# Matching DSGE models,VARs, and state space models 

Fabio Canova<br>EUI and CEPR

September 2012

## Outline

- Alternative representations of the solution of a DSGE model.
- Fundamentalness and finite VAR representation of DSGE solutions.
- Linking DSGE and VARs with sign restrictions.
- Some empirical concerns.
- Dealing with non-fundamentalness


## References

Canova, F. and Paustian, M. (2011). Business cycle measurement with some theory, Journal of Monetary Economics, 48, 345-361.

Christiano, L., Eichenbaum, M. and Viggfusson, R. 2006, Assessing structural VARs, NBER Macroeconomic Annual, 1-125.

Chari, V.V., Kehoe, P. and McGrattan, E. (2008) Are structural VARs with long run restrictions useful in developing business cycle theory?, Journal of Monetary Economics, 55, 1337-1352.

Dupor, W. and Han, J., (2011) Handling non-invertibility: theory and applications, Ohio State manuscript.

Kascha, C. and K. Martens (2009) Business cycle analysis and VARMA models, Journal of Economic Dynamics and Control, 33, 267-282

Ravenna, F.(2007) Vector autoregressions and reduced Form Representations of DSGE models, Journal of Monetary Economics, 54, 2048-2064.

Ramey, V. (2011) , Identifying Goverment Spending shocks: It is all in the timing!, Quarterly Journal of Economics, .

Gambetti, L. and Forni, M. (2011) Fiscal foresignt and the effect of goverment spending, UAB mqanuscript..

Merten, K. and Ravn, M. (2010). Measuring the impact of fiscal policy in the face of anticipation, Economic Journal, 120, 395-413.

Fernandez Villaverde, J., Rubio Ramirez, J, Sargent , T.and Watson, M.(2007) The A,B,C and D for understanding VARs, American Economic Review, 97, 1021-1026.

Sims, E.(2011) News, Non-invertibilities and structural VARs, university of Notre Dame, manuscript

## 1 Solution of DSGE models

- Typical (Log-) linearized solution of a DSGE model is of the form:

$$
\begin{align*}
& y_{2 t}=\mathcal{A}_{22}(\theta) y_{2 t-1}+\mathcal{A}_{21}(\theta) y_{3 t}  \tag{1}\\
& y_{1 t}=\mathcal{A}_{11}(\theta) y_{2 t-1}+\mathcal{A}_{12}(\theta) y_{3 t} \tag{2}
\end{align*}
$$

$y_{2 t}=$ states (endogenous and exogenous), $y_{1 t}=$ controls, $y_{3 t}$ shocks.

- $\mathcal{A}_{i j}(\theta), i, j=1,2$ are time invariant (reduced form) matrices and depend on $\theta$, the structural parameters of technologies, policies, etc.
- There are cross equation restrictions since $\theta_{i}, i=1, \ldots, n$ appears in more than one entry of these matrices.
- (1)-(2) is a state space model, with (1) being the transition equation and (2) the measurement equation.
- If both $y_{2 t}$ and $y_{1 t}$ are observables (1)-(2) is also a restricted $\operatorname{VAR}(1)$.
- Restrictions are on the lag length and on the entries of the coefficient matrix.

Letting $W_{t}=\left[y_{2 t}, y_{1 t}\right]^{\prime}, u_{t}=y_{3 t}$

$$
\mathcal{A}_{0}=\left[\begin{array}{cc}
A_{21} & 0 \\
0 & A_{12}
\end{array}\right]^{-1}, \quad \mathcal{A}(\ell)=\left[\begin{array}{cc}
A_{21} & 0 \\
0 & A_{12}
\end{array}\right]^{-1}\left[\begin{array}{ll}
A_{22} & 0 \\
A_{11} & 0
\end{array}\right]
$$

Then the system is

$$
\begin{equation*}
W_{t}=\mathcal{A}(\ell) W_{t-1}+u_{t} \tag{3}
\end{equation*}
$$

- Singular system: $\operatorname{dim}\left(y_{3 t}\right)<\operatorname{dim}\left(y_{1 t}+y_{2 t}\right)$.


## 2 Alternative representations for $y_{1 t}$

1) If $\mathcal{A}_{12}$ is invertible

$$
\begin{gathered}
y_{3 t}=\mathcal{A}_{12}^{-1}\left(y_{1 t}-\mathcal{A}_{11} y_{2 t-1}\right) \\
y_{2 t}=\mathcal{A}_{22} y_{2 t-1}+\mathcal{A}_{21} \mathcal{A}_{12}^{-1}\left(y_{1 t}-\mathcal{A}_{11} y_{2 t-1}\right) \\
\left(1-\left(\mathcal{A}_{22}-\mathcal{A}_{21} \mathcal{A}_{12}^{-1} \mathcal{A}_{11}\right) \ell\right) y_{2 t}=\mathcal{A}_{21} \mathcal{A}_{12}^{-1} y_{1 t}
\end{gathered}
$$

If $\mathcal{A}_{22}-\mathcal{A}_{21} \mathcal{A}_{12}^{-1} \mathcal{A}_{11}$ has all eigenvalues less than 1

$$
y_{2 t}=\left(1-\left(\mathcal{A}_{22}-\mathcal{A}_{21} \mathcal{A}_{12}^{-1} \mathcal{A}_{11}\right) \ell\right)^{-1} \mathcal{A}_{21} \mathcal{A}_{12}^{-1} y_{1 t}
$$

and

$$
\begin{equation*}
y_{1 t}=\mathcal{A}_{11}\left(1-\left(\mathcal{A}_{22}-\mathcal{A}_{21} \mathcal{A}_{12}^{-1} \mathcal{A}_{11}\right) L\right)^{-1} \mathcal{A}_{21} \mathcal{A}_{12}^{-1} y_{1 t-1}+\mathcal{A}_{12} y_{3 t} \tag{4}
\end{equation*}
$$

- If only $y_{1 t}$ is observable, the solution solution is a $\operatorname{VAR}(\infty)$.
- Does it make sense to assume that $\mathcal{A}_{12}$ is invertible? We need to have as many controls as shocks. Typically, this condition is not satisfied. Add measurement error to make this condition work.
- Need invertibility conditions on the eigenvalues of $\mathcal{A}_{22}-\mathcal{A}_{21} \mathcal{A}_{12}^{-1} \mathcal{A}_{11}$.

2) If $\mathcal{A}_{11}$ is invertible

$$
\begin{gathered}
y_{2 t-1}=\mathcal{A}_{11}^{-1}\left(y_{1 t}-\mathcal{A}_{12} y_{3 t}\right) \\
\mathcal{A}_{11}^{-1}\left(y_{1 t+1}-\mathcal{A}_{12} y_{3 t+1}\right)=\mathcal{A}_{22} \mathcal{A}_{11}^{-1}\left(y_{1 t}-\mathcal{A}_{12} y_{3 t}\right)+\mathcal{A}_{21} y_{3 t} \\
\left.y_{1 t+1}=\mathcal{A}_{11} \mathcal{A}_{22} \mathcal{A}_{11}^{-1} y_{1 t}+\left(\mathcal{A}_{11} \mathcal{A}_{21}-\mathcal{A}_{11} \mathcal{A}_{22} \mathcal{A}_{11}^{-1} \mathcal{A}_{12}\right) y_{3 t}+\mathcal{A}_{12} y_{3 t+1}\right) \\
y_{1 t+1}=\mathcal{A}_{11} \mathcal{A}_{22} \mathcal{A}_{11}^{-1} y_{1 t}+\left(I+\left(\mathcal{A}_{11} \mathcal{A}_{21} \mathcal{A}_{12}^{-1}-\mathcal{A}_{11} \mathcal{A}_{22} \mathcal{A}_{11}^{-1}\right) \ell\right) y_{4 t+1}
\end{gathered}
$$

where $y_{4 t} \equiv \mathcal{A}_{12} y_{3 t}$.

- If only $y_{1 t}$ is observable solution is a $\operatorname{VARMA}(1,1)$
- Does it make sense to assume that $\mathcal{A}_{11}$ is invertible? Need as many controls as states.
- No need to impose restrictions on the eigenvalues of $\mathcal{A}_{22}-\mathcal{A}_{21} \mathcal{A}_{12}^{-1} \mathcal{A}_{11}$.

3) Final form computations

From 2) $y_{1 t}=\mu_{1} y_{1 t-1}+u_{t}+\nu_{1} u_{t-1}$. Then for any $\left(y_{1 t}^{1}, y_{1 t}^{2}\right)$

$$
\left[\begin{array}{cc}
1-\mu_{11} \ell & -\mu_{12} \ell \\
-\mu_{21} \ell & 1-\mu_{22} \ell
\end{array}\right]\left[\begin{array}{l}
y_{1 t}^{1} \\
y_{1 t}^{2}
\end{array}\right]=\left[\begin{array}{cc}
1-\nu_{11} \ell & \nu_{12} \ell \\
\nu_{21} \ell & 1+\nu_{22} \ell
\end{array}\right]\left[\begin{array}{l}
u_{t}^{1} \\
u_{t}^{2}
\end{array}\right]
$$

Note that $\operatorname{det}(\nu(\ell))=\left(1+\nu_{11} \ell\right)\left(1+\nu_{22} e l l\right)-\nu_{12} \nu_{21} \ell^{2}$. Then:

$$
\left[\begin{array}{cc}
1+\nu_{22} \ell & -\nu_{21} \ell \\
-\nu_{12} \ell & 1+\nu_{11} \ell
\end{array}\right]\left[\begin{array}{cc}
1-\mu_{11} \ell & -\mu_{12} \ell \\
-\mu_{21} \ell & 1-\mu_{22} \ell
\end{array}\right]\left[\begin{array}{l}
y_{2 t}^{1} \\
y_{2 t}^{2}
\end{array}\right]=\operatorname{det}(\nu(\ell))\left[\begin{array}{l}
u_{t}^{1} \\
u_{t}^{2}
\end{array}\right]
$$

Under this alternative representation, the solution for $y_{1 t}$ is a $\operatorname{VARMA}(2,2)$.

- If $\nu(\ell)$ is of reduced rank, write $\nu(\ell)=\left[\begin{array}{cc}1+\nu_{11} \ell & \alpha \nu_{11} \ell \\ \nu_{21} \ell & 1+\alpha \nu_{21} \ell\end{array}\right]$ some
$\alpha$. Then $\operatorname{det}(\nu(\ell))=1+\left(\nu_{11}+\nu_{21}\right) \ell$.
In this case the solution for $y_{1 t}$ is a $\operatorname{VARMA}(2,1)$.


### 2.1 An important digression

The above representations assume that the state $y_{2 t}$ is observable. What if it is not?

- Typically, one needs to compute Kalman filter forecast of it, i.e. $\hat{y}_{2 t}=$ $\left(A_{22}-K A_{11}\right) \hat{y}_{2 t-1}+K \hat{y}_{1 t}$ where $K$ is the Kalman gain which depends on the observables $y_{1 t}$.
- Adding and subtracting $A_{11} \hat{y}_{2 t}$ to (2) we have

$$
\begin{align*}
y_{1 t} & =A_{11} \hat{y}_{2 t-1}+u_{t}  \tag{5}\\
u_{t} & =A_{11}\left(y_{2 t-1}-\hat{y}_{2 t-1}\right)+A_{12} y_{3 t} \tag{6}
\end{align*}
$$

- If the states are not observables and the observable $y_{1 t}$ fail to perfectly reveal them, i.e. $y_{2 t-1} \neq \hat{y}_{2 t-1}$, the innovations of the model include not only the structural shocks but also the forecast errors make in predicting the non-observable states.

Why is this relevant? If, for example, news shocks are present, they will be state variables, but they are not observable. Thus an econometrician needs to forecast them using the observables.

- The condition that the eigenvalues of $\left(A_{22}-A_{21} A_{12}^{-1} A_{11}\right)$ are all less than one in modulus implies that $\Sigma_{u}=A_{12} \Sigma_{y_{3}} A_{12}^{\prime}$ and that the forecast errors disappear
- This condition is equivalent to saying that with the observables variables we can perfectly reconstruct the unobservable states.

Example 1 Suppose that the capital stock is a state and it is unobservable. Then the effect of capital as a state can be recovered from a VAR if he variables included allow to perfectly predict it. Thus, for example, VARs without investment will not satisfy the condition on the eigenvalues of $\left(A_{22}-A_{21} A_{12}^{-1} A_{11}\right)$

- Think about the potential structural model which has generated the data before choosing the variables of a VAR!

When the eigenvalue condition is not satisfied the covariance matrix of the VAR innovations is strictly larger than the covariance matrix of the structural innovations.

- $A_{11} \Sigma_{y_{2}} A_{11}^{\prime}$ could be small if $y_{1 t}$ reveals well the information about $y_{2 t}$. Thus it is possible to find models which are non-invertible and still VARs approximately well their dynamics.

Example 2 Sims (2011) consider a standard medium scale New Keynesian model of the type used in central banks. The model has a technology shock and a news shock which enter in the model via

$$
\begin{equation*}
\ln T F P_{t}=a+\ln T F P_{t-1}+e_{t}+v_{t-q} \tag{7}
\end{equation*}
$$

where $v_{t-q}$ is a news shocks about the level of the technology known $q$ periods in advance. This process has the following state space representation

$$
\begin{align*}
\ln T F P_{t} & =a+\ln T F P_{t-1}+e_{t}+z_{1, t-1} \\
z_{1, t} & =z_{2, t-1} \\
z_{2, t} & =z_{3, t-1} \\
& \vdots  \tag{8}\\
z_{q-1, t} & =u_{t-q}
\end{align*}
$$

Thus, we have trasformed (7) into a vector of process with a "fake"" state vector $z_{t}$.
In the model, agents keep track of the z's when computing their optimal decisions. Econometricians do not know or observe the z's. Thus the state vector cannot, in general, be observed based on a history of observables TFP.

Given a standard parameterization and $v=4$ periods of foresight, the maximum eigenvalues of the $\left(A_{22}-A_{21} A_{12}^{-1} A_{11}\right)$ is 1.32. Thus the invertibility conditions fails. Nevertheless
the distortions that an econometrician would generate estimating a VAR on the data simulated by the model would be small (see below). Here we assume that the identification of the shocks is corrrect.


The solid lines are the theoretical impulse responses to news and surprise technology shocks in the "Full Model" using the parameterization as described in the text. For the Monte Carlo exercises the VARs feature $\ln \hat{a}_{t}$ and $\ln \hat{y}_{t}$ and are estimated with $p=8$ lags. In the left panel, labeled "Finite Sample," the dashed lines are the mean responses averaged across 500 simulations of data sets with 200 observations each, while the shaded gray regions depict the middle 68 percent of the distribution of estimated responses across the 500 simulations. In the right panel, labeled "Large Sample," the dashed lines are the estimated impulse responses from the estimation on one sample with 100,000 observations.

## Conclusions

- If variables are omitted, the solution of the model has a VAR representation only in certain (restricted) conditions.
- Even if a VAR representation exists, it may not allow us to recover the true shocks, if observables do not the information about missing states.
- Invertibility is not a either or propositions. There are non-invertible systems which could be recovered well with VARs.
- Can we use a finite VAR to approximate a DSGE model? If the eigenvalue condition holds, generally yes.


## 3 An alternative setup

Here we use the specification:

$$
\begin{align*}
y_{t} & =P x_{t-1}+Q z_{t}  \tag{9}\\
x_{t} & =R x_{t-1}+S z_{t}  \tag{10}\\
Z(\ell) z_{t} & =\epsilon_{t} \tag{11}
\end{align*}
$$

where $y_{t}$ is $r \times 1, z_{t}$ is $m \times 1, x_{t}$ is $n \times 1$. The difference with the previous setup is that here endogenous and exogenous states are differentiated.

We want to solve out the $z_{t}$ in both cases below. In addition, in one case $x_{t}$ is observable, in the other it is not.

Case 1). Suppose both $x_{t}$ and $y_{t}$ are observable. Let $Y_{t}=\left[x_{t}, y_{t}\right], A=$

$$
\begin{align*}
& {\left[\begin{array}{ll}
R & 0 \\
P & 0
\end{array}\right], B=\left[\begin{array}{c}
S \\
Q
\end{array}\right] \text { and let } Z(\ell)=I-z_{1} \ell . \text { Then }} \\
& \qquad \begin{aligned}
Y_{t} & =A Y_{t-1}+B z_{t} \\
z_{t} & =z_{1} z_{t-1}+\epsilon_{t}
\end{aligned} \tag{12}
\end{align*}
$$

This is an ARMA $(1,1)$ representation (it is used for example in Kommunjer and Ng (2011)).

Letting $B^{G}$ be the generalized inverse of B , we have

$$
\begin{align*}
z_{t} & =B^{G}\left(Y_{t}-A Y_{t-1}\right) \\
& \equiv z_{1}\left(B^{G} Y_{t-1}-B^{G} A Y_{t-2}\right)+\epsilon_{t} \tag{14}
\end{align*}
$$

Thus

$$
\begin{align*}
Y_{t} & =A Y_{t-1}+B\left[z_{1} B^{G} Y_{t-1}-B z_{1} B^{G} A Y_{t-2}\right]+B \epsilon_{t} \\
& =\left(A+B z_{1} B^{G}\right) Y_{t-1}-B^{2} z_{1} B^{G} A Y_{t-2}+B \epsilon_{t} \tag{15}
\end{align*}
$$

Note: $Y_{t}$ is $n+r \times 1$ while $\epsilon_{t}$ is $m \times 1$.

- If $m=n+r, B^{G}=B^{-1}$ and $Y_{t}=\left(A+z_{1}\right) Y_{t-1}-B z_{1} A Y_{t-2}+B \epsilon_{t}$.
- If $B$ is invertible, the representation is a $\operatorname{VAR}(2)$.
- If $m<n+r, B^{G} \neq B^{-1}$ system driven by a small number of shocks relative to the observables.
- If $B$ admits a generalized inverse, the representation is still a $\operatorname{VAR}(2)$ ! This means that we can delete rows and using any combinations of element $\hat{Y}_{t}$ of $Y_{t}$ we can make a non-singular $\operatorname{VAR}(2)$.
- There will be problems using maximum likelihood to estimate the parameters since covariance matrix is singular. If we want to use it need to drop endogenous variables until $m=n+r$. Which ones to drop? How do you do this? (see Canova, Ferroni and Matthes, 2012).

Case 2). Suppose that only $y_{t}$ are observables and assume $m=r$.

## Results:

1) $\hat{Y}_{t}$ is a VARMA $(\mathrm{n}+\mathrm{pm}, \mathrm{n}+\mathrm{p}(\mathrm{m}-1))$, where $p$ is the number of lags in the representation of $z_{t}$.
2) A finite order VAR representation for $\hat{Y}_{t}$ exists iff either $\operatorname{det}[|I-R \ell|+$ $\left.\operatorname{Padj}(I-R \ell) S Q^{-1} \ell\right]$ is of degree zero in $\ell$ or $\operatorname{det}\left[\mid I-\left(R S Q^{-1} P\right) \ell\right]$ is of degree zero in $\ell$.
3) 2) holds also if $Y_{t}=\left[x_{t}, y_{t}\right]$ where $x_{t}$ is $n_{1}<n \times 1$ and $y_{t}$ is $r-n_{1} \times 1$.

## Implications

1) Eliminating endogenous states from the VAR leads to a much more complicated VARMA representation. The length of the AR and MA depend on the size of the $x_{t}$ and $y_{t}$ vectors, the assumed process for the exogenous states. In general, a VAR must have very long lags to approximate the solution.
2) Need a particular configuration of parameters to generate a finite order VAR representation. Need restrictions on $P, R, Q, S$ matrices.
3) A necessary and sufficient condition for the existence of a finite order VAR is that the determinant of $\left(I-\left(R S Q^{-1} P\right) L\right]$ is of degree zero in L (compare with the eigenvalue restrictions we had before).
4) We need a particular setup where the number of observable (controls) $y_{t}$ matches exactly the number of exogenous shocks.

- Conditions appear to be restrictive in general.


## 4 VAR misspecification and sign restrictions

- Consider the first setup.
- If use robust sign restriction are used to identify shocks, misrepresentation (with a finite order VAR) of the decision rules less important.

DSGE models:

$$
\begin{align*}
& y_{2 t}=\mathcal{A}_{22}(\theta) y_{2 t-1}+\mathcal{A}_{21}(\theta) y_{3 t}  \tag{16}\\
& y_{1 t}=\mathcal{A}_{11}(\theta) y_{2 t-1}+\mathcal{A}_{12}(\theta) y_{3 t} \tag{17}
\end{align*}
$$

This model fits into the general VAR setup

$$
\left[\begin{array}{cc}
I-F_{11} \ell & F_{12} \ell \\
F_{21} \ell & I-F_{22} \ell
\end{array}\right]\left[\begin{array}{l}
y_{1 t} \\
y_{2 t}
\end{array}\right]=\left[\begin{array}{l}
G_{1} \\
G_{2}
\end{array}\right] e_{t}
$$

Representation for $y_{1 t}$ (integrating out $y_{2 t}$ ):

$$
\begin{equation*}
\left(I-F_{11} \ell-F_{12} F_{21}\left(1-F_{22} \ell\right)^{-1} \ell^{2}\right) y_{1 t}=\left[G_{1}-\left(F_{12}\left(1-F_{22} \ell\right)^{-1} G_{2} \ell\right] e_{t}\right. \tag{18}
\end{equation*}
$$

- This is, in general, a $\operatorname{ARMA}(\infty, \infty)$.
- Impact effects of $e_{t}$ on $y_{1 t}$ has the correct sign and the right magnitude.

Still there could be problems:

- Shocks in the reduced system will be time aggregated.
- Representation is not one-sided (future values of the shocks enter the representation).


## 5 Practical implications for SVAR analyses.

Let $y_{t}=\sum_{j=0}^{\infty} a_{j} e_{t-j}$ be the reduced form MA representation of the observable data where $a_{0}=I, e_{t} \sim(0, \Sigma)$ are the innovations in the process, that is, $e_{t}=y_{t}-P\left(y_{t} \mid I_{t-1}\right)$ where $I_{t-1}$ is the information set available at time $t-1$.

Take the first representation of the (log-) linearized solution of a DSGE as your structural model. Solving the state equation we get

$$
\begin{equation*}
y_{2 t}=\mathcal{A}_{21}(\theta) \sum_{j=0}^{\infty} \mathcal{A}_{22}(\theta)^{j} y_{3 t-j} \tag{19}
\end{equation*}
$$

Plugging this equation in the equation for $y_{1 t}$ we have

$$
\begin{equation*}
y_{1 t}=\mathcal{A}_{11}(\theta) \mathcal{A}_{21}(\theta) \sum_{j=1}^{\infty} \mathcal{A}_{22}(\theta)^{j} y_{3 t-j}+\mathcal{A}_{12}(\theta) y_{3 t} \tag{20}
\end{equation*}
$$

Let $y_{t}=S\left[y_{2 t}, y_{1 t}\right]^{\prime}$ where $S$ is a selection matrix. Then

$$
\begin{align*}
e_{t}= & S\left[y_{2 t}, y_{1 t}\right]^{\prime}-E_{t-1} S\left[y_{2 t}, y_{1 t}\right]^{\prime}  \tag{21}\\
= & S\left[\mathcal{A}_{21}(\theta) \sum_{j=0}^{\infty} \mathcal{A}_{22}(\theta)^{j} y_{3 t-j},\right. \\
& \left.\mathcal{A}_{11}(\theta) \mathcal{A}_{21}(\theta) \sum_{j=1}^{\infty} \mathcal{A}_{22}(\theta)^{j} y_{3 t-j}+\mathcal{A}_{12}(\theta) y_{3 t}\right]^{\prime}  \tag{22}\\
- & E_{t-1} S\left[\mathcal{A}_{21}(\theta) \sum_{j=0}^{\infty} \mathcal{A}_{22}(\theta)^{j} y_{3 t-j},\right. \\
& \left.\mathcal{A}_{11}(\theta) \mathcal{A}_{21}(\theta) \sum_{j=1}^{\infty} \mathcal{A}_{22}(\theta)^{j} y_{3 t-j}+\mathcal{A}_{12}(\theta) y_{3 t}\right]^{\prime} \tag{23}
\end{align*}
$$

- If $y_{t}$ is a vector with the same dimensions as $y_{3 t}$ and $I_{t-1}$ includes $y_{2 t-1}$
then

$$
\begin{equation*}
e_{t}=A_{12} y_{3 t} \tag{24}
\end{equation*}
$$

and we are in the standard setup. If some or all $y_{2 t}$ are omitted $y_{2 t}-$ $E_{t-1} y_{2 t}$ will produce a forecast error. Then

$$
\begin{equation*}
e_{t}=A_{12} y_{3 t}+S\left[y_{2 t}-E_{t-1} y_{2 t}, \mathcal{A}_{11}(\theta)\left(y_{2 t}-E_{t-1} y_{2 t}\right)\right]^{\prime} \tag{25}
\end{equation*}
$$

so that the structural errors will be contaminated by the forecast errors. Thus

1) $e_{t}$ will not be a white noise. There could be autocorrelation and heteroschedasticity if the forecast errors are autocorrelated and heteroschedastic.
2) The size of the distortion due to omitted variables depends on the entries of $\mathcal{A}_{11}(\theta)$ and how distant is the info set $I_{t-1}$ from $y_{2 t}$; that is, whether or not the variables in $I_{t-1}$ help to forecast $y_{2 t}$ or not.
3) Structural innovations will be determined with error.

- The dynamic responses will not be necessarily wrong.
- Variance decomposition generally wrong.
- Counterfactuals generally wrong.


## 6 Potential problems: Permanent income

Here is a simple example of how inference could be wrong when some variables are non-obserables.

Assume the income process is $y_{t}=y_{t-1}+\sigma_{w} w_{t}$.
If utility depends on consumption and it is quadratic, the consumption Euler equation is

$$
\begin{equation*}
c_{t+1}=\beta c_{t}+\sigma_{w}\left(1-R^{-1}\right) w_{t+1} \tag{26}
\end{equation*}
$$

where $\beta$ is the discount factor and $R>1$ the constant interest rate. Suppose that $c_{t}$ is non-observable and instead we observe $s_{t}=y_{t}-c_{t}$. The optimality condition for saving is:

$$
\begin{equation*}
s_{t+1}=-c_{t}+\sigma_{w} R^{-1} w_{t} \tag{27}
\end{equation*}
$$

The vector $\left(c_{t}, s_{t}\right)$ forms a state space model. The process for $s_{t}$ is non-fundamental (non-invertible) since $\beta+R-1>1$ and this violate the eigenvalue condition. The fundamental representation for savings is

$$
\begin{equation*}
s_{t+1}=-\hat{c}_{t}+\sigma_{w}\left(\frac{\beta-1+R}{R}\right) e_{t+1} \tag{28}
\end{equation*}
$$

where $\hat{c}_{t+1}=\beta \hat{c}_{t}+\sigma_{w}\left(\frac{\beta-\beta^{2}+1}{R}-\beta\right) e_{t+1}$ and $e_{t}$ is the fundamental innovation (i.e. the forecast error made in forecasting $s_{t+1}$ using current and past information).

To see what the two representations imply compute the $\operatorname{ARMA}(1,1)$ representations for saving in the model and in the fundamental representation:

$$
\begin{aligned}
& s_{t+1}=\beta s_{t}+\sigma_{w} R^{-1} w_{t+1}-\sigma_{w}\left(1-R^{-1}+\beta R^{-1}\right) w_{t} \\
& s_{t+1}=\beta s_{t}+\sigma_{w}\left(\frac{\beta-1+R}{R}\right) e_{t+1}-\sigma_{w} R^{-1} e_{t}
\end{aligned}
$$

Conclusions:

1) In the model the immediate effect is $\sigma_{w} R^{-1}$; in the fundamental representation it is $\sigma_{w}\left(\frac{\beta-1+R}{R}\right)$ Since $\beta-1+R>1$, the impact effect is overestimated.
2) By the same toke, the lagged effect is underestimated.
3) In the model the lagged effect is larger $1-R^{-1}+\beta R^{-1}>R^{-1}$. In the fundamental model the opposite is true.
4) $e_{t} \neq w_{t}$. In particular $w_{t}$ are linear combinations of current and future $e_{t}^{\prime}$ 's. Thus, using current and past values of $s_{t}$ allows us to recover current values of $e_{t}$ but not current values of $w_{t}$.

## 7 Dealing with non-invertibility problems

Not clear how to solve non-fundamental problems in practice. many alternatives have been suggested:

- Expand the observable vector (hoping to capture the effect of future shocks) (See e.g. Ramey (2011)). How do we know that this is enough?
- Use a factor model setup (see Forni and Gambetti (2010)). It can deal with non-fundamentalness due to omitted variables but it has not much to say about model based non-fundamentalness.
- Estimate directly the structural model (Fernandez et al. (2007)). Full information likelihood methods will have problems: likelihood has multiple peaks with non-invertibility. Dispense completely with VAR setup.
- Ravn and Mertens (2010). Use theory (calibrated DSGE) to tell us about the root that create non-fundamentalness. Fix it based on theory and proceed in VAR analysis from there.
- Dupor and Han (2011). Four step approach:
i) Estimate an $\operatorname{ARMA}(1,1)$ model from the data.
ii) Compute all possible covariance equivalent representation of the estimated model.
iii) Identify shocks and compute impulse responses for all possible representations.
iv) Eliminate representations which fail to satisfy additional restrictions (e.g. variance decomposition restrictions (size of variance explained), sign or shape restrictions, etc.).

Problem is step iv). Restrictions have to be determined by investigator. Potentially debatable.

True question is whether fundamental representation and one or more nonfundamental representations differ and in what way; if differences are small issue minor.

Procedure somewhat cumbersome for medium-large scale VAR.

Requires somewhat stringent assumptions about

- the number of observables and the number of shocks (they must be the same). Need to add measurement errors most of the time.
- the number of states and the number of controls must be the same. Need to have non-minimal state space representation.

Idea: Start from

$$
\begin{align*}
x_{t+1} & =Q x_{t}+U e_{t+1} \\
y_{t+1} & =W x_{t}+Z e_{t+1} \tag{29}
\end{align*}
$$

where $e_{t}$ is $k \times 1$. Assume that only $y_{t}$ is observable.

Assume:

1) $\operatorname{dim}\left(y_{t}\right)=\operatorname{dim}\left(x_{t}\right)$ (alternatively, assume that $W$ has a left-inverse).
2) All eigenvalues of $Q$ and $W Q W^{T}$ are inside the unit circle.
3) $\operatorname{dim}\left(y_{t}\right)=\operatorname{dim}\left(e_{t}\right)$ (alternatively, assume that $Z$ is invertible).

Let the covariance structure of the observable $y_{t}$ be $C_{i}=E\left(y_{t} y_{t-i}\right)$, for
all $i . C_{i}$ can be computed from the state space system (29) as

$$
\begin{align*}
C_{0} & =W Q W^{T} C_{0}\left(W Q W^{T}\right)^{T}+Z Z^{T}+W U U^{T} W^{T} \\
& -W Q W^{T} C_{0}\left(W Q W^{T}\right)^{T}  \tag{30}\\
C_{1} & =W Q W^{T} C_{0}+W U Z^{T}-W Q W^{T} Z Z^{T}  \tag{31}\\
C_{i} & =\left(W Q W^{T}\right)^{i-1} C_{1}, \quad \forall i>1 \tag{32}
\end{align*}
$$

The MA representation for $y_{t}$ from the model is

$$
\begin{equation*}
y_{t+1}=Z e_{t+1}+W \sum_{i=0}^{\infty} Q^{i} U e_{t-1} \tag{33}
\end{equation*}
$$

- Given the assumptions made, there are $2^{k}$ infinite order covariance equivalent MA representations for $y_{t}$ with innovations $e_{t}^{i}$ satisfying $E_{t}\left(e_{t}^{i}\left(e_{t}^{i}\right)^{T}\right)=$ $I_{k}$.
- The representation $j$ is given by

$$
y_{t+1}=(I-A L)^{-1}\left(D_{j}+B_{j} L\right) e_{t+1}^{j}
$$

where $A, B_{j}, D_{j}$ are given by

$$
\begin{align*}
A & =C_{2} C_{1}^{-1}  \tag{34}\\
B_{j} & =\bar{C}_{1}\left(D_{j}^{T}\right)^{-1}  \tag{35}\\
\bar{C}_{1} & =C_{1}-A C_{0}  \tag{36}\\
\bar{C}_{0} & =C_{0}-A C_{0} A^{T}-A \bar{C}_{1}-\bar{C}_{1} A^{T}  \tag{37}\\
0 & =\left(D_{j} D_{j}^{T}\right)\left(\bar{C}_{1}^{T}\right)^{-1}\left(D_{j} D_{j}^{T}\right)-\bar{C}_{0}\left(\bar{C}_{1}^{T}\right)^{-1}\left(D D_{j}^{T}\right)+\bar{C}_{1} \tag{38}
\end{align*}
$$

where $C_{i}$ are the autocovariances of $y_{t}$.

Note 1) $D_{j}+D_{J}^{c} H_{j}$, where $D_{j}^{c}$ is the Choleski decomposition of $D_{j} D_{j}^{T}$ and $H_{j}$ an orthonormal matrix. With Choleski restrictions $H_{j}=I$, with long run and sign restrictions $H_{j} \neq H_{j^{\prime}}, j \neq j^{\prime}$.

Note 2) in the above set of equivalent MA, one is fundamental, the rest are not.

How do we choose among the $2^{k}$ representations we produce?
i) Drop all the representations with complex roots (not clear why).
ii) Drop all representations that do not satisfy certain additional restrictions.

Example 3 Permanent in come of section 6. Here there is only one obsrevables so there are only two MA representations, the fundamental one and the non-fundamental one. The impulse responses of the two representations are below. How do we choose between the two repesentations?

Figure 2: Covariance-equivalent impulse responses to a positive savings shock



Notes: From the permanent income model with $r=0.2$. Impulse responses to a one unit shock from step three and before application of step four.

Why do we save? For future consumption! Thus, at some point in the future savings should reverse sign (addition theoretical restriction). Of the two repesentations only the
non-fundamental one implies that savings will drop at some horisons in the future, i.e. eliminate the fundamental one because it is inconsistent with theory.

- Look like an ex-post criteria !! Maybe people have utility from wealth. Wrong to choose the non-fundamental representation.

Example 4 Anticipated tax shocks (Leeper, et al. 2009). This model does not have unobservable controls but has a tax shock which is anticipated $T_{t}=\phi a_{t}+\epsilon_{t-2}$, where $a_{t}$ is a technology shock. The solution to the model has the state space representation like (29) where $Q=$ $\left[\begin{array}{ccc}\alpha & -\frac{\tau(1-\theta)}{1-\tau} & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0\end{array}\right], U=\left[\begin{array}{ccc}1 & -\frac{\tau \theta(1-\theta)}{1-\tau} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0\end{array}\right], W=\left[\begin{array}{ccc}0 & 0 & 1 \\ \alpha & -\frac{\tau(1-\theta)}{1-\tau} & 0 \\ \alpha & \frac{\theta \tau}{1-\tau} & 0\end{array}\right]$, $Z=\left[\begin{array}{ccc}\psi & 0 & 0 \\ 1 & -\frac{\tau(\theta 1-\theta)}{1-\tau} & 0 \\ 0 & \frac{\tau \theta^{2}}{1-\tau} & 0\end{array}\right]$, and $x_{t}=\left[k_{t}, \epsilon_{t}, \epsilon_{t-1}\right]^{\prime}, e_{t}=\left(a_{t}, \epsilon_{t}, u_{t}\right)^{\prime}, y_{t}=$
( $\tau_{t}, k_{t}, c_{t}$ ) where $u_{u}$ is a an iid measurement error, $(\tau, \theta, \alpha, \psi)$ are parameters.

The $\operatorname{VARMA}(1,1)$ representation for $y_{t}$ is

$$
\begin{equation*}
y_{t+1}=W Q W^{T} y_{t}+Z e_{t+1}+\left(W U-W Q W^{T} Z\right) e_{t} \tag{39}
\end{equation*}
$$

With 3 observable variables there are $2^{3}=8$ covariance equivalent $M A$ representations. The pairs $B_{j}, D_{j}$ associated with these euivalent MA representations satisfy:
$D_{j} D_{j}^{T}+B_{j} B_{j}^{T}=$

$$
\left[\begin{array}{ccc}
\psi^{2} \sigma_{a}^{2}+\theta^{2}+\left(\frac{\sigma_{u}}{\kappa}\right)^{2} & \psi \sigma_{a}^{2}+\kappa \theta(1-\theta) & \phi \sigma_{a}^{2}-\kappa \theta^{2} \\
\psi^{2} \sigma_{a}^{2}+\kappa \theta(1-\theta) & \sigma_{a}^{2}+\kappa^{2}\left(1+\theta^{2}\right)(1-\theta)^{2} & \sigma_{a}^{2}-\kappa^{2} \theta(1-\theta)\left(1+\theta^{2}\right) \\
\psi \sigma_{a}^{2}-\kappa \theta^{2} & \sigma_{a}^{2}-\kappa^{2} \theta(1-\theta)\left(1+\theta^{2}\right) & \sigma_{a}^{2}+\kappa^{2} \theta^{2}\left(1+\theta^{2}\right)+\sigma_{u}^{2}
\end{array}\right]
$$

$B_{j} D_{j}^{T}=\left[\begin{array}{ccc}0 & \kappa \theta^{2}(1-\theta) & -\kappa \theta^{3}-\frac{\sigma_{u}^{2}}{\kappa} \\ 0 & \kappa^{2} \theta(1-\theta)^{2} & -\kappa^{2} \theta^{2}(1-\theta) \\ 0 & -\kappa^{2} \theta^{2}(1-\theta) & \kappa^{2} \theta^{3}\end{array}\right]$, where $\kappa=\frac{\tau}{1-\tau}$.
With a standard parameterization, there are only 4 real values structural responses since $D_{j} D_{j}^{T}$ has a pair of complex eigenvalues.

Identify tax shocks and techology shocks using lower triangular Choleski (as detailed by the entries of the matrix $Z$ ).

Figure 3: Response To Tax and Technology Shocks (after step three)


Notes: upper panel responses to a positive tax shock; lower panels responses to a positive technology shock. PS $i$ : the $i$ th solution based on the Potter equation

How do you choose between the four remaining repesentations?

Note that responses to tax shocks are all similar in sign (except for consumption), differences is only in magnitude. Fundamental representation
not that different from the three non-fundamental representations. Differences occur in response to technology shocks.

What kind of restrictions can we impose to eliminate representations?

1) Contribution of the identified measurement error shocks to variance decomposition of any variables should not exceed 30 percent (why 30 percent?)
2) Technology shocks should not be the main source of volatility of tax rates at long horizons. Exclude all representations generating decompositions where they account for more than 50 percent (Why shouldn't technology shocks explain more than 50 percent of the variance of tax rates?).

Table 1: Identification Based on Short-Term Variance Decomposition

|  | Model One | Model Two | Model Three | Model Four |
| :---: | :---: | :---: | :---: | :---: |
| The average contributions on different horizons of identified measurement errors on variables |  |  |  |  |
| tax rate | 0 | 34.82 | 0 | 14.78 |
| capital | 0 | 39.32 | 0 | 0.51 |
| consumption | 7.84 | 39.45 | 7.84 | 70.51 |
| The average contributions of technology on tax rate at different horizons |  |  |  |  |
|  | 1.42 | 35.05 | 1.42 | 53.24 |
| The contribution of technology shocks on capital and consumption when $h=1$ |  |  |  |  |
| capital | 0 | 37.55 | 79.11 | 71.01 |
| consumption | 0 | 48.01 | 83.23 | 0.09 |

From table decompositions 2 and 4 are eliminated. Note model 4 corresponds to the fundamental representation.

If we add:
3)The effect of techology shocks should be fast enough - so that technology shocks explain most of the variability of consumption and capital in short run (say larger than 30 percent). We are left with just one model.

- Interesting way to proceed, but difficult to justify restrictions in step 4 (choose representations that satisfy a-priori ideas about how the economy works).
- Important assumptions needed to derive all the equivalent MA representations.
- Cumbersome to apply in medium scale systems.

Aside: How do you estimate VARMA models?

$$
\begin{equation*}
A_{0} y_{t}=A(\ell) y_{t-1}+M_{0} u_{t}+M(\ell) u_{t-1} \tag{40}
\end{equation*}
$$

$A(\ell)$ is of order $p_{1}, M(\ell)$ is of order $p_{2}$. The standard way is to use ML (see e.g. Hamilton (1994)).

Problems:

- difficult to identify correctly the model.
- near cancellation of roots possible.
- likelihood has multiple peaks in correspondence of the multiplicity of covariance equivalent MA representations.

Alternatives: all based on Durbin (1960 and Hannan and Rissanen (1982)).

- Two step approach: step 1 running a AR and estimate the residuals; plug the estimated residuals in and estimate the resulting VARMA by OLS.

Model (stage 1):

$$
y_{t}=\sum_{i}^{q} \Pi_{i} y_{t-i}+u_{t}
$$

- q should be long enough to make $u_{t}$ white noise.
- Typically set $q=0.5 * T^{0.5}$ ( to avoid non-invertible representations)

Model (stage 2):

$$
y_{t}=\left(I-A_{0}\right)\left(y_{t}-\hat{u}_{t}\right)+A(\ell) y_{t-1}+M(\ell) \hat{u}_{t-1}+u_{t}
$$

where $M_{0}=A_{0}$, and $\hat{u}_{t}$ are obtained at the first stage.

- All the regressors are observables, can use OLS.

Dufour and Jouini (2005) if $q$ grows at the rate $T^{0.5}$ as $T \rightarrow \infty$ parameter estimation is consistent and convergence to asymptotic normality at a rate $T^{0.25}$ (slow convergence).

Other approach add a third step to improve the efficiency of the estimation (see Hannan and Kavalieris (1984) or iterate on the two steps (see Kapetanios (2003)) to improve the convergence rate.

