Asymptotic Distribution of Variance Decompositions

Anders Warne

February 27, 2008

ABSTRACT: This note discusses how to compute the asymptotic covariance matrix for a forecast error variance decomposition. The theory relies on having an estimate of the asymptotic covariance matrix for the impulse response function and on the variance of structural shocks being normalized to unity. The results apply to a wide range of identification schemes, including contemporaneous and long run restrictions.

KEYWORDS: Impulse Response Function, Variance Decomposition.

JEL CLASSIFICATION NUMBERS: C32.

1. SETUP

Let η_t be a p dimensional i.i.d. shock with zero mean and identity covariance matrix. The impulse response function for some p dimensional variable vector x at horizon h after a unit shock to η is assumed to be respresented by:

$$\text{resp}\big(x_{t+h}|\eta_t = I_p, \eta_{t+1} = 0, \ldots, \eta_{t+h} = 0\big) = R_h, \quad h = 0, 1, \ldots, \tag{1}$$

where R_h is a $p \times p$ matrix. In other words, it is assumed that the impulse response function is invariant with respect to t, the point in time when the shock occurs. In linear models, such as VAR models with or without cointegration relations, the impulse responses satisfy this assumption.

Furthermore, it is assumed that $R_h = R_h(\theta)$ is a differentiable function of the vector $\theta \in \mathbb{R}_k$, and that we have an estimator of θ for a sample size of T observations ($\hat{\theta}$) which satisfies:

$$\sqrt{T}\big(\hat{\theta} - \theta\big) \xrightarrow{\text{d}} N_k\big(0, \Sigma_\theta\big), \tag{2}$$

with N_k being a k-dimensional Gaussian distribution, $\xrightarrow{\text{d}}$ denoting convergence in distribution, and Σ_θ being positive semidefinite. Models which satisfy these assumptions include VARs with an upper bound for the lag length, Gaussian error terms, and variables that are integrated of order d (with d being some finite integer) and potentially cointegrated. Given such assumptions it follows that the estimated impulse response function $\hat{R}_h = R_h(\hat{\theta})$ satisfies

$$\sqrt{T}\big(\text{vec}\big(\hat{R}_h\big) - \text{vec}\big(R_h\big)\big) \xrightarrow{\text{d}} N_{p^2}\big(0, \Sigma_{R_h}\big), \tag{3}$$

where vec is the column stacking operator, and

$$\Sigma_{R_h} = \frac{\partial \text{vec}(R_h)}{\partial \theta'} \Sigma_\theta \left(\frac{\partial \text{vec}(R_h)}{\partial \theta'} \right)'.$$

The interested reader is referred to Lütkepohl (1990), Lütkepohl and Reimers (1992), Warne (1993), Vlaar (2004), and references therein, for examples of parameterizations of the partial derivatives $\partial \text{vec}(R_h) / \partial \theta'$.

---

2. Variance Decompositions

Variance decompositions measure the fraction of the forecast error variance of some variable x_i that can be explained by shock η_j at the forecast horizon h. Denoting this variance decomposition by $v_{ij,h}$, it can be expressed relative to the impulse response function as:

$$v_{ij,h} = e_i' v_h e_j = \left[e_j' \otimes e_i' \right] \text{vec}(v_h), \tag{4}$$

where e_i is the i:th column of I_p, $\otimes$ is the Kronecker product,

$$v_h = \left[\sum_{i=0}^{h-1} \left(R_i R_i' \odot I_p \right) \right]^{-1} \left[\sum_{i=0}^{h-1} \left(R_i \odot R_i \right) \right], \tag{5}$$

and $\odot$ is the Hadamard (element-by-element) product.

Given that $0 < v_{ij,h} < 1$ for all $i, j = 1, \ldots, p$, it follows that v_h is a differentiable function of $R_0, \ldots, R_{h-1}$. Hence, the asymptotic distribution of v_h can be directly obtained by differentiating v_h with respect to the impulse responses. This means that:

$$\sqrt{T} \left(\hat{v}_{ij,h} - v_{ij,h} \right) \xrightarrow{d} N(0, \sigma_{ij,h}). \tag{6}$$

An expression for $\sigma_{ij,h}$ is, e.g., given by Lütkepohl (1990), but the computationally simplest formula is found in Warne (1993) as it relies on differentiating v_h directly.

Warne (1993) shows that

$$\frac{\partial \text{vec}(v_h)}{\partial \theta'} = 2 \left\{ I_p \otimes \left[\sum_{i=0}^{h-1} \left(R_i R_i' \odot I_p \right) \right]^{-1} \right\} \times$$

$$\sum_{j=0}^{h-1} \left\{ \text{diag}\left[\text{vec}(R_j) \right] - \left[v_h' \otimes I_p \right] \text{diag}\left[\text{vec}(I_p) \right] \left[R_j \otimes I_p \right] \right\} \frac{\partial \text{vec}(R_j)}{\partial \theta'}, \tag{7}$$

where $\text{diag}(a)$ is a diagonal matrix with the vector a in its diagonal. Hence, under the assumption that each element of v_h is inside the 0-1 interval, it follows that

$$\sqrt{T} \left(\text{vec}(\hat{v}_h) - \text{vec}(v_h) \right) \xrightarrow{d} N_{p^2}(0, \Sigma_{v_h}), \tag{8}$$

where

$$\Sigma_{v_h} = \frac{\partial \text{vec}(v_h)}{\partial \theta'} \Sigma_\theta \left(\frac{\partial \text{vec}(v_h)}{\partial \theta'} \right)',$$

whereas

$$\sigma_{ij,h} = \left[e_j' \otimes e_i' \right] \Sigma_{v_h} \left[e_j \otimes e_i \right].$$

These expressions make it relatively easy to write computer code for estimating variance decompositions and their asymptotic standard errors.

3. Remarks

It is worth pointing out that if $v_{ij,h}$ is zero or unity for some i, j and h, then $\sigma_{ij,h} = 0$ by construction. The reasons for this are that (i) $\left(R_i R_i' \odot I_p \right)$ is a diagonal matrix, and (ii)

$$\left[e_j' \otimes e_i' \right] \left\{ \text{diag}\left[\text{vec}(R_l) \right] - \left[v_h' \otimes I_p \right] \text{diag}\left[\text{vec}(I_p) \right] \left[R_l \otimes I_p \right] \right\} = 0,$$

for all $l \in \{0, 1, \ldots, h-1\}$ when $v_{ij,h} = 0$ or $v_{ij,h} = 1$. For example, let $p = 2$, $i = 2$, and $j = 1$ so that we are considering $v_{21,h}$. The left hand side of the above relation is now:

$$\left[0 \quad (1 - v_{21,h}) R_{21,l} \quad 0 \quad -v_{21,h} R_{22,l} \right],$$

where $R_{ij,l}$ is the (i, j):th element of R_l. If $v_{21,h} = 0$, then $R_{21,l} = 0$ and, thus, the result is a zero vector. Similarly, if $v_{21,h} = 1$, then $R_{22,l} = 0$ and again we obtain a zero vector. This means that we cannot test if a variance decomposition is zero or unity using a t or a Wald test based on the above results.

In fact, the asymptotic covariance matrices Σ_{v_h} takes into account that the rows of v_h sum to unity. Letting ι_p be a vector with ones, we will show that

$$\left[\iota_p' \otimes I_p\right] \Sigma_{v_h} \left[\iota_p \otimes I_p\right] = 0.$$

The results in the previous paragraph are simple consequences of this property of the asymptotic covariance matrices. Hence, we know that the rank of Σ_{v_h} is at most $p(p-1)$. To prove this we note that $v_h \iota_p \equiv \iota_p$. Accordingly, the differential $dv_h \iota_p = 0$, i.e.,

$$\left[\iota_p' \otimes I_p\right] d\text{vec}\left(v_h\right) = 0.$$

Based on this differential we find that the partial derivatives satisfy:

$$\left[\iota_p' \otimes I_p\right] \frac{\partial \text{vec}(v_h)}{\partial \theta'} = 0.$$

Hence, premultiplication of Σ_{v_h} by $\left[\iota_p' \otimes I_p\right]$ and postmultiplication by its transpose gives us a $p \times p$ zero matrix.

While this is rather trivial, it leads us to an additional property of the asymptotic covariance matrix. Namely, that two linear combinations of the shocks, denoted by the vectors a and b, which satisfy $a + b \equiv \iota_p$, will always have the same asymptotic covariance matrix. Post-multiplying v_h by a we have that $v_h a \equiv \iota_p - v_h b$. The differentials of $v_h a$ therefore satisfies $dv_h a = -dv_h b$. Accordingly,

$$\left[a' \otimes I_p\right] \frac{\partial \text{vec}(v_h)}{\partial \theta'} = -\left[b' \otimes I_p\right] \frac{\partial \text{vec}(v_h)}{\partial \theta'},$$

so that

$$\left[a' \otimes I_p\right] \Sigma_{v_h} \left[a \otimes I_p\right] = \left[b' \otimes I_p\right] \Sigma_{v_h} \left[b \otimes I_p\right].$$

The simplest example is when $p = 2$ and we consider $a = e_1$ and $b = e_2$. This gives us $v_{i1,h} + v_{i2,h} = 1$ for $i = 1, 2$, and thus we find that $\sigma_{i1,h} = \sigma_{i2,h}$ for $i = 1, 2$.

Generally, we expect estimates of the asymptotic variance $\sigma_{ij,h}$ to represent the unknown small-sample uncertainty of the variance decomposition $\hat{v}_{ij,h}$ better when the variance decomposition is sufficiently far away from zero or unity. What far away means depends on the size of the t-ratio $\hat{v}_{ij,h} / \sqrt{(\hat{\sigma}_{ij,h}/T)}$. The smaller this ratio is, the more likely it is that, e.g., a confidence band for $\hat{v}_{ij,h}$ based on the asymptotic distribution includes either zero or unity. Since a variance decomposition, by construction, cannot be less than zero or greater than unity, a confidence band which includes those limits cannot be correct and is bound to be a poor estimate of the unknown small-sample confidence band. In practise, estimated asymptotic standard errors for $\hat{v}_{ij,h}$ should therefore be treated with great caution.

An alternative to applying the asymptotic results to variance decompositions is to use boot-strapping. Theoretically, bootstraps often provide an asymptotic refinement over asymptotics, but the requirement is usually that we bootstrap statistics that are (asymptotically) pivotal; see, e.g., Horowitz (2001). That is, bootstrap theory tells us that bootstrapping is more reliable than asymptotics when we evaluate a statistic which (asymptotically) does not depend on nuisance parameters. Applied to variance decompositions, this tells us that we should not bootstrap the variance decompositions themselves since these are not asymptotically pivotal. Rather, we should bootstrap, e.g., the t-ratios. But the t-ratios do not exist if the true value of the variance decomposition is zero or unity. From this perspective, bootstrapping variance decompositions may not help us (relative to asymptotics) either when attempting to evaluate confidence bands for a variance decomposition based on bootstraps of $\hat{v}_{ij,h}$ directly (although the bootstrap intervals, by construction, will always lie between zero and unity) or when attempting to make inference about the hypothesis $v_{ij,h} = 0$ (or $v_{ij,h} = 1$) using the variance decompositions.

The hypothesis $v_{ij,h} = 0$ (or $v_{ij,h} = 1$) may instead be addressed directly from the impulse responses. When $v_{ij,h} = 0$, then an algebraically equivalent hypothesis is:

$$\sum_{l=0}^{h-1} e_i'\left(R_l \odot R_l\right)e_j = 0.$$

Similarly, the hypothesis $v_{ij,h} = 1$ is algebraically equivalent to:

$$\sum_{l=0}^{h-1} e_i'(R_l \odot R_l)e_j - \sum_{l=0}^{h-1} e_i'R_lR_l'e_i = 0.$$

When testing such nonlinear restrictions (in terms of θ) it is usually better to apply LM or LR tests, since unlike the Wald (and t) test they do not suffer from the numerical issue regarding nonlinear restrictions discussed by, e.g., Gregory and Veall (1985).[1] Moreover, since, e.g., $v_{ij,h} = 0$ may be implied by $v_{ij,h^*} = 0$ for all $h > h^*$ (as in the case of an unrestricted VAR model with k lags, where $h^* = (p-1)k$), the complexity of the testing (and estimation) problem can, for certain forecast horizons, be greatly simplified. Nevertheless, impulse responses are typically nonlinear functions of θ and since both LM and LR tests require that we have estimates of the impulse responses under the null hypothesis, the problem of testing, e.g., $v_{ij,h} = 0$ remains nontrivial (at least from a computational aspect).

REFERENCES

Gregory, A. W. and M. R. Veall (1985), "Formulating Wald Tests of Nonlinear Restrictions", *Econometrica*, 53, 1465–1468.

Horowitz, J. L. (2001), "The Bootstrap", in J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics*, Vol. 5, 3159–3228, Amsterdam: North-Holland.

Lütkepohl, H. (1990), "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models", *Review of Economic & Statistics*, 72, 116–125.

Lütkepohl, H. and H.-E. Reimers (1992), "Impulse Response Analysis of Cointegrated Systems", *Journal of Economic Dynamics and Control*, 16, 53–78.

Warne, A. (1993), "A Common Trends Model: Identification, Estimation and Inference", Seminar Paper No. 555, Institute for International Economic Studies, Stockholm University.

Vlaar, P. J. G. (2004), "On the Asymptotic Distribution of Impulse Response Functions with Long Run Restrictions", *Econometric Theory*, 20, 891–903.

---

[1] That is, algebraically equivalent forms of writing a nonlinear restriction, lead to different numerical values for the Wald (and t) test. For example, suppose that we have estimated a parameter θ, given by $\hat{\theta}$, that $\hat{\theta}$ satisfies (2), while the estimated asymptotic variance of $\hat{\theta}$ is $\hat{\sigma}_\theta$. A Wald test of the hypothesis $\theta = 0$ is then given by $W_1 = T\hat{\theta}^2/\hat{\sigma}_\theta$. An algebraically equivalent hypothesis is $f(\theta) = \theta^2 = 0$. The Wald test of this restriction is given by $W_2 = Tf(\hat{\theta})/(df/d\theta|_{\theta=\hat{\theta}})^2\hat{\sigma}_\theta = T\hat{\theta}^2/4\hat{\sigma}_\theta$. In the limit, W_1 and W_2 are both $\chi^2(1)$ distributed under the null hypothesis. The former statistic is, however, (4 times) greater than the latter for finite T. Although this example is trivial, it illustrates an important shortcoming of Wald (and t) tests. Namely, the nonlinearity in the numerator and the denominator of the statistic do not cancel and the numerical value therefore changes. It follows that a test based on W_1 may suggest that we should reject the null hypothesis, while a test based on W_2 may not.