11-1-1994

Error Bands for Impulse Responses

Christopher A. Sims

Tao Zha

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1085

ERROR BANDS FOR IMPULSE RESPONSES

Christopher A. Sims and Tao Zha

November 1994

Error Bands for Impulse Responses

Christopher A. Sims

Yale University

and

Tao Zha

Federal Reserve Bank of Atlanta

July 1995

We examine the theory and behavior in practice of Bayesian and bootstrap methods for generating error bands on impulse responses in dynamic linear models. The Bayesian intervals have a firmer theoretical foundation in small samples, are easier to compute, and are about as good in small samples by classical criteria as are the best bootstrap intervals. Bootstrap intervals based directly on the simulated small-sample distribution of an estimator, without bias correction, perform very badly. We show that a method that has been used to extend to the overidentified case standard algorithms for Bayesian intervals in reduced form models is incorrect, and we show how to obtain correct Bayesian intervals for this case.

Support for this research was provided in part by National Science Foundation grant number SES91-22355. The second author acknowledges research grant support from the University of Saskatchewan.

1. Introduction

In interpreting dynamic multivariate linear models, impulse response functions are of central interest. Presenting measures of the statistical reliability of estimated impulse responses is therefore important. On the other hand, such models have three features that raise difficulties for construction of classical confidence intervals:

i) Estimates of the underlying autoregressive reduced form parameters of such models have sampling distributions that depend strongly in shape as well as location on the true value of the parameters, especially in the neighborhood of parameters that imply non-stationarity;

ii) Impulse responses are highly nonlinear functions of the underlying autoregressive reduced form parameters; and

iii) The distribution of the estimate of a particular response at a particular horizon depends strongly on the true values of other impulse responses at other time horizons, with no apparent good pivotal quantity available to dampen such dependence on nuisance parameters.

Most applied researchers are familiar with forming and interpreting confidence intervals in situations without these features, particularly in standard cases of estimators with asymptotic normal distributions and an asymptotic "pivot". More precisely, these standard cases are those in which we have an estimator $\hat{q}(X)$ based on data X for a parameter q and an estimator $\hat{s}(X)$ for the variability of $\hat{q}(X)$ such that $\left(\hat{q}(X)-q\right)\big/\hat{s}(X)$ (the pivot) is asymptotically N(0,1). Confidence intervals based on such an asymptotic approximation are convenient because they are symmetric and can be formed using standard tables of the normal or t distribution. More importantly, under mild regularity conditions their Bayesian posterior probability is asymptotically well approximated by their confidence level, regardless of the Bayesian prior. Because these cases are so common, applied work seldom carefully articulates the difference between confidence levels and posterior probabilities. Decision makers without statistical training probably usually regard an econometrician's statement that "with confidence level 90%, q is between a and b" as meaning that, having observed the data, the econometrician regards .90 as a reasonable value to give, for decision-making purposes, to the probability that q lies in (a,b). But this is the Bayesian interpretation of the interval, not the classical one.

An exact 90% confidence interval does "contain the true parameter value with probability .90", but this probability statement is only appropriate before data is observed. That is, *before* we see X, it is reasonable to say that the probability that the confidence interval contains q is .90. This does not mean in general that it is still reasonable *after* we see X to say that the probability that the confidence interval contains q is .90. The fundamental difference between a Bayesian approach to inference and a classical approach is that the latter restricts itself to probability statements based on pre-observation probability distributions, while the former aims at aiding the formation of reasonable probability statements about the parameter conditional on the observed data.

Example 1

To remind oneself of why confidence levels are not decision-making probabilities given the data, consider a case where we know that μ lies in (0,1) and observe two i.i.d. N(μ, σ) random variables X_1 and X_2. It would be natural (though arguably not optimal) to form confidence intervals from the pivot $(\hat{\mu} - \mu)/\hat{\sigma}$, where the "hats" refer to the standard sample mean and sample variance estimators. This is a $t(1)$ variable. If the true value of μ is .5 and of σ is 1, a not terribly unlikely observation would be $X_1 = 2.1$, $X_2 = 2.2$. This leaves the entire (0,1) interval more than 16 standard deviations from the sample mean, so a 95% confidence interval is empty. Also not unlikely would be $X_1 = 1.3$, $X_2 = -.3$, which leaves the entire (0,1) interval within even a 50% confidence interval. Obviously it makes no sense, since we knew to start with that μ lay in (0,1), to say after seeing the data that we believe in the first case that the probability that μ lies in (0,1) is .01 or in the second case that the probability is .5. But it is precisely justifiable to say with *confidence level* 95% in the first case that μ is nowhere in (0,1) and with *confidence level* 50% in the second case that μ is somewhere in (0,1).

Asymptotic pivotal quantities with normal distributions are available in stationary linear time series models, but not in nonstationary models, generally, because of the property listed as (i) above. Many econometricians would be willing to assume stationarity in practice if models arbitrarily close to non-stationarity were allowed. However, in a finite sample the accuracy of the asymptotic normal approximation begins to break down as the boundary of the stationary region of the parameter space is approached. In practice, with economic data, it seems to be the usual case that we cannot rule out parameter values close enough to nonstationarity to make the usual normal asymptotic distribution theory unreliable. Time series modeling is therefore a rare instance in which Bayesian posterior probabilities and classical confidence levels can be in substantial conflict. Our view is that when the two conflict, posterior probabilities are almost always more useful and more important to report than confidence levels.

We begin this paper with a discussion of the theory of classical confidence intervals and regions, explaining why in this context they are harder to construct than Bayesian posterior probability regions. We explain why bootstrap methods that use computer simulation to determine the sampling distribution of an estimator conditional on a single true parameter value can produce correct confidence intervals only under strong auxiliary assumptions that are likely to conflict with the model at hand. We discuss the scope for divergence between the classical confidence level for an interval and its Bayesian posterior probability, and the sense in which each can be "biased" from the point of view of the other.

Some implementations of classical bootstrap methods for impulse responses in the existing literature have resulted in confidence intervals that are misleading or mistaken, and we point out some of the sources of error. Practical methods for computing Bayesian posteriors on overidentified linear models are not obvious, and some existing work computes them by methods that have no justification, classical or Bayesian, in small samples. We show how to construct correct Bayesian error bands for such models.

We undertake a number of computational exercises. We show that bias-corrections to bootstrap methods that have been suggested, but not yet widely used, in the literature are capable of producing classical confidence intervals whose coverage probability is in most, but not all, of the models we consider less divergent from their nominal coverage probabilities. Similar results have been obtained recently by Killian [1995], though that paper differs from our work in ruling out nonstationary parameter values. We show that Bayesian intervals, which are less difficult to compute than the bootstrap ones, are competitive with the corrected classical ones even on classical criteria. And we document the bias and imprecision in "corrected" classical bootstrap intervals as summaries of the implications of the data for the unknown true impulse responses (i.e. as summaries of the likelihood shape).

The paper's objective is not mainly to advance the state of the art of constructing error bands with good classical coverage probabilities.[1] We regard classical coverage probabilities as of secondary interest and leave to others the daunting task of finding random intervals with better coverage probabilities in dynamic models. We document the fairly good performance of Bayesian intervals by classical criteria partly as a way to reassure classically trained econometricians that Bayesian intervals do not misbehave badly by the criteria they usually study. We examine the performance by Bayesian criteria of "bias-corrected bootstrap intervals" motivated by classical reasoning because these methods may be used in practice, and econometricians who share with us the view that the Bayesian performance criteria are primary will want to know how deficient such bootstrap methods are. Conscientious classical econometricians will be interested in these results also, because they will recognize that confidence levels are not decision-making probabilities and will want to warn readers or clients of cases where the discrepancy is likely to be large. Our main objective is to show that Bayesian intervals are relatively straightforward to compute and well-behaved and to show how to construct them for overidentified models, where there are some non-trivial computational difficulties.

2. Confidence Interval Fundamentals

Suppose we have a parameter space Θ whose elements q index probability distributions for observable data X that lies in the sample space Ω. Classical confidence levels and Bayesian posterior probabilities are both ways of associating probabilities with subsets of Θ. Bayesians postulate a joint probability distribution over $\Theta \times \Omega$ and associate with $S \subset \Theta$, after observing data X, its posterior probability $P[S|X]$. Classical inference avoids putting probability distributions on Θ. It insists that S be regarded as varying randomly with X. If $P[q \in S(X)|q] = (1-a)$ for all q, $S(\cdot)$ is said to be a (1-a)% confidence region for q. Sometimes the definition is weakened to requiring only that $P[q \in S(X)|q] \geq (1-a)$. With the inequality, we

[1] We try to follow recent suggestions in the literature for improving performance of bootstrapped error bands in dynamic models, and our computations include consideration of some models larger than those previously analyzed in Monte Carlo studies of such bands.

will refer to the interval as a *conservative* (1-α)% confidence interval, while with the equality we will refer to it as an *exact* (1-α)% confidence interval.

A (1-α)% confidence region is equivalent to a collection of hypothesis tests with significance level α:[2] Given $S(X)$, form a test of $H_0{:}\theta = \theta_0$ for each $\theta_0 \in \Theta$ by rejecting when $\theta_0 \notin S(X)$. Given a set of tests with significance level α for $H_0{:}\theta = \theta_0$, one for each $\theta_0 \in \Theta$, and with rejection regions $R(\theta_0)$, form $S(X)$ as $\{\theta | X \notin R(\theta)\}$.

Confidence levels and posterior probabilities are related. Suppose $S(\cdot)$ is a $(1-\alpha)$% confidence region and consider the subset of $\Theta \times \Omega$ defined as $V = \{\langle \theta, X \rangle | \theta \in S(X)\}$. Denoting the Bayesian marginal distribution on Θ (usually called the prior) by p, we can write

$$P[V] = \int_\Theta P[V|\theta]p(d\theta) = \int_\Theta P[\theta \in S(X)|\theta]p(d\theta) \ . \tag{1}$$

From (1) it is clear that $P[V] = 1-\alpha$ if $S(\cdot)$ is exact and $P[V] \geq 1-\alpha$ if $S(\cdot)$ is conservative. But we can also write

$$P[V] = \int_\Omega P[V|X]m(dX) = \int_\Omega P[\theta \in S(X)|X]m(dX) \ , \tag{2}$$

where m is the marginal distribution over Ω. Equations (1) and (2) together imply that an exact confidence level is the mean over X of Bayesian posterior probabilities. Thus Bayesian posterior probabilities for $S(X)$ cannot differ from the classical confidence level in the same direction for all X. Any tendency for posterior probabilities to be smaller than the confidence level for some X must be balanced by their being larger for other X. Also, if a conservative confidence level is close to one, Bayesian posterior probabilities for $S(X)$ must be as high as the confidence level with high probability. (These results depend on the assumption of a proper prior probability on Θ and may not carry over to Bayesian posterior probabilities generated from improper prior p.d.f.'s that do not have a finite integral.)

Bayesian posterior probabilities depend in general on the prior distribution p over Θ. Since p must come from a source other than the data at hand, econometricians and statisticians rightly resist making their analysis dependent on such an arbitrary and possibly subjective element. Bayesians reporting data analyses for wide audiences, though, do not make their analysis dependent on a subjectively chosen p (see Hildreth [1963] for an explicit discussion). Instead they try to summarize the form of the likelihood in a way that may be useful to a wide class of readers who may have differing priors and loss functions. This may involve treating the normalized likelihood itself as a p.d.f. on Θ (i.e., using a "flat prior"), or it may involve using some transparent and standard "reference prior" that will allow readers to adjust conclusions to reflect their own prior

[2] A significance level can be conservative or exact, just as with a confidence level.

beliefs. In this paper we concentrate on Bayesian analysis with flat priors.[3] Besides their usefulness as a transparent reporting device, flat priors can also be justified by the fact that in large samples posterior distributions under flat priors will generally approximately match those under any prior that has non-zero, continuous density in the neighborhood of the true parameter value.

To verify that $S(X)$ has a confidence level $(1-\alpha)\%$ we need to know how it is distributed for each θ in Θ. The idea of the bootstrap is to approximate the distribution of X at the true value of θ by the distribution of X at some estimated point $\hat{\theta}$ in Θ. It may seem that, because the bootstrap generates an accurate estimates of the shape of the finite-sample distribution of $\hat{\theta}$ at one point in the parameter space, that it is generating exact small-sample confidence intervals. In fact, though, this method produces exact small-sample confidence intervals only under strong assumptions, rarely met in practice, about the nature of the dependence of the distribution of X on θ, as the following two examples illustrate.

Example 2

Suppose the p.d.f. of a scalar X is

$$f(X;m) = \frac{\exp\left(-\frac{1}{2}\log^2(X-m+1)\right)}{(X-m+1)\sqrt{2\pi}}$$

on the half-line where $X > m-1$. That is, $X-m+1$ is lognormal with $\log(X-m+1)$ distributed as N(0,1). The p.d.f. is plotted in Figure 1 for two different values of μ. If we observe X=1, the likelihood has the same shape as the p.d.f. for X, but with the reverse orientation, as shown in Figure 2. A Bayesian posterior under a flat prior would produce a "highest posterior density" region with probability .95 as $[-1.97, 1.9925]$. This interval leaves much more probability in the left than in the right tail. An equal-tailed interval is $[-2.72, 1.86]$. Because this problem is one with the parameter affecting the distribution of X purely through a location shift, Bayesian posterior probabilities and confidence levels coincide exactly. The fact that the intervals spread out over the area below X+1 while the distribution of X spreads out over the area above m-1 makes perfect sense. After all, X cannot be below m-1, so having seen X we know that m cannot be above X+1.

In constructing the intervals discussed in example 2, we used knowledge of the entire structure of the dependence of the distribution of X on m. What if instead we had used computer simulation to find the distribution of X with m set equal to 1, the observed value of X? This is what the "parametric bootstrap" would do. What Hall [1992] (and we, henceforth) call the equal-tail "percentile interval" is then constructed as follows. Let $F(X;m)$ denote the c.d.f. of X when

---

[3] In simple linear models posterior probabilities and confidence levels coincide under certain forms of flat prior. This is not true in general, however, and is very far from true in dynamic models near the boundary of the stationary region.

μ is the true value of the parameter. We suppose that simulation gives a perfectly accurate estimate of $F(X;1)$. We form our interval for μ by finding a and b such that $F(1-a;1)=.025$, $F(1+b;1)=.975$. Then the interval is [1-b, 1+a]. But this bootstrap interval coincides with the flat-prior Bayesian equal-tailed .95 probability interval. We seem to have arrived at an interval that should satisfy everyone, Bayesian or classical, on the basis of calculating only $F(\cdot,1)$, without having to know the distribution of X for any other value of μ.

The bootstrap percentile interval of example 2 can be contrasted with the other-percentile interval, which would be [1-a, 1+b] in example 2. It is obviously a terrible choice in example 2, as it includes μ's greater than 2, while it is in fact impossible for μ to be above 2. But as we now see in example 3, it is possible that $F(\cdot,1)$ is exactly what was computed in example 2, and for the other-percentile interval to have correct coverage probability and to coincide with a Bayesian interval with posterior probability matching the confidence level.

Example 3

Suppose that $\log(X)$ is distributed N(log(μ),1). When $\mu=1$, this gives X the same distribution as for Example 2 when $\mu=1$ there. However, as can be seen by comparing Figures 1 and 3, the distribution of X changes as μ deviates from 1 in quite different ways in the two examples. Here $Z=\log(X)\sim N(\eta,1)$, where $\eta=\log(\mu)$. It seems natural to use the normally distributed $Z-\eta$ as a pivot, then convert our confidence interval on η into a confidence interval on μ. This produces as a 95% confidence interval on μ, [.141,7.10]. The equal-tailed flat prior Bayesian posterior 95% probability interval is not the same, but it is similarly skewed, at [.37,17.4]. The likelihood, as can be seen in Figure 4, is similar in shape to the p.d.f., but damps more slowly as its argument increases. The flat-prior posterior probability of the classical interval is about .83. If the prior is flat on $\log(\mu)$ rather than on μ itself, though, so that the prior density is $1/\mu$, the posterior equal-tailed 95% interval is identical to the classical 95% equal-tail confidence interval[4].

As these examples illustrate, the fact that the bootstrap gives us accurate information about the small-sample distribution of an estimator or statistic at a particular point in the sample space does not in itself suggest that it is giving us small-sample accuracy in confidence intervals. Asymptotic distribution theory does tell us that under broad regularity conditions many estimators $\hat{\theta}$ have distributions in large samples that are close to normal and depend on the true parameter θ approximately via a pure location shift. This situation gives asymptotic justification to bootstrap confidence intervals, but not in general to any asymmetries they may show. With stronger regularity conditions, it is possible to show that bootstrap intervals of certain types, or with

[4] The flat prior on $\log(\mu)$ is the Jeffreys prior for this example. A prime motivation for the Jeffreys prior is to achieve invariance of a standardized "flat" prior to parameter transformations, so it is natural that it should lead us to the same conclusion as would transforming the problem into Z,η space. In Example2 the Jeffreys prior is an ordinary flat prior.

certain corrections[5], are asymptotically second-order accurate in a certain sense. However second-order accurate intervals often misbehave in finite samples, and in any case the regularity conditions they invoke appear to be violated in the neighborhood of unit roots for dynamic models.

3. Dimensionality

In this paper we are concerned primarily with confidence intervals on dynamic models' impulse response functions. The point that an exact $(1-a)\%$ confidence set $S(X)$ is equivalent to a collection of tests of point null hypotheses, all with exact significance level a, suggests a general approach to finding confidence regions in a multi-dimensional parameter space, but it does not provide help in finding confidence intervals for individual elements of the parameter vector. If the parameter space is R^k and we want a confidence interval on q_1, the first element of the parameter vector, then we are looking for a confidence region in Θ of the form $S_1(X)\times R^{k-1}$. Confidence regions assembled from a collection of point hypothesis tests will not take this form unless the hypothesis tests are all based on statistics whose distributions do not depend on q except through q_1. Though in the leading special case of the Gaussian linear regression model such statistics (the usual *t*-statistics on individual coefficients) exist, they are not in general easy to find. Of course it is possible to find a conservative confidence level for an arbitrary random set $S(X)$ (since it is just $\max_q P[S(X)|q]$), but if $P[S(X)|q]$ varies a great deal with q, conservative confidence levels may be misleading.

A Bayesian approach has no more difficulty with evaluating the posterior probability of a region with the form $S_1(X)\times R^{k-1}$ than with a region of any other form.

More generally we may be interested in a subvector of q, with the remainder of q regarded as nuisance parameters. For a Bayesian approach, this simply requires integrating over the nuisance parameters in forming probability statements, a straightforward, if possibly difficult, task. For a classical approach, the difficulties are more fundamental, because all probability statements must condition on the full parameter vector, whether they are "nuisances" or not.

When an analytic expression for the likelihood function is available, a Bayesian approach to generating a posterior probability interval or region is much easier than bootstrap approaches to generating confidence intervals or regions. With a given data set, the Bayesian problem is to characterize the shape of the posterior, which is a single distribution whose form is given by the likelihood, or by the likelihood multiplied by the prior p.d.f. Sampling from this distribution may be a non-trivial problem if the likelihood has a non-standard form, but there are well-understood computational techniques for attacking it. But the existence of an analytic form for the p.d.f. does not save a classical approach from the need to analyze the form of many distributions -- the

[5] What Hall [1992] calls the percentile-*t* interval and the BCA interval of Efron and Tibshirani [1993] are examples.

distributions of the data for each possible value of the parameter. Of course some parameter values may be so far from fitting well that we can ignore them, and continuity may allow us to approximate the behavior of $P[X|q]$ over a continuum of possible values for q from knowledge of its behavior for a finite number of values of q, but in multivariate models the required number of values of q is likely to be large.

4. Error Bands for Multivariate Dynamic Models

Consider a model of the form

$$y(t) = g\big(e(t), y(t-1), \ldots, y(t-k) \big| b\big) \tag{3}$$

with $e(t)$ independent of $y(t-s)$, all $s>0$, and having p.d.f. $h(\cdot\,|\Omega)$. If $e(t)$ is of the same dimension as $y(t)$, and if we have an analytic expression for the jacobian $|dy(t)/de(t)|$,[6] then (3) allows us to write a p.d.f. for $y(1)$, ..., $y(T)$ conditional on $y(0)$, $y(-1)$, ..., $y(-k+1)$, b, and Ω, which we will label

$$p\big(y(1), \ldots, y(T) \big| y(0), \ldots, y(-k+1), b, \Omega\big) \ . \tag{4}$$

It is common but not universal practice to treat (4) as the likelihood function. It is actually the likelihood function only if the distribution of $y(0), \ldots, y(-k+1)$ does not depend on unknown parameters, or depends only on unknown parameters unrelated to b and Ω. If (3) is consistent with y being ergodic, then it is natural to take the ergodic marginal distribution of $y(t), \ldots, y(t-k)$, which of course in general depends on b and Ω, as the distribution for $y(0), \ldots, y(-k+1)$. This is then combined with (4) to produce a p.d.f. for the full vector of observations $y(-k+1), y(-k+2), \ldots y(T)$. There are three reasons this is not often done: it often makes computations much more difficult; it requires that we rule out, or treat as a separate special case, non-stationary (and thus non-ergodic) versions of the model; and it may not be plausible that the dynamic mechanism described in (3) has been in operation long enough, in unchanged form, to give $y(0), \ldots, y(-k+1)$ the ergodic distribution. The last two points are related. A non-stationary model has no ergodic distribution. A near-non-stationary model may have an ergodic distribution, yet it may imply that the time required to arrive at the ergodic distribution from arbitrary initial conditions is so long that imposing the ergodic distribution on $y(0), \ldots, y(-k+1)$ may be unreasonable.

---

[6] If e is of higher dimension than y, we in effect have a latent variable, and integration over part of the e vector is required to form the likelihood. If e is of lower dimension than y, the implied distribution for y is singular, which ordinarily makes the model conflict too sharply with the data to be usable for inference.

Bayesian inference would ideally use the ergodic distribution for the initial conditions at parameter values well within the stationary region of the parameter space, then shift smoothly over to a distribution less connected to b and Ω as the non-stationary region is approached. Such a model is likely to be application-dependent, however, and in the remainder of this paper we treat (4) as the likelihood. We also hold initial conditions fixed in generating Monte Carlo samples of $y(1), \ldots, y(T)$ when we evaluate classical coverage probabilities.[7]

Models in the class we consider here, then, have an analytic likelihood function and therefore make exact Bayesian posterior probability calculations much easier than exact classical confidence level calculations. In fact much of our analysis will focus on the case where g is linear in b and h is Gaussian, which makes the log likelihood quadratic in b, i.e. Gaussian in shape, in small and large samples, for stationary and non-stationary models.

For a general model of the form (3), there is ambiguity about how "impulse responses" ought to be defined. Here, though, we consider only models such that g is linear in $e(t)$, so that the response $a_{ij}(s)$ of $y_i(t+s)$ to $e_j(t)$ is easily and unambiguously defined. By applying (3) recursively, we can solve for $y(t+s)$ as a function of $e(t+s), e(t+s-1), \ldots, e(t)$ and $y(t-1), \ldots, y(t-k)$. Then

$$a_{ij}(s) = \frac{\partial y_i(t+s)}{\partial e_j(t)} \tag{5}$$

depends only on b, not y or e, and can be found by elementary matrix operations.[8]

We focus our attention entirely on confidence or probability intervals for individual responses at particular horizons, i.e. for single $a_{ij}(s)$ values. Usually our interest is actually in the shape of the function $a_{ij}(\cdot)$, i.e. of the whole pattern of the response over time, or even in an interrelated

---

[7] On this latter point we differ from Killian [1995]. Conditioning on initial data values in forming the likelihood or in constructing estimators (on which point Killian's practice matches ours) amounts to ignoring potential information in the initial observations. Conditioning on them in doing Monte Carlo simulations of the data-generation process amounts to recognizing the distinction between initial conditions that generate more and less informative samples. If the initial y's show unusually large deviations from their steady-state values, the sample is likely to generate unusually sharp information about the parameters. It does not make sense to calculate coverage probabilities that average across informative and uninformative samples when it is easy to take account of the fact that we have been lucky (or unlucky) in the initial conditions of the particular sample at hand.

[8] Note that, though the impulse responses as defined here coincide with the coefficients of the moving average representation (MAR) for stationary, linearly regular models, they are also defined for nonstationary models where no MAR exists.

set of such patterns of response. Confidence intervals or probability bands are often displayed in graphs showing an estimated $a_{ij}(\cdot)$ plotted as a function of time horizon *s* along with the upper and lower limits of confidence intervals or probability bands on the individual $a_{ij}(s)$ values. But posterior distributions on $a_{ij}(s)$ values or sampling distributions of estimates of $a_{ij}(s)$ values are likely to be dependent across *s*, *i*, and *j*, and the nature of the dependence will differ across applications or samples. This raises problems in reporting results similar to those arising in regression models, where the collection of confidence intervals on individual coefficients may not give an accurate picture of the nature of uncertainty about the whole coefficient vector. We leave to future research the problem of how to provide more thorough characterizations of uncertainty about the *a*'s.[9]

5. Bias-Correcting the Bootstrap

The usual parametric bootstrap procedure begins with an estimate $\hat{q}(X)$ based on the observed sample *X*. It then generates the distribution function $F\left(\cdot\middle|\hat{q}(X)\right)$ for $\hat{q}(X^*)$, where X^* is a Monte Carlo random variable generated from the distribution *X* would have if $\hat{q}(X)$ were the true value of q. If it turns out that the distribution of $\hat{q}(X^*)$ is not centered at $\hat{q}(X)$, this suggests that the original estimate $\hat{q}(X)$ was biased as an estimate of q. Does this mean we should adjust $\hat{q}$ for bias before proceeding?

As should be clear from our examples 2 and 3 above, there is no prima facie case for bias-adjusting if the percentile interval is used in generating the bootstrap confidence interval. This interval, because it is based on reflecting the bootstrap sampling distribution of $\hat{q}$ about the pseudo-true-value $\hat{q}(X)$, automatically accounts for bias. Bias-adjustment does not in general make an estimator more accurate in the sense of lowered root mean squared error (RMSE), and indeed when applied as here to a maximum likelihood estimate it is more likely than not to worsen RMSE over some range of parameter values. But most applied work has used the other-percentile interval, which, as is clear from example 2 above, is badly affected by bias. The main intuitive appeal of the other-percentile interval is its transformation invariance, and Efron and

---

[9] Though we have some ideas. If the *a*'s were jointly Gaussian, one could treat the coefficients in $a_{ij}(\cdot)$ as a jointly normal vector and display the largest few and smallest few principle axes of their p.d.f. contour ellipsoids. When plotted, these would show which kinds of variation in $a_{ij}(\cdot)$ are ill-determined by the likelihood function and which kinds are more sharply determined. This idea might be helpful despite the fact that the posterior on the $a_{ij}(\cdot)$'s is not Gaussian. We do not think it very helpful to construct bands such that the posterior probability that responses are contained entirely within the band is 1-a, any more than it is useful to inflate confidence intervals in a regression model until the rectangular region defined by their intersection has posterior probability 1a.

Tibshirani [1993] discuss ways to adjust it for bias in such a way that it retains this property. Their adjustment works directly with the confidence interval and in effect adjusts for median bias, not expectational bias.

Many applied studies have used the RATS Monte Carlo procedure for constructing Bayesian error bands for VAR impulse responses, usually with little or no discussion of underlying methodological issues. Blanchard and Quah [1989], Lastrapes and Selgin [1994], and Runkle [1987] use bootstrap methods, relying basically on other-percentile methods without bias-correction. Runkle uses it without modification; Blanchard and Quah modify it to avoid such strong bias that the interval fails to include the point estimate; and Lastrapes and Selgin follow Blanchard and Quah in this latter respect. Koop [1992] computes Bayesian intervals for the Blanchard-Quah impulse responses, noting how different they are,[10] and also provides an algorithm to handle models with differing lists of right-hand-side variables across equations. It is possible simply to derive analytical expressions for asymptotic standard errors of impulse response estimates and to use them, together with the normal asymptotic approximation, to arrive at confidence intervals, as shown by Lutkepohl. [1990] and Mittnik and Zadrozny [1993], and this approach has been used in practice, e.g. by Poterba, Rotemberg and Summers [1986] and Runkle [1987]. Bands constructed this way are always symmetric about the estimates and shrink to zero with the time horizon in stationary models. Bootstrap methods are motivated by a desire to avoid these unrealistic properties of the asymptotic intervals. Killian [1995] finds in a Monte Carlo study that the symmetric intervals arrived at this way have poor small-sample properties compared to other alternatives.

Killian proposes to adjust the estimator $\hat{q}(\cdot)$ itself for expectational bias. This is not at all the same thing as what Efron and Tibshirani [1993] call bias adjustment,[11] and we see no basis in existing asymptotic distribution theory for thinking that this will work generally to produce better confidence intervals. It casts aside the transformation invariance of the other-percentile interval. Nonetheless, this form of bias adjustment seems to help. Killian reports trying other methods with firmer foundations in asymptotic theory without good result, and we have ourselves found that some other approaches to generating bootstrap intervals do not perform systematically better, in terms of coverage probability, than Killian's intervals.[12] Because our aim is to compare the

---

[10] Though, as we will see below, most of the stark asymmetry found by Blanchard and Quah arose from an error in their calculations.

[11] See Hall [1992], p.129 for some elaboration of the point that what is called "bias correction" in the literature of bootstrapped confidence intervals is not correction of the bootstrapped estimator for bias.

[12] In some cases we examined, the percentile method produced small deviations from its nominal coverage probability more often than did the other-percentile method with Killian's correction, while it deviated by more than .20 much less often. In others (our version of) Killian's method produced strictly better results than the percentile method. We also tried applying Killian's type of bias-correction directly to the confidence intervals on impulse responses rather

Bayesian intervals we prefer with bootstrap alternatives actually in use, not to advance further into the dense and (in our view) fruitless thickets of classical bootstrap theory, we use a variant of Killian's approach as representative of bootstrap methods.

Killian suggests bias correction by what he calls "bootstrap after bootstrap." Its motivation is as follows. We begin by forming a point estimate $\hat{q}(X)$ and using it as a pseudo-true value in generating an artificial sample of draws X^* from the distribution of X. Denoting the mean of the Monte Carlo sample of $\hat{q}(X^*)$ values by $\bar{q}^*$, we estimate the bias in $\hat{q}(X)$ as $\hat{q}(X)-\bar{q}^*$ and arrive at a bias-corrected estimate

$$\hat{\hat{q}}(X)=\hat{q}(X)+\left(\hat{q}(X)-\bar{q}^*\right)=2\hat{q}(X)-\bar{q}^* \ . \tag{6}$$

Killian's basic idea is that $\hat{\hat{q}}(X)$ should be a better estimate than $\hat{q}(X)$, so we will get better confidence intervals if we bootstrap it rather than $\hat{q}(X)$.

But to bootstrap $\hat{\hat{q}}(X)$ requires what Killian calls "bootstrap within bootstrap" calculations. For each Monte Carlo draw X^* we have to draw a new Monte Carlo sample to find the bias correction. The number of Monte Carlo draws is thus squared, which is likely to be infeasible. Killian points out that there is asymptotic justification for simply using the original bias adjustment, $\hat{q}(X)-\bar{q}^*$, for all the Monte Carlo draws rather than calculating a new adjustment for each draw. This means that the confidence interval is constructed in two steps -- an initial bootstrap calculation to find the bias correction, then an additional bootstrap calculation to find the bias-corrected confidence interval. He calls this "bootstrap after bootstrap," and it doubles the number of Monte Carlo draws instead of squaring it.

Our procedures for computing coverage probabilities and bias corrections do differ from Killian's in certain respects. He follows Runkle [1987] in computing coverage probabilities based on the unconditional distribution of the time series sample implied by the model's stationary ergodic distribution. As we explained above, our view is that there is seldom good a priori reason to exclude nonstationary models, and that in any case coverage probabilities using the distribution of the data conditional on the initial observations are more useful than unconditional coverage probabilities.[13] We therefore use the conditional distribution, given initial conditions, when we compute coverage probabilities. Because Killian must restrict attention to stationary versions of

---

than to the autoregressive parameter estimates, confirming Killian's experience that this did not work as well.

[13] Killian apparently follows Runkle [1987] in drawing initial conditions as random subsequences of the observed sample values. Whenever the model contains roots on the order of (T-2)/T (where T is sample size) or larger in absolute value, though, this procedure will give a seriously distorted picture of the true unconditional distribution of initial conditions.

models to make his procedures internally consistent, he uses ad hoc procedures to map parameter values implying nonstationary models back into the stationary region of the parameter space when he is generating Monte Carlo draws. We include nonstationary draws of the parameter vector without alteration. We believe this better reflects actual practice, since reports of estimates that seem to imply slightly explosive behavior are not uncommon in the literature.

We follow Killian in using the semi-parametric bootstrap. That is, we draw with replacement from the sample distribution of residual vectors in generating artificial bootstrap samples. The alternative would be to use the fully parametric bootstrap, drawing from the joint normal distribution with covariance matrix equal to the sample estimate. When we evaluate coverage probabilities, we of course draw from the hypothesized true model's distribution, which involves drawing from the normal distribution for the residuals.

6. Some Simple Bivariate Cases

Bayesian and classical error bands for impulse responses may be similar or different. When they differ, the classical interval may be misleading by Bayesian criteria or not so misleading, and vice versa. And classical intervals generated by bootstrap computations, which are rigorously justified only asymptotically, may perform well or badly in finite samples by classical criteria. All these cases actually occur, and we begin by displaying some simple models that may give some insight into the range of behaviors of error band calculations as the parameter values of the model change.

We will examine a model of the form

$$\begin{aligned} y(t) &= c + Ay(t-1) + e(t) \\ e(t) &\sim N(0, \Omega) \end{aligned} \tag{7}$$

with $e(t)$ independent of past y's. First suppose

$$A = \begin{bmatrix} .9 & .1 \\ .1 & .9 \end{bmatrix}, \; c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; \Omega = \begin{bmatrix} 1.16 & 1 \\ 1 & 1 \end{bmatrix}. \tag{8}$$

This model gives y a single unit root and a co-integrating vector [1 -1]. The stable root in the system is .8, and the constant term feeds in to the nonstationary part of the system to generate a linear trend component. Such a model is known to have asymptotic Gaussian distribution theory for the OLS estimates of A,[14] so we might expect it to be easy to construct intervals with good classical coverage properties and expect classical coverage probabilities to be close to flat-prior Bayesian posteriors. To a considerable extent, this is what we find.

---

[14] Sims, Stock and Watson [1989], for example, show that since in this model each variable is a linear combination of a nonstationary variable dominated by linear trend and a stationary variable, the coefficients estimated will be jointly normal, asymptotically.

In all the results reported below for reduced-form models, the responses are triangularly orthogonalized, with the first variable's innovation producing an immediate response in the second variable, but not vice versa. This means that the first-period response of variable 1 to variable 2 is always identically zero so that coverage probabilities are not meaningful. We report zeros in these positions in tables as placeholders. All the calculations for these models involve data sample sizes of 100, 1000 Monte Carlo draws in the construction of the Bayesian or bootstrap interval for each data set, and 600 Monte Carlo draws of data sets.

We deviate from the procedure suggested in the RATS manual (and from early drafts of this paper) in using bands based on quantiles of simulated distributions rather than on second moments of the simulated distributions. For some cases (the Y-M model below, for example) one-standard error bands about means of Monte Carlo distributions are practically identical by eye to 68% quantile intervals. But for other cases, particularly at long horizons, the differences are substantial. Quantile-based intervals make more demands on computer storage, because they require storing at least some fraction of draws for sorting, while moment-based intervals can simply cumulate sums of products of simulated values, but our experience is that the quantile-based intervals are sometimes different and more informative.

The right-hand panel of Table 1 shows that Bayesian flat-prior .68 posterior probability[15] regions for this model have essentially exactly .68 classical coverage probabilities, to within Monte Carlo sampling error. Just one response, the t=2 response of y_2 to y_2, is slightly outside a band of two Monte-Carlo standard errors about .68. The left panel shows that the bootstrap intervals have modest, but definite, coverage errors, all at the short time horizons and all in the direction of undercoverage. Table 2 shows that, for a random sample of four time series drawn for this process, the bootstrap interval often, but not always, showed substantial error as a measure of posterior probability. Between intervals with equal posterior probability or coverage probability, the one that tends to be shorter is best, because it better concentrates attention on high-probability density regions. As we can see from Table 3, the Bayesian intervals for this model tend to be shorter on average, especially at longer horizons.

Having thus begun with a case where by every criterion, Bayesian or classical, Bayesian intervals are better, we now turn to a case in which the Bayesian intervals are bad by classical criteria. Here we use the parameter setting

---

[15] Through most of this paper we concentrate on intervals with coverage or posterior probability .68 (one standard error in the Gaussian case). In much applied work .90, .95 or .99 probability intervals are used. We think that as econometricians move away from the practice of pretesting and data-mining to arrive at small models with inflated t-statistics, it would be a good idea to make one-standard-error intervals the norm, as they are likely to be closer to the relevant range of uncertainty. In examining "error" in intervals, use of high-probability intervals camouflages the occurrence of large errors of over-coverage. A ridiculously large .95 interval may have coverage probability of only .995, wrong by "only" .045. Finally, accurate Monte-Carlo calculation of the end points of high-probability intervals generally requires considerably larger numbers of Monte Carlo draws.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; \Omega = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{9}$$

This parameter setting produces two unit roots, thus no cointegration, and makes the two components of *y* two independent random walks with the same drift. As can be seen from Table 4, both Bayesian and bootstrap intervals deviate substantially from the nominal .68 coverage probability at most time horizons, with the deviations uniformly in the direction of undercoverage. Furthermore, with the exception of a few low values of *t*, the undercoverage is worse for the Bayesian intervals.

Of course this extreme undercoverage for the Bayesian intervals reflects the fact that stationary models are likely to be estimated with larger sampling error than non-stationary models.[16] Bayesian inference produces intervals that are biased according to pre-sample probabilities because such inference takes appropriate account of the asymmetries in sampling distribution dispersion between stationary and non-stationary models.[17] We might expect then, that classical intervals would do badly in this situation in the sense that their posterior probabilities would deviate substantially from their nominal confidence levels. Table 5 shows that this is indeed the case, with the bootstrap intervals tending to have probability exceeding .68 at long time horizons and probability below .68 at some short horizons. We show only one representative random draw of a coefficient estimate for this case, because there is not much qualitative variation across samples in the results. Though the deviation from .68 is sometimes a little less, sometimes quite a bit worse, the pattern of excessive probability at long time horizons and deficient probability at some short horizons is pervasive. As can be seen from Table 6, the tendency of Bayesian intervals to have both lower coverage probability and lower posterior probability than the bootstrap intervals corresponds to a pronounced tendency for the Bayesian intervals to be narrower, particularly at long time horizons. Note the contrast with the cointegrated model, where the shorter length for the Bayesian intervals did not correspond to any general tendency for them to have less probability, either pre- or post-sample.

Next we consider two examples of bivariate models based on economic time series. For both these models we use 1000 Monte Carlo draws in generating Bayesian or bootstrap intervals for a given data set, and 600 draws of data sets in constructing coverage probabilities.

---

[16] See Sims and Uhlig [1991] for a more extended discussion of this point.

[17] The performance of the Bayesian interval by classical criteria *for this particular parameter value* could of course be improved by use of a prior favoring this parameter value. In fact we have verified that, for example, a standard Litterman prior, with standard error .045 on own lag coefficients and .022 on cross lag coefficients produces .68 posterior probability intervals whose coverage probability is uniformly better than that of the bootstrap model for this parameter setting. Of course this improved classical performance at this point in the parameter space is probably offset by poorer performance for strongly stationary models.

First we consider a triangularly orthogonalized reduced-form VAR with four lags fitted to quarterly data on real GNP and M1 over 1948:1-1989:3. Figure 5 shows 68% other-percentile intervals. In all of Figures 5-10 the dotted lines are 95% bands, and the heavy lines are 68% bands. Note that for the responses of GNP and M1 to their own innovations, the responses are quite asymmetric, shifted down toward 0. This reflects bias in the estimator, and it does not make sense to treat these intervals, which accurately reflect the shape of the distribution of the estimates about the truth, as confidence intervals for the true parameters. Figure 6 shows that the Bayesian posterior probability intervals correct for the bias, with the .68 probability bands skewed slightly *above* the point estimates for the own responses, as accords with common sense given the bias apparent in Figure 5. Compared to the Bayesian intervals, bias-adjusted bootstrap intervals, displayed in Figure 7, are more skewed upward, somewhat wider, and with a tendency to splay out at the longer time horizons.

Treating the point estimates of this model as true parameter values, we can examine coverage probabilities. As can be seen in Table 7, Bayesian .68 intervals have classical coverage probability systematically lower than .68, though never lower than .50. The bootstrap intervals have coverage probabilities less than .68 for short horizons, above .68 for long horizons, but the discrepancies are smaller than for the Bayesian intervals at most horizons. Table 8, computed for the posterior implied by the actual data, shows that the bootstrap intervals have posterior probability that exceeds .68, generally by about the same amounts by which Bayesian intervals' coverage probabilities fall below .68.

We turn now to the bivariate VAR with 8 lags fitted by Blanchard and Quah [1989] to quarterly data on real GNP growth and the unemployment rate for males 20 and over, for 1948:1-1987:4. We construct error bands for the structural impulse responses under the Blanchard-Quah identification, which relies on long-run restrictions. The original article presented what were meant to be other-percentile intervals, with an ad hoc correction to prevent their lying entirely to one side of the estimated responses. We were able to duplicate those results, but realized in the process that there were some errors in the paper's implementation of the bootstrap. The very strong asymmetry in the intervals displayed in the article resulted mostly from the errors.[18] We first show, in Figure 8, uncorrected other-percentile bootstrap intervals. They are modestly asymmetric, and Killian-style bias correction, whose results appear in Figure 10, has correspondingly modest effect on them. The corrected bands are somewhat wider and shifted away from the axis, especially at longer horizons. It is interesting to note that the bias correction for the estimated parameters has not much affected the strong skewness toward zero of the intervals on responses to demand at short horizons. The Bayesian intervals in Figure 9 are remarkably similar to the bias-corrected bootstrap intervals, tending to be slightly narrower in most cases, and largely reproducing the skewness toward zero at short horizons of the intervals on responses to demand.

---

[18] See the more detailed discussion in an earlier draft of this paper, Cowles Foundation Discussion Paper Number 1085.

If we take the estimated coefficients of the Blanchard-Quah model as the truth, we can check coverage probabilities for bootstrap and Bayesian intervals. Table 10 shows similar behavior for Bayesian and bootstrap intervals by classical criteria. Both intervals tend to undercover for the GNP-to-demand response at short horizons, though the undercoverage is considerably worse for the bootstrap interval. Both tend to overcover at long time horizons, with the tendency slightly worse for the Bayesian intervals. Table 11, computed using the posterior implied by the actual data, shows deficiency of the bootstrap interval by Bayesian criteria: intervals for responses to demand shocks at short time horizons have probability well under the nominal .68. This is not surprising, given Figure 12, which shows that for these short horizons the bootstrap 68% interval lies entirely or almost entirely to one side of the point estimate of the response. Though the Bayesian intervals are also skewed at these horizons, their somewhat more modest skewness results in a substantially different posterior probability. Table 12 shows that the mean interval lengths are very close for the two types, with the Bayesian intervals being slightly shorter at long horizons, but by amounts that could be due to Monte Carlo sampling error.

With these simple models we hope to have made our case that Bayesian error bands are about as reliable as bootstrap bands by classical criteria, while being by construction better by Bayesian criteria and in practice often very substantially better by Bayesian criteria. We now turn to considering a model closer to the scale now in use in analyzing macroeconomic policy. For this larger model, use of simulation methods to compute bootstrap intervals and Bayesian intervals is possible, but Monte Carlo study of the sampling properties of the intervals is computationally too ambitious for us. Also, the larger model involves overidentifying restrictions. It turns out that this makes computing Bayesian intervals more demanding, in ways that have not been recognized in some of the existing literature.

7. Monte Carlo Methods for Posterior Probabilities in Overidentified Models

We consider linear simultaneous equations models of the form

$$\Gamma(L)y(t) = e(t) \ . \tag{10}$$

We take

$$e(t)\big|y(s), s<t \sim N(0,\Lambda), \tag{11}$$

with Λ diagonal. We assume Γ_0 to be non-singular, so that (10) provides a complete description of the conditional distribution of $y(t)$ given $y(s), s<t$ and can be solved by multiplying through on the left by Γ_0^{-1} to produce the reduced form

$$B(L)y(t) = u(t) \ , \tag{12}$$

in which $B_0 = I$ and $u(t)$, while still uncorrelated with past y's, has a covariance matrix which is not in general diagonal, being given by

$$\Sigma = \Gamma_0^{-1}\Lambda\Gamma_0'^{-1} \ . \tag{13}$$

We assume the system is a finite-order autoregression, meaning that there is a $k < \infty$ such that $\Gamma_j = B_j = 0$ for all $j > k$.

The p.d.f. for the data $y(1),\ldots,y(T)$, conditional on the initial observations $y(-k+1),\ldots,y(0)$, is proportional to q as defined by

$$q(B,\Sigma) = |\Sigma|^{-\frac{T}{2}} \exp\left[-\tfrac{1}{2}\operatorname{trace}\left(S(B)\Sigma^{-1}\right)\right] \tag{14}$$

$$\hat{u}(t;B) = B(L)y(t) \tag{15}$$

$$S(B) = \sum_{t=1}^{T} \hat{u}(t,B)\hat{u}(t,B)' \ . \tag{16}$$

For a given sample, (14) treated as a function of the parameters B and Σ is the likelihood function. Its form is exactly that of the likelihood for a regression with Gaussian disturbances and strictly exogenous regressors, a classic model for which Bayesian calculations are well-discussed in the literature.[19] The RATS program includes routines to implement Monte Carlo drawing from the joint distribution of B and Σ and use of those draws to generate a Monte Carlo sample from the posterior distribution of impulse responses.[20]

The impulse responses for the model , defined by (5) above, are in this case the coefficients of

$$B^{-1}(L)\Gamma_0^{-1}\Lambda^{\frac{1}{2}} \ , \tag{17}$$

where the $\Lambda^{\frac{1}{2}}$ factor scales the structural disturbances to have unit variance, or equivalently converts the responses so they have the scale of a response to a disturbance of "typical" (one-standard-deviation) size. Equation (13) gives us a relation among Σ, Γ, and Λ. Because Σ is symmetric, Γ_0 and Λ have more unrestricted coefficients than Σ. An exactly identified VAR

---

[19] See, e.g., Box and Tiao [1973] Chapter 8 for the theory.

[20] Box and Tiao [1973] recommend using a Jeffreys prior on Σ, which turns out to be proportional to $|\Sigma|^{-\frac{m+1}{2}}$. The packaged RATS procedure uses instead $|\Sigma|^{-\frac{m+n+1}{2}}$, where n is the number of estimated coefficients per equation. Phillips [1992] suggests using the joint Jeffreys prior on B and Σ, which in time series models (unlike models with exogenous regressors) is not flat in B. The Phillips suggestion has the drawback that the joint Jeffreys prior is computationally inconvenient and changes drastically with sample size, making it difficult for readers to compare results across data sets. We therefore prefer the Box and Tiao suggestion in principle, though they point out (p. 44) that even in models with exogenous regressors mechanical use of Jeffreys priors can lead to anomalies. In this paper, to keep our results as comparable as possible to the existing applied literature, we have followed the RATS procedure's choice of prior.

model is one in which we have just enough restrictions available to make (13) a one-one mapping from Σ to Γ_0 and Λ. In this case, sampling from the impulse responses defined by (17) is straightforward: sample from the joint distribution of B and Σ by standard methods, then use the mapping defined by (13) and the restrictions to convert these draws to draws from the distribution of impulse responses. The most common use of this procedure restricts Γ_0 to be triangular, solving for $\Gamma_0^{-1}\Lambda^{\frac{1}{2}}$ by taking a Choleski decomposition of Σ.

When the model is not exactly identified, however, reliance on the standard methods and programs that generate draws from the joint distribution of the reduced form parameters is no longer possible. A procedure with no small-sample rationale that does use the standard methods has occurred independently to a number of researchers (including ourselves) and been used in at least one published paper (Gordon and Leeper [1994]). We will call it the naive procedure. Because the method has a misleading intuitive appeal and may sometimes be easier to implement than the correct method we describe below, we begin by describing it and explaining why it produces neither a Bayesian posterior nor a classical sampling distribution.

In an overidentified model, (13) restricts the behavior of the true reduced-form innovation variance matrix Σ. It remains true, though, that the OLS estimates $\hat{B}$ and $\hat{\Sigma}$ are sufficient statistics, meaning that the likelihood depends on the data only through them. Thus maximum likelihood estimation of B, Γ_0, and Λ implies an algorithm for mapping reduced form $\left(\hat{B},\hat{\Sigma}\right)$ estimates into structural estimates $\left(B^*,\Gamma_0^*,\Lambda^*\right)$ that satisfy the restrictions Often there are no restrictions that affect B, so that $\hat{B}=B^*$. The naive method proceeds by drawing from the unrestricted reduced form's posterior p.d.f. for (B, Σ), then mapping these draws into values of $\left(B,\Gamma_0,\Lambda\right)$ via the maximum likelihood procedure, as if the parameter values drawn from the unrestricted posterior on (B, Σ) were instead reduced form parameter estimates. The resulting distribution is of course concentrated on the part of the parameter space satisfying the restrictions, but is not a parametric bootstrap classical distribution for the parameter estimates, because the posterior distribution for (B, Σ) is not a sampling distribution for $\left(\hat{B},\hat{\Sigma}\right)$. It is not a true posterior distribution because the unrestricted posterior distribution for (B, Σ) is not the restricted posterior distribution, and mapping it into the restricted parameter space via the estimation procedure does not convert it into a restricted posterior distribution.

The procedure does have the same sort of asymptotic justification that makes nearly all bootstrap and Bayesian methods of generating error bands asymptotically equivalent from a classical point of view for stationary models, and it is probably asymptotically justified from a Bayesian viewpoint as a normal approximation even for non-stationary models. To see this, consider a simple normal linear estimation problem, where we have a true parameter b, an unrestricted estimate distributed as $N\left(b,\Omega\right)$, and a restriction $Rb=g$ with R $k\times m$. The restricted maximum likelihood estimate is then the projection on the $Rb=g$ manifold of the unrestricted ML estimate $\hat{b}$, under the metric defined by Ω, i.e.

$$\hat{b}^* = \Phi\left(\Phi'\Omega^{-1}\Phi\right)^{-1}\Phi'\Omega^{-1}\hat{b} + M\left(M'\Omega^{-1}M\right)^{-1}g \ , \tag{18}$$

where $M = \Omega R'$ and Φ is chosen to be of full column rank *m-k* and to satisfy $R\Phi = 0$. The sampling distribution of $\hat{b}^*$ is then in turn normal, since it is a linear transformation of the normal $\hat{b}$. In this symmetrically distributed, pure location-shift problem, the unrestricted posterior on b has the same normal p.d.f., centered at $\hat{b}$, as the sampling p.d.f. of $\hat{b}$ about b. We could make Monte Carlo draws from the sampling distribution of $\hat{b}^*$ by drawing from the sampling distribution of $\hat{b}$, the unrestricted estimate, and projecting these unrestricted estimates on the restricted parameter space using the formula (18). But since in this case the posterior distribution of $\hat{b}$ and its sampling distribution are the same, drawing from the posterior distribution in the first step would give the same correct result. And since in this case the restricted posterior has the same normal shape about $\hat{b}^*$ that the sampling distribution of $\hat{b}^*$ has about b, the simulated distribution matches the posterior as well as the sampling distribution of the restricted estimate.

The naive method for sampling from the distribution of impulse responses rests on confusing sampling distributions with posterior distributions, but in the case of the preceding paragraph this would cause no harm, because the two kinds of distribution have the same shape. For stationary models, distribution theory for $\hat{\Sigma}$ and $\hat{B}$ is asymptotically normal, and differentiable restrictions will behave asymptotically as if they were linear. So the case considered in the previous paragraph becomes a good approximation in large samples. For stationary or non-stationary models, the posterior on Σ is asymptotically normal, so the naive method is asymptotically justified from a Bayesian point of view.

But in this paper we are focusing on methods that produce error bands whose possible asymmetries are justifiably interpreted as informative about asymmetry in the posterior distribution of the impulse responses or in the sampling distribution of their estimates. Any asymmetries that appear in bands generated by the naive method are interpretable only as evidence that the asymptotic approximations that might justify the method are not holding in the sample at hand.

To describe a correct procedure for generating Monte Carlo draws from the Bayesian posterior for the parameters in (10), we begin by introducing a reparameterization. In place of (10) we use

$$A(L)y(t) = h(t) \ , \tag{19}$$

where $A = \Lambda^{-\frac{1}{2}}\Gamma$ and $h(t) = \Lambda^{-\frac{1}{2}}e(t)$ so that $\text{var}\left(h(t)\right) = I$. There is no real reduction in the number of free parameters, because the diagonal of Γ_0 is always normalized to a vector of ones. so that an unrestricted A_0 has the same number of parameters as a diagonal Λ together with a normalized Γ_0. There are a number of reasons to prefer this parameterization. It simplifies the mapping (17) between reduced form and structural parameters. The usual parameterization can

give a misleading impression of imprecision in the estimate of a row of Γ_0 if the normalization happens to set to 1 a particularly ill-determined coefficient in the corresponding row of A_0. But the main reason for introducing this parameterization here is that the likelihood is not in general integrable when written as a function of (B, Γ_0, Λ), requiring some adjustment of the flat prior to allow formation of a posterior distribution, whereas it is integrable as a function of (B, A_0).[21]

We can rewrite the likelihood(14) as

$$|A_0|^T \exp\left\{-\frac{1}{2}\text{trace}\left(A_0' A_0 S(\hat{B})\right) - \frac{1}{2}\text{trace}\left((B-\hat{B})' X' X (B-\hat{B}) A_0' A_0\right)\right\} . \tag{20}$$

Taking the prior as flat in B and A_0, we can integrate over B to obtain the marginal posterior on A_0,

$$p(A_0) \propto |A_0|^{T-n} \exp\left[-\frac{1}{2}\text{trace}\left(A_0 S(\hat{B}) A_0'\right)\right] . \tag{21}$$

Here as with the reduced-form model we have followed the widely used RATS code in dropping the "degrees of freedom correction" -n in (21), in effect using $|A_0|^n$ as an improper prior. As can be seen by comparing (20) with (21), this has the effect of making the marginal posterior on A_0 proportional to the concentrated likelihood and thereby eliminating possible discrepancies between posterior modes and maximum likelihood estimates.

Expression (21) is not in the form of any standard p.d.f. To generate Monte Carlo samples from it, we first take a second-order Taylor expansion of it about its peak, which produces the usual Gaussian approximation to the asymptotic distribution of the elements of A_0. We draw a sample from this distribution, but weight the draws by the ratio of (21) to the Gaussian p.d.f. from

---

[21] In general, the likelihood, with B and Λ integrated out, is $O(1)$ in individual elements of Γ_0 at all sample sizes. This does not mean that it fails to be asymptotically normal -- the likelihood does become small outside a region around the true parameter values. There are also special cases where the likelihood is integrable, e.g. where Γ_0 is normalized with ones down the diagonal and restricted to be triangular. For details, see the earlier draft of this paper, Cowles Foundation Discussion Paper No. 1085. Of course, we could use the standard parameterization and retain integrability if we made the prior flat in A_0, then transformed back to the (Γ_0, Λ) parameter space, including the appropriate Jacobian term $|\Lambda|^{-\frac{m+1}{2}}$. But the easiest way to explain the appearance of this term would be to derive it from the A_0 parameterization, which is in any case more well-behaved.

which we draw. The weighted sample c.d.f. then approximates the c.d.f. corresponding to (21).[22] The weights often vary rather widely, so that a given degree of Monte Carlo sampling error in impulse response bands computed this way generally requires several times as many Monte Carlo draws as for a reduced form model where weighting is not required. Note that it also possible to compute the error bands without any weighting. This is yet another example of a method for computing error bands that is asymptotically justified, but admits no rationale for interpreting its asymmetries as providing information about small-sample deviations from normality.

Note that, though in switching to the A_0 parameterization we have eliminated the need to choose a normalization in the usual sense, there is still a discrete redundancy in the parameterization: the sign of a row of A_0 can be reversed without changing the likelihood function. We eliminate the redundancy by considering only A_0's with non-negative diagonal elements. Since the asymptotic approximation to the posterior on A_0 is Gaussian, when we draw from it we generate some A_0's with negative diagonal elements. We simply discard such draws.[23]

8. Results for a Six-Variable Model

An earlier paper by one of us (Sims [1986]) contains an example of an overidentified VAR model with an interesting interpretation. (It identifies a money supply shock that produces both a liquidity effect -- an initial decline in interest rate -- and a correct-signed price effect -- inflation following a monetary expansion.) The paper contains two sets of identifying restrictions, and we have computed error bands for both. One set is not displayed here, because it gave the uninteresting outcome that all non-bootstrap methods of computing error bands gave about the same, symmetric, results, whether by correct Bayesian calculations, unweighted Bayesian Monte Carlo, or the naive method.[24] It is worth noting that this can happen; one should not be satisfied generally with the simpler methods of calculating error bands on the basis of a few test cases where more sophisticated calculations appear not to make a difference. The other version of the model, called identification 2 in the original paper, produced substantial differences across methods.

---

[22] This idea, importance sampling, has seen wide use since its introduction into econometrics by Kloek and Van Dijk [1978].

[23] If we retained such draws, flipping the sign on deviant rows of A_0, we would have to recognize that even when we have drawn an A_0 with positive diagonal, the density of the distribution from which we are drawing is not simply the Gaussian density at that point, but the sum of densities over all A_0's that map into that point via sign-reversals of diagonal elements. This seemed likely to slow down calculations by as much as does discarding the deviant draws.

[24] Other-percentile method bootstrap intervals gave different results. We did not compute bands by Killian's method for this model.

Figure 11 shows impulse responses for this model, with .68 and .95 probability bands computed by the importance-sampling Bayesian Monte Carlo method. In all of Figures 11-13, longer-dashed lines are 95% bands and dotted lines are 68% bands. The .68 bands are tight in Figure 11, which makes the .95 bands useful.. The bands do not show much asymmetry, though there is some, particularly for longer horizons and the .95 bands. For the model's interpretation, the second column of the figure, which is identified as the responses to a money demand shock, are particularly interesting. Though most of them look like what would be expected of a money demand shock, the fourth, showing the response of the price level, may not be. It shows immediate and strong inflation following the money demand increase.

This is not the place to discuss whether this discredits the model. (We are not sure it does. Equilibrium models can generate responses something like this to money demand shocks.) But it is clear that the price response is estimated as sