Lecture 2:

Bayesian Vector Autoregressions: Priors and Identification

James M. Nason

Centre for Applied Macroeconomic Analysis
and
Virginia Center for Economic Policy

February 11, 2024

Priors for Unrestricted Bayesian VARs

- The Minnesota Prior
- Conjugate and Non-Informative Priors

Structural BVARs, Priors, and Identification

- Short Run Restrictions
- Long Run Restrictions
- Sign Restrictions
- Proxy VARs

Critiques of SVARs

- Rubio-Ramírez, Waggoner, and Zha (REStud, 2010)
- Faust and Leeper (JBES, 1997)
- Baumeister and Hamilton (Econometrica, 2015)
- Fundamentalness and Misspecification of SVARs

INTRODUCTION

- BVARs were first developed to solve a forecasting problem.
- The problem is an unrestricted VAR aims to describe the auto- and cross-covariance functions, which define the dynamics of y_t.
- This goal is often achieved by "mining" the data.
  1. The process only stops when, for example, the marginal significance of adding more lags to the VAR is small.
  2. $\Longrightarrow$ There is no more of the dynamics of y_t to be explained.
  3. Over-fitting (*i.e.*, maximizing R^2) provides a good reduced form depiction of the data, but is neither necessary nor sufficient attributes of a good forecasting model.
- Methods that eliminate or "shrink" the number of VAR parameters to be estimated are often found to have improved forecasting performance.
- Those interested in using BVARs for forecasting should see Karlsson, S. (2013, "Forecasting with Bayesian vector autoregressions," in Elliott, G., and A. Timmermann (eds.) HANDBOOK OF ECONOMIC FORECASTING, VOL. 2 (PART B), pp. 791–897, New York, NY: Elsevier).

Introduction

- The Minnesota prior was developed in the late 1970s and early 1980s by researchers at the FRB-Minneapolis to improve forecasting with VARs.
- At that time, the unit root or $I(1)$ revolution was sweeping across macro.
  1. Many aggregate time series were perceived to be random walks.
  2. See Nelson and Plosser (1982, "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics* 10, 139–162).
- The Minnesota prior (MP) is the belief that macro and financial variables are dominated by own $I(1)$ dynamics.
  1. This restricts the diagonal elements of $\mathbf{B}_1$ to one,
  2. the off diagonal elements of $\mathbf{B}_1$, the elements of $\mathbf{c}$, and all of the elements of $\mathbf{B}_j$, $j = 2, \ldots, p$, to zero of an unrestricted VARs.
- These restrictions shrink the dimension of the parameter vector that needs to be estimated $\Longrightarrow$ a more efficient estimator.

Unit Root Beliefs

- The MP is grounded on the belief the elements of $\mathbf{B}_j$, $j = 1, \ldots, p$, (and $\mathbf{c}$) are distributed normal and $\mathbf{\Omega}^{-1}$ is fixed or known.
  1. Let $\overline{\mathbf{B}}_j$ (or $\overline{\mathbf{b}}$) denote the priors of $\mathbf{B}_j$ ($\mathbf{b}$).
  2. MP: $\mathbf{E}\{B_{ii,1}\} = \overline{B}_{ii,1} = 1$, $i = 1, \ldots, n$, $\mathbf{E}\{B_{i\ell,1}\} = \overline{B}_{i\ell,1} = 0$, $i \neq \ell$, $\mathbf{E}\{B_{j,1}\} = \overline{\mathbf{B}}_j = \mathbf{0}_{n\times n}$, $j = 2, \ldots, p$, $\mathbf{E}\{\mathbf{c}\} = \overline{\mathbf{c}} = \mathbf{0}_{n\times 1}$, and $\mathbf{\Omega}_{\mathbf{B}}$ is diagonal.
  3. $\Longrightarrow \mathcal{Y}_t$ consists of n independent (driftless) random walks.
  4. Under the MP, the BVAR is estimated on levels data because $\mathcal{Y}_t \sim I(1)$.

- MP: $\mathbf{\Omega}_{\mathbf{B}}$ depends on three hyperparameters (*i.e.*, parameters of a prior).
  1. $\sigma_{ii,1} = \lambda_0$, $\sigma_{i\ell,j} = \dfrac{\lambda_0\lambda_1}{j^{\lambda_3}}\dfrac{\sigma_\ell}{\sigma_i}$, $\lambda_1 \leq 1$, and $\sigma_{i,j} = \lambda_0\lambda_2$.
  2. λ_0 = tightness (*i.e.*, precision) of the variances of own first lags.
  3. λ_1 = tightness of the variances of other variables relative to λ_0.
  4. λ_2 = tightness of variances of “exogenous” variables (*i.e.*, intercepts).
  5. λ_3 = tightness of the variances of lags $j = 2, \ldots, p$.
  6. Priors on the standard deviations of the VAR coefficients.

- The MP priors are not the posterior of the BVAR $\Longrightarrow$ the posterior gives the BVAR more complicated dynamics than a multivariate random walk.

Intuition for the Minnesota Prior

- Assume a diagonal $\mathbf{\Omega_B} \Longrightarrow$ VAR coefficients are independent.
- The decay rate of own and other variable lags is set by the prior on λ_3 $\Longrightarrow$ as $j \uparrow$, rate of decay is faster.
  1. Low order lags matter more for explaining dynamics of $\mathcal{Y}_t$.
  2. $\Longrightarrow$ Coefficients on higher order lags are shrunk to zero.
- The prior on λ_0 affects the variances of all VAR coefficients $\Longrightarrow$ reflects information in the sample and researcher's priors.
- Whether intercepts matter (relative to own first lag) relies on the prior λ_2.
- The impact of lags of other variables is controlled by $\lambda_1 \leq 1 \Longrightarrow$ prior is other variables have little role, but $\lambda_1 = 1$ is equivalent to OLS.
- The MP treats $\mathbf{\Omega}$ as if its precision is perfect $\Longrightarrow$ there is no uncertainty about variation in regression errors.

Pros and Cons of the Minnesota Prior

- Suppose a VAR estimated on quarterly data has $n = 6$, $p = 8$, and intercepts $\Longrightarrow$ 49 coefficients to be estimated per regression $\Longrightarrow$ 294 coefficients to be estimated, that suggests the need for nearly 80 years of quarterly data for any confidence OLS estimates mimic asymptotic properties.

- The MP solves this problem by shrinking the dimension of the VAR coefficient vector $\Longrightarrow$ more efficient or precise coefficient estimates.
  1. Shrinkage restricts VAR coefficients using sample information
  2. $\Longrightarrow$ standard deviations of regression errors σ_i, $i = 1, \ldots, n$),
  3. and researcher's beliefs about multivariate unit roots in y_t.

- The issue is the researcher's beliefs about $\lambda_{MP} = [\lambda_0 \;\; \lambda_1 \;\; \lambda_2 \;\; \lambda_3]$.
  1. A "loose" prior $\Longrightarrow$ OLS or noninformative prior suggests a large number of VAR coefficients to estimate and these over-fit the data.
  2. A "tight" prior can produce a posterior that too closely resembles the MP, which ignores information in the data.

How to Set λ_{MP}

- There are several ways to set λ_{MP}.
- A benchmark prior sets $\lambda_{MP,T} = [0.2\ \ 0.5\ \ 1\ \ 10^5]$, which is tighter compared with $\lambda_{MP,L} = [0.2\ \ 0.5\ \ 2\ \ 10^5]$.
  1. $\lambda_2 = 2 \Longrightarrow$ a looser prior on the slope coefficients and a non-informative prior on the intercepts.
  2. $\lambda_3 = 10^5$ yields a (damped) harmonic decay rate (*i.e.*, decays smoothly to zero) on higher order lags in the BVAR.
  3. BVARs are estimated equation by equation under these MPs.
- Or estimate λ_{MP} as part of the BVAR on the data $\Longrightarrow$ as if λ_{MP} becomes part of the moment matrix of the data $\Longrightarrow$ employ a system MLE and a "training sample."
- This "empirical Bayes" approach to the MP is similar to giving $\mathbf{\Omega} \sim \mathcal{IW}$, which also involves a system estimator.

Extensions to the MP

- Suppose one or more (but $< n$) elements of $\mathcal{Y}_t$ are stationary $\Longrightarrow \mathbf{E}\{B_{kk,1}\} = \overline{B}_{kk,1} \in (-1,\, 1)$, for some $k \in [1, 2, \ldots, n]$.

- Another way to express the unit root assumption is with a prior that restricts the sums of own lags to one, $\sum_{j=1}^{p} B_{ii,j} = 1$, $i = 1, \ldots, n$.
  1. Similarly, the prior of zero on the off-diagonal elements can be conveyed as $\sum_{j=1}^{p} B_{i\ell,j} = 0$, $i \neq \ell$.
  2. $\Longrightarrow$ an implication of these two priors is the average of the recent observations of the ith variable is the forecast; see Doan, Litterman, and Sims (1984), "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews* 3, 1–144.

- Seasonality in the data can also be dealt with as a prior.
  1. The prior can be deterministic $\Longrightarrow$ seasonality occurs the same time each year, say, in the fourth quarter.
  2. Or there is randomness in the seasonal variation during the year.
  3. This randomness can be interpreted as additional observations on which to estimate the BVAR $\Longrightarrow$ fake or "dummy" observations that reflect the researcher's belief about seasonality.

Priors and Dummy Observations, I

- Dummy observations represent "uncertain prior knowledge about the model's parameters;" see Sims, C.A. ("Dummy observation priors revisited," manuscript, Department of Economics, Princeton University).

- The idea is to add observations to the data that contain information about the prior of interest $\Longrightarrow$ the dummy observations induce or impose the prior on the model.

- Prior information is placed in a square matrix $\mathbf{R}$ conformable with $\mathbb{B}$.
  1. The dummy observations are generated with $\mathbf{R}\mathbb{B} = \mathbb{B}_R + \boldsymbol{\upsilon}_{\mathbb{B}}$, where
  2. $\mathbb{B}_R$ contains the prior restrictions and
  3. the tightness restrictions on the prior is embedded in $\boldsymbol{\upsilon}_{\mathbb{B}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_R)$.

- The unconditional distribution of $\mathbb{B}$ becomes $\mathbb{B} \sim \mathcal{N}(\mathbf{R}^{-1}\mathbb{B}_R, \mathbf{R}^{-1}\boldsymbol{\Omega}_R\mathbf{R}^{-1\prime})$.
  1. The estimator $\widetilde{\mathbb{B}} = \left[\mathbf{R}^{-1}\boldsymbol{\Omega}_R\mathbf{R}^{-1\prime} + \boldsymbol{\Omega}^{-1} \otimes \mathbb{X}'\mathbb{X}\right]^{-1}\left[\mathbf{R}^{-1}\mathbb{B}_R + (\boldsymbol{\Omega}^{-1} \otimes \mathbb{X}')\mathbb{Y}\right]$.
  2. Any sort of restrictions can be imposed on $\mathbb{B}$ in this way $\Longrightarrow$ from zero restrictions to sum of coefficients, but there are only linear restrictions.
  3. Also, note that as $T \longrightarrow \infty$ the dummy observations receive less weight in the estimator of $\widetilde{\mathbb{B}} \longrightarrow \widehat{\mathbb{B}} \Longrightarrow$ OLS.

PRIORS AND DUMMY OBSERVATIONS, II

- The previous example gives $\mathbb{B}$ a normal prior under which $\mathbb{B}_R$ and $\mathbf{\Omega}_R$ are known $\Longrightarrow$ these assumptions matter for several reasons.
- Suppose $\mathbf{R} = \mathbf{I} \Longrightarrow \mathbb{B} = \mathbb{B}_R + \boldsymbol{\upsilon}_{\mathbb{B}} \Longrightarrow$ formally the prior is
  1. $g(\mathbb{B}) \propto |\mathbf{\Omega}_R|^{-0.5} \exp\{-0.5(\mathbb{B} - \mathbb{B}_R)'\mathbf{\Omega}_R^{-1}(\mathbb{B} - \mathbb{B}_R)\}$
     $= |\mathbf{\Omega}_R|^{-0.5} \exp\{-0.5\,\mathbf{\Omega}_R^{-0.5}(\mathbb{B} - \mathbb{B}_R)'\mathbf{\Omega}_R^{-0.5}(\mathbb{B} - \mathbb{B}_R)\}$.
  2. $\Longrightarrow$ The posterior is $g(\mathbb{B}|\mathbb{Y}) \propto g(\mathbb{B})\mathscr{L}\left(\mathbb{B}, \mathbf{\Omega}\middle|\mathbb{Y}\right) = g(\mathbb{B})$
     $\times \exp\{[(\mathbf{\Omega}_R^{-0.5} \otimes \mathbf{I})\mathbb{Y} - (\mathbf{\Omega}_R^{-0.5} \otimes \mathbb{X})\mathbb{B}]'[(\mathbf{\Omega}_R^{-0.5} \otimes \mathbf{I})\mathbb{Y} - (\mathbf{\Omega}_R^{-0.5} \otimes \mathbb{X})\mathbb{B}]\}$
     $= \exp\{-0.5(\boldsymbol{z} - \mathbb{Z}\mathbb{B})'(\boldsymbol{z} - \mathbb{Z}\mathbb{B})\} = \exp\{-0.5(\mathbb{B} - \tilde{\mathbb{B}})'\mathbb{Z}'\mathbb{Z}(\mathbb{B} - \tilde{\mathbb{B}})$
     $+ \exp\{-0.5(\boldsymbol{z} - \mathbb{Z}\tilde{\mathbb{B}})'(\boldsymbol{z} - \mathbb{Z}\tilde{\mathbb{B}})\}$, where $\boldsymbol{z} = [\mathbf{\Omega}_R^{-0.5}\mathbb{B}_R \;\; (\mathbf{\Omega}_R^{-0.5} \otimes \mathbf{I})\mathbb{Y}]'$,
     $\mathbb{Z} = [\mathbf{\Omega}_R^{-0.5} \;\; (\mathbf{\Omega}_R^{-0.5} \otimes \mathbb{X})]'$, and $\mathbb{B}_R = (\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\boldsymbol{z}$.
  3. The last term of the posterior, $\exp\{-0.5(\boldsymbol{z} - \mathbb{Z}\tilde{\mathbb{B}})'(\boldsymbol{z} - \mathbb{Z}\tilde{\mathbb{B}})\}$, is known because of the assumptions on $\mathbb{B}_R$ and $\mathbf{\Omega}_R$ $\Longrightarrow$ the posterior becomes
     $g(\mathbb{B}|\mathbb{Y}) \propto \exp\{-0.5(\mathbb{B}-\tilde{\mathbb{B}})'\mathbb{Z}'\mathbb{Z}(\mathbb{B}-\tilde{\mathbb{B}})\} = \exp\{-0.5(\mathbb{B}-\tilde{\mathbb{B}})'\tilde{\mathbf{\Omega}}_R^{-1}(\mathbb{B}-\tilde{\mathbb{B}})\}$,
     where $\tilde{\mathbf{\Omega}}_R = [\mathbf{\Omega}_R^{-1} + (\mathbf{\Omega}^{-1} \otimes \mathbb{X}'\mathbb{X})]^{-1} \Longrightarrow g(\mathbb{B}|\mathbb{Y}) \sim \mathcal{N}(\tilde{\mathbb{B}}, \tilde{\mathbf{\Omega}}_R)$.
  4. $\tilde{\mathbb{B}}$ is a function of $\hat{\mathbb{B}}$ when $\mathbf{\Omega} = \hat{\mathbf{\Omega}} \Longrightarrow$ OLS.

Priors and Dummy Observations, III

- Remember a VAR(p) can be written as a static regression $\mathbb{Y} = \mathbb{X}\mathbb{B} + \Xi$, where $\Xi \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}) \implies$ GLS estimator of $\mathbb{B}$ is $(\mathbb{X}'\mathbf{\Omega}^{-1}\mathbb{X})^{-1}\mathbb{X}'\mathbf{\Omega}^{-1}\mathbb{Y}$, which is also $\widetilde{\mathbb{B}}$.

- This is the Theil-Goldberger mixed estimator developed in Theil and Goldberger (1961, "On pure and mixed statistical estimation in economics," *International Economic Review* 2, 65–78).

- This suggests treating priors as fake or "dummy" observations.

- Consider the prior $\mathbb{B} = \mathbb{B}_R + \boldsymbol{\upsilon}_{\mathbb{B}}$, which appends the dummy observations
  1. $\mathbb{Y}_0 = \mathbf{\Omega}\mathbf{\Omega}_R^{-0.5}\mathbb{B}$, $\mathbb{X}_0 = \mathbf{\Omega}\mathbf{\Omega}_R^{-0.5}$, and $\Xi_0 = \mathbf{\Omega}\mathbf{\Omega}_R^{-0.5}\boldsymbol{\upsilon}_{\mathbb{B}}$
  2. to the static regression to install the prior in the data.

- An example is the MP $\implies \mathbb{B}_{i\ell,j,r} = 0$ except for $i = \ell$, $j = 1$, and $\mathbf{\Omega}_R = \mathbf{\Omega}_R(\lambda_{MP})$.

Introduction

- The MP is a forecasting tool rather than a prior on coefficients that help identify a structural VAR.

- SVARs are useful for studying business cycle and monetary theories using the responses of macro and financial variables to identified shocks (*i.e.*, IRFs and FEVDs) $\Longrightarrow$ draw from the SVAR's posterior.

- The task is to construct a posterior simulator that respects the restrictions identifying a SVAR.

- As we have already seen, Gaussian VARs have
  1. a natural conjugate prior and its cousin the non-informative prior,
  2. which yield posteriors for just-identified recursive identifications
  3. from which confidence bands of IRFs and FEVDS are straightforward to compute.

The Conjugate Prior of a Gaussian VAR

- A review of Gaussian VARs begins with the model explaining the dynamics of the n-dimensional time series y_t

$$y_t \;=\; \mathbf{c} \;+\; \sum_{j=1}^{p} \mathbf{B}_j y_{t-j} \;+\; \varepsilon_t,\;\; \varepsilon_t \sim \mathcal{N}\left(\mathbf{0}_{n\times 1},\, \mathbf{\Omega}_{n\times n}\right).$$

- Since the likelihood of a Gaussian VAR can be decomposed into the conditional distribution of the VAR coefficients and the moment matrix of the VAR residuals, we know

$$\mathcal{H}\left(\mathbf{b},\, \mathbf{\Omega}_{\boldsymbol{\xi}} \,\middle|\, \mathbf{y}\right) \propto \mathcal{N}\left(\mathbf{b} \,\middle|\, \hat{\mathbf{b}},\, \mathbf{\Omega}_{\boldsymbol{\xi}},\, \mathbb{X},\, \mathbf{y}\right) \times \mathcal{W}\left(\mathbf{\Omega}_{\boldsymbol{\xi}}^{-1} \,\middle|\, \mathbf{y},\, \mathbb{X},\, \hat{\mathbf{b}},\, T-np-1\right).$$

- The implications are
  1. the VAR intercept and slope coefficients are normal conditional on the OLS VAR estimates and the data and
  2. the precision of the covariance matrix of the VAR residuals is distributed Wishart conditional on its degrees of freedom.

The Gaussian Conjugate Prior and Posterior Estimator of a BVAR

- The decomposition of a Gaussian VAR's likelihood suggests
  1. the priors $\mathbb{B}\big|\boldsymbol{\Omega} \sim \mathcal{N}\left(\underline{\mathbb{B}}, \boldsymbol{\Omega} \bigotimes \underline{\mathbb{Z}}\right)$ and $\boldsymbol{\Omega}^{-1} \sim \mathcal{W}\left(\underline{\mathbb{S}}^{-1}, \underline{\boldsymbol{\nu}}\right)$, which
  2. gives the posterior $\mathbb{B}\big|\boldsymbol{\Omega}, \mathbb{Y} \sim \mathcal{N}\left(\overline{\mathbb{B}}, \boldsymbol{\Omega} \bigotimes \overline{\mathbb{Z}}\right)$ and $\boldsymbol{\Omega}^{-1}\big|\mathbb{Y} \sim \mathcal{W}\left(\overline{\mathbb{S}}^{-1}, \overline{\boldsymbol{\nu}}\right)$,
  3. where $\overline{\mathbb{Z}} = [\underline{\mathbb{Z}}^{-1} + \mathbb{X}'\mathbb{X}]^{-1}$, $\overline{\mathbb{B}} = \overline{\mathbb{Z}}[\underline{\mathbb{Z}}^{-1}\underline{\mathbb{B}} + \mathbb{X}'\mathbb{X}\widehat{\mathbb{B}}]$, $\overline{\boldsymbol{\nu}} = T + \underline{\boldsymbol{\nu}}$,
  4. and $\overline{\mathbb{S}} = \widehat{\boldsymbol{\Omega}} + \underline{\mathbb{S}} + \widehat{\mathbb{B}}'\mathbb{X}'\mathbb{X}\widehat{\mathbb{B}} + \underline{\mathbb{B}}'\underline{\mathbb{Z}}^{-1}\underline{\mathbb{B}} - \overline{\mathbb{B}}'[\underline{\mathbb{Z}}^{-1} + \mathbb{X}'\mathbb{X}]\overline{\mathbb{B}}$.

- The researcher sets the priors for the hyperparameters $\underline{\mathbb{B}}$, $\underline{\mathbb{S}}$, $\underline{\mathbb{Z}}$, and $\underline{\boldsymbol{\nu}}$.

- There are (almost) no constrains on a researcher when choosing these hyperparameters to any values (*i.e.*, your priors are your priors).

- The Gaussian conjugate prior is more restrictive compared with the MP.
  1. The lag length is the same for every regression and every regression includes the same regressors $\Longrightarrow$ only just-identified recursive orderings are consistent with the Gaussian conjugate prior.
  2. The prior standard errors of the VAR coefficients are proportional, $\boldsymbol{\Omega} \bigotimes \underline{\mathbb{Z}} \Longrightarrow$ the MP loosens this restriction making the choice an explicit part of the researcher's prior.

