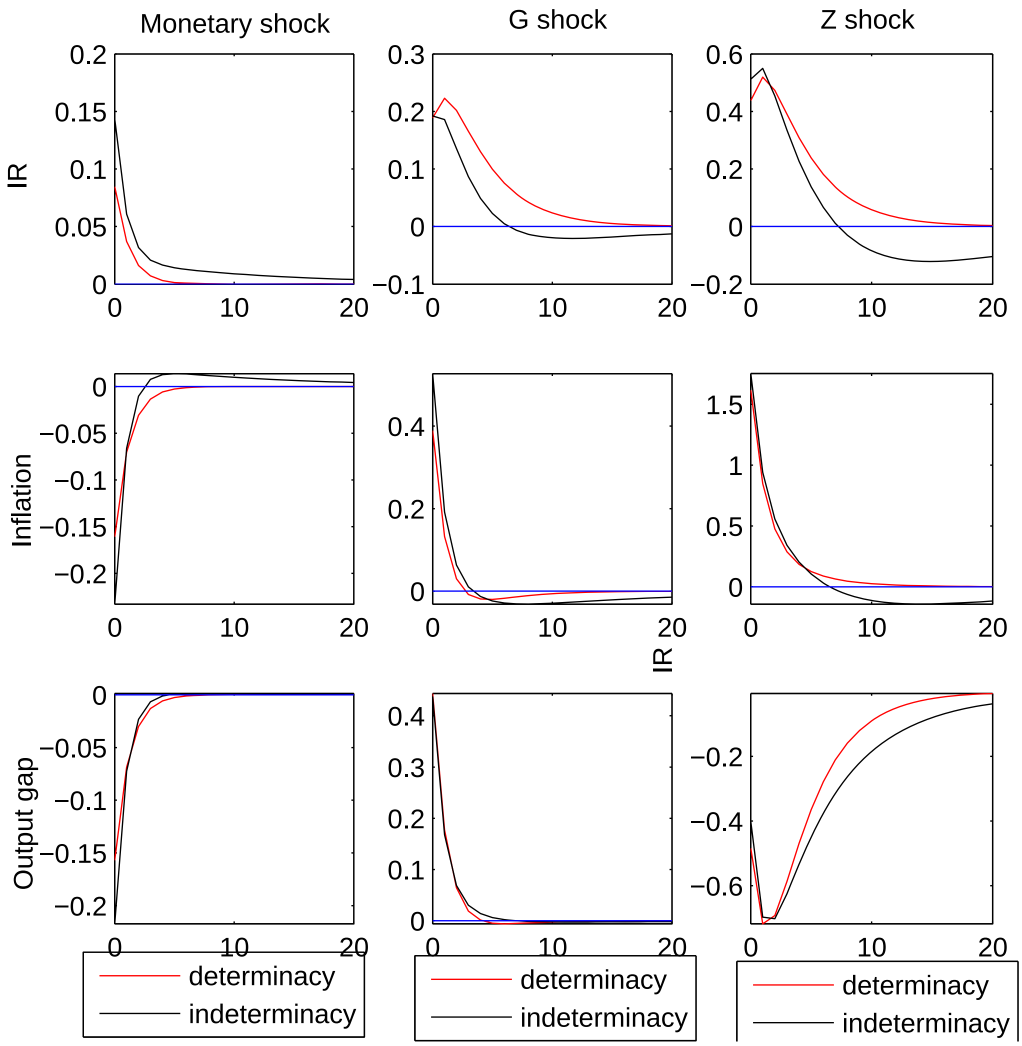

Figure 2: Alternative dynamics for regime 1

What can one conclude from this exercise? First, that structural evidence is not very informative about the existence of two possible regimes in the data. It is possible to fit the actual data without any need to assume the existence of indeterminacy (as shown in Canova (2004)) and the dynamics produced under indeterminacy are reproducible with a model where only a determinate equilibrium is present. Second, structural evidence is not necessarily more informative than VAR evidence. When all the parameters of the model are well identified and no misspecification of the structural relationship is present, structural methods have an edge. Otherwise, and this seems to be the norm, the opposite may be true. Third, since no method is uniformly superior, progress in understanding the Great Moderation episode can be made integrating various methods of analysis.

Table 3: Parameterization producing figure 2

Parameter	Indeterminate Estimate	Determinate
β	0.99	0.99
τ	0.68	0.57
κ	0.77	0.58
ρ_g	0.68	0.74
ρ_z	0.82	0.74
σ_g	0.27	0.29
σ_z	1.13	1.05
σ_R	0.23	0.15
ϕ_π	0.77	1.75
ϕ_x	0.17	0.82
ϕ_R	0.6	0.81

5 Conclusions

There is very little to add to what I have already said. VAR methods when properly used can inform us about the nature of the true DGP. But to do so, statistical methods should be carefully employed to avoid possible misspecifications; judgement should play an important role in choosing the variables of the system and the bridge with economic theory should be made more solid and overreaching. Benati and Surico, as other have done in the literature, have shown that when VARs are mechanically run, they may lead to incorrect and unreliable conclusions. I don't think anybody needs to be convinced about this but I question the fact that such a point is relevant for carefully conducted VAR analyses. In theory structural methods have an edge over VAR methods; in practice, they probably don't have any. Since it very common to see very "unstructural" structural estimation, one should realistically approach the questions of interest and avoid dogmatic positions. There is always the possibility of making mistakes, but one should try to minimize their occurrence.

