

SVAR IDENTIFICATION THROUGH HETEROSKEDASTICITY WITH MISSPECIFIED REGIMES

CHRISTOPHER A. SIMS

ABSTRACT. It is well known that if the relative variances of structural shocks change across time spans we label “regimes”, then the coefficients of a structural VAR (SVAR) are identified. If we assume that relative variances are constant within regimes, but in fact they change within as well as across regimes, possibly because the dating of the regimes is inaccurate, the coefficients are nonetheless usually estimated consistently.

I. IDENTIFICATION THROUGH HETEROSKEDASTICITY

A structural VAR model for an n -dimensional vector time series y_t is written

$$A(L)y_t = \varepsilon_t, \quad (1)$$

where A is a finite-order matrix polynomial in non-negative powers of the lag operator and ε_t is a vector of structural shocks, independent (or at least uncorrelated) across time and with a diagonal covariance matrix Λ_t at each date t . If A_0 is non-singular, we can multiply the system through by A_0^{-1} to obtain its reduced form representation

$$B(L)y_t = \eta_t. \quad (2)$$

where

$$\text{Var}(\eta_t) = \Sigma_t = A_0^{-1} \Lambda_t (A_0')^{-1}. \quad (3)$$

If $\Lambda_t \equiv I$, so there is no time variation in Σ_t , the reduced form model, because $B_0 = I$ by construction, has $(n^2 - n)/2$ fewer parameters than the original SVAR form (1). Identification through heteroskedasticity recognizes that if Λ_t varies enough over time, it may be possible to identify all the parameters of (1) without restrictions on $A(L)$.

In the simplest case, data can be split into two “regimes”, one period in which $\Lambda_t \equiv \Lambda^0$ and one in which it is Λ^1 . So long as $A(L)$ is constant across the regimes, $B(L)$ will also be constant across the regimes. If for each regime we estimate $B(L)$ by maximum likelihood under Gaussianity assumptions or by least squares, we will arrive at the maximum likelihood estimates within each regime. It is a standard “seemingly unrelated regressions” result that we can get more efficient estimates of

Date: July 5, 2020.

This document is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

$B(L)$ by simply running OLS on the whole sample, since with the same variables on the right-hand side of every equation, accounting for the cross-correlation of residuals has no effect on the Gaussian MLE. Regardless of how the residual covariance matrix behaves or of whether the residuals are normally distributed, $B(L)$, estimated by OLS, will converge as sample size increases to the best linear predictor of y given y 's history. As the number of observations within each regime gets large, therefore,

$$\hat{\Sigma}^0 \rightarrow A_0^{-1}\Lambda^0(A_0^{-1})' \quad \text{and} \quad \hat{\Sigma}^1 \rightarrow A_0^{-1}\Lambda^1(A_0^{-1})', \quad (4)$$

where the “ $\hat{\cdot}$ ” indicates an ML estimate under Gaussian assumptions. But

$$((\Sigma^0)')^{-1}\Sigma_1 = A_0'(\Lambda^0)^{-1}\Lambda^1(A_0')^{-1}. \quad (5)$$

which implies that if all the diagonal elements λ_i^1/λ_i^0 of $(\Lambda^0)^{-1}\Lambda^1$ are distinct, The right eigenvectors of the product in (5) are the columns of A_0' . This determines A_0 up to scale factors on those columns and the ordering of the columns. Thus it is clear that $A(L)$ can be estimated consistently under standard assumptions, so long as it is constant across regimes and the ratios of structural shock variances across regimes are all distinct.

II. CONSEQUENCES OF MISSPECIFIED REGIMES

What if we model the data as generated by two regimes with Λ^0 and Λ^1 constant within each regime, as in the previous section, but in fact the first regime actually has two distinct sub-regimes, with Λ^3 and Λ^4 prevailing in the two subregimes? Then the estimated reduced form residual covariance matrix for the first regime will converge in large samples to

$$A_0^{-1}(\Lambda^3\pi + \Lambda^4(1 - \pi))(A_0')^{-1}, \quad (6)$$

where π is the proportion of the first regime in which Λ^3 prevails. But then it will still be true that $(\hat{\Sigma}^0)^{-1}\hat{\Sigma}^1$ converges to a limit that has the columns of A_0' as right eigenvectors, so $A(L)$ will still be consistently estimated by the same procedures we used for a constant Λ^0 .

There is likely to be a loss of efficiency from mis-specifying regimes, because the averaging of diagonal λ_i values from distinct regimes reduces their variability. In a given sample, though, if Λ^3 and Λ^4 , say, differed only slightly, while both were quite different from Λ^1 , The increased precision in the estimation of Σ^0 from assuming constant Λ^0 might offset the gains from recognizing $\Lambda^3 \neq \Lambda^4$.