The Non-Informative Prior

- Suppose a researcher's priors are $\underline{\mathbb{S}} = \underline{\mathbb{Z}} = c\,\mathbf{I}_{n\times n}$, and $\underline{\boldsymbol{\nu}} = c$, which can take any values.
- However, as $c \longrightarrow 0$, the posterior consists of $\overline{\mathbb{B}} = \widehat{\mathbb{B}}$ and $\overline{\mathbb{S}} = \widehat{\Omega}$.
- Since the hyperparameters play no role in the estimators of $\overline{\mathbb{B}}$ and $\overline{\mathbb{S}}$, the prior is "non-informative" $\Longrightarrow$ the non-informative prior yields a posterior that is the OLS estimator of a Gaussian VAR.
- The non-informative prior is easy to construct, but the efficiency gains tied to shrinkage are lost.

Several More Thoughts on Priors for BVARs

- Priors are beliefs about models that reflect a researcher's subjective judgment about the uncertainty surrounding the models.
  1. The uncertainty is subjective because the researcher selects the probability distribution that gives this judgment content.
  2. "Classical" Bayesians set priors before looking at the data.
  3. However, there are Bayesian methods that "peek" at the data first $\Longrightarrow$ empirical Bayesian priors.
  4. The important point is that priors belong to the researcher $\Longrightarrow$ your priors are your priors.
  5. Message: A researcher should not accept priors for a model used by other analysts without asking do these reflect her beliefs.

- Some priors are incompatible with the way in which $\mathcal{Y}_t$ is constructed.
  1. Ex: The MP is built on the belief that $\mathcal{Y}_t$ is a multivariate random walk.
  2. $\Longrightarrow$ Estimate BVARs on macro and financial data in levels.
  3. This part of the MP is often employed to estimate structural BVARs.

- Do not confuse the restrictions that render a BVAR a structural macro model with the priors expressing a researcher's uncertainty about random variables $\Longrightarrow$ these variables are the coefficients of the structural BVAR.

Introduction

- As discussed previously, SVARs impose restrictions on the responses of $\mathcal{Y}_t$ to identified shocks.
- Shocks are identified in SVARs using short-run or long-run restrictions.
- A Cholesky decomposition of the covariance matrix of a BVAR's residuals
  1. is an example of a short-run identification scheme,
  2. which is just-identified and recursive.
- Short-run identifications are located in the impact or lag zero responses of $\mathcal{Y}_t$ to η_t $(= \mathbf{D}^{-1}\varepsilon_t)$.
  1. Several short-run schemes have been developed, which require different estimators and interpretations.
  2. These are often zero restrictions that denote predetermined variables (*i.e.* weakly exogenous), variables excluded from "structural" relationships (*i.e.* a production function or Phillips curve), or variables responding "slugglishly" w/r/t the identified shocks.
- Long-run restrictions are about the response of $\mathbf{E}_t\mathcal{Y}_{t+j}$ to η_t, as $j \longrightarrow \infty$ $\Longrightarrow$ zero restrictions indicating long-run neutrality $\Longrightarrow$ labor market variables w/r/t a TFP shock or real variables w/r/t a nominal shock.

A Typology of Short-Run VAR Restrictions, I

- Return to the unrestricted Gaussian VAR

$$y_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{B}_j y_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{\Omega}_{n\times n}).$$

- A SVAR is constructed using $\mathbf{DD}' = \mathbf{\Omega} \Longrightarrow$ the structural shocks $\eta_t = \mathbf{D}^{-1}\varepsilon_t$.
  1. Set $\mathbf{A}_0 \equiv \mathbf{D}^{-1}$, $\mathbf{a} = \mathbf{D}^{-1}\mathbf{c}$, and $\mathbf{A}_j = \mathbf{D}^{-1}\mathbf{B}_j \Longrightarrow$ the SVAR

$$\mathbf{A}_0 y_t = \mathbf{a} + \sum_{j=1}^{p} \mathbf{A}_j y_{t-j} + \eta_t, \quad \eta_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{I}_{n\times n}).$$

  2. This SVAR is motivated by a recursive ordering $\Longrightarrow$ $\mathbf{A}_0$ is lower triangular, which obeys the recursive just-identified scheme of $\mathbf{\Omega}^{0.5}$.
  3. Since $\mathbf{A}_0\mathbf{\Omega}\mathbf{A}_0' = \mathbf{I}_n$, $\mathbf{A}_0$ has $0.5(n^2 - n)$ free parameters $\Longrightarrow$ given $\mathbf{\Omega}$, $\mathbf{A}_0$ has $0.5(n^2 + n)$ nonlinear restrictions pinned down by the equality.
  4. This identification is only one of many possible set of restrictions that can be imposed on $\mathbf{A}_0$ $\Longrightarrow$ non-recursive schemes that are just- or over-identified (as long as $\mathbf{A}_0$ is non-singular).
  5. Amisano & Giannini (1997) label this SVAR the $\mathbf{K}$-model, where $\mathbf{K} = \mathbf{A}_0$.

A Typology of Short-Run VAR Restrictions, II

- The **K**-model has an explicit simultaneous equations interpretation $\Longrightarrow$ restrict the interactions of the elements of $\mathcal{Y}_t$ at impact (*i.e.*, lag zero).
- Instead, as shown previously, restrictions can be placed on the linear combinations of the elements ε_t that produce the elements of η_t.
- The unrestricted Gaussian VAR lends itself to identifying the Amisano and Giannini (1997) **C**-model by setting $\varepsilon_t = \mathbf{D}\eta_t$, where $\mathbf{C} = \mathbf{D} \Longrightarrow$ the SVAR is

$$\mathcal{Y}_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{B}_j \mathcal{Y}_{t-j} + \mathbf{D}\eta_t, \;\; \eta_t \sim \mathcal{N}\left(\mathbf{0}_{n\times 1}, \mathbf{I}_n\right).$$

- Since $\varepsilon_t = \mathbf{D}\eta_t \Longrightarrow \mathbf{\Omega} = \mathbf{D}\mathbf{D}'$, there are $0.5\left(n^2 - n\right)$ free parameters in **D**.
- The **C**-model employs these free parameters to create linear combinations of orthogonalized shocks, η_t, that produce just-identified recursive and non-recursive schemes and over-identified non-recursive schemes.

A Typology of Short-Run VAR Restrictions, III

- Amisano & Giannini (1997) combine the **K**-model and the **C**-model to produce the **AB**-model $\Longrightarrow$ the SVAR is

$$\mathbf{A}_0 y_t = \mathbf{a} + \sum_{j=1}^{p} \mathbf{A}_j y_{t-j} + \mathbf{Q}\eta_t, \quad \eta_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{I}_n),$$

  where $\mathbf{A} = \mathbf{A}_0$ and $\mathbf{B} = \mathbf{Q} \Longrightarrow \mathbf{A}_0\varepsilon_t = \mathbf{Q}\eta_t \Longrightarrow \mathbf{A}_0\mathbf{\Omega}\mathbf{A}_0' = \mathbf{Q}\mathbf{Q}'$.

- The **AB**-model generalizes or nests the **K**- and **C**-models.
  1. $\Longrightarrow \mathbf{A}_0$ contains simultaneous equation restrictions.
  2. Restrictions on linear combinations of orthogonalized shocks are located in **Q**.
  3. The nesting involves $\mathbf{A}_0 = \mathbf{I}_n \Longrightarrow$ the **C**-model and the **K**-model is obtained by making **Q** diagonal.

- Thus, $\mathbf{A}_0\mathbf{\Omega}\mathbf{A}_0' = \mathbf{Q}\mathbf{Q}' \Longrightarrow 2n^2 - 0.5n(n+1)$ free parameters in $\mathbf{A}_0$ and **Q**.

A Typology of Short-Run VAR Restrictions, IV

- There are mappings that move between the **K**-, **C**-, and **AB**-models.
- These mapping are nonlinear.
  1. The **K**-model and the **AB**-model are equivalent if **Q** is diagonal.
  2. When $\mathbf{A}_0 = \mathbf{I}_n$, the **C**-model and the **AB**-model are equivalent.
- Suggests examining the nonlinearities inherent in the likelihood of the **AB**-model

$$\mathcal{L}\Big(\mathcal{Y}_{1:T}\,\Big|\,\mathbf{A}_0,\,\mathbf{Q},\,\mathbb{B}\Big) = -0.5T\left[n\ln(2\pi) + \ln\left|\mathbf{A}_0^{-1}\mathcal{Q}\left(\mathbf{A}_0^{-1}\right)'\right|\right]$$

$$-\ 0.5\sum_{t=1}^{T}\left[\mathcal{Y}_t - \mathbb{B}\mathbf{X}_t\right]'\left[\mathbf{A}_0^{-1}\mathcal{Q}\left(\mathbf{A}_0^{-1}\right)'\right]^{-1}\left[\mathcal{Y}_t - \mathbb{B}\mathbf{X}_t\right],$$

where $\mathbb{A} = \left[\mathbf{a}\ \ \mathbf{A}_1\ \ldots\ \mathbf{A}_p\right]$, $\mathcal{Q} = \mathbf{Q}\mathbf{Q}'$, and $\mathbb{B} = \mathbf{A}_0^{-1}\left[\mathbf{a}\ \ \mathbf{A}_1\ \ldots\ \mathbf{A}_p\right]$.

A Typology of Short-Run VAR Restrictions, V

- Maximize the likelihood of the **AB**-model w/r/t to $\mathbb{B}$ $\Longrightarrow$ substitute for the OLS estimates of these reduced form intercepts and slope coefficients to find

$$\mathscr{L}\left(\mathcal{Y}_{1:T}\,\middle|\,\mathbf{A}_0, \mathbf{Q}, \hat{\mathbb{B}}\right) = -\frac{1}{2}T\left[n\ln(2\pi) + \ln\left|\mathbf{A}_0^{-1}\mathcal{Q}\left(\mathbf{A}_0^{-1}\right)'\right| + \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t'\left[\mathbf{A}_0^{-1}\mathcal{Q}\left(\mathbf{A}_0^{-1}\right)'\right]^{-1}\hat{\varepsilon}_t\right].$$

- Since $\sum_{t=1}^{T}\hat{\varepsilon}_t'\left[\mathbf{A}_0^{-1}\mathcal{Q}\left(\mathbf{A}_0^{-1}\right)'\right]^{-1}\hat{\varepsilon}_t = \sum_{t=1}^{T}\operatorname{trace}\left(\left[\mathbf{A}_0^{-1}\mathcal{Q}\left(\mathbf{A}_0^{-1}\right)'\right]^{-1}\hat{\varepsilon}_t\hat{\varepsilon}_t'\right)$
$= T\operatorname{trace}\left(\left[\mathbf{A}_0^{-1}\mathcal{Q}\left(\mathbf{A}_0^{-1}\right)'\right]^{-1}\hat{\mathbf{\Omega}}\right) = T\operatorname{trace}\left(\left[\mathbf{A}_0\mathcal{Q}^{-1}\left(\mathbf{A}_0\right)'\right]\hat{\mathbf{\Omega}}\right).$

- Also, note $\ln\left|\mathbf{A}_0^{-1}\mathcal{Q}\left(\mathbf{A}_0^{-1}\right)'\right| = \ln\left|\mathbf{A}_0^{-1}\right|\left|\mathcal{Q}\right|\left|\mathbf{A}_0^{-1}\right| = -\ln\left|\mathbf{A}_0\right|^2 + \ln\left|\mathcal{Q}\right|$, which gives

$$\mathscr{L}\left(\mathcal{Y}_{1:T}\,\middle|\,\mathbf{A}_0, \mathbf{Q}, \hat{\mathbb{B}}\right) = -\frac{T}{2}\left[n\ln(2\pi) - \ln\left|\mathbf{A}_0\right|^2 + \ln\left|\mathcal{Q}\right| + \operatorname{trace}\left(\left[\mathbf{A}_0\mathcal{Q}^{-1}\left(\mathbf{A}_0\right)'\right]\hat{\mathbf{\Omega}}\right)\right].$$

- Nonlinear restrictions are difficult to implement, but Amisano and Giannini develop a general class of affine restrictions that are easy to apply to the **K**-, **C**-, and **AB**-models.

A Typology of Short-Run VAR Restrictions, VI

- Amisano and Giannini propose linear restrictions on $\mathbf{D}$, $\mathbf{A}_0$, and $\mathbf{A}_0$ and $\mathbf{Q}$ within the $\mathbf{C}$-, $\mathbf{K}$-, and $\mathbf{AB}$-models, respectively.
- Linear restrictions are embedded in the matrices $\mathbf{R_D}$, $\mathbf{R_{A_0}}$, and $\mathbf{R_Q}$, where
  1. $\mathbf{R_D}\,\text{vec}(\mathbf{D}) = \mathbf{r_D}$ for the $\mathbf{C}$-model,
  2. $\mathbf{R_{A_0}}\text{vec}(\mathbf{A}_0) = \mathbf{r_{A_0}}$ for the $\mathbf{K}$-model, and
  3. $\mathbf{R_Q}\,\text{vec}(\mathbf{Q}) = \mathbf{r_Q}$ and $\mathbf{R_{A_0}}\text{vec}(\mathbf{A}_0) = \mathbf{r_{A_0}}$ for the $\mathbf{AB}$-model.
  4. Restrictions are in "implicit" form, where $\mathbf{R_D}$, $\mathbf{R_{A_0}}$, and $\mathbf{R_Q}$ have n^2 columns and full row rank.
- Another class of linear restrictions is
  1. $\text{vec}(\mathbf{D}) = \mathbf{S_D d} + \mathbf{s_D}$ for the $\mathbf{C}$-model, where $\mathbf{d}$ is a column vector of the structural parameters in $\mathbf{D}$,
  2. $\text{vec}(\mathbf{A}_0) = \mathbf{S_{A_0}}\mathbf{a}_0 + \mathbf{s_{A_0}}$ for the $\mathbf{K}$-model, where $\mathbf{a}_0$ is a column vector of the structural parameters in $\mathbf{A}_0$ and
  3. $\text{vec}(\mathbf{Q}) = \mathbf{S_Q q} + \mathbf{s_Q}$ and $\text{vec}(\mathbf{A}_0) = \mathbf{S_{A_0}}\mathbf{a}_0 + \mathbf{s_{A_0}}$ for the $\mathbf{AB}$-model, where $\mathbf{q}$ is a column vector of the structural parameters in $\mathbf{Q}$.
  4. Restrictions have "explicit" form, where $\mathbf{S_D}$, $\mathbf{S_{A_0}}$, and $\mathbf{S_Q}$ have n^2 rows, full column rank, and column dimension $\leq 0.5(n^2 - n) \Longrightarrow$ maximum number of free parameters in $\mathbf{D}$, $\mathbf{A}_0$, or $\mathbf{A}_0$ and $\mathbf{Q}$.
- Other restrictions are $\mathbf{R_D S_D} = \mathbf{0}$, $\mathbf{R_{A_0} S_{A_0}} = \mathbf{0}$, $\mathbf{R_Q S_Q} = \mathbf{0}$, $\mathbf{R_D s_D} = \mathbf{r_D}$, $\mathbf{R_{A_0} s_{A_0}} = \mathbf{r_{A_0}}$ and $\mathbf{R_Q s_Q} = \mathbf{r_Q}$.

A Typology of Short-Run VAR Restrictions, VII: An Example

- The example is the SVAR estimated by Galí (1992, "How well does the IS-LM model fit the post-war data," *Quarterly Journal of Economics* 107, 709–738).

- The IS-LM model consists of
  1. IS schedule $\Longrightarrow y_t = y_{t-1} - \sigma(r_t - \pi_t) + \upsilon_{s,t} + \upsilon_{d,t}$,
  2. LM schedule $\Longrightarrow \Delta m_t - \Delta\pi_t = \phi\Delta y_t - \lambda\Delta r_t + \upsilon_{md,t}$,
  3. Money Supply $\Longrightarrow \Delta m_t = \upsilon_{ms,t}$,
  4. Phillips curve $\Longrightarrow \pi_t = \pi_{t-1} + \beta\Delta y_t - \gamma\upsilon_{s,t}$,
  5. where y_t, r_t, π_t, m_t, $\upsilon_{s,t}$, $\upsilon_{d,t}$, $\upsilon_{md,t}$, $\upsilon_{ms,t}$, and Δ are output, the nominal rate, inflation, the money stock, a supply shock, and a demand shock, a money demand shock, a money supply shock, and the first difference operator.

- Let $\mathcal{Y}_t = [\Delta y_t \;\; \Delta r_t \;\; \Delta m_t \;\; \Delta\pi_t]'$, $\eta_t = [\upsilon_{d,t} \;\; \upsilon_{md,t} \;\; \upsilon_{ms,t} \;\; \upsilon_{s,t}]'$, and $\varepsilon_t = [\varepsilon_{y,t} \;\; \varepsilon_{r,t} \;\; \varepsilon_{m,t} \;\; \varepsilon_{\pi,t}]' \Longrightarrow$ the **AB**-model

$$\mathbf{A}_0\mathcal{Y}_t = \mathbf{a} + \sum_{j=1}^{p}\mathbf{A}_j\mathcal{Y}_{t-j} + \mathbf{Q}\eta_t, \;\; \mathbf{A}_0\varepsilon_t = \mathbf{Q}\eta_t.$$

A Typology of Short-Run VAR Restrictions, VII

- The IS-LM model restricts the **AB**-model by

$$\mathbf{A}_0 = \begin{bmatrix} \mathbf{A}_{0,11} & \mathbf{A}_{0,12} & 0 & \mathbf{A}_{0,14} \\ \mathbf{A}_{0,21} & \mathbf{A}_{0,22} & \mathbf{A}_{0,23} & \mathbf{A}_{0,24} \\ 0 & 0 & \mathbf{A}_{0,33} & 0 \\ \mathbf{A}_{0,41} & 0 & 0 & \mathbf{A}_{0,44} \end{bmatrix} = \begin{bmatrix} 1 & \sigma & 0 & -\sigma \\ \phi & -\lambda & -1 & 1 \\ 0 & 0 & 1 & 0 \\ -\beta & 0 & 0 & 1 \end{bmatrix}$$

$$\text{and} \qquad \mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{11} & 0 & 0 & \mathbf{Q}_{14} \\ 0 & \mathbf{Q}_{22} & 0 & 0 \\ 0 & 0 & \mathbf{Q}_{33} & 0 \\ 0 & 0 & 0 & \mathbf{Q}_{44} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\gamma \end{bmatrix}.$$

- The impact matrices $\mathbf{A}_0$ and $\mathbf{Q}$ impose 6 and 11 short-run restrictions on the **AB**-model (total = 17), but there are 22 $\left(= 2n^2 - 0.5n(n+1),\ n = 4\right)$ free parameters $\Longrightarrow$ the SVAR is over-identified.

A Typology of Short-Run VAR Restrictions, VII

- Next, apply the vec$(\cdot)$ operator to $\mathbf{A}_0$ and $\mathbf{Q}$, define

$$\mathbf{S}_{\mathbf{Q}} = \begin{bmatrix} \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ 1 & 0 \\ \mathbf{0}_{11\times1} & \mathbf{0}_{11\times1} \\ 0 & 1 \end{bmatrix}, \quad \mathbf{S}_{\mathbf{A}_0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ & & \mathbf{0}_{4\times5} & & \\ 0 & 0 & 0 & 0 & 1 \\ & & \mathbf{0}_{3\times5} & & \end{bmatrix},$$

$\mathbf{q} = [1 \;\; -(\gamma+1)]'$, $\mathbf{a}_0 = [\sigma \;\; \phi \;\; -(\lambda+1) \;\; 1 \;\; -\beta]'$, $\mathbf{s}_{\mathbf{Q}} = \text{vec}(\mathbf{I}_n)$, and $\mathbf{s}_{\mathbf{A}_0} = \text{vec}(\mathbf{I}_n)$.

- Use these matrices to form the linear restrictions $\text{vec}(\mathbf{Q}) = \mathbf{S}_{\mathbf{Q}}\mathbf{q} + \mathbf{s}_{\mathbf{Q}}$ and $\text{vec}(\mathbf{A}_0) = \mathbf{S}_{\mathbf{A}_0}\mathbf{a}_0 + \mathbf{s}_{\mathbf{A}_0}$ $\Longrightarrow$ restrictions are useful when constructing the likelihood of the SVAR.

The Sims and Zha (1998) SVAR Prior and Estimator: Introduction

- Sims and Zha (1998) were not the first to estimate SVARs with non-recursive identifications.
  1. See Gordon and Leeper (1994) and Cushman and Zha (1997).
  2. Gordon and Leeper (1994) incorporate equilibrium money market outcomes into a monetary policy VAR.
  3. Cushman and Zha (1997) add an equation determining the equilibrium exchange rate to a money supply-demand system to estimate the impact of monetary policy shocks on a small open economy.

- Gordon and Leeper (1994) and Cushman and Zha (1997) impose non-recursive identification schemes on their BVARs, but employ the normal-Wishart estimation-similation technology developed for unrestricted (or recursively identified) BVARs.
  1. Consider the **K**-model $\Longrightarrow$ draw $\mathbf{b}$ and $\mathbf{\Omega}^{-1}$, impose the non-recursive identification on $\mathbf{A}_0$, and draw these structural BVAR parameters.
  2. $\Longrightarrow$ At the kth draw $\mathbf{\Omega}_k^{-1} = \mathbf{A}_{0,k}^{-1}\mathbf{D}^{-1}\mathbf{A}_{0,k}^{-1\prime}$ yields $\mathbf{G}_{j,k} = \mathbf{A}_{0,k}^{-1}\mathbb{B}_{j,k}$.

- Sims and Zha (1998) innovate on this "naive" Bayesian approach to non-recursive identifications of structural BVARs with priors for and a Bayesian estimator of this class of SBVARs.

The Sims and Zha (1998) SVAR Prior and Estimator, I

- Consider the $\mathbf{K}$-model

$$\mathbf{A}_0 \mathcal{Y}_t = \mathbf{a} + \sum_{j=1}^{p} \mathbf{A}_j \mathcal{Y}_{t-j} + \eta_t, \quad \eta_t \sim \mathcal{N}\left(\mathbf{0}_{n\times 1}, \mathbf{I}_{n\times n}\right).$$

- This SVAR in simultaneous (static) equations system is $\mathbb{Y}\mathbf{A}_0 - \mathbb{X}\mathbb{A} = \boldsymbol{\Xi}$.
  1. Define $\mathbb{Z} = [\mathbb{Y} \ \ -\mathbb{X}]$ and $\boldsymbol{\mathcal{A}} = [\mathbf{A}_0 \ \ \mathbb{A}]'$, where $\mathbb{A}$ collects the intercept vector $\mathbf{a}$ and p slope coefficient matrices into a $n \times n(p+1)$ matrix.
  2. $\Longrightarrow$ The likelihood function of the SVAR is

$$\begin{aligned}
\mathscr{H}\left(\boldsymbol{\mathcal{A}} \middle| \mathbb{Y}\right) &\propto \left|\mathbf{A}_0\right|^T \exp\left\{-\frac{1}{2}\left(\mathbf{A}_0\mathcal{Y}_t - \mathbf{a} - \mathbf{A}(\mathbf{L})\mathcal{Y}_{t-1}\right)'\left(\mathbf{A}_0\mathcal{Y}_t - \mathbf{a} - \mathbf{A}(\mathbf{L})\mathcal{Y}_{t-1}\right)\right\} \\
&\propto \left|\mathbf{A}_0\right|^T \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\mathcal{A}}'\mathbb{Z}'\mathbb{Z}\boldsymbol{\mathcal{A}}\right)\right\} \\
&= \left|\mathbf{A}_0\right|^T \exp\left\{-\frac{1}{2}\boldsymbol{a}_{\mathcal{A}}'\left(\mathbf{I}_{nk} \bigotimes \mathbb{Z}'\mathbb{Z}\right)\boldsymbol{a}_{\mathcal{A}}\right\},
\end{aligned}$$

where $k = np$ and $\boldsymbol{a}_{\mathcal{A}} = \mathrm{vec}(\boldsymbol{\mathcal{A}})$ is a $n(k+n) \times 1$ column vector.

The Sims and Zha (1998) SVAR Prior and Estimator, II

- Consider the prior for $\mathcal{A}$, $g(\boldsymbol{a}_{\mathcal{A}}) = g(\boldsymbol{a}_0)g(\boldsymbol{a}_{\mathbb{A}} | \boldsymbol{a}_0)$, where
  1. $\boldsymbol{a}_0 = \text{vec}(\mathbf{A}_0)$, $\boldsymbol{a}_{\mathbb{A}} = \text{vec}(\mathbb{A})$, $g(\boldsymbol{a}_{\mathbb{A}} | \boldsymbol{a}_0) \sim \mathcal{N}\left(\boldsymbol{a}_{\mathbb{A}} - \overline{\boldsymbol{a}}_0, \overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0)\right)$,
  2. and the zero restrictions on $\boldsymbol{a}_0$ can impose singularities on the distribution (*i.e.*, there are points in the distribution with zero probability) of $g(\boldsymbol{a}_0)$, but
  3. this places no restrictions on the form $g(\boldsymbol{a}_0)$ can take.

- The result is the posterior distribution of $\boldsymbol{a}_{\mathcal{A}}$

$$g(\boldsymbol{a}_{\mathcal{A}} | \mathbb{Y}) \propto \left|\mathrm{A}_0\right|^T \exp\left\{-\frac{1}{2}\boldsymbol{a}'_{\mathcal{A}}\left(\mathbf{I}_{nk} \bigotimes \mathbb{Z}'\mathbb{Z}\right)\boldsymbol{a}_{\mathcal{A}}\right\}\left|\boldsymbol{\Omega}(\boldsymbol{a}_0)\right|^{-0.5}$$
$$\times \exp\left\{-\frac{1}{2}\left[\boldsymbol{a}_{\mathbb{A}} - \overline{\boldsymbol{a}}_0\right]\overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0)^{-1}\left[\boldsymbol{a}_{\mathbb{A}} - \overline{\boldsymbol{a}}_0\right]'\right\} g(\boldsymbol{a}_0),$$

where

$$\boldsymbol{a}'_{\mathcal{A}}\left(\mathbf{I}_{nk} \bigotimes \mathbb{Z}'\mathbb{Z}\right)\boldsymbol{a}_{\mathcal{A}} = \boldsymbol{a}'_0\left(\mathbf{I}_{nk} \bigotimes \mathbb{Y}'\mathbb{Y}\right)\boldsymbol{a}_0 + \boldsymbol{a}'_{\mathbb{A}}\left(\mathbf{I}_{nk} \bigotimes \mathbb{X}'\mathbb{X}\right)\boldsymbol{a}_{\mathbb{A}} - \boldsymbol{a}'_{\mathbb{A}}\left(\mathbf{I}_{nk} \bigotimes \mathbb{X}'\mathbb{Y}\right)\boldsymbol{a}_0.$$

The Sims and Zha (1998) SVAR Prior and Estimator, III

- The quadratic $\boldsymbol{a}'_{\mathcal{A}}\left(\mathbf{I}_{nk} \otimes \mathbb{Z}'\mathbb{Z}\right)\boldsymbol{a}_{\mathcal{A}}$ is conditional on $\boldsymbol{a}_0$ and $\exp\left\{-\frac{1}{2}\left[\boldsymbol{a}_{\mathbb{A}} - \overline{\boldsymbol{a}}_0\right]\overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0)^{-1}\left[\boldsymbol{a}_{\mathbb{A}} - \overline{\boldsymbol{a}}_0\right]'\right\}$ is quadratic in $\boldsymbol{a}_{\mathbb{A}}$.
  1. $\Longrightarrow$ the posterior of $\boldsymbol{a}_{\mathbb{A}}$ is $g(\boldsymbol{a}_{\mathbb{A}} \,|\, \boldsymbol{a}_0, \mathbb{Y}) \sim \mathcal{N}\left(\tilde{h}(\boldsymbol{a}_0), \tilde{\boldsymbol{\Omega}}(\boldsymbol{a}_0)\right)$, where
  2. the posterior variance and mean are $\tilde{\boldsymbol{\Omega}}(a_0) = \left[(\mathbf{I}_{nk} \otimes \mathbb{X}'\mathbb{X}) + \overline{\boldsymbol{\Omega}}(a_0)^{-1}\right]$ and $\tilde{h}(\boldsymbol{a}_0) = \tilde{\boldsymbol{\Omega}}(a_0)\left[(\mathbf{I}_{nk} \otimes \mathbb{X}'\mathbb{Y})\,\hat{h}(\boldsymbol{a}_0) + \overline{\boldsymbol{\Omega}}(a_0)^{-1}\overline{h}(\boldsymbol{a}_0)\right]$.
  3. Note the length of the column vector $\boldsymbol{a}_{\mathbb{A}} = nk = n^2p$, where $n = 6$ and $p = 6 \Longrightarrow \dim(\boldsymbol{a}_{\mathbb{A}}) = 216$.

- Similarly the posterior distribution of $\boldsymbol{a}_0$ is

$$
\begin{aligned}
g(\boldsymbol{a}_0 \,|\, \mathbb{Y}) \quad \propto \quad & \left|\mathrm{A}_0\right|^T \left|\left(\mathbf{I}_{nk} \bigotimes \mathbb{X}'\mathbb{X}\right)\overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0) \;+\; \mathbf{I}_{n^2}\right|^{-0.5} \\
& \times \exp\left\{-\frac{1}{2}\left[\boldsymbol{a}_0\left(\mathbf{I}_{nk} \bigotimes \mathbb{Y}'\mathbb{Y}\right)\boldsymbol{a}_0' \right.\right. \\
& \left.\left. + \, h(\boldsymbol{a}_0)'\overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0)^{-1}h(\boldsymbol{a}_0) - \tilde{h}(\boldsymbol{a}_0)'\tilde{\boldsymbol{\Omega}}(\boldsymbol{a}_0)^{-1}\tilde{h}(\boldsymbol{a}_0)\right]\right\}.
\end{aligned}
$$

The Sims and Zha (1998) SVAR Prior and Estimator, IV

- The dimension of $\boldsymbol{a}_{\mathbb{A}}$ is problematic for Monte Carlo sampling $\Longrightarrow$ draw n^2p coefficients from $g(\boldsymbol{a}_{\mathbb{A}} \,|\, \boldsymbol{a}_0)$ to update and construct $g(\boldsymbol{a}_{\mathbb{A}} \,|\, \boldsymbol{a}_0, \mathbb{Y})$.
- Sims and Zha show the problem can be decomposed into n LS regressions.
- The idea is
  1. the SBVAR needs to have a SUR format and $\overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0)$ is chosen to induce scalar multiples of $\mathbf{I}_n$ and/or $\mathbb{X}'\mathbb{X} \Longrightarrow \overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0) = c\mathbf{I} \otimes \overline{\boldsymbol{\Omega}}_{\mathbb{A}} \;\; \left(\overline{\boldsymbol{\Omega}}_{\mathbb{A}} \equiv \overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0)\right)$,
  2. where $c > 0$ and $\overline{\boldsymbol{\Omega}}_{\mathbb{A}}$ need not be symmetric in the sense $\overline{\boldsymbol{\Omega}}_{\mathbb{A},\ell} \neq \overline{\boldsymbol{\Omega}}_{\mathbb{A},j}$.
  3. Still, there is independence across the n regressions $\Longrightarrow \left(\mathbf{I}_{nk} \otimes \mathbb{X}'\mathbb{X}\right) + \overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0) \propto \left(\mathbf{I}_{nk} \otimes \mathbb{X}'\mathbb{X}\right) + \mathrm{diag}\left\{\overline{\boldsymbol{\Omega}}_{\mathbb{A},1}, \overline{\boldsymbol{\Omega}}_{\mathbb{A},2}, \ldots, \overline{\boldsymbol{\Omega}}_{\mathbb{A},n}\right\}$ $= \mathrm{diag}\left\{\mathbb{X}'\mathbb{X} + \overline{\boldsymbol{\Omega}}_{\mathbb{A},1}, \mathbb{X}'\mathbb{X} + \overline{\boldsymbol{\Omega}}_{\mathbb{A},2}, \ldots, \mathbb{X}'\mathbb{X} + \overline{\boldsymbol{\Omega}}_{\mathbb{A},n}\right\}$.
  4. $\Longrightarrow$ There is correlation in the slope coefficients across equations, but of a special form $\Longrightarrow$ slope coefficients have a block diagonal structure.
- Beliefs about this correlation need to be embedded in $g(\boldsymbol{a}_{\mathbb{A}} \,|\, \boldsymbol{a}_0)$ and $g(\boldsymbol{a}_0)$.

The Sims and Zha (1998) SVAR Prior and Estimator, V

- Sims and Zha start from the MP prior $\Longrightarrow$ the reduced form VAR(p) is

$$\mathcal{Y}_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{B}_j \mathcal{Y}_{t-j} + \varepsilon_t, \;\; \varepsilon_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{\Omega}).$$

  where $\mathbf{c} = \mathbf{A}_0^{-1}\mathbf{a}$, $\mathbf{B}_j = \mathbf{A}_0^{-1}\mathbf{A}_j$, $j = 1, \ldots, p$, and $\varepsilon_t = \mathbf{A}_0^{-1}\eta_t$.

- Under the MP, $\mathbf{E}\{\mathbf{b}\} = \text{vec}([\mathbf{I}_n \;\; \mathbf{0}_{n\times n} \; \ldots \; \mathbf{0}_{n\times n}])$ and $\overline{\mathbf{\Omega}}_{\mathbf{b}}$ is given by the MP.
  1. $\Longrightarrow \mathbf{E}\{\mathbb{A}|\mathbf{A}_0\} = [\mathbf{A}_0 \;\; \mathbf{0}_{n\times n} \; \ldots \; \mathbf{0}_{n\times n}]'$ and $\mathbf{\Omega}_{\mathbb{A}}|\mathbf{A}_0 = \text{diag}(\mathbb{A}_{i\ell,j})$, where $\text{diag}(\mathbb{A}_{i\ell,j})$ $= \dfrac{\lambda_0\lambda_1}{j^{\lambda_3}\sigma_\ell^2}$, $i = \ell$, $j = 1, \ldots, p$, and $= \lambda_0\lambda_2$ otherwise.
  2. $\Longrightarrow \lambda_0$, λ_1, λ_2, and λ_3 denote the tightness priors on the variances of $\mathbf{A}_0$, $\mathbb{A}$, $\mathbf{a}$, and the decay rate of lags j = 2, ..., p, respectively.
  3. Sims-Zha treat a SBVAR as a SEM $\Longrightarrow$ there is no a prior normalization along the diagonal of $\mathbf{A}_0 \Longrightarrow \lambda_0$ and λ_1 do not impose separate prior beliefs on own lags and lags of the other $n-1$ elements of $\mathcal{Y}_t$.
  4. Sims-Zha scale factors differ from the MP $\Longrightarrow \mathbf{E}\{\eta_t\eta_t'\} = \mathbf{I}_n$ not $\mathbf{\Omega}$.
  5. Given priors on elements of $\mathbf{A}_0$ are correlated $\Longrightarrow \mathbf{E}\{\mathbb{A}|\mathbf{A}_0\}$ gives $\mathbf{b} = \text{vec}(\mathbf{A}_0^{-1}\mathbb{A})$ correlated priors $\Longrightarrow$ a prior on $\mathbf{A}_{0,ij}$, $i \neq j$, affects the prior of $\mathbf{A}_{1,ij}$.

The Sims and Zha (1998) SVAR Prior and Estimator, VI

- There are several more priors invoked by Sims and Zha to estimate the **K**-model SBVAR that can be added as dummy observations.
  1. λ_4 = the random walk prior applied to the sum of own coefficients in a regression $\Longrightarrow \overline{\mathbb{Y}}\mathbf{A}_0 - \overline{\mathbb{X}}\mathbb{A} = \overline{\Xi}$ if $\overline{\mathbb{Y}}_i = \lambda_4 p^{-1}\sum_{\ell=1}^{p} y_{ii,\ell,t}$ and $\overline{\mathbb{X}}_i = \lambda_4 m^{-1}\sum_{\ell=1}^{m} y_{ii,\ell,t}$, $m < p$ the best forecast of $y_{i,t}$ is its own lags with other variables having no role $\Longrightarrow$ when $\lambda_4 \longrightarrow \infty$, y_t is a full rank multivariate unit root $\Longrightarrow$ can difference the data Δy_t, and
  2. λ_5 = the cointegration prior implying stationary relationships among the elements of $y_t \Longrightarrow$ duplicate the prior λ_4 using λ_5, but $m < p-1$, and $\overline{\mathbb{X}}_i = \lambda_5$ for $m = p$.
  3. λ_3 and λ_4 control the tightness of the prior smoothing distributed lags of a regression $\Longrightarrow$ prevents long own lag structures from acting as if a deterministic function dominates $\mathbf{E}_t y_{i,t+j}$,
  4. When $\lambda_5 \longrightarrow \infty$, y_t is a reduced rank multivariate random walk $\Longrightarrow [\mathbf{I}_n - \sum_{j=1}^{p}\mathbf{A}_j]\overline{\mathbb{Y}} = \mathbf{A}_0^{-1}\mathbf{c}$ and if $\mathbf{c} = \mathbf{0}_{n\times 1}$ implying less than n random walks driving $y_t \Longrightarrow$ common trends or cointegration.
  5. Collect the priors in $\Lambda = [\lambda_0\ \ \lambda_1\ \ \lambda_2\ \ \lambda_3\ \ \lambda_4\ \ \lambda_5]'$.

The Sims and Zha (1998) SVAR Prior and Estimator, VII

- In summary, generating a posterior distribution for $\mathbb{A}$, $g(\boldsymbol{a}_{\mathbb{A}} \,|\, \boldsymbol{a}_0, \mathbb{Y})$, is easy.
  1. $\Longrightarrow$ The prior $g(\boldsymbol{a}_{\mathbb{A}} \,|\, \boldsymbol{a}_0)$ is multivariate normal with variance $\overline{\boldsymbol{\Omega}}(\boldsymbol{a}_0)$,
  2. $\Longrightarrow$ $\boldsymbol{a}_{\mathbb{A}}$ is $g(\boldsymbol{a}_{\mathbb{A}} \,|\, \boldsymbol{a}_0, \mathbb{Y}) \sim \mathcal{N}\left(\tilde{h}(\boldsymbol{a}_0), \tilde{\boldsymbol{\Omega}}(\boldsymbol{a}_0)\right)$, where
  3. the n regressions of the SBVAR can be estimated using LS, given the SBVAR can be mapped into a SUR model and Λ.

- Constructing the posterior of $\mathbf{A}_0$, $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$, is more difficult.
  1. Given the zeros or hard restrictions imposed on $\mathbf{A}_0$ $\Longrightarrow$ identification.
  2. Need priors for the non-zero elements, $\mathbf{A}_{nz,ij,0}$.

- The problem is $g(\mathbf{A}_{nz,ij,0})$ depends on $\mathbf{A}_0$ being just- or over-identified,
  1. the size of the sample relative to $n = \dim\,(\mathcal{Y}_t)$, and that
  2. prior restrictions on $\boldsymbol{a}_0$ apply to the structural slope coefficients through $\mathbf{A}_0$, which imposes prior restrictions on the reduced form regressions.

The Sims and Zha (1998) SVAR Prior and Estimator, VIII

- Sims and Zha suggest the priors on these non-zero parameters, which are soft restrictions, should be non-informative $\Longrightarrow \mathbf{A}_{nz,ij,0} \propto 1$.
- A often used non-informative prior is $g\left(\mathbf{A}_{nz,ij,0}\right) \sim \mathcal{N}\left(0,\, \sigma^2\left(\mathbf{A}_{nz,ij,0}\right)\right)$ $\Longrightarrow$ with T large enough, the effect of n on $g\left(\mathbf{A}_{nz,ij,0}\right) \longrightarrow 0$.
- Can impose a prior assuming independence across
  1. the structural regressions $\Longrightarrow g\left(\mathbf{A}_{nz,ij,0}\right) \sim \mathcal{N}\left(0,\, \lambda_6^2/\sigma_i^2\right)$, where σ_i^2 is the prior variance of the error of the ith structural regression, which fixes the variance of the non-zero elements of the ith row of $\mathbf{A}_0$ and $\mathbf{E}\left\{\mathbf{A}_{nz,ij,0}\mathbf{A}_{nz,\ell s,0}\right\} = 0$ or
  2. the reduced form regressions $\Longrightarrow g\left(\mathbf{A}_{nz,ij,0}\right) \sim \mathcal{N}\left(0,\, \lambda_6^2/\mathbf{\Omega}_{ii}\right)$, where $\mathbf{\Omega}_{ii}$ is the prior variance of the error of the ith reduced form regression.

The Sims and Zha (1998) SVAR Prior and Estimator, IX

- Still, numerical methods are needed to compute $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$, given $g(\mathbf{A}_{nz,ij,0}) \sim \mathcal{N}\Big(0,\, \sigma^2(\lambda_6)\Big)$.

- Have a choice of importance sampling (IS), Gibbs sampling, or Metropolis-Hasting (MH) Markov chain Monte Carlo (MH-MCMC) simulation.

- Waggoner and Zha (JEDC, 2003) advise using Gibbs sampling because IS needs a large number of iterations to cover the entire posterior density.
  1. If $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$ is non-normal $\Longrightarrow$ IS simulator is inefficient and needs many, many steps to produce a reasonable approximation of this posterior.
  2. However, computing $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$ using a Gibbs sampler involves constructing an algorithm that iterates between $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$ and $\boldsymbol{a}_{\mathbb{A}}$ is $g(\boldsymbol{a}_{\mathbb{A}} \,|\, \boldsymbol{a}_0,\, \mathbb{Y})$.

- IS relies on a finite number $\mathcal{M}$ of *iid* random draws from an arbitrary density $\overline{f}(\boldsymbol{a}_0)$ to approximate $g(\boldsymbol{a}_0 \,|\, \mathbb{Y}) \Longrightarrow$ the goal is to compute $\mathbf{E}\Big\{\overline{f}(\boldsymbol{a}_0)\Big\} = \int \overline{f}(\boldsymbol{a}_0) g(\boldsymbol{a}_0 \,|\, \mathbb{Y}) d\boldsymbol{a}_0 \Big/ \int g(\boldsymbol{a}_0 \,|\, \mathbb{Y}) d\boldsymbol{a}_0$.

- The approximation depends on the weight or importance ratio (IR), $\omega(\boldsymbol{a}_0)$,
  1. smooths the approximation by giving less (greater) mass to posterior draws of $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$ that occur frequently (infrequently),
  2. where $\omega_j(\boldsymbol{a}_0) = g(\boldsymbol{a}_{0,j} \,|\, \mathbb{Y}) \Big/ \overline{f}(\boldsymbol{a}_{0,j})$, $j = 1, \ldots, \mathcal{M}$.

The Sims and Zha (1998) SVAR Prior and Estimator, X

- An IS algorithm to calculate $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$ consists of
  1. compute the mode, $\boldsymbol{a}_{0,MD}$, of $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$ and the Hessian at $\boldsymbol{a}_{0,MD}$ $\Longrightarrow$ maximize the likelihood of the non-recursive SBVAR using classical optimization methods to compute these moments of $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$ (may need to "calibrate" $\mathbb{A}$ to run the optimizer),
  2. the computed Hessian of $\boldsymbol{a}_{0,MD}$, $\boldsymbol{\Omega}_{\boldsymbol{a}_0,MD}$, is the covariance matrix of the prior $g(\boldsymbol{a}_0) = \mathcal{N}(\boldsymbol{a}_{0,MD}, \boldsymbol{\Omega}_{\boldsymbol{a}_0,MD})$ or $g(\boldsymbol{a}_0) = t\Big(\boldsymbol{a}_{0,MD}, n+1\Big)$, conditional on $\boldsymbol{\Omega}_{\boldsymbol{a}_0,MD}$, when drawing $\boldsymbol{a}_0$, in the IS, which
  3. sample $\overline{f}(\boldsymbol{a}_{0,j})$ and $g(\boldsymbol{a}_{0,j} \,|\, \mathbb{Y})$ $j = 1, \ldots, \mathcal{M}$ times to generate the IR $\omega_j(\boldsymbol{a}_0)$ $\Longrightarrow$ compute $g(\boldsymbol{a}_0 \,|\, \mathbb{Y})$ and its $\boldsymbol{\Omega}_{\boldsymbol{a}_0}$ as the weighted average $\widetilde{f}(\boldsymbol{a}_0) = \sum_{j=1}^{\mathcal{M}} \overline{f}(\boldsymbol{a}_{0,j})\omega_j(\boldsymbol{a}_0) \Big/ \sum_{j=1}^{\mathcal{M}} \omega_j(\boldsymbol{a}_0)$, and $\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{a}_0} = \mathcal{M}^{-1} \sum_{j=1}^{\mathcal{M}} \Big[\overline{f}(\boldsymbol{a}_{0,j}) - \widetilde{f}(\boldsymbol{a}_0)\Big]^2 \Big[\omega_j(\boldsymbol{a}_0)\Big]^2 \Big/ \Big[\sum_{j=1}^{\mathcal{M}} \omega_j(\boldsymbol{a}_0)\Big]^2$, where $\widetilde{f}(\boldsymbol{a}_0) \xrightarrow{Pr} \mathbf{E}\Big\{g(\boldsymbol{a}_0 \,|\, \mathbb{Y})\Big\}$ and $\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{a}_0} \xrightarrow{Pr} \boldsymbol{\Omega}_{\boldsymbol{a}_0}$.

The Sims and Zha (1998) SVAR Prior and Estimator, XI

- Sims, Waggoner, and Zha (JofE, 2008) develop a Metropolis-Hastings (MH) within a Gibbs sampler algorithm to estimate a MS-BVAR that can be applied to a fixed coefficient **K**-model BVAR.
  1. Repeat step 1 of the IS algorithm to compute $\boldsymbol{a}_{0,MD}$, $\boldsymbol{a}_{\mathbb{A},MD}$ and the Hessians.
  2. Priors for $\mathbf{A}_0$ and $\mathbb{A}$ are constructed after imposing identifying restrictions developed by Waggoner and Zha (JEDC, 2003) on the likelihood of the SBVAR.
  3. The restrictions are $\mathbb{A} = \mathbf{S} + \overline{S}\mathbf{A}_0$, $\mathbf{A}_{0,i} = \mathbf{U}_j\mathbb{a}_i$ and $\mathbb{A}_j = \mathbf{V}_j\mathbb{b}_i - \mathbf{W}_j\mathbf{U}_j\mathbb{a}_i$, $\mathbb{b}_i = \mathbf{\Psi}_i\mathbb{s}_i$, where $i = 1, \ldots, n$, $\mathbf{E}\left\{\mathbf{S}\right\} = \mathbf{0}$ implies a random walk prior because $\overline{S} = \mathbf{I}_n$, $\mathbf{U}_j$, $\mathbf{V}_j$, and $\mathbf{W}_j$, and $\mathbf{\Psi}_i$, which is an orthonormal matrix, $\mathbf{\Psi}_i\mathbf{\Psi}_i' = \mathbf{I}$.
  4. $\Longrightarrow$ The priors of $\mathbb{a}_i$ and $\mathbb{s}_i$ are normal centered on $\boldsymbol{a}_{0,MD}$ and $\boldsymbol{a}_{\mathbb{A},MD}$ with covariances $\left(\mathbf{U}_j\,\mathbf{\Omega}^{-1}_{\boldsymbol{a}_{0,j},MD}\mathbf{U}_j'\right)^{-1}$ and $\left(\mathbf{\Psi}_j\,\mathbf{\Omega}^{-1}_{\mathbb{s}_j}\mathbf{\Psi}_j'\right)^{-1}$.
  5. The Gibbs sampler draws from $g\left(\mathbb{s}_i \,|\, \mathbb{Y}, \mathbf{\Omega}, \mathbb{a}_1, \ldots, \mathbb{a}_n\right)$, $i = 1, \ldots, n \Longrightarrow$ prior assumes independence across the structural regressions, which include the intercepts of the SVAR.
  6. The MH step samples $g\left(\mathbb{a}_i \,|\, \mathbb{Y}, \mathbf{\Omega}, \mathbf{S}, \mathbb{a}_1, \ldots, \mathbb{a}_{j\neq i}, \ldots, \mathbb{a}_n\right)$ to generate the posterior of $\mathbb{a}_i$, $i = 1, \ldots, n$, and evaluates whether these or the previous draws of these parameters raise the likelihood of the SBVAR.

- Similarly, Canova and Pérez-Forero (QE, 2015) propose a Metropolis inside a Gibbs sampler to estimate a time-varying parameter SBVAR that can be applied to a fixed coefficient **K**-class SVAR.

SVAR Estimation: A Metropolis in Gibbs Algorithm, I

- The Canova and Pérez-Forero estimator exploits several facts about a **K**-class SVAR,

$$\mathbf{A}_0 y_t = \mathbf{a} + \sum_{j=1}^{p} \mathbf{A}_j y_{t-j} + \eta_t = \mathbf{A}_0^{-1}\mathbf{c} + \mathbf{A}_0^{-1} \sum_{j=1}^{p} \mathbf{B}_j y_{t-j} + \eta_t, \;\; \eta_t \sim \mathcal{N}\left(\mathbf{0}_{n\times 1}, \mathbf{I}_{n\times n}\right).$$

- Reparameterize the K-class SVAR as a system of “static” regressions.
  1. Remember $\boldsymbol{\mathcal{X}}_t' = \mathbf{I}_n \otimes \left[y_{t-1}' \; \dots \; y_{t-p}' \;\; 1\right]$ and $\mathbb{B} = \text{vec}\left(\left[\mathbf{B}_1 \; \dots \; \mathbf{B}_p \;\; \mathbf{c}\right]\right) \Longrightarrow$ write the SVAR in concentrated form, $\mathbf{A}_0\left(y_t - \boldsymbol{\mathcal{X}}_t'\overline{\mathbb{B}}\right) = \eta_t$, where $\overline{\mathbb{B}}$ is a draw from the posterior of the reduced from intercept and slope coefficients.
  2. Define $\hat{y}_t \equiv y_t - \boldsymbol{\mathcal{X}}_t'\overline{\mathbb{B}}$ and recall $\text{vec}\left(\mathbf{A}_0\right) = \mathbf{S}_{\mathbf{A}_0}\boldsymbol{a}_0 + \mathbf{s}_{\mathbf{A}_0}$, where $\mathbf{s}_{\mathbf{A}_0} = \text{vec}(\mathbf{I}_n)$ and $\mathbf{S}_{\mathbf{A}_0}$ is $n^2 \times \dim(\boldsymbol{a}_0) \Longrightarrow \left(\hat{y}_t' \otimes \mathbf{I}_{n^2}\right)\left[\mathbf{S}_{\mathbf{A}_0}\boldsymbol{a}_0 + \mathbf{s}_{\mathbf{A}_0}\right] = \eta_t \otimes \mathbf{I}_{n^2}$.
  3. Let $\overline{y}_t' \equiv \left(\hat{y}_t' \otimes \mathbf{I}_{n^2}\right)\mathbf{s}_{\mathbf{A}_0}$ and $\boldsymbol{z}_t = -\left(\hat{y}_t' \otimes \mathbf{I}_{n^2}\right)\mathbf{S}_{\mathbf{A}_0} \Longrightarrow \overline{y}_t' = \boldsymbol{z}_t\boldsymbol{a}_0 + \eta_t \otimes \mathbf{I}_{n^2}$, which is a system of static regressions $\Longrightarrow$ estimate $\mathbf{A}_0$ conditional on $\overline{\mathbb{B}}$.

SVAR Estimation: A Metropolis in Gibbs Algorithm, II

- The posterior $\mathcal{P}\left(\overline{\boldsymbol{a}}_0 \middle| \overline{\mathbb{Y}},\ \overline{\mathbb{B}},\ \overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_0}\right)$ samples $\overline{\boldsymbol{a}}_0$ conditional on $\overline{\mathbb{B}}$, where $\overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_0}$ is the posterior covariance matrix of $\boldsymbol{a}_0$.

- The Gibbs step samples $\overline{\mathbb{B}}$ conditional on a normal prior conditional on $\underline{\mathbb{B}}$, $\mathbb{Y}$, and $\overline{\boldsymbol{a}}_0$, where $\underline{\mathbb{B}}$ is the prior mean $\Longrightarrow$ the normal prior suggests the posterior $\mathcal{P}\left(\overline{\mathbb{B}} \middle| \mathbb{Y},\ \underline{\mathbb{B}},\ \overline{\boldsymbol{\Omega}}_{\varepsilon}\right) \sim \mathcal{N}\left(\overline{\mathbb{B}},\ \overline{\boldsymbol{\Omega}}_{\mathbb{B}}\right)$ that is not conjugate, where $\boldsymbol{\Omega}_{\varepsilon} = \mathbf{A}_0^{-1}\mathbf{A}_0^{-1\,\prime}$.

- The Metropolis step decides whether to keep the existing draw of $\boldsymbol{a}_0$ or accept a new draw from $\mathcal{P}\left(\overline{\boldsymbol{a}}_0 \middle| \overline{\mathbb{Y}},\ \overline{\mathbb{B}},\ \overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_0}\right)$ using a criterion comparing the likelihoods of the SVAR evaluated at the existing and new draws of the impact coefficients.

- $\Longrightarrow$ Update $\boldsymbol{\Omega}_{\varepsilon}$, the covariance matrix of reduced form errors, using $\overline{\mathbf{A}}_0^{-1}$, which is a nonlinear function of the impact coefficients, $\boldsymbol{a}_0$.

SVAR Estimation: A Metropolis in Gibbs Algorithm, III

- The source for the Gibbs sampler part of the algorithm that generates draws of $\widehat{\mathbb{B}}$ is Koop and Korobilis (2010, "Bayesian multivariate time series methods for empirical macroeconomics," Boston, MA: Now Publishers). The Metropolis part of the algorithm that draws $\boldsymbol{a}_0$ is adapted from Canova and Pérez-Forero (QE, 2015).

- The Gibbs sampler rests on the prior $\overline{\mathbb{B}} \sim \mathcal{N}\left(\widetilde{\mathbb{B}},\, \widetilde{\boldsymbol{\Omega}}_{\mathbb{B}}\right)$, $\overline{\boldsymbol{a}}_{0,j}$ $\left(\Longrightarrow \overline{\boldsymbol{\Omega}}_{\varepsilon,j},\, j = 1, \ldots, \mathcal{J}\right)$, and updating equations

$$\widetilde{\boldsymbol{\Omega}}_{\mathbb{B}} = \left[\underline{\boldsymbol{\Omega}}_{\mathbb{B}}^{-1} + \sum_{t=1}^{T} \mathcal{X}_t' \overline{\boldsymbol{\Omega}}_{\varepsilon,j-1}^{-1} \mathcal{X}_t\right]^{-1} \quad \text{and} \quad \widetilde{\mathbb{B}} = \widetilde{\boldsymbol{\Omega}}_{\mathbb{B}} \left[\underline{\boldsymbol{\Omega}}_{\mathbb{B}}^{-1} \underline{\mathbb{B}} + \sum_{t=1}^{T} \mathcal{X}_t' \overline{\boldsymbol{\Omega}}_{\varepsilon,j-1}^{-1} \mathcal{Y}_t\right]$$

to calculate $\widetilde{\mathbb{B}}_{j+1}$ and $\widetilde{\boldsymbol{\Omega}}_{\mathbb{B},j+1}$.

- The Gibbs draw is $\overline{\mathbb{B}}_j \sim \mathcal{N}\left(\widetilde{\mathbb{B}},\, \widetilde{\boldsymbol{\Omega}}_{\mathbb{B}}\right)$ at iteration j.
  1. The prior mean $\underline{\mathbb{B}}$ is set to the OLS estimates, $\widehat{\mathbb{B}}$, of the reduced form intercept and slope coefficients while
  2. the prior on the covariance matrix $\underline{\boldsymbol{\Omega}}_{\mathbb{B}} = r_{\mathbb{B}} \widehat{\boldsymbol{\Omega}}_{\mathbb{B}}$, where $r_{\mathbb{B}}$ is a strictly positive tuning parameter and $\widehat{\boldsymbol{\Omega}}_{\mathbb{B}}$ is the OLS estimate of the covariance matrix of $\mathbb{B}$.

- However, if the largest eigenvalue of $\overline{\mathbb{B}}_j \geq 1$, toss out the jth draw of the reduced form intercept and slope coefficients and set $\overline{\mathbb{B}}_j = \overline{\mathbb{B}}_{j-1} \Longrightarrow$ this part of the prior maintains the posterior is consistent with a stationary SVAR.

SVAR Estimation: A Metropolis in Gibbs Algorithm, IV

- The Metropolis step needs a prior on $\boldsymbol{a}_0$ to generate the proposal.
  1. Draw $\overline{\boldsymbol{a}}_{0,j}$ from the normal distribution $\mathcal{N}\left(\overline{\boldsymbol{a}}_{0,j-1},\ r_{\boldsymbol{a}_0}\,\overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_0,j-1}\right)$
  2. or if the distribution of $\boldsymbol{a}_0$ is believed to have fat tails, draw from a prior with a t–distribution, $\overline{\boldsymbol{a}}_{0,j} \sim t\left(\overline{\boldsymbol{a}}_{0,j-1},\ r_{\boldsymbol{a}_0}\,\overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_0,j-1},\ \tau\right)$,
  3. where $\overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_0,j} = \left[\mathcal{Z}_t'\left(\eta_{\boldsymbol{a}_0,j,t}'\eta_{\boldsymbol{a}_0,j,t}\right)^{-1}\mathcal{Z}_t\right]^{-1}$, $\eta_{\boldsymbol{a}_0,j,t} = \overline{\mathcal{Y}}_t' - \mathcal{Z}_t\boldsymbol{a}_{0,j}$, $r_{\boldsymbol{a}_0}$ is a strictly positive tuning parameter, and the prior on the degrees of freedom of the t-distribution is $\tau \geq 4$.

SVAR Estimation: A Metropolis in Gibbs Algorithm, V

- Next, the Metropolis step in the Gibbs sampler decides whether to update the posterior of $\boldsymbol{a}_0$ to the new proposal $\overline{\boldsymbol{a}}_{0,j}$ or keep the previous draw $\overline{\boldsymbol{a}}_{0,j-1}$.

- This decision relies on the Metropolis criterion, which compares the likelihood of the SVAR under the new proposal $\overline{\boldsymbol{a}}_{0,j}$ and the previous draw $\overline{\boldsymbol{a}}_{0,j-1}$.
  1. The decision rule compares the likelihood of the SVAR evaluated at $\overline{\boldsymbol{a}}_{0,j}$ with the likelihood produced by $\overline{\boldsymbol{a}}_{0,j-1}$, where the likelihood of the SVAR is

$$\mathcal{H}\Big(\mathbb{Y}\,\Big|\,\boldsymbol{a}_0,\ \mathbb{B}\Big) \;=\; (2\pi)^{-0.5nT}\,\Big|\mathbf{A}_0\Big|^{T}\exp\left\{-\frac{1}{2}\Big(\overline{y}_t' - \boldsymbol{z}_t\boldsymbol{a}_0\Big)'\Big(\overline{y}_t' - \boldsymbol{z}_t\boldsymbol{a}_0\Big)\right\}.$$

  2. The Metropolis step compares $\omega_{j,j-1} = \mathcal{H}\Big(\mathbb{Y}\,\Big|\,\overline{\boldsymbol{a}}_{0,j},\ \overline{\mathbb{B}}_j\Big)\Big/\mathcal{H}\Big(\mathbb{Y}\,\Big|\,\overline{\boldsymbol{a}}_{0,j-1},\ \overline{\mathbb{B}}_j\Big)$ with $u_j \sim \mathcal{U}(0,\,1) \Longrightarrow$ if $u_j < \omega_{j,j-1}$ update to $\overline{\boldsymbol{a}}_{0,j}$; otherwise keep $\overline{\boldsymbol{a}}_{0,j-1}$.
  3. The decision criterion updates to the new proposal for $\boldsymbol{a}_0$, $\overline{\boldsymbol{a}}_{0,j}$, if the "probability" it increases $\mathcal{H}\Big(\mathbb{Y}\,\Big|\,\boldsymbol{a}_0,\ \mathbb{B}\Big)$ is greater than 50 percent.

- Run the Metropolis within Gibbs sampler for $\mathcal{J}$ steps, $j = 1\,\ldots,\,\mathcal{J}$, to estimate the fixed coefficient non-recursive K-class SVAR.

SVAR Estimation: A Summary of the Metropolis in Gibbs Algorithm

- Canova and Pérez-Forero develop a Metropolis in Gibbs sampler to calculate the posterior of a K-class SVAR consisting of the following steps.
    1. Given the initial condition $\boldsymbol{a}_{0,0}$ ($\Longrightarrow \boldsymbol{\Omega}_{\varepsilon,0}$) and priors $\underline{\boldsymbol{\Omega}}_{\boldsymbol{a}_0}$, $\underline{\mathbb{B}}$, and $\underline{\boldsymbol{\Omega}}_{\mathbb{B}}$, draw $\overline{\mathbb{B}}_1 \sim \mathcal{N}\left(\widetilde{\mathbb{B}},\ \widetilde{\boldsymbol{\Omega}}_{\mathbb{B}}\right)$ using the updating equations to calculate $\widetilde{\mathbb{B}}$ and $\widetilde{\boldsymbol{\Omega}}_{\mathbb{B}}$.
    2. If the largest eigenvalue of $\overline{\mathbb{B}}_1 \geq 1$, toss out this draw and repeat step 1.
    3. Given $\boldsymbol{a}_{0,0}$ and $\overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_0,0}$, draw $\overline{\boldsymbol{a}}_{0,1}$ from the normal (or t-)distribution.
    4. Conditional on $\overline{\mathbb{B}}_1$, construct $\overline{\mathcal{Y}}_t' - \mathcal{Z}_t\boldsymbol{a}_0$ to calculate $\mathcal{H}\left(\mathbf{Y}\,\middle|\,\overline{\boldsymbol{a}}_{0,1},\ \overline{\mathbb{B}}_1\right)$ and $\mathcal{H}\left(\mathbf{Y}\,\middle|\,\overline{\boldsymbol{a}}_{0,0},\ \overline{\mathbb{B}}_1\right) \Longrightarrow$ draw u_1 from $\mathcal{U}(0,\,1)$, compute the log likelihood ratio $\omega_{1,0}$, and engage the Metropolis criterion to decide whether the proposed update $\overline{\boldsymbol{a}}_{0,1}$ improves on $\boldsymbol{a}_{0,0}$ or not.
    5. If $u_1 \geq \omega_{1,0}$, repeat steps 1 through 4; otherwise repeat steps 1 through 4 replacing $\boldsymbol{a}_{0,0}$ with $\overline{\boldsymbol{a}}_{0,1}$ to generate realizations of $\overline{\boldsymbol{a}}_{0,2}$ and $\overline{\mathbb{B}}_2$.
- Run the sampler for $\mathcal{J}$ steps, $j = 1\ \ldots,\ \mathcal{J}$, to calculate the posterior of the fixed coefficient non-recursive K-class SVAR.
- The initial conditions $\boldsymbol{a}_{0,0}$ and $\overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_{0,0}}$ can be set equal
    1. to appropriately sized identity matrices with diagonal elements multiplied by some scalar on the open interval between zero and one
    2. or apply classical optimization tools to estimate $\boldsymbol{a}_{0,0}$ by maximizing $\mathcal{H}\left(\mathbf{Y}\,\middle|\,\boldsymbol{a}_0,\ \widehat{\mathbb{B}}\right) \Longrightarrow \overline{\boldsymbol{\Omega}}_{\boldsymbol{a}_{0,0}}$ is the inverse of the Hessian of $\mathcal{H}\left(\mathbf{Y}\,\middle|\,\boldsymbol{a}_{0,0},\ \widehat{\mathbb{B}}\right)$.

Introduction

- Macro theory also provides long-run identifying restrictions.
  1. A one sector growth model in which the exogenous trend is a unit root TFP process has a balanced growth path $\Longrightarrow$ predicts the long run level of output is driven only by shocks to the trend, which are TFP shocks.
  2. A monetary growth model that obeys the classical dichotomy allows real variables respond to nominal shocks only in the short-run and medium-run and not in the long-run $\Longrightarrow$ long-run monetary neutrality.
  3. The Fisher equation suggests that in the long run r_t and π_t move proportionally in response to nominal shocks to keep the real interest rate stable $\Longrightarrow$ in the long-run the ex post real rate $r_t - \pi_t$ is independent of nominal shocks.

- These long-run identifying assumptions imply three zero restrictions on the long-run responses of the level of output to aggregate demand and money market shocks and inflation to money market shocks given there are these two shocks plus an aggregate supply shock in the economy.

- The supply, demand, and money market shocks are tied to y_t, π_t, and the ex post real rate $r_t - \pi_t$, respectively.

Long-Run SVAR Identification

- Unit roots in y_t and π_t imply three long-run restrictions.
  1. $\mathbf{E}_t y_{t+j}$ is driven only by the TFP (or supply) shock, $\upsilon_{s,t}$, as $j \longrightarrow \infty$,
  2. $\mathbf{E}_t \pi_{t+j}$ responds to $\upsilon_{s,t}$, and the demand shock, $\upsilon_{d,t}$, as $j \longrightarrow \infty$.
  3. Or $\partial \mathbf{E}_t y_{t+j} / \partial \upsilon_{d,t} = \partial \mathbf{E}_t y_{t+j} / \partial \upsilon_{m,t} = \partial \mathbf{E}_t \pi_{t+j} / \partial \upsilon_{m,t} = 0$, where the money market shock $\upsilon_{m,t} = \upsilon_{md,t} - \upsilon_{ms,t}$.
  4. $\Longrightarrow \mathbf{E}_t \{ r_{t+j} - \pi_{t+j} \}$ responds to $\upsilon_{s,t}$, $\upsilon_{d,t}$, and $\upsilon_{m,t}$, as $j \longrightarrow \infty$.
  5. This suggests recursive long-run restrictions to identify a SVAR.

- Start with a reduced-form VAR(p) for $\mathcal{Y}_t = [\Delta y_t \;\; \Delta \pi_t \;\; (r_t - \pi_t)]'$

$$\mathcal{Y}_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{B}_j \mathcal{Y}_{t-j} + \varepsilon_t, \;\; \varepsilon_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \boldsymbol{\Omega}_{n\times n}),$$

  where the reduced-form errors are $\varepsilon_t = [\varepsilon_{y,t} \;\; \varepsilon_{\pi,t} \;\; \varepsilon_{r-\pi,t}]'$.

- Its reduced form VMA(∞) is $\mathcal{Y}_t = \mu_{\mathcal{Y}} + \sum_{\ell=0}^{\infty} \mathbf{C}_\ell \varepsilon_{t-\ell}$, where $\mathbf{C}_0 \equiv \mathbf{I}_n$ and $\mathbf{C}(\mathbf{L}) = [\mathbf{I}_n - \mathbf{B}(\mathbf{L})]^{-1} \Longrightarrow$ Wold representation.

Mapping of Long-Run SVAR Identification

- A structural VMA(∞) also exists, $y_t = \mu_y + \sum_{\ell=0}^{\infty} \mathbf{\Gamma}_\ell \eta_{t-\ell}$, $\eta_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$.
- Recover the SVAR from the structural VMA(∞) using the map from the structural to reduced-form errors $\eta_t = \mathbf{\Gamma}_0^{-1}\varepsilon_t$, where $\eta_t = [\upsilon_{s,t} \;\; \upsilon_{d,t} \;\; \upsilon_{m,t}]'$.
- Mapping relies on $\mathbf{B}(\mathbf{L})$ and $\mathbf{\Omega}$ to recover the $n^2 = 9$ elements of $\mathbf{\Gamma}_0$.
  1. $\mathbf{\Omega} = \mathbf{\Gamma}_0\mathbf{\Gamma}_0'$, which gives $0.5(n^2+n)$ ($= 6$ when $n = 3$) nonlinear equations to solve for 6 of the 9 elements of $\mathbf{\Gamma}_0$.
  2. Three other elements of $\mathbf{\Gamma}_0$ are found from the three zero restrictions on the long-run responses of y_t to $\upsilon_{d,t}$ and $\upsilon_{m,t}$ and π_t to $\upsilon_{m,t}$.
  3. The zero restrictions are embedded in the $(1,2)$, $(1,3)$, and $(2,3)$ elements of $\sum_{\ell=0}^{\infty}\mathbf{\Gamma}_\ell \Longrightarrow$ accumulated IRFs at the infinite horizon.
- Maintained assumption of the SVAR identification is that y_t and π_t are $I(1)$ $\Longrightarrow$ some shocks to the levels of output and inflation have permanent effects (*i.e.*, y_t and π_t have unit roots).
- This long-run identification is the Blanchard and Quah (1989) version of the Beveridge-Nelson decomposition.

An Aside: The Beveridge-Nelson Decomposition

- Times series econometrics treats the long-run behavior of a $I(1)$ variable, which is restricted by its trend, as atheoretic.
- There are several notions of what a trend is in macro, but in times series econometrics the two polar cases are trend stationary (TS) and difference stationary (DS) models.
- These models impose no economically relevant restrictions on the trend (or permanent) and transitory components of a time series $\Longrightarrow$ limits their usefulness for macro and finance.
- For example, consider the hypotheses that the natural log of real GDP possesses a unit root $\Longrightarrow$ what is the stochastic trend of output?
- This assumpion begs an important question $\Longrightarrow$ since the trend of real GDP is unobservable, how can it be constructed?
- Beveridge and Nelson (1981, *Journal of Monetary Economics*) develop a decomposition to measure the trend (or permanent) component of a time series that is empirically relevant for macro and finance.

TS and DS Models

- Consider the TS model, $x_t = \gamma_0 + \gamma_1 t + \alpha(\mathbf{L})u_t$, where $\alpha_0 \equiv 1$.
- The TS trend and transitory components are $\gamma_1 t$ and $\alpha(\mathbf{L})u_t$.
- The IRF of x_{t+j} to u_t consist of the coefficients of the MA(∞) of the Wold representation, which are $\alpha_j,\ j = 1, 2, \ldots$.
- For the DS model, $(1 - \mathbf{L})x_t = \gamma_0 + \beta(\mathbf{L})e_t$, where $\beta_0 \equiv 1$.
- The IRF of $(1 - \mathbf{L})x_{t+j}$ with respect to a unit change in e_t are the coefficients of the MA(∞), $\beta_j,\ j = 1, 2, \ldots$.
- In the case of the DS model, the IRF of x_{t+j} is the sum of the MA coefficients, $\sum_{i=0}^{j} \beta_j,\ j = 1, 2, \ldots \Longrightarrow$ as $j \longrightarrow \infty,\ \sum_{i=0}^{\infty} \beta_j \equiv \beta(\mathbf{1})$.
- In the limit, this is the IRF of the level response of x_t in the DS model.
- This generates the DS trend and transitory components $\beta(\mathbf{1})$ and $\beta(\mathbf{L})e_t$.

A Map from a TS Model to a DS Model, I

- Suppose x_t is the natural log of real GDP and the choice is between the TS model and DS model.
- If truth is the TS model, the DS model remains valid.
- The DS model is legitimate because $(1 - \mathbf{L})\alpha(\mathbf{L}) = \beta(\mathbf{L})$.
- In this case, the DS model is covariance stationary, but a non-invertible unit MA root is induced in x_t.
- The non-invertible unit MA root shows the IRF of $(1 - \mathbf{L})x_t$ at the infinite horizon is $(1 - \mathbf{1})\alpha(\mathbf{1}) = \beta(\mathbf{1}) = 0$.
- This result is consistent with the TS model because its IRF equals zero at the infinite horizon $\Longrightarrow$ a change in e_t has no long-run effects on x_t under the TS model.

A Map from a TS Model to a DS Model, II

- The important point is the only way to distinguish between TS and DS models is the behavior of these models at the infinite horizon.
- Serial correlation in x_t does not matter (except its impact on the small sample properties of unit root tests).
- Nonetheless, the DS model allows for a rich set of dynamics in the IRF of x_t not possible in TS models.
- For example, given $x_t \sim I(1)$, the IRF of x_t can exhibit decay or mean reversion to the long run trend at long horizons (*i.e.*, ten years or more).
- Long-horizon mean or trend reversion is observed in stock prices and other asset prices and returns.
- Real GDP and other macro aggregates tend to show mean reversion at the business cycle horizons of one to, as long as, five years.

The Beveridge-Nelson (BN) Decomposition, I

- Let x_t be the natural log of real GDP and truth is the DS model.
- **Proposition**: If x_t is a stochastic DS process, x_t can always be decomposed into a covariance stationary process, ε_t, and a random walk trend, τ_t,

$$x_t = \tau_t + \varepsilon_t.$$

- x_t is an unobserved components (UC) model $\Longrightarrow$ τ_t and ε_t are latent or hidden state variables driving movements in x_t.
- Use the Beveridge-Nelson (BN) decomposition to show
The BN trend is the conditional expectation of the random walk component for any UC representation of an $I(1)$ process.
- See Watson (1986, *Journal of Monetary Economics*) and Morley, Nelson, and Zivot (2003, *Review of Economics and Statistic*).

The Beveridge-Nelson (BN) Decomposition, II

- The BN decomposition begins with the Wold representation of the DS model

$$(1 - \mathbf{L})x_t = \gamma + \phi(\mathbf{L})\eta_t, \quad \phi_0 \equiv 1.$$

- Adopt the hypothesis the Wold representation decomposes x_t into its
  1. trend, $\tau_{t+1} = \gamma + \tau_t + \phi(\mathbf{1})\eta_{t+1}$, $\eta_{t+1} \sim \mathcal{N}(0, \sigma_\eta^2)$, which is a random walk with drift, and
  2. a transitory component, $\varepsilon_t = \tilde{\phi}(\mathbf{L})\eta_t$, which is a MA($\infty$), where $\tilde{\phi}_\ell = -\sum_{i=\ell+1}^{\infty}\phi_i, \;\ell = 0, 1, 2, \ldots$.

- The latent covariance stationary ε_t induces serial correlation in x_t and a non-invertible MA unit root in the first difference of x_t.

- Example: Let $\phi_0 = 1$ and $\phi_j = 0, \; j = 1, 2, \ldots,$ which implies $x_t = \tau_t$, or

$$x_t = \gamma + x_{t-1} + \eta_t.$$

- $\Longrightarrow x_t$ is a random walk with drift that lacks serial correlation.

The BN Factorization, I

- Factor a DS model with the BN decomposition with the lag polynomial $\phi(\mathbf{L})$.
- Pass the first difference operator through the UC model of x_t

$$(1 - \mathbf{L})x_t = (1 - \mathbf{L})\tau_t + (1 - \mathbf{L})\varepsilon_t = \gamma + \phi(\mathbf{1})\eta_t + (1 - \mathbf{L})\widetilde{\phi}(\mathbf{L})\eta_t.$$

- Also, remember $\phi(\mathbf{1}) = \sum_{i=0}^{\infty} \phi_i = \phi_0 + \phi_1 + \phi_2 + \ldots.$

The BN Factorization, II

- Beginning at $\ell = 0$, accumulate the MA coefficients

$$\tilde{\phi}_0 = -\sum_{i=1}^{\infty} \phi_i, \;\; \tilde{\phi}_1 = -\sum_{i=2}^{\infty} \phi_i, \; \ldots, \;\; \tilde{\phi}_j = -\sum_{i=j+1}^{\infty} \phi_i, \; \ldots.$$

- All that remains is to unwind $\phi(\mathbf{1})$ and $(1 - \mathbf{L})\tilde{\phi}(\mathbf{L}) \Longrightarrow$

$$\phi(\mathbf{1}) + (1 - \mathbf{L})\tilde{\phi}(\mathbf{L}) = \sum_{i=0}^{\infty} \phi_i + \sum_{i=0}^{\infty} \tilde{\phi}_i \mathbf{L}^i - \sum_{i=0}^{\infty} \tilde{\phi}_i \mathbf{L}^{i+1}$$

$$= \sum_{i=0}^{\infty} \phi_i - \sum_{i=1}^{\infty} \phi_i - \sum_{i=2}^{\infty} \phi_i \mathbf{L} - \sum_{i=3}^{\infty} \phi_i \mathbf{L}^2 - \ldots + \sum_{i=1}^{\infty} \phi_i \mathbf{L} + \sum_{i=2}^{\infty} \phi_i \mathbf{L}^2 + \sum_{i=3}^{\infty} \phi_i \mathbf{L}^3$$

$$= \sum_{i=0}^{\infty} (\phi_i - \phi_{i-1}) - \sum_{i=2}^{\infty} (\phi_i - \phi_{i-1})\mathbf{L} - \sum_{i=2}^{\infty} (\phi_i - \phi_{i-1})\mathbf{L}^2$$

$$- \sum_{i=3}^{\infty} (\phi_i - \phi_{i-1})\mathbf{L}^3 - \ldots = \phi_0 + \phi_1 \mathbf{L} + \phi_2 \mathbf{L}^2 + \phi_3 \mathbf{L}^3 + \ldots.$$

The BN Factorization, III

- As the BN decomposition predicts, the infinite sum is $\sum_{i=0}^{\infty} \phi_i \mathbf{L}^i = \phi(\mathbf{L})$.
- The BN decomposition is only one way to break x_t into trend and transitory elements.
- Its usefulness stems from the properties of the trend, τ_t.
- Forecasts of τ_t and x_t are equivalent at the infinite horizon $\Longrightarrow$ at $j \longrightarrow \infty$, forecast of x_t equals x_t plus its expected changes from $j = t+1$ to $j = \infty$

$$\lim_{j \longrightarrow \infty} \mathbf{E}_t \left\{ x_{t+j} - j\gamma \right\} = x_t + \sum_{i=1}^{\infty} \left[\mathbf{E}_t (1 - \mathbf{L}) x_{t+i} - \gamma \right].$$

- This implies $\lim_{j \longrightarrow \infty} \mathbf{E}_t \left\{ x_{t+j} - j\gamma \right\} = \lim_{j \longrightarrow \infty} \mathbf{E}_t \left\{ \tau_{t+j} + \varepsilon_{t+j} - j\gamma \right\} = \tau_t$ $\Longrightarrow$ the BN trend because it is a random walk with drift and ε_t is transitory.
- $\Longrightarrow$ x_t below (above) τ_t predicts x_t is expected to rise (fall) in the future.
- The economic interpretation of τ_t is that it measures the trend in x_t $\Longrightarrow$ its long run forecast.

The BN Factorization, IV

- What is (are) the difference(s) between TS and DS models?
- Remember the IRF of x_{t+j} is $\lim_{j \longrightarrow \infty} \left[\mathbf{E}_t x_{t+j} - \mathbf{E}_{t-1} x_{t+j} \right] = \phi(\mathbf{1})\eta_t$, it follows

$$\tau_t - \tau_{t-1} = \lim_{j \longrightarrow \infty} \left[\mathbf{E}_t x_{t+j} - \mathbf{E}_{t-1} x_{t+j} - \gamma \right].$$

- This is the BN definition of the trend, $\tau_t \Longrightarrow$ the $\text{var}(\tau_t - \tau_{t-1}) = \phi(\mathbf{1})^2 \sigma_\eta^2$, which is the innovation variance in the trend of x_t.
- If x_t is actually TS, the innovation variance of τ_t is zero by definition.
- Nonetheless, adding a tiny $\phi(\mathbf{1})$ to a TS process produces mean reversion in x_t because of the presence of a small random walk component, τ_t, where variation in the trend can be made arbitrarily small.

An Alternative BN Decomposition, I

- The BN decomposition studied thus far does not yield estimable models.
- Innovations of the transitory component, ε_t, and trend component, τ_t, are perfectly correlated $\Longrightarrow$ share the innovation η_t.
- An alternative BN decomposition gives τ_t and ε_t different innovations.
  1. Trend remains a random walk with drift, $\tau_{t+1} = \gamma + \tau_t + \eta_{t+1}$, $\eta_t \sim \mathcal{N}(0, \sigma^2_\eta)$.
  2. The MA lives in the transitory component $\varepsilon_t = \psi(\mathbf{L})\upsilon_t$, $\upsilon_t \sim \mathcal{N}(0, \sigma^2_\upsilon)$.
  3. Key assumption is whether $\mathbf{E}\left\{\eta_{t+j}\upsilon_{t+\ell}\right\} = 0$, for $j, \ell = -\infty, \ldots, \text{-1}, 0, 1, \ldots, \infty$, or only for $j, \ell = -\infty, \ldots, \text{-1}, 1, \ldots, \infty \Longrightarrow \mathbf{E}\left\{\eta_t \upsilon_t\right\} \neq 0$
- The orthogonality restriction is key to understanding predictions of the BN decomposition; see Morley, Nelson, and Zivot (2003, *Review of Economics and Statistic*).

An Alternative BN Decomposition, II

- Since $(1 - \mathbf{L})x_t = (1 - \mathbf{L})\tau_t + (1 - \mathbf{L})\varepsilon_t, \Longrightarrow$

$$(1 - \mathbf{L})x_t \;=\; \gamma + \eta_t + (1 - \mathbf{L})\psi(\mathbf{L})\upsilon_t.$$

- However, the Wold representation of the DS model restricts the right hand side of the first difference of x_t to equal $\eta_t + (1 - \mathbf{L})\psi(\mathbf{L})\upsilon_t = \phi(\mathbf{L})\upsilon_t$.
- Hence, the long run forecast of x_t equals $\eta_t + (1 - 1)\psi(\mathbf{1})\upsilon_t = \phi(\mathbf{1})\upsilon_t$ or $\eta_t = \phi(\mathbf{1})\upsilon_t$.
- This implies the variance of the random walk component is $\sigma_\eta^2 = \phi(\mathbf{1})^2\sigma_\upsilon^2$.

An Alternative BN Decomposition, II

- Although there are differences in the two BN decompositions, the variance of $(1 - \mathbf{L})\tau_t$ is the same in both.
- This is true for any decomposition of x_t into trend and transitory elements.
- The important point is the random walk component of x_t is unrestricted with respect to its innovation variance.
- The upshot is, add an arbitrarily small innovation variance, $\phi(\mathbf{1})$, to a TS model, and it becomes a DS model $\Longrightarrow x_t$ has long-horizon mean reversion.
- This a reminder that the only difference between a TS model and a DS model resides in $\phi(\mathbf{1})$.

Identifying BN Decompositions

- A problem still exists with the BN decomposition $\Longrightarrow$ it is always possible to create a TS process that is arbitrarily close to a DS process.
- For example, consider a TS model in which $\phi(\mathbf{1}) = 0.999 \Longrightarrow$ the TS process acts as if or is observationally equivalent to a DS model in finite samples, say, $T < 700$.
- Add an infinitesimal random walk to the TS model and create a DS model.
- In finite samples, unit root tests often lack the power to distinguish between unit root and stationary processes.
- For example, the null of a unit root is a joint test of a unit root process and restrictions on $\phi(\mathbf{L})$.
- This motivates *In finite samples, $T < \infty$, the failure to reject a unit root only indicates x_t is observationally equivalent to a unit root process.*
- Restrictions on $\phi(\mathbf{L})$ may be relaxed as the sample size grows $\Longrightarrow$ if $T \longrightarrow \infty$, can approximate $\phi(\mathbf{1})$ arbitrarily well.

Summary of the BN Decomposition

- The BN decomposition is an easy and intuitive tool for computing the trend of a $I(1)$ variable.

- Behavior of $I(1)$ and $I(0)$ processes differ only at the infinite horizon.

- The sampling uncertainty around estimates at the long-horizon and inference on those estimates can be substantial $\Longrightarrow$ may not have much confidence in the estimates and inference.

- Have only studied univariate processes, but BN intuition carries over to multivariate models $\Longrightarrow$ multivariate random walks and cointegration.
  1. For the former see Stock and Watson (1998, “Testing for common trends,” *Journal of the American Statistical Association* 83, 1097–1107).
  2. Engle and Granger (1987, “Co-Integration and error correction: Representation, estimation, and testing,” *Econometrica* 55, 251–276) introduce cointegration.
  3. Multivariate BN trends are referred to as BNSW models; early example in macro is King, Plosser, Stock, and Watson (1991, “Stochastic trends and economic fluctuations,” *American Economic Review* 81, 819–840).
  4. Surveys are Watson (1994, “Vector autoregressions and cointegration,” in Engle and McFadden (eds.), Handbook of Econometrics, v. 4, ch. 47, 2843–2915, New York, NY: Elsevier B.V.) and chs. 6–9 in Lütkepohl (2007, New Introduction to Multiple Time Series Analysis, New York, NY: Springer.

The BNSW Decomposition: A Problem

- Estimate a VAR(p) on growth rates of $x_{1,t}$ and $x_{2,t}$, which are $I(0)$.
- If the covariance matrix of the VAR residuals is positive definite, trends driving $x_{1,t}$ and $x_{2,t}$ are independent $\Longrightarrow$ bivariate random walk.
  1. $(1-\mathbf{L})x_{1,t} = (1-\mathbf{L})\tau_{1,t} + (1-\mathbf{L})\varepsilon_{1,t}$ and $(1-\mathbf{L})x_{2,t} = (1-\mathbf{L})\tau_{2,t} + (1-\mathbf{L})\varepsilon_{2,t}$
  2. The bivariate random walk is $\begin{bmatrix} \tau_{1,t} \\ \tau_{2,t} \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \tau_{1,t-1} \\ \tau_{2,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix}$.
- If not, the VAR on $(1-\mathbf{L})x_{1,t}$ and $(1-\mathbf{L})x_{2,t}$ is not fundamental.
  1. $(1-\mathbf{L})x_{1,t} = (1-\mathbf{L})\tau_t + (1-\mathbf{L})\varepsilon_{1,t}$ and $(1-\mathbf{L})x_{2,t} = (1-\mathbf{L})\tau_t + (1-\mathbf{L})\varepsilon_{2,t}$
  2. $\Longrightarrow \begin{bmatrix} (1-\mathbf{L})x_{1,t} \\ (1-\mathbf{L})x_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\eta_{t-1} + \begin{bmatrix} (1-\mathbf{L})\psi_1(\mathbf{L})\upsilon_{1,t} \\ (1-\mathbf{L})\psi_2(\mathbf{L})\upsilon_{2,t} \end{bmatrix}$.
  3. Cannot retrieve the fundamental shocks because of unit MA roots in $(\upsilon_{1,t},\ \upsilon_{2,t})$.

The BQ Decomposition: A Fix of the Problem

- The problem is $x_{1,t}$ and $x_{2,t}$ share a common trend or cointegrate.
  1. $\Longrightarrow$ A stochastic singularity in the growth rates VAR.
  2. Its reduced-form errors are linear combinations of the same shocks $\Longrightarrow$ $\mathbf{\Omega}$ is not positive definite.

- A solution is to note that there exists a stationary linear combination,
  1. $\ln x_{2,t} - \vartheta \ln x_{1,t} = \psi_2(\mathbf{L})\upsilon_{2,t} - \vartheta\psi_1(\mathbf{L})\upsilon_{1,t}$, which integrates out the common trend τ_t.
  2. This is a cointegrating relation, $z_t = [1 \;\; -\vartheta]\,[\ln x_{2,t} \;\; \ln x_{1,t}]'$, where $[1 \;\; -\vartheta]$ is the cointegrating vector.
  3. Use economic theory or econometrics of cointegration to obtain ϑ.

- Estimate a VAR on $(1-\mathbf{L})\ln x_{1,t}$ and z_t to obtain fundamental shocks.

- Satisfy the BQ decomposition because as $j \longrightarrow \infty$, $\mathbf{E}_t x_{1,t+1}$ equals zero in response to the transitory shock(s) by construction.

Frequentist Estimation of a SVAR with LR Restrictions, I

- Step 1: Estimate the unrestricted VAR by OLS

$$y_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{B}_j y_{t-j} + \varepsilon_t, \ \ \varepsilon_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{\Omega}_{n\times n}).$$

- Step 2: Construct the reduced form VMA(∞), $y_t = \mu_y + \sum_{\ell=0}^{\infty} \mathbf{C}_\ell \varepsilon_{t-\ell}$, where
  - A. $\mathbf{C}(\mathbf{L}) = [\mathbf{I}_n - \mathbf{B}(\mathbf{L})]^{-1}$ and $\mathbf{C}_0 \equiv \mathbf{I}_n \Longrightarrow$ Wold representation.
  - B. The structural VMA(∞) is $y_t = \mu_y + \sum_{\ell=0}^{\infty} \mathbf{\Gamma}_\ell \eta_{t-\ell}, \ \eta_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$.
  - C. $\Longrightarrow \varepsilon_t = \mathbf{\Gamma}_0 \eta_t$ and $\mathbf{\Omega} = \mathbf{\Gamma}_0 \mathbf{\Gamma}_0'$, which gives $0.5(n^2 + n)$ (= 6 when $n = 3$) nonlinear equations to solve for as many of the n^2 unknown elements of $\mathbf{\Gamma}_0$.

- Step 3: Solve for the remaining $0.5(n^2 - n)$ unknown elements of $\mathbf{\Gamma}_0$ using the $0.5(n^2 - n)$ (= 3 for $n = 3$) long-run neutrality restrictions.
  - A. $\sum_{\ell=0}^{\infty} \mathbf{\Gamma}_{12,\ell} = 0, \ \sum_{\ell=0}^{\infty} \mathbf{\Gamma}_{13,\ell} = 0$, and $\sum_{\ell=0}^{\infty} \mathbf{\Gamma}_{23,\ell} = 0$.
  - B. Since $\mathbf{C}(\mathbf{L})\varepsilon_t = \mathbf{\Gamma}(\mathbf{L})\eta_t$ and $\varepsilon_t = \mathbf{\Gamma}_0 \eta_t \Longrightarrow \sum_{\ell=0}^{\infty} \mathbf{C}_\ell \mathbf{\Gamma}_0 = \sum_{\ell=0}^{\infty} \mathbf{\Gamma}_\ell$.
  - C. Define $\mathbf{C}(\mathbf{1}) = \sum_{\ell=0}^{\infty} \mathbf{C}_\ell$ and $\mathbf{\Gamma}(\mathbf{1}) = \sum_{\ell=0}^{\infty} \mathbf{\Gamma}_\ell \Longrightarrow$ solve the $0.5(n^2 - n) = 3$ equations that equal zero in the system $\mathbf{C}(\mathbf{1})\mathbf{\Gamma}_0 = \mathbf{\Gamma}(\mathbf{1})$.

- Step 4: Given $\mathbf{\Gamma}_0$, compute $\eta_t = \mathbf{\Gamma}_0^{-1} \varepsilon_t$ and the IRFs as $\mathbf{\Gamma}_\ell = \mathbf{C}_\ell \mathbf{\Gamma}_0$, $\ell = 1, 2, \ldots$.

An Example of a SVAR Identified on LR Restrictions, II

- Estimate a VAR on $\mathcal{Y}_t = [\Delta y_t \;\; \Delta\pi_t \;\; (r_t - \pi_t)]'$ by OLS $\Longrightarrow$ obtain $\widehat{\mathbf{B}}(\mathbf{L})$ and $\widehat{\boldsymbol{\Omega}}$.
- Since $\dim(\mathcal{Y}_t) = 3$, there are $0.5(n^2 - n) = 3$ long-run restrictions.
  1. The restrictions are $\dfrac{\partial \mathbf{E}_t y_{t+j}}{\partial \upsilon_{d,t}} = \dfrac{\partial \mathbf{E}_t y_{t+j}}{\partial \upsilon_{m,t}} = \dfrac{\partial \mathbf{E}_t \pi_{t+j}}{\partial \upsilon_{m,t}} = 0$, as $j \longrightarrow \infty$,
  2. which appear in $\boldsymbol{\Gamma}(\mathbf{1}) = \begin{bmatrix} \Gamma_{11}(\mathbf{1}) & 0 & 0 \\ \Gamma_{21}(\mathbf{1}) & \Gamma_{22}(\mathbf{1}) & 0 \\ \Gamma_{31}(\mathbf{1}) & \Gamma_{32}(\mathbf{1}) & \Gamma_{33}(\mathbf{1}) \end{bmatrix}$.
- The nine unknown elements of $\boldsymbol{\Gamma}_0$ are solved for using the
  1. three equations in $\mathbf{C}(\mathbf{1})\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}(\mathbf{1})$ in the upper triangle of $\boldsymbol{\Gamma}(\mathbf{1})$ equal to zero
  2. together with the six nonlinear equations in the lower triangle of $\boldsymbol{\Omega} = \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0'$.

Frequentist Estimation of a SVAR with LR Restrictions, II

- Another method for computing the BQ decomposition uses the Cholesky decomposition.
- Since $\mathbf{\Gamma}(\mathbf{L})\eta_t = \mathbf{C}(\mathbf{L})\varepsilon_t \Longrightarrow \mathbf{\Gamma}(\mathbf{1})\mathbf{\Gamma}(\mathbf{1})' = \mathbf{C}(\mathbf{1})\mathbf{\Omega}\mathbf{C}(\mathbf{1})'$.
- Also, $\mathbf{\Gamma}(\mathbf{1})\mathbf{\Gamma}(\mathbf{1})' = \mathbf{C}(\mathbf{1})\mathbf{\Gamma}_0\mathbf{\Gamma}_0'\mathbf{C}(\mathbf{1})' \Longrightarrow$ calculate the Cholesky decomposition, $\left[\mathbf{\Gamma}(\mathbf{1})\mathbf{\Gamma}(\mathbf{1})'\right]^{0.5} = \mathbf{C}(\mathbf{1})\mathbf{\Gamma}_0 = \left[\mathbf{C}(\mathbf{1})\mathbf{\Omega}\mathbf{C}(\mathbf{1})'\right]^{0.5} \Longrightarrow \mathbf{\Gamma}_0 = \mathbf{C}(\mathbf{1})^{-1}\left[\mathbf{C}(\mathbf{1})\mathbf{\Omega}\mathbf{C}(\mathbf{1})'\right]^{0.5}$.
- Repeat steps 1, 2, and 4 of the previous slide, but

  ALT-3. replace step 3 by computing $\mathbf{\Gamma}_0 = \mathbf{C}(\mathbf{1})^{-1}\left[\mathbf{C}(\mathbf{1})\mathbf{\Omega}\mathbf{C}(\mathbf{1})'\right]^{0.5}$.
- A final note is the procedures for computing the BQ decomposition
  1. do not depend on a recursive ordering of the long-run restrictions.
  2. There are neither necessary nor sufficient conditions restricting the LR identification to be recursive $\Longrightarrow \mathbf{\Gamma}(\mathbf{1})$ can have non-recursive restrictions.
  3. However, the BQ decomposition requires $0.5(n^2 - n)$ zero restrictions on $\mathbf{\Gamma}(\mathbf{1})$ for just-identification of the SVAR.

Identification of SVARs, V: Sign Restrictions, 1

- **Sign Restrictions**: Instead of forcing point restrictions on elements of $\mathbf{A}_0$ and/or $\mathbf{D}$,
  1. only examine IRFs that are consistent with $\mathbf{IRF}_{1:n,\cdot}(h)$ that reside in a pre-selected set,
  2. where h is a small integer, say, H in quarterly data $\Longrightarrow$ from impact, $h = 0$, to a horizon of one year horizon, $H = 5$.
  3. Sign restrictions can also be applied to FEVDs.

- Faust (CRCSPP, 1998) and Uhlig (JME, 2005) are the earliest examples of sign restrictions in the VAR literature, but their goals and approaches differ.
  1. From $h = 0$ to H, monetary policy shocks contribute $< x_y\%$ to the FEVD of output and $> x_\pi\%$ to the FEVD of inflation, which involve imposing a penalty function on the VAR; see Faust (CRCSPP, 1998).
  2. Uhlig (JME, 2005) proposes a sign restriction tied to a contractionary monetary policy shock lowers y and π from $h = 0$ to H.
  3. Placing sign restrictions on IRFs is a far more popular identification strategy than is the penalty function approach of Faust (1998).

Identification of SVARs, V: Sign Restrictions, 2

- The original use of sign restrictions to identify VARs is Faust (CRCSPP, 1998).

- However, Faust's goal is evaluate the **robustness** of identifying restrictions, especially with respect to output and inflation responses to identified monetary policy shocks.
  1. Are SVAR results robust to small changes in identification restrictions?
  2. The problem is SVARs often rely on implicit zero restrictions because $\mathcal{Y}_t$ may lack variables important for explaining the dynamics and variation of variables that are included $\Longrightarrow$ omitted variable problem.
  3. However, a large $\dim(\mathcal{Y}_t)$ almost always leads to a SVAR that is only partially identified $\Longrightarrow$ inference on partially identified models is possible but the degree of econometric difficulty is much greater.

- Faust: Consider the vector of sign restrictions μ applied to $\mathbf{IRF}_{i,j}(\ell, \ell+s)$, $\ell = 0, 1, \ldots, H \Longrightarrow$ the sign restrictions $\mathbf{IRF}_{i,j}(\ell, \ell+s)\mu \geq \mathbf{0}$.
  1. Choose the vector of sign restrictions μ to maximize $\mu'\mathbf{FEVD}_{i,j}\mu \Longrightarrow$ for every "admissible" SVAR identification, the share of the jth shock in explaining $\mathcal{Y}_{i,t}$ is μ s.t. $\mathbf{IRF}_{i,j}(\ell, \ell+s)\mu \geq \mathbf{0}$ and the normalization $\mu'\mu = 1$.
  2. Fix $\tilde{\mu}$, compute $\tilde{\mu}'\mathbf{FEVD}_{i,j}\tilde{\mu}$ s.t. the constraints, ask if the solution satisfies $x_y\%$,
  3. if yes stop, otherwise check if the IRFs driven by $\eta_{j,t}$ are a priori "reasonable."
  4. $\Longrightarrow$ if these IRFs are reasonable, "reject" the $x_y\%$ hypothesis.
  5. Otherwise, add restrictions to penalize these unreasonable IRFs on μ and restart Faust's procedure from step 1.

Identification of SVARs, V: Sign Restrictions, 3

- **Sign Restrictions**: The most widely used method for estimating SVARs on sign restrictions differs from the approach Faust (1998) advocates.

- The standard approach identifies SVARs using methods developed by Uhlig (2005).
  1. Impose sign restrictions from impact to some finite H on the responses of variables either controlled or targeted by the central bank to a contractionary monetary policy shock.
  2. For example, the response of the policy rate is non-negative and increases, say, by no more than 100 basis points
  3. while inflation's response is not positive (*i.e.*, throw out the price puzzle) and falls, say, by no more than two percent.

- Uhlig proposes to create the IRFs of a sign restricted SVAR by
  1. estimating an unrestricted VAR and covariance matrix, $\mathbf{\Omega}$, of the OLS residuals.
  2. Could be a **K**- or **C**-model, but most often the SVAR is just-identified and recursive $\Longrightarrow \mathbf{D} = \mathbf{\Omega}^{1/2}$.
  3. Draw K random orthonormal matrices $\mathcal{U}_k$ conformable with $\mathbf{D}$, where $k = 1, \ldots, K$ and $\mathcal{U}_k\mathcal{U}_k' = \mathbf{I}_n \Longrightarrow \mathcal{U}_k$s rotate the impact matrix, $\mathbf{D}\mathcal{U}_k$, K ways.
  4. Construct K IRFs initialized by $\mathbf{D}\mathcal{U}_k$, but if an IRF violates a sign restriction toss out this draw of IRFs.
  5. Adjust K to obtain the desired number of IRFs, which satisfy the sign restrictions and report error bands.

Sign Restrictions: Preliminary Summary

- Sign restrictions involve several under discussed issues.
- Sign restrictions generate posterior sets of IRFs not point valued posterior distributions.
- Sign restrictions represent the beliefs or priors of the analyst estimating the SVAR.
- Sign restrictions impose nonlinear restrictions on the impact, $\mathbf{A}_0$, and slope coefficients, $\mathbf{A}_j s$, of a SVAR.
- The nonlinear restrictions are left unknown by the analyst.
- Hence, the implications of the sign restrictions for the economic interpretation of the SVAR are hidden from view.

Sign Restrictions: Inoue and Kilian (2013)

- Sign restrictions are popular in the SVAR literature, but have several problems.
  1. The inference problem is sign restrictions are set defined rather than being restricted to a point as are zero (or hard) restrictions.
  2. Inoue and Kilian (JofE, 2013) argue the way to conduct this inference is to find the SVAR that generates the highest posterior density (HPD) of $\mathbf{A}_0$ and $\mathbb{A}$.

- Define the HPD set $\left\{\widetilde{\mathbf{IRF}}_{i,j}(0:H) \in \mathbb{IRF}_{i,j}(0:H) : g\left(\widetilde{\mathbf{IRF}}_{i,j}(0:H)\right) \geq x_i\%\right\}$

  $\Longrightarrow$ measures the posterior uncertainty of $\widetilde{\mathbf{IRF}}_{i,j}(0:H) \in \mathbb{IRF}_{i,j}(0:H)$.
  1. Estimate the unrestricted BVAR conditional on the normal conjugate prior generating the posteriors $\mathbb{B}\big|\boldsymbol{\Omega}, \mathbb{Y} \sim \mathcal{N}\left(\overline{\mathbb{B}}\big|\boldsymbol{\Omega} \otimes \overline{\mathbb{Z}}\right)$ and $\boldsymbol{\Omega}^{-1}\big|\mathbb{Y} \sim \mathcal{W}\left(\overline{\mathbb{S}}^{-1}, \overline{\boldsymbol{\nu}}\right)$.
  2. Draw from the posteriors and form a space of $n \times n$ orthonormal matrices $\boldsymbol{\mathcal{U}}$, $\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{U}}' = \mathbf{I}_n$ using a uniform distribution K times.
  3. Apply each $\boldsymbol{\mathcal{U}}_k$, $k = 1, \ldots, K$, to rotate $\mathbf{D} \Longrightarrow$ compute $\widetilde{\mathbf{IRF}}_{i,j}(0:H)$ starting from $\mathbf{D}\boldsymbol{\mathcal{U}}$ and if $\mathbf{IRF}_{i,j}(\ell, \ell+s)\mu \geq \mathbf{0}$ keep this draw, otherwise toss it.
  4. Steps 3 and 4 are repeated $\mathbf{M}$ times that ends with sorting the IRFs in descending order using $g\left(\widetilde{\mathbf{IRF}}_{i,j}(0:H)\right)$ to form $\mathbb{IRF}_{i,j}(0:H)$.
  5. The first $1-\alpha$ elements of $g\left(\widetilde{\mathbf{IRF}}_{i,j}(0:H)\right)$ form the HPD $\Longrightarrow$ use these $\mathbf{IRF}_{i,j}(0:H)$ to construct posterior credible sets that are used to conduct inference on the SVAR identified by sign restrictions.

Sign Restrictions: Giacomini and Kitagawa (2018)

- There are two issues not often confronted when conducting inference on SVARs identified by sign restrictions.
  1. The information content of the prior dominates inference asymptotically $\Longrightarrow$ the prior is never updated in the posterior, and
  2. the information content of the priors dominate $\Longrightarrow$ Bayesian credible sets are always covered by truth.

- Giacomini and Kitagawa (2018) propose a "robust" Bayes approach to assess SVARs endowed with sign restrictions.
  1. Robust Bayes places a prior on the reduced-form VAR parameters not the elements of the structural IRFs.
  2. Sign restrictions are constraints on the reduced-form VAR parameters.

Sign Restrictions: Giacomini and Kitagawa (2018), cont.

- Follow usual practice and impose a single prior on the reduced-form VAR parameters, but
  1. invoke a class of "arbitrary" priors for the structural parameters $\Longrightarrow \boldsymbol{\mathcal{U}}$.
  2. Use Bayes rule to update or mix priors over $\boldsymbol{\mathcal{U}}$, which robustifies inference of the sign restricted SVAR.
  3. This procedure generates posterior mean bounds on the set, which GK "interpret as an estimator of the identified set" $\Longrightarrow$ a robust measure of the posterior uncertainty surrounding the identified set.

- A robust prior over all potential rotations of $\boldsymbol{\mathcal{U}} \Leftrightarrow$ data are uninformative w/r/t $\boldsymbol{\mathcal{U}}$.
  1. The analyst has uninformative or ambiguous beliefs about which priors are most "credible" in terms of the posterior of the SVAR $\Longrightarrow$ no prior information about the sign restrictions, which impose nonlinear restrictions on $\mathbf{A}_0$ and $\mathbb{A}$.
  2. $\mathbb{Y}$ are only informative about the reduced form VAR $\Longrightarrow \mathbb{B}\big|\mathbf{\Omega}, \mathbb{Y}$ and $\mathbf{\Omega}^{-1}\big|\mathbb{Y}$.
  3. Assess sign restrictions on IRFs across the priors on $\boldsymbol{\mathcal{U}} \Longrightarrow$ varying $\boldsymbol{\mathcal{U}}$ generates information about the impact of the sign restrictions on the likelihood of the SVAR $\Longrightarrow$ robustify or produce more efficient uncertainty intervals for the IRFs.

Sign Restrictions: Arias, Rubio-Ramírez, and Waggoner (2018)

- Arias, Rubio-Ramírez, and Waggoner (2018) construct three algorithms to estimate SVARs identified, at least, in part by sign restrictions.
- Two algorithms produce posteriors of SVARs that have sign restrictions and hard or zero short-run restrictions on $\mathbf{A}_0$.
- As is standard, a uniform prior is given to $\mathcal{U}$ by Arias, Rubio-Ramírez, and Waggoner (ARRW).
- The algorithms also use the normal-inverse Wishart density as the prior on the reduced-form VAR parameters.
- ARRW prove the uniform-normal-inverse Wishart (UNIW) prior yields a posterior (that they call) the normal-generalized-normal density.
  1. Normal-generalized-normal (MGN) posterior is conjugate to the UNIW prior.
  2. The generalized-normal distribution adds a shape parameter that can yield either fat tails or skew to the posterior.
- Algorithm 1 of ARRW is the canonical sampler of sign restricted SVARs $\Longrightarrow$ show posterior is drawn from the NGN density.

Sign Restrictions: Arias, Rubio-Ramírez, and Waggoner (2018), cont.

- Algorithm 2 adapts Algorithm 1 to handle the point restrictions by placing linear restrictions on the columns of $\mathcal{U}$.

- The problem is the sign and linear restrictions are only satisfied by the reduced-form VAR parameters and $\mathcal{U}$.

- There are no structural parameters (*i.e.*, $\mathbf{A}_0$ and $\mathbf{A}_j$s) that are consistent with the sign and hard restrictions.
  1. An implication is the posterior of a SVAR with sign and point restrictions does not exist (except on sets of measure zero).
  2. Algorithm 2 is not a valid procedure to generate the posterior of the SVAR.

- ARRW propose to solve the problem by including importance sampling (IS) with replacement steps in Algorithm 2, which is their Algorithm 3.
  1. The importance sampler is necessary because the likelihood of the SVAR with respect to $\mathbf{A}_0$ and $\mathbf{A}_j$s is not analytic.
  2. IS weights equal zero if $\mathbf{A}_0$ and $\mathbf{A}_j$ do not satisfy the sign and point restrictions.
  3. Otherwise, the importance sampling weights are the ratio of the $\left|\det(\mathbf{A}_0)\right|$ to the *pdf* of the importance density draw of $\mathbf{A}_0$ and $\mathbf{A}_j$s.
  4. Algorithm 3 needs to make (many) more IS draws (*i.e.* effective sample size) than the desired number of draws of $\mathbf{A}_0$ and $\mathbf{A}_j$ to be valid.
  5. The resampling step produces unweighted and independent draws that are needed for the posterior to give Bayesian credible sets.

Introduction

- Frequentist structural VAR estimators can always be interpreted as an instrumental variables (IV) problem.

- An early example of a SVAR estimated by IV is Shapiro and Watson (1988, "Sources of business cycle fluctuations," in Fisher, S. (ed.), NBER Macroeconomic Annual, Cambridge, MA: MIT Press).

- There are many papers interpreting SVARs as IV estimators.
  1. Long-run restrictions: Pagan (1994, "Introduction, Calibration and econometric Research: An overview," *Journal of Applied Econometrics* 9, S1–S10), Sarte (1997, "On the identification of structural vector autoregressions," *Economic Quarterly*, Federal Reserve Bank of Richmond 83(3), 45–67), and King and Watson (1997, "Testing long-run neutrality," *Economic Quarterly*, Federal Reserve Bank of Richmond 83(3), 69–101).
  2. Short-run restrictions: Sarte (1997), King and Watson (1997), and Pagan and Robertson (1998).

- The label "Proxy VARs" is a way to rebrand SVARs estimated by IV.

An Example of a Proxy VAR

- Ramey (2016, pp. 79–80) suggests proxy VARs are identified using external information (*i.e.*, instruments).

- As an example, consider estimating a SVAR to obtain the responses of $y_{1,t}$ to $y_{2,t}$ and $y_{3,t}$.

- IV teaches that if there is a variable, x_t, "external" to the SVAR that is correlated with $y_{1,t}$ and not $y_{2,t}$ and $y_{3,t}$, x_t is a valid instrument.
  1. Estimate the reduced form VAR and save the residuals $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and $\varepsilon_{3,t}$.
  2. Rank condition requires $\mathbf{E}\{x_t\varepsilon_{1,t}\} = \vartheta \neq 0$ and $\mathbf{E}\{x_t\varepsilon_{2,t}\} = \mathbf{E}\{x_t\varepsilon_{3,t}\} = 0$ for x_t to be a valid instrument. (Since the example is just-identified, the order condition is satisfied.)
  3. The first-step regressions are $\varepsilon_{2,t}$ on $\varepsilon_{1,t}$ and $\varepsilon_{3,t}$ on $\varepsilon_{1,t}$ employing x_t as the instrument in both cases, where the residuals are labeled $\nu_{2,t}$ and $\nu_{3,t}$.
  4. A second-step regress $\varepsilon_{1,t}$ on $\varepsilon_{2,t}$ and $\varepsilon_{3,t}$ using $\nu_{2,t}$ and $\nu_{3,t}$ as instruments.
  5. External instruments often pose as dummy variables $\Longrightarrow$ subjective readings of the history of shifts in fiscal or monetary policy, as in Romer and Romer (1994, "Monetary policy matters," *Journal of Monetary Economics* 34, 75–88), but whether "narrative" variables are legitimate instruments is questioned by Leeper (1997, "Narrative and VAR approaches to monetary policy: Common identification problems," *Journal of Monetary Economics* 40, 641–657).

IV and Weak Instruments

- Weak instruments create problems when applying IV to aggregate and financial data.

- The problem is the correlation(s) of the instrument(s) and the variable needing an instrument is small, or $\left|\mathbf{E}\{x_t\varepsilon_{1,t}\}\right| \ll \epsilon$, where ϵ is some small positive number.
  1. This is not a small sample problem, but a problem in population!!!
  2. The first modern treatments of weak instruments are Nelson and Startz (1990, "Some further results on the exact small sample properties of the instrumental variable estimator," *Econometrica* 58, 967–976) and (1990, "The distribution of the instrumental variables estimator and its *t*-ratio when the instrument is a poor one," *Journal of Business* 63, S125–S140); also see Stock, Wright, and Yogo (2002, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics* 20, 518–529) and Andrews, Stock, and Sun (2019, "Weak instruments in IV regression: Theory and practice," *Annual Review of Economics* 11, 727–753).

- Nelson and Startz show the weak instrument problem appears as
  1. asymptotic bias in the IV estimator that approaches OLS in the limit.
  2. Load more and more weak instruments in $X_t \implies$ IV estimator collapses to OLS.
  3. The asymptotic distribution of the IV estimator is bimodal $\implies$ the probability is zero that the estimator falls between the modes of the two peaks, which
  4. implies a large asymptotic variance in the IV estimator yields empty confidence intervals because the region of zero probability cannot be ruled out,
  5. and the quality of the asymptotic distribution of the IV estimator is sensitive to sample size and the constraints placed on the rank condition.

Weak Instrument-Robust Proxy VARs

- Proxy VARs face a non-standard inference problem when there are weak instruments $\Longrightarrow$ ϑ is within an ϵ-neighborhood of zero.

- Assume the first element of the structural shock vector η_t needs an external instrument for identification.
  1. Since the mapping from the reduced-from innovations to the structural shocks is $\varepsilon_t = \mathbf{\Omega}^{-0.5}\eta_t$, the rank condition is $\mathbf{E}\{\varepsilon_t \mathcal{X}_t'\} = \mathbf{E}\{\mathbf{\Omega}^{-0.5}\eta_t \mathcal{X}_t'\} = \vartheta\,\mathbf{\Omega}_1^{-0.5}$, where $\mathbf{\Omega}_1^{-0.5}$ is the first column of the inverse of the Cholesky decomposition of $\mathbf{\Omega}$.
  2. Given $|\vartheta| < \epsilon$, Olea, Stock, and Watson (2018) show the IRFs are asymptotically the same applying a recursive ordering to $\mathbf{\Omega}$ $\Longrightarrow$ estimate the reduced-form VAR by OLS and apply a Cholesky decomposition to $\mathbf{\Omega}$.

Weak Instrument-Robust Proxy VARs, cont.

- Olea, Stock, and Watson (2018) propose a solution to the inference problem of proxy VARs affected by weak instruments.

- The solution relies on tests developed in two classic papers.
  1. A test of the equality of the means of two variables (or distributions) that are measured in different units is in Fieller (1944, "A Fundamental formula in the statistics of biological assay, and some applications," *Quarterly Journal of Pharmacy and Pharmacology* 17, 117–123).
  2. A test that $\vartheta = \vartheta_0$ is constructed by Anderson and Rubin (1949, "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics* 20, 46–63).

- The Fieller (1944) test is applied to the means of the unrestricted MA(k) and the impact restriction at $\vartheta = \vartheta_0 \Longrightarrow$ the IRF of interest.

- The difference of these means is used by Olea, Stock, and Watson to form
  1. a Wald test using elements of $\mathbf{\Omega}$ that is inverted to construct
  2. a Anderson and Rubin (1949) confidence set for the IRFs that
  3. is robust to weak instruments over a sequence of ϑ_0s.

How and Why Does a Proxy VAR Identify Shocks?

- Proxy VARs were introduced to avoid the problem that the reduced-form VAR lag polynomial is not invertible $\Longrightarrow$ violate the Wold decomposition theorem.
  1. The VAR is not fundamental $\Longrightarrow$ cannot recover fundamental shocks from the data and the converse.
  2. Overcome lack of fundamentalness because the external instruments identify
  3. the structural responses without having to recover the structural shocks from fundamental reduced-form innovations.

- Stock and Watson (2018) discuss that identification of proxy VARs rest on more assumptions than are often stated.
  1. Besides the rank condition, $\mathbf{E}\{\varepsilon_{1,t}\mathcal{X}_t'\} = \vartheta \neq 0$, and instrument exogeneity, $\mathbf{E}\{\varepsilon_{i\neq1,t}\mathcal{X}_t'\} = 0$, where $\varepsilon_{i\neq1,t}$ are reduced-form innovations that are not $\varepsilon_{1,t}$,
  2. instruments are uncorrelated at all leads and lags with all the elements of ε_t, $\mathbf{E}\{\varepsilon_{t+\ell}\mathcal{X}_t'\} = 0$, $\ell \neq 0 \Longrightarrow \mathcal{X}_t$ is unpredictability projected on the history of ε_t.
  3. The exogeneity of instruments at all leads and lags is necessary because identification is about the impact responses of $\mathcal{Y}_t$ w/r/t $\varepsilon_{1,t}$.
  4. Key restrictions on the lags not the leads $\Longrightarrow$ otherwise invertibility matters.
  5. If the assumption is violated, need to identify impact and lags responses of $\mathcal{Y}_t$ to the history of $\varepsilon_{1,t}$.

Proxy VARs are (almost) SVARs

- Suppose the lead-lag exogeneity assumption is violated for some subset of the lags of the reduced-form VAR, $\mathbf{E}\left\{\varepsilon_{t-s}\mathcal{X}_t'\right\} \neq 0$, $s = 1, \ldots, k < p$.

- Stock and Watson (2018) show that "control" variables can restore exogeneity.
  1. Project $\mathcal{X}_t$ onto a vector of control variables, $\mathcal{Z}_t$, and $u_{\mathcal{X},t} = \mathcal{X}_t - \mathbf{E}\left\{\mathcal{X}_t \middle| \mathcal{Z}_t\right\}$
  2. to purge $\mathcal{X}_t$ of the correlation with ε_{t-s}, $s = 1, \ldots, k$, where the linear VAR yields conditional expectations equivalent to the projections operator.

- Where to find control variables? Stock and Watson (2018) suggest
  1. leading candidates are lags of $\mathcal{Y}_t$, $\mathcal{Z}_t = \left[\mathcal{Y}_{t-1}' \ldots \mathcal{Y}_{t-k}'\right]'$.
  2. Equivalent to adding $\mathcal{X}_t$ to the reduced-form VAR (almost).
  3. $\mathcal{X}_t$ is regressed on a restricted subset of lags of $\mathcal{Y}_t$ and no own lags.
  4. This approach is more restrictive than proxy VARs $\Longrightarrow$ need to satisfy the invertibility of the VAR.
  5. Will be more efficient than proxy VARs, given the instruments are not weak $\Longrightarrow$ one-step estimation instead of a two-step IV estimator.

Proxy VARs can be Sign Restricted

- Another issue is proxy VARs produce can produce set identified IRFs; for example, see Antolín-Díaz and Rubio-Ramírez (2018, "Narrative Sign Restrictions for SVARs," *American Economic Review* 108, 2802–2829).
- Combining narrative identification with sign restrictions restricts, say, a QE policy episode to produce impact and h-step ahead positive responses.
- There at least two problems that a sign restricted-proxy VAR estimator has to address, according to Giacomini, Kitagawa, and Read (2021, "Identification and inference under narrative restrictions," manuscript, Department of Economics, University College London).
  1. Priors on the reduced-form VAR are not updated under narrative-sign restrictions $\Longrightarrow$ its likelihood is flat given the restrictions set by the orthonormal rotation matrix.
  2. This matrix controls the probability the narrative-sign restrictions are true.
  3. Bayesian inference often rests on the priors of the narrative-sign restrictions.
  4. $\Longrightarrow$ Prior yields a SVAR with a flat likelihood in the direction of the restrictions.
  5. This is the Baumeister and Hamilton (2015) result that the posterior of the SVAR is proportional to the prior of the sign restrictions.
- Giacomini, Kitagawa, and Read (2021) propose Bayesian methods to solve the problems of estimating and evaluating sign restricted-Proxy VARs.

Introduction

- Identification of a SVAR is about its properties **in population**.
  1. These are the "algebraic" properties of the SVAR, which are the restrictions imposed on the unrestricted VAR.
  2. Do not confuse identification with the small sample properties of an estimator
  3. This is true for identifying any and all econometric models.

- Identification of SVARs is more than counting the number of restrictions.
  1. The **K**- and **C**-models have $0.5n(n-1)$ free parameters while there are $0.5n(3n-1)$ free parameters in the **AB**-model.
  2. Have to restrict at least this many elements in $\mathbf{A}_0$ (and $\mathbf{Q}$ for the **AB**-model) to achieve just-identification $\Longrightarrow$ this is the order condition, which should be familiar from GMM and IV estimators in general; see Rothenberg (1971, "Identification in parametric models," *Econometrica* 39, 577–591).

- There is more to identifying SVARs than checking the order condition.
  1. The order condition is only necessary. Are there sufficient conditions that yield global identification? What are the implications of local identification?
  2. Are identified shocks of a SVAR represent fundamental economic disturbances impinging on an economy?
  3. Are the short- and long-run restrictions identifying a SVAR robust? $\Longrightarrow$ Does a small perturbation to these restrictions generate large changes in a SVAR's predictions (*i.e.*, IRFs, FEVDs, and the historical error decomposition).

Identification of SVARs, I

- Rubio-Ramírez, Waggoner, and Zha (REStud, 2010) provide necessary and sufficient conditions to identify locally and globally recursive and non-recursive just- and over-identified **K**-model SVARs.
  1. Study identification of SR restrictions on $\mathbf{A}_0$ or restrictions on IRFs.
  2. **Observational equivalence**: Given $\mathbf{A}_0$, $\mathbb{A}$, and $\mathbb{Y}$, no combination of $\mathbf{A}_0^{\ddagger}$ and $\mathbb{A}^{\ddagger}$ gives $\mathbb{Y}$ the same distribution (*i.e.*, likelihood of SVAR) $\Rightarrow \mathbf{A}_0 = \mathbf{A}_0^{\ddagger}\mathbf{P}$ and $\mathbb{A} = \mathbb{A}^{\ddagger}\mathbf{P}$, where $\mathbf{P}$ is an orthogonal matrix, $\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I} \Rightarrow \mathbf{P}' = \mathbf{P}^{-1}$ $(\det(\mathbf{P}) = 1$ or $-1)$.
  3. **Local identification**: $\mathbf{A}_0$ and $\mathbb{A}$ are locally identified *iff* no other SVARs exist in an open neighborhood around $(\mathbf{A}_0, \mathbb{A})$ that are observationally equivalent.
  4. **Global identification**: $\mathbf{A}_0$ and $\mathbb{A}$ are globally identified *iff* there are no other SVAR coefficients that are observationally equivalent.

- **Necessary condition**: The number of identifying restrictions $\geq$ free parameters $\Longrightarrow$ order condition of Rothenberg (*Econometrica*, 1971).
  1. If identifying restrictions = free parameters, SVAR is just-identified.
  2. SVAR is over-identified when the inequality is strong.

- **Sufficient condition**: SVAR restrictions are embedded in a sequence of matrices with rank $= n-j \Longrightarrow \dim(\mathcal{Y}_t)$ net of the location of a structural shock $\Longrightarrow$ rank condition.

- Although ordering $\mathcal{Y}_t$ is arbitrary, the necessary and sufficient conditions imply the ordering is "unique" for the identifying restrictions (almost everywhere).

Identification of SVARs, II

- IV estimators require every instrument to be correlated with the right hand side variables $\Longrightarrow$ relevance condition analogous to RRWZ's rank condition.
- RRWZ's sufficient condition for global identification of a SVAR is satisfied when there is a *measurable* response by the n elements of $\mathcal{Y}_t$ to the j+1st structural shock in η_t given this holds for jth shock, $j = 1, \ldots, n$.
- These responses place restrictions on $\mathbf{A}_0$ of the form $\mathbf{R}_j\mathbf{A}_0'\iota_j = 0$, where $\mathbf{R}_j$ is a $n \times n$ matrix and ι_j is a $n \times 1$ vector with a one in its jth position and zeros otherwise $\Longrightarrow$ $\mathbf{R}_j$ contains linear restrictions on the columns of $\mathbf{A}_0'$ w/r/t $\eta_{j,t}$ and ι_j picks off these responses.
- Form $\mathbf{M}_j$ by stacking $\mathbf{R}_j\mathbf{A}_0'$ on top of $\left[\mathbf{I}_{j\times j} \;\; \mathbf{0}_{j\times(n-j)}\right]$, $j = 1, \ldots, n$.
- Global identification of a SVAR rests on
  1. necessary condition, $\sum_{j=1}^{n} r_j \geq$ number of free parameters, $0.5n(n-1)$, where $\text{rank}\left(\mathbf{R}_j\right) = r_j$, $\Longrightarrow$ number of restrictions embedded in $\mathbf{R}_j$.
  2. and the $\text{rank}\left(\mathbf{M}_j\right) = n$ for $j = 1, \ldots, n \Longrightarrow$ sufficient conditions.
  3. The sequence of $r_1 \geq r_2 \geq \ldots \geq r_n$ is available because the ordering of $\mathcal{Y}_t$ is arbitrary under linear restrictions.

An Example of a SVAR Identified on SR Restrictions

- Suppose a SVAR is estimated on $\mathcal{Y}_t = [\Delta y_t \;\; \Delta\pi_t \;\; (r_t - \pi_t)]'$ using a short-run recursive identification on $\mathbf{A}_0 = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

- Since $\dim(\mathcal{Y}_t) = 3$, there are $0.5(n^2 - n) = 3$ short-run restrictions $\Longrightarrow$ will satisfy the order condition for identification with the upper triangle of $\mathbf{A}_0$ full of zeros.

- The sufficient conditions involve the linear restrictions $\mathbf{R}_j\mathbf{A}_0'$.

  1. which appear in $\mathbf{R}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, $\mathbf{R}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and $\mathbf{R}_3 = \mathbf{0}_{3\times 3}$.

  2. $\Longrightarrow \mathbf{M}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $\mathbf{M}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, and $\mathbf{M}_3 = \begin{bmatrix} \mathbf{0}_{3\times 3} \\ \mathbf{I}_3 \end{bmatrix}$.

Necessary and Sufficient Conditions to Identify a SVAR on SR Restrictions

- The $\text{rank}(\mathbf{R}_1) = 2$, $\text{rank}(\mathbf{R}_2) = 1$, $\text{rank}(\mathbf{R}_3) = 0 \Longrightarrow \text{rank}(\mathbf{R}_1) > \text{rank}(\mathbf{R}_2) > \text{rank}(\mathbf{R}_3)$, which satisfies the order condition.
- Sufficient condition is satisfied because $\text{rank}(\mathbf{M}_1) = \text{rank}(\mathbf{M}_2) = \text{rank}(\mathbf{M}_3) = 3$.
- Do the results change for $\mathbf{A}_0 = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$?
- In this case, $\mathbf{R}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, $\mathbf{R}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and $\mathbf{R}_3 = \mathbf{0}_{3\times 3}$, which gives

$$\mathbf{M}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{M}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \text{and } \mathbf{M}_3 = \begin{bmatrix} \mathbf{0}_{3\times 3} \\ \mathbf{I}_3 \end{bmatrix}.$$

- Third column of $\mathbf{M}_2$ full of zeros $\Longrightarrow \text{rank}(\mathbf{M}_2) = 2$ and SVAR is not globally identified.

Identification of SVARs, III

- RRWZ's necessary and sufficient conditions for global identification are germane for linear restrictions on $\mathbb{A}$, (*i.e.*, lagged SVAR coefficients) and nonlinear restrictions on IRFs $\Longrightarrow$ long-run restrictions.

- Linear restrictions are imposed on $\mathbf{A}_0$ and $\mathbb{A}$ by $\mathbf{R}_j\boldsymbol{\mathcal{A}}\iota_j = 0$, where $\boldsymbol{\mathcal{A}}$ stacks $\mathbf{A}_0'$ on top of $\mathbb{A}'$ $\Longrightarrow$ form the $\mathbf{M}_j$s and check the necessary and sufficient conditions for global identification.

- Linear restrictions on IRFs are nonlinear restrictions on $\mathbf{A}_0$ and $\mathbb{A}$.
  1. Define $\mathbf{IRF}_\infty \equiv \left[\mathbf{A}_0 - \sum_{\ell=1}^{p}\mathbf{A}_\ell\right]^{-1}$ and $\mathbf{IRF}_{1:n,\,1:n}(h) \equiv \left[\mathbf{F}^h\right]_{1:n,\,1:n}$.
  2. $\Longrightarrow$ $\mathbf{R}_j$ can place linear restrictions on a $\mathbf{IRF}_\infty$ (*i.e.*, LR identifying restrictions), $\mathbf{IRF}_{1:n,\,1:n}(h)$ (*i.e.*, identifying restrictions at forecast horizon h), a matrix that stacks $\mathbf{A}_0$ on top of $\mathbf{IRF}_\infty$ or stacking $\mathbf{IRF}_{1:n,\,1:n}(h)$ beneath $\mathbf{A}_0$ $\Longrightarrow$ can mix SR, medium-run, and LR identifying restrictions.
  3. Ask if the resulting $\mathbf{M}_j$ matrices satisfy the RRWZ necessary and sufficient conditions.

- Check necessary and sufficient conditions before estimating a SVAR.

Identification of SVARs, IV: RRWZ's Other Results

- **Partial identification**: SVARs are often used to recover only the IRFs of y_t to $\eta_{j,t}$ $\Longrightarrow$ responses to an identified monetary policy shock.
  1. Global identification of only one structural shock or a subset of shocks.
  2. $\Longrightarrow$ "A subset of equations is identified if each equation in the subset is identified." (RRWZ, p. 674)

- **Exact identification**: RRWZ's rank condition is necessary and sufficient for global identification of just-identified SVARs *iff*
  1. the weak inequality restrictions on the number of identifying restrictions in $\mathbf{R}_j$, $r_1 \geq r_2 \geq \ldots \geq r_n$, become
  2. strong inequality restrictions, $r_1 > r_2 > \ldots > r_n$, where $r_j = n-j$, which satisfies Rothenberg's order condition $\Longrightarrow$ previous recursive just-identified example.

- **Local identification**: Suppose $\mathbf{A}_0 = \begin{bmatrix} a_{11,0} & 0 & a_{13,0} \\ a_{21,0} & a_{22,0} & 0 \\ 0 & a_{32,0} & a_{33,0} \end{bmatrix} \Longrightarrow r_1 = r_2 = r_3 = 1.$
  1. $\Longrightarrow$ Satisfies Rothenberg's order condition $0.5n(n-1) = 3$.
  2. Let $a_{11,0} = a_{22,0} = a_{33,0} = 1$ and $a_{13,0} = a_{21,0} = a_{32,0} = 2$, which is equivalent to $\tilde{a}_{11,0} = \tilde{a}_{22,0} = \tilde{a}_{33,0} = 2$ and $\tilde{a}_{13,0} = \tilde{a}_{21,0} = \tilde{a}_{32,0} = 1$
  $$\Longrightarrow \mathbf{A}_0 = \mathbf{P}\tilde{\mathbf{A}}_0, \text{ where } \mathbf{P} = \begin{bmatrix} 2/3 & 2/3 & -1/3 \\ -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \end{bmatrix}.$$
  3. $\mathbf{A}_0$ is locally identified, but not globally.

SVARs and Long Run Identification: The Problems

- The BQ decomposition relies on assumptions that are not innocuous.
  1. Assume estimation of long run economic relationships is possible $\Longrightarrow$ as $j \longrightarrow \infty$, estimate $\mathbf{E}_t y_{t+j}$ precisely on a finite span of data.
  2. The LR identification is assumed to separate transitory shocks from permanent shocks $\Longrightarrow$ the lack of infinitely long samples suggests the potential for the BQ decomposition to confound the LRN hypothesis.

- Blanchard and Quah (AER, 1989) have a theorem giving necessary and sufficient conditions under which a bivariate AR ($n = 2$) identifies the correct responses to a permanent shock and a transitory shock when truth is that there are several of these shocks.

- Faust and Leeper (JBES, 1997) extend BQ's theorem to account for
  1. uncertainty surrounding estimates of $\mathbf{C}(\mathbf{1})$ and hypothesis tests of $\mathbf{\Gamma}(\mathbf{1})_{i,j} = 0$ and
  2. temporal and cross-section aggregation of data.

SVARs and Long Run Identification: Faust and Leeper (JBES, 1997)

- Faust and Leeper include several propositions that provide a framework for analyzing the BQ decomposition.

- **Proposition 1**: The hypothesis $\mathbf{\Gamma}(\mathbf{1})_{i,j} = 0$ has no power (*i.e.*, type II error) $\Longrightarrow$ if the size of the test (*i.e.*, type I error) is $x\%$, the power is $\leq x\%$.
  1. When LRN is false, its rejection rate $\leq x\% \Longrightarrow$ the test has no power.
  2. This is frequentist analysis, but it indicates that in small sample the potential for biased inference is severe $\Longrightarrow$ $\mathbf{C}(\mathbf{1})$ is estimated imprecisely, which suggests SVAR identified by LR restrictions need priors that reflect this uncertainty.

- **Proposition 2**: Restricts analysis to bivariate ARs, but there are more than $n = 2$ fundamental shocks.
  1. Truth is $\mathcal{Y}_t = \sum_{\ell=1}^{\infty} \widetilde{\mathbf{C}}_\ell \tilde{\varepsilon}_t$, where $\mathcal{Y}_t$ is $n\times 1$, $\widetilde{\mathbf{C}}_\ell$ is $n\times\tilde{n}$, and $\tilde{\varepsilon}_t$ is $\tilde{n}\times 1$.
  2. Researcher estimates $\mathcal{Y}_t = \sum_{\ell=1}^{\infty} \mathbf{C}_\ell \varepsilon_t$, where $\mathbf{C}_\ell$ is $n\times n$, and ε_t is $n\times 1$.
  3. These VMA(∞) $\Longrightarrow \sum_{\ell=1}^{\infty} \widetilde{\mathbf{C}}_\ell = \sum_{\ell=1}^{\infty} \mathbf{C}_\ell \mathbf{W}_\ell$, where $\mathbf{W}_\ell$ is $n\times\tilde{n}$.
  4. $\Longrightarrow \mathbf{W}_\ell$ has to be block diagonal to separate transitory and permanent shocks (given $\mathbf{\Gamma}(\mathbf{1})$ is lower triangular).
  5. $\Longrightarrow$ The transitory components of $\tilde{\varepsilon}_t$ from a linear combination that is the transitory component of ε_t and same for the permanent shocks.

Sign Restrictions and Identification of SVARs

- Signs restrictions give potential inferences on a SVAR, but do not identify the SVAR. *"an assumption about signs is not enough by itself to identify structural parameters . . . procedure . . . delivers is a set of possible inferences, each of which is equally consistent with both the observed data and the underlying restrictions," Baumeister and Hamilton (Econometrica, 2015, p. 1963).*

- Baumeister and Hamilton argue the standard approach to estimating sign restricted SVARs is Bayesian, Giacomini and Kitagawa (2018) make similar arguments, but analysts often do not state their priors.

- Well known standard methods for estimating SVARs on sign restrictions are only consistent with the Haar prior; see Rubio-Ramírez, Waggoner, and Zha (REStud, 2010)
    1. Haar prior: Related to the Haar measure that "assigns an 'invariant volume' to subsets of locally compact topological groups, consequently defining an integral for functions on those groups."
    2. Loosely, the idea is that the measure of a closed set of functions cannot be changed as more functions are added.
    3. An example is the Jeffreys prior, which is non-informative and proportional to the square root of determinant of Fisher's information (*i.e.*, the covariance matrix of the IRFs implied by the sign restrictions).

- The problem is the Haar prior can lead to posteriors that are identical to the prior unless subjective beliefs are included in the prior.

Sign Restrictions, Priors, and Posteriors

- Focus is on **AB**-class SVARs in Baumeister and Hamilton conditional on the true covariance matrix, $\mathbf{\Omega}$.

- Baumeister and Hamilton show asymptotically $\mathbf{A}_0$ and $\mathbf{Q}$ reside in a set defined by $\mathbf{\Omega}$ with probability $\longrightarrow 1$.
  1. The posterior of the SVAR is in this set, but the posterior of $\mathbf{A}_0$ is proportional to $g(\boldsymbol{a}_0)$, where $g(\boldsymbol{a}_0)$ is the prior on $\mathbf{A}_0$ and the constant of proportionality normalizies $g(\boldsymbol{a}_0 \,|\, \mathbf{Y})$, which is the posterior of $\mathbf{A}_0$ conditional only on the data.
  2. The standard method of estimating SVARs on sign restrictions yields a posterior that is (dominated by) the prior.

- The Haar prior has odd effects on the implicit priors of the SVAR.
  1. $\dim(y_t)$ affects location of structural shocks in the implicit prior distribution.
  2. In many cases, implicit priors on $\boldsymbol{a}_0$ are Cauchy $\Longrightarrow$ prior on a (log) probability.

- The standard method for generating IRFs under sign restrictions reflect the set valued prior, not the data, and not the ordering.

- Baumeister and Hamilton suggest mixing sign restrictions with priors on the impact, slope, and volatility coefficients of a SVAR $\Longrightarrow$ Sims and Zha prior along with priors as dummy observations.

Fundamentals, Misspecification, and Identification of SVARs

- A critique of SVARs is that linearized DSGE models do not predict VARs with finite lags except in a few special cases.
- Instead, the reduced form of linearized DSGE models predict vector autoregressive-moving averages $\Longrightarrow$ VARMA(p, q) models.
- There is a claim that this invalidates using SVARs to study business cycle fluctuations and conducting policy evaluation $\Longrightarrow$ SVARs do not recover fundamental shocks, which are found in DSGE models.
- This critique of SVAR identification conflates several disparate issues.
  1. Is the claim that SVARs can never recover fundamental shocks?
  2. Is the claim that SVARs estimators fail to control for expectation formation by economic agents, which produce biased estimates of fundamental shocks?
  3. Or is it SVAR specifications are fragile $\Longrightarrow$ change the elements of y_t and alter the identification of the SVAR.

Fundamentals and SVARs

- Lippi and Reichlin (JofE, 1994) combine two seemingly unrelated ideas.
  1. The reduced form of a linearized DSGE model is a VARMA$(p,\, q)$.
  2. VARMA$(p,\, q)$s often are associated with VMA(∞)s that are not fundamental.

- Let $\mathcal{Y}_t = \mathbf{C}(\mathbf{L})\varepsilon_t$, where the matrix polynomial $\mathbf{C}(\mathbf{L}) = \sum_{j=0}^{\infty} \mathbf{C}_j$, $\mathbf{C}_0 = \mathbf{I}$, $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \Omega)$, and $\mathbf{C}(\mathbf{L})$ has roots with modulus on or outside the unit circle.
  1. ε_t is fundamental *iff* $\mathbf{C}(\mathbf{L})$ has roots with modulus outside the unit circle.
  2. $\Longrightarrow$ the history $\mathcal{Y}_{t-j}$, $j = 0, 1, 2, \dots$, recovers $\varepsilon_t \Longrightarrow$ these shocks reside in a *linear* space spanned by the history of $\mathcal{Y}_t$.
  3. Otherwise, $\mathbf{C}(\mathbf{L})$ has roots with modulus on the unit circle $\Longrightarrow \varepsilon_t$ is not fundamental.

- **Fundamentalness** of a MA(1): Consider $x_t = e_t - \alpha e_{t-1}$, $e_t \sim iid\mathcal{N}\left(0,\, \sigma_e^2\right)$.
  1. Suppose $\alpha \geq 1 \Longrightarrow \alpha$ can be estimated using MLE, but e_t cannot be recovered from the history of $x_t \Longrightarrow e_t$ is not a fundamental shock for x_t.
  2. x_t is forward-looking in $e_t \Longrightarrow (1 - \alpha\mathbf{L}) = \left(\frac{1}{\alpha}\mathbf{L}^{-1} - 1\right)\alpha\mathbf{L} = -\left(1 - \frac{1}{\alpha}\mathbf{L}^{-1}\right)\alpha\mathbf{L}$
  $\Longrightarrow \left(1 - \frac{1}{\alpha}\mathbf{L}^{-1}\right)^{-1} x_t = \alpha e_{t-1}$ if $\alpha \geq 1$.
  3. Define $\sigma_u^2 = \alpha^2\sigma_e^2 \Longrightarrow$ the MA(1) becomes $x_t = u_t - \dfrac{1}{\alpha}u_{t-1} \Longrightarrow 1\,/\alpha$ is a "discount factor" on lags of $x_t = \sum_{j=1}^{\infty} \alpha^{-j}x_{t-j} + u_t \Longrightarrow x_t$ is backward-looking in u_t and u_t is fundamental for x_t.

Lippi and Reichlin: Fundamentals, DSGE Models, and SVARs

- Linearized DSGE models predict $\mathbf{\Theta}(\mathbf{L})y_t = \mathbf{\Psi}(\mathbf{L})\eta_t$, where $\mathbf{\Theta}$, $\mathbf{\Psi}$, and η_t are
  1. finite-order matrix lag polynomials that are nonlinear functions of the DSGE model parameters and structural shocks with unit variances.
  2. The roots of $\mathbf{\Theta}$ and $\mathbf{\Psi}$ have modulus on or outside the unit circle.
  3. An econometrician's problem is that η_t is known only to the agents populating the DSGE model.

- An econometrician only observes y_t and its history.
  1. $\Longrightarrow$ no a priori reasons for the fundamentalness of $y_t = \mathbf{\Theta}(\mathbf{L})^{-1}\mathbf{\Psi}(\mathbf{L})\eta_t$.
  2. $\Longrightarrow$ DSGE theory does not give necessary and/or sufficient conditions that predict $\mathbf{\Theta}(\mathbf{L})$ has roots strictly outside the unit circle.
  3. $\Longrightarrow$ η_t is not fundamental for y_t.

- Consider $[\mathbf{A}_0 - \mathbf{A}(\mathbf{L})]y_t = \eta_t$, where $y_t \sim I(1)$ and $\eta_t \sim iid\mathcal{N}(\mathbf{0}, \mathbf{I})$.
  1. The SVMA(∞) is $y_t = [\mathbf{I} - \mathbf{A}_0^{-1}\mathbf{A}(\mathbf{L})]^{-1}\mathbf{A}_0^{-1}\eta_t$, which exists *iff* $[\mathbf{I} - \mathbf{A}_0^{-1}\mathbf{A}(\mathbf{L})]^{-1}\mathbf{A}_0^{-1}$ has roots outside the unit circle.
  2. Since the SVAR is the economic model, an econometrician observes the same information as the economic agents implied by the SVAR.
  3. An implication of SVARs that is stronger than the standard RE assumption implicit in DSGE models.
  4. Claim: The SVAR is an approximation to a VARMA, which is the reduced-form of a DSGE model.

Lippi and Reichlin: Fundamentalness and Blaschke Factors

- Consider $\mathcal{Y}_t = \mathbf{S}(\mathbf{L})\nu_t$, which is not fundamental.
  1. The space generated by the history of ν_t is not known to the econometrician.
  2. $\Longrightarrow \nu_t$ are not forecast errors, but $\mathcal{Y}_t = \mathbf{C}(\mathbf{L})\varepsilon_t = \mathbf{S}(\mathbf{L})\nu_t$.

- There is a mapping which creates a fundamental VMA(∞) using $\mathbf{S}(\mathbf{L})$.
  1. The mapping relies on the Blaschke matrix, $\mathbf{M}(\mathbf{L})$, which has no poles (*i.e.*, roots or eigenvalues) in and on the *complex* unit circle and $\mathbf{M}(\mathbf{z})^{-1} = \mathbf{M}^*(\mathbf{z}^{-1})$, where $\mathbf{M}^*(\cdot)$ is the conjugate transpose of $\mathbf{M}(\cdot)$.
  2. A leading example of a Blaschke matrix is $\mathbf{M}(\mathbf{z}) = \begin{bmatrix} \dfrac{\mathbf{z}-\alpha}{1-\overline{\alpha}\mathbf{z}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$, where $\alpha \in (-1, 1)$ and $\overline{\alpha}$ is its complex conjugate.
  3. A useful result is that $\nu_t = \mathbf{M}(\mathbf{L})^{-1}\varepsilon_t \Longrightarrow$ given $\varepsilon_t \sim \mathcal{WN}$, ν_t is as well.
  4. The upper left block of $\mathbf{M}^*(\mathbf{L}^{-1})$ is $\dfrac{\mathbf{L}^{-1}-\alpha}{1-\overline{\alpha}\mathbf{L}^{-1}} = \dfrac{1-\alpha\mathbf{L}}{1-\overline{\alpha}\mathbf{L}^{-1}}\mathbf{L}^{-1} \Longrightarrow \nu_t$ is backward-looking in and fundamental for ε_t.
  5. **Theorem 2**: $\mathbf{S}(\mathbf{L}) = \mathbf{C}(\mathbf{L})\mathbf{M}(\mathbf{L}) \Longrightarrow \mathcal{Y}_t = \mathbf{C}(\mathbf{L})\mathbf{M}(\mathbf{L})\nu_t$ or $\mathbf{C}(\mathbf{L})^{-1}\mathcal{Y}_t = \mathbf{M}(\mathbf{L})\nu_t$ $\Longrightarrow$ to recover ν_t need the history of $\mathcal{Y}_t$ and its future.
  6. However, $\mathcal{Y}_t = \mathbf{S}(\mathbf{L})\nu_t = \mathbf{S}(\mathbf{L})\mathbf{M}(\mathbf{L})^{-1}\mathbf{M}(\mathbf{L})\nu_t = \mathbf{C}(\mathbf{L})\varepsilon_t \Longrightarrow$ obtain an “estimate” of α to recover the fundamental VAR representation from $\mathbf{S}(\mathbf{L})$.
  7. Given $\mathbf{M}(\mathbf{L})$, compute $\widehat{\mathbf{C}}(\mathbf{L})$ and $\widehat{\varepsilon}_t$ using $\widehat{\mathbf{S}}(\mathbf{L})$ and $\widehat{\nu}_t$, which solves the MA invertibility problem by discounting future ν_t by α.

Fundamentalness and VARs

- Lippi and Reichlin locate the problem of non-fundamentalness of VARs in the non-invertible MA component of underlying VARMA.
- News or anticipated shocks are a source of non-fundamentalness in VARs $\Longrightarrow$ create non-invertible MA processes.
    1. Examples are news about technology innovations, lags in legislation that change fiscal policy, and forward guidance statements about monetary policy.
    2. See Barsky and Sims (2011. "News shocks and business cycles," *Journal of Monetary Economics* 58, 273–289) and Mertens and Ravn (2010, "Measuring the impact of fiscal policy in the face of anticipation: A structural VAR approach," *The Economic Journal* 120, 393–413).
- Leeper, Walker and Yang (2013) show non-fundamentalness is tied to anticipated shocks is a symptom of the structure that transmits this "news" into the economy.
    1. Blaschke matrices are a tool to model the information flow of anticipated shocks, (*i.e.*, the ν_ts), which are non-fundamental and is the source of the MA non-invertibility.
    2. They show that adding variables to $\mathcal{Y}_t$ which contain information about news shocks can in some cases negate the problem of non-fundamentalness.
    3. Also see Lütkepohl (2012, "Fundamental problems with nonfundamental shocks," Discussion Paper 1230, DIW-German Institute for Economic Research, Berlin, Germany).
- This suggests another source of non-fundamentalness in VARs is model misspecification w/r/t to which elements of $\mathcal{Y}_t$ should be included and excluded.

Fundamentalness and The ABCs and Ds of VARs

- DSGE models provide information about the variables to include $\mathcal{Y}_t$ $\Longrightarrow$ defines the shocks the SVAR meant to recover.

- This restricts the VAR in several ways.
  1. The $\dim(\mathcal{Y}_t)$ = the number of shocks driving the DSGE model.
  2. If this is true, the MA invertibility problem may still exist.
  3. However, DSGE models often contain dynamic predictions about more variables than there are structural shocks.
  4. This suggests there are combinations of variables defining a $\mathcal{Y}_t$ that yield fundamental VARs and others that do not.
  5. Some specifications of $\mathcal{Y}_t$ have VAR dynamics that are fundamental, but this is not true for all $\mathcal{Y}_t$ $\Longrightarrow$ omitted variables problem.

- Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007) develop an invertibility condition for the reduced form VAR of a linearized DSGE model that is conditional on $\mathcal{Y}_t$ $\Longrightarrow$ a necessary condition for a VAR specification to be fundamental.

The State Space of a Linearized DSGE model and a VAR(∞), I

- The solution of many linearized DSGE models has the state space representation

$$\begin{array}{rrcccc} \text{System of State Equations:} & \mathcal{S}_{t+1} & = & \boldsymbol{\mathcal{A}}\mathcal{S}_t & + & \boldsymbol{\mathcal{B}}\eta_t \\ \text{System of Observation Equations:} & \mathcal{Y}_t & = & \boldsymbol{\mathcal{C}}\mathcal{S}_t & + & \boldsymbol{\mathcal{D}}\eta_t \end{array}$$

  where $\mathcal{S}_t$ is a $m \times 1$ $(n \leq m)$ vector of endogenous and exogenous state variables some of which are unobserved and $\eta_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$.

- The system of observation equations imply $\eta_t = \boldsymbol{\mathcal{D}}^{-1}(\mathcal{Y}_t - \boldsymbol{\mathcal{C}}\mathcal{S}_t)$, given $\boldsymbol{\mathcal{D}}$ is full rank.

- Use the expression to substitute for η_t in the system of state equations to find $\mathcal{S}_{t+1} = (\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{\mathcal{C}})\mathcal{S}_t + \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\mathcal{Y}_t$ or $\mathcal{S}_{t+1} = \sum_{j=0}^{\infty}(\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{\mathcal{C}})^j\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\mathcal{Y}_{t-j}$.
  1. As long as $[\mathbf{I} - (\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{\mathcal{C}})\mathbf{L}]$ is invertible $\Longrightarrow$ has roots with modulus outside the unit circle, recover $\mathcal{S}_{t+1}$ from the history of $\mathcal{Y}_t$.
  2. $\Longrightarrow$ "Poor man's invertibility condition" is $\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{\mathcal{C}}$, which is a square summable sequence in the matrix power j.

- Lag this VMA one period and substitute for $\mathcal{S}_t$ in the system of observation equations

$$\mathcal{Y}_t = \boldsymbol{\mathcal{C}}\sum_{j=0}^{\infty}\left(\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{\mathcal{C}}\right)^j\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\mathcal{Y}_{t-1-j} + \boldsymbol{\mathcal{D}}\eta_t.$$

- The DSGE model has a VAR(∞) as its reduced form dynamic representation, given $\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{\mathcal{C}}$ has roots with modulus outside the unit circle.

The State Space of a Linearized DSGE model and a VAR(∞), II

- The invertibility restriction on $\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C}$, which is conditional on $\mathcal{D}^{-1}$, equates η_t with the forecast innovation of $\mathcal{Y}_t$.
- The system observation equations restrict the forecast innovation of $\mathcal{Y}_t$

$$\mathcal{Y}_{t+1} - \mathbf{E}_t \mathcal{Y}_{t+1} = \mathcal{C}\left(\mathcal{S}_{t+1} - \mathbf{E}_t \mathcal{S}_{t+1}\right) + \mathcal{D}\eta_{t+1}.$$

- $\Longrightarrow$ innovations in the state vector create a "wedge" between forecast innovations in $\mathcal{Y}_{t+1}$ and the structural errors η_{t+1}.
- Wedge disappears when the invertibility condition $\mathcal{A} - \mathcal{B}\mathcal{D}^{-1}\mathcal{C}$ is satisfied $\Longrightarrow$ $\mathcal{S}_{t+1} - \mathbf{E}_t \mathcal{S}_{t+1}$ is in the linear space spanned by the history of $\mathcal{Y}_{t+1}$.
- Searching for a specification of $\mathcal{Y}_t$ that equates its forecast errors with the structural errors of interest is another way to solve the problem of VAR fundamentalness.

The State Space of a Linearized DSGE model and a VAR(∞), III

- The notion the specification of $\mathcal{Y}_t$ matters for identification of a SVAR has other implications.

- The identification issue is whether η_t is in the space spanned by $\mathcal{Y}_{t-j}$, $j \geq 0$ $\Longrightarrow$ the history of $\mathcal{Y}_t$.
    1. The first order issues are the choice of $\mathcal{Y}_t$ and restrictions imposed on $\mathbf{A}_0$ (and $\mathbf{Q}$ in the $\mathbf{AB}$-model) $\Longrightarrow$ goal is to recover structural shocks that map into macro theory (*i.e.*, DSGE model shocks).
    2. $\Longrightarrow$ Claims the true data generating process is not a VAR misses the point.

- Same is true for estimators $\Longrightarrow$ the choice of the estimator does not alter the identification problem of recovering structural errors in time series models.

- The exception is the research question can drive the choice of estimator.
    1. If the question is whether there are shifts in policy regimes, which drive responses by economic agents, estimators need to include regime switching or time-varying parameters to "test" the hypothesis $\Longrightarrow$ Markov-switching (MS) and/or time-varying parameter (TVP) BVARs.
    2. SVARs also have difficulties recovering latent elements of the state vector S_t (*i.e.*, level of TFP, inflation expectations or output, inflation, and unemployment gaps) $\Longrightarrow$ apply ML estimators to unobserved components models.