References

Benati, L. and Surico, P. (2006) VAR analysis and the Great Moderation, manuscript, ECB and the Bank of England.

Benati, L. and Mumtaz, H. (forthcoming) "The great stability in the UK: Good policy or good luck?", *Journal of Money, Banking and Credit.*

Benati, L. and Surico, P. (2006) "VARS and the great moderation", Bank of England manuscript.

Canova, F. (2002) " Validating two DSGE monetary models with VARs", CEPR working paper.

Canova, F. (2004) "Monetary Policy and the evolution of the US economy: 1948-2002", available at www.econ.upf.edu/crei/people/Canova.

Canova, F. (2006) Methods for Applied Macroeconomic Research, Princeton University Press, Princeton, NJ.

Canova, F. and Gambetti, L. (2004) " Structural changes in the US economy: Bad luck or bad policy?", CEPR Working Paper No. 5457.

Canova, F. Gambetti, L: and Pappa, E. (2006) " The structural dynamics of output growth and inflation: some international evidence", forthcoming, *Economic Journal*.

Canova, F. and Pina, J. (2005) "What VAR tells about DSGE models? ", in Deibolt, C. and Krytsou. C. (eds.) *New Trends and Macroeconomics*, Springer Verlag.

Canova, F. and Sala, L (2006) Back to square one: identification issues in VAR models, ECB working paper.

Chari, VV, Kehoe, P. and E. McGrattan (2004) "Are structural VAR useful for developing Business Cycle Theories ", Fed of Minneapolis, working paper 631.

Christiano, L., Eichenbaum, M. and Vigfusson, R. (2005) VAR as a guide to estimating DSGE models, available at www.faculty.econ.northwestern.edu/faculty/christiano/reserach.html

Clarida, R., Gali, J. and Gertler, M. (2000) "Monetary policy rules and macroeconomic stability: Evidence and some theory", *Quarterly Journal of Economics*, CXV, 147-180.

Cogley, T. and Sargent,T., J. (2001) "Evolving post-world war II U.S. inflation dynamics", *NBER Macroeconomic Annual*, 16, 331-373

Cogley, T. and Sargent, T., J. (2005) "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. ", *Review of Economic Dynamics*, 8, 262-302.

Cogley, T. and Sbordone, A. (2005), " A search for a structural Phillips curve" , UC Davis manuscript.

Collard, F., Dellas, H., Neusser, K. (2006) "On the causes of the Great Inflation", University of Bern, manuscript.

Cooley, T. and LeRoy, S. (1985), "Atheoretical macroeconomics: A critique, *Journal of Monetary Economics*, 16, 283-308.

Cooley, T. and Dwyer, M. (1998), "Business Cycle Analysis without much Theory: A Look at Structural VARs, *Journal of Econometrics*, 83, 57-88.

Del Negro, M. and Schorfheide, F (2004), " Priors from General Equilibrium Models for VARs", *International economic Review*

Del Negro, M, Schorfheide, F., Smets, F. and Wouters, R (2005) "On the fit and forecasting performance of New-Keynesian Models", forthcoming, *Journal of Business and Economic Statistics*.

Faust, J. and Leeper, E. (1997) Do Long Run Restrictions Really Identify Anything?, *Journal of Business and Economic Statistics*, 15, 345-353.

Fernandez Villaverde, J. Rubio Ramirez , J. and Sargent, T. (2005) The ABC and D for understanding VAR, available at www.econ.upenn.edu/jesusfv/reserach.htm

´ Gambetti, L., Pappa, E. and Canova, F. (2005), "Structural changes in the US economy: What explains the changes?", CEPR working paper 5879

Giordani, P. (2004) " An Alternative Explanation of the Price Puzzle", *Journal of Monetary Economics*, 51, 1271-1296.

Lubik, T. and Schorfheide, F. (2003) " Computing sunspot equlibria in linear rational expectations models" *Journal of the Economic Dynamic and Control*, 28, 273-285.

Lubik, T. and Schorfheide, F. (2004) "Testing for Indeterminacy: An application to US monetary Policy", *American Economic Review*, 94, 190-217

McConnell, M. and Perez Quiroz, G. (2001) "Output fluctuations in the US: What has changed since the early 1980s? ", *American Economic Review*, 90, 1464-1476.

Mumtaz, H. and Surico, P. (2006) "Evolving International Inflation Dynamics: World and country specific factors", Bank of England, manuscript.

Primiceri, G. (2005), "Time varying structural VAR and monetary policy, *Review of Economic Studies*, 72, 821-852.

Sims, C., A. and Zha, T. (2006), "Were there regime switches in US monetary policy?", *American Economic Review,* 96, 54-81.

Stock, J. and Watson, M. (2003) "Understanding changes in Business Cycle dynamics", NBER Working Paper, No. 9859.