As a special case of this general argument, errors in specifying the ends or beginning of regimes do not necessarily undermine identification, as they simply result in regimes with non-constant Λ . Of course if the regimes are very badly specified, they might end up with little cross-regime variation in $\hat{\Sigma}$, despite large within-regime variation and thus lead to inaccurate or even inconsistent estimates.

III. t -DISTRIBUTED ERRORS

Brunnermeier, Palia, Sastry, and Sims (2018) use identification through heteroskedasticity in a model where they also assume structural shocks are distributed as mutually independent univariate t . A perhaps surprising fact is that if this assumption on the distribution of residuals is correct, identification is formally possible without specifying any regimes at all. This is true because a vector of independent univariate t random variables does not retain that distribution if linearly transformed, even by an orthonormal matrix. The individual elements of the ε_t vector have a more dispersed distribution than any non-trivial linear combination of them. Likelihood based estimation using an assumption of independent univariate t shocks will look for A_0 matrices that make large shocks occur in isolation, in individual ε_{it} 's, rather than in several ε_{it} values at the same t .

Another approach to identification that has similarities to use of the independent t assumption is a narrative approach that labels particular dates as likely to have been dominated by particular structural shocks, and in effect chooses A_0 to make individual large ε_{it} values line up with those dates.

However in the t -errors case we can no longer use the argument in the previous section that with a Gaussian-error specification, variation in Λ within regimes leaves estimates consistent even though constant- Λ within regimes is assumed. That argument relied on the specific structure of the Gaussian likelihood.

Suppose we use the independent- t likelihood, even though this is not the true distribution of the structural shocks. So long as the shocks are in fact independent of each other and of past values of y , each distributed symmetrically around zero, and fat-tailed, estimates of the system dynamics are in general consistent. Consistency will in the absence of further restrictions fail because of lack of identification, of course, when the shocks have a joint normal distribution. Estimates will be inconsistent if the structural shocks have distributions thinner-tailed than the normal. In what follows we sketch arguments that support these claims.

Under the symmetrically distributed errors assumption, the reduced form coefficients, and hence individual reduced form shocks η_{it} are consistently estimated. The log likelihood under the independent- t distribution with degrees of freedom ν is

$$-\frac{\nu+1}{2} \sum_{i,t} \log \left(1 + \frac{(A_i \eta_t(B))^2}{\nu} \right) + T \log |A|, \quad (7)$$

where A_i is the i 'th row of A_0 in (1) and B represents all the coefficients in $B(L)$ from (2). The first-order conditions with respect to B for a maximum of this likelihood are

$$-\frac{\nu+1}{\nu} \sum_{i,t} \frac{A_i \eta_t \frac{\partial \eta_t}{\partial B}}{1 + \frac{1}{\nu} (A_i \eta_t)^2} = 0. \quad (8)$$

The $\partial\eta_t/\partial B$ terms in this expression are just lagged dependent variables, which we assume are independent of the true reduced form shocks. Because of the symmetry assumption,

$$E \left[\frac{A_i\eta_t(B) \frac{\partial\eta}{\partial B}}{1 + \frac{1}{\nu}(A_i\eta_t(B))^2} \right] = 0 \quad (9)$$

when evaluated at the true value of the reduced form coefficients B , regardless of what values the A_i coefficients are set at. At other value of B , the residuals $\eta_t(B)$ would be correlated with lagged data, so (8) would fail. In other words, using this likelihood function, together with our symmetry independence assumptions, implies consistency of maximum likelihood estimates of B by standard GMM arguments.

Now what about A_i . We derive the FOC's with respect to the elements of A_i , with the results for different i values stacked up. We evaluate the expression with B at its true value, letting us use η_t without its B argument in the FOC expression, The result is

$$-\frac{\nu+1}{\nu} \sum_t \Lambda_t A \eta_t \eta_t' = T(A')^{-1}, \quad (10)$$

where Λ_t is a diagonal matrix with typical diagonal element

$$\frac{1}{1 + \frac{1}{\nu}(A_i\eta_t)^2} \quad (11)$$

and A (without subscript) is the matrix formed by stacking the A_i vectors. When A is set to the true value, the A_0 coefficient of the $A(L)$ polynomial in (1), $A\eta_t = \varepsilon_t$, so multiplying (11) on the right by A' , we can write it as

$$\left[-\frac{\nu+1}{\nu} \frac{1}{T} \sum_{i,t} \frac{\varepsilon_{it}\varepsilon_{jt}}{(1+\varepsilon_{it}^2/\nu)(1+\varepsilon_{jt}^2/\nu)} \right] = I. \quad (12)$$

With the structural shocks independent across i and t and symmetrically distributed about zero, the expectation of the elements in the sum on the left-hand-side of this equality is a diagonal matrix, — not in general an identity — matrix. It is a scalar matrix if the ε_{it} 's are in fact independent t variates with ν degrees of freedom. In that case it will be the identity if we have normalized $A(L)$ by scaling it so that the matrix is the identity.

But if all we know about the distribution of the structural shocks is that they are symmetrically distributed and independent, the solution to (11) is not a simple scalar multiple of the true A_0 . It is, though, of the form MA_0 , where M is a diagonal matrix. So by rescaling the rows of A by premultiplying it by a diagonal matrix Ψ , we are likely to be able to find an $A^* = \Psi A_0$ that solves (11).¹ But multiplying A_0 by a diagonal matrix only changes the implied variances of the structural shocks. In

¹Likely, but not certainly. The expectation of the left-hand side of (11) is bounded above as we scale rows of A , while the right-hand side is not, so it is possible that the equation has no solution.

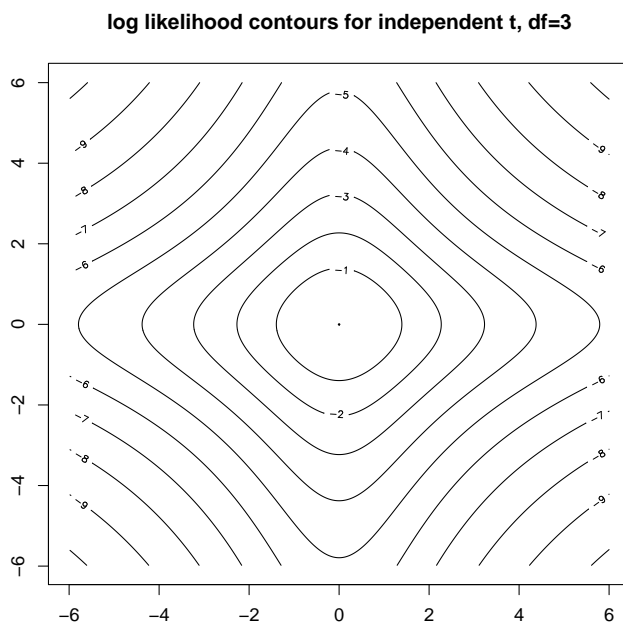


FIGURE 1

other words, if we plot impulse responses of the system to structural shocks, this kind of distortion in A only changes the scale of the impulse responses to any given structural shock. This could distort estimated variance decompositions, but it does not distort the shape of the time paths in the impulse responses or the relative sizes of responses to a given shock.

We know that $A(L)$ is not identified without further restrictions if the ε 's are normally distributed. Normal shocks are certainly symmetrically distributed. And indeed the FOC's for the MLE are satisfied under the t likelihood when the truth is that shocks are normal. The inconsistency arises because in the Gaussian case they are also satisfied under any orthogonal rotation of the A_0 matrix. We may get consistency in non-Gaussian cases, because for non-Gaussian ε , orthonormal rotations of ε generally change the expected value of the left-hand side of (12). This produces identification when the distributions of the ε_{it} 's are fat-tailed, but lack of identification if they are normal and inconsistency if they are thinner-tailed than the normal distribution.

To illustrate how this works, Figure 1 shows a contour plot of the log likelihood element for a pair of independent t variates with 3 degrees of freedom. It is clear that the likelihood declines more slowly along the axes than along other rays through zero. Therefore when ε_{1t} and ε_{2t} are independent and have fat-tailed distributions (even if not $t(3)$), multiplying the ε vector by an orthonormal matrix, which would

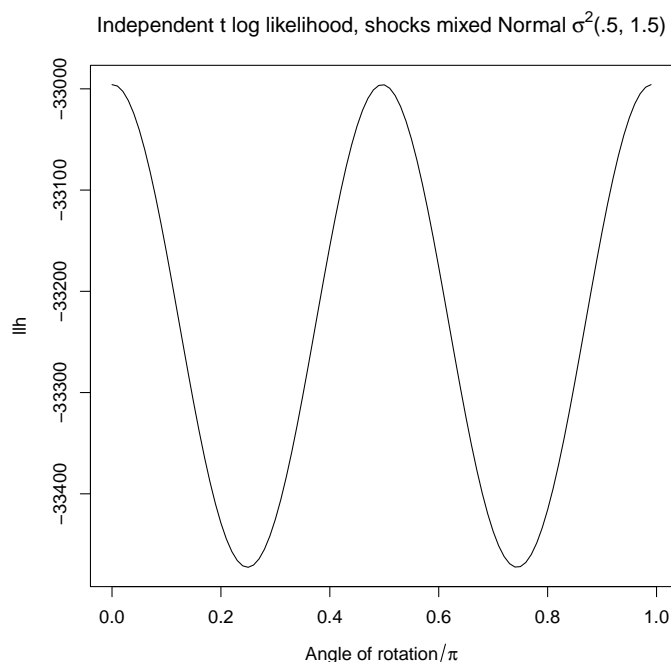


FIGURE 2

just rotate the scatter of (x, y) points on this likelihood plot, must reduce the likelihood. Of course as the degrees of freedom in the t likelihood increase, the shape of these contours converges to the circular shape of a $N(0, I)$ pdf, so likelihood is unchanged by orthonormal transforms of the data. And distributions of ε thinner-tailed than the normal would generate inconsistency of t -likelihood-based estimation. We illustrate this with Figures 2 and 3. They show how likelihood changes as A is rotated, by orthonormal transformation, away from its true value. In Figure 2, the data is drawn from two independent mixed normal distributions, which makes it fatter-tailed than a normal, and one can see that likelihood is maximized at $0, \pi/2$, and π . These rotation angles correspond to simply permuting the rows of A , which does not affect the economic interpretation of the model. In Figure 3, the data are drawn from independent uniform distributions. Here the true A solves the FOC's of likelihood maximization by being a local *minimum* of the likelihood.²

²Figures 2 and 3 were generated by creating a 10,000 by 2 array of draws from the mixed-normal or the uniform distribution, respectively. Then for each value of the rotation angle θ , the data were rotated by that angle and the sum of $t(5.7)$ log likelihood elements was calculated. This is Monte Carlo integration, of course, so it is important that the range of variation in the log likelihood shown is large relative to the Monte Carlo standard errors. With data generated by draws from a $t(40)$ pdf, a sample of this size is not enough to accurately estimate the true pattern of variation in expected log likelihood with θ , because the variation is very small.

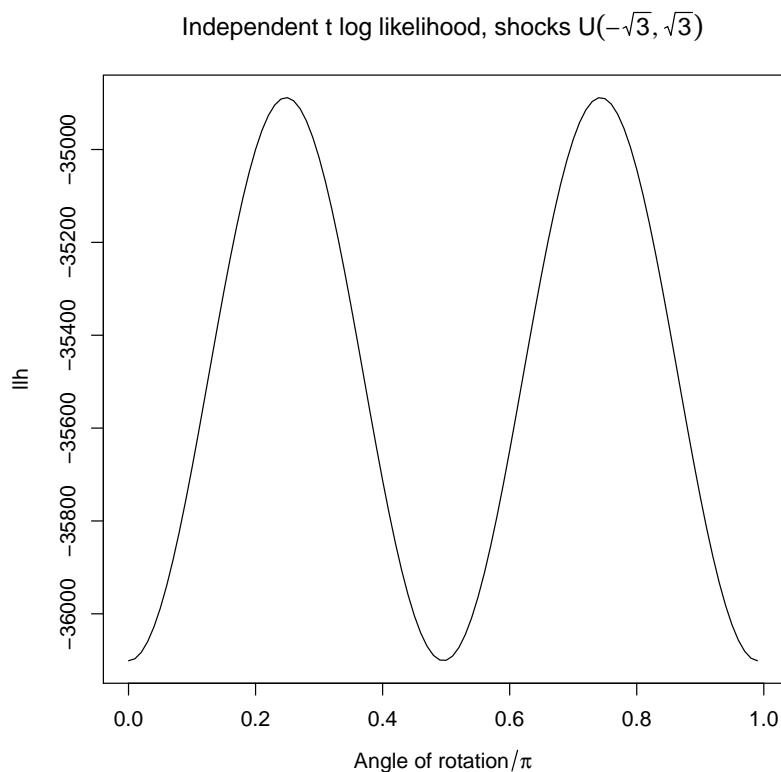


FIGURE 3

IV. REGIME SHIFTS AND t -LIKELIHOOD

Identification through regime shifts in relative variances and through fat-tailed distributions interacting with t likelihood reinforce each other. There is a possible downside to the combination, however, because with a thin-tailed true distribution for the structural shocks, the use of the t likelihood could undermine consistency, even though with the Gaussian likelihood the regime shifts would give consistent estimates. It is easy to check for this possibility, though. If the t assumption on the structural shock distributions is correct, the consistently estimated structural shocks based on the Gaussian likelihood should have a fat-tailed empirical distribution. If not, it would make sense to stick with the Gaussian-likelihood estimates.

REFERENCES

BRUNNERMEIER, M., D. PALIA, K. A. SASTRY, AND C. A. SIMS (2018): "Feedbacks: Financial Markets and Economic Activity," Discussion paper, Princeton University, <http://sims.princeton.edu/yftp/bpss/>.

DEPARTMENT OF ECONOMICS, PRINCETON UNIVERSITY
Email address: sims@princeton.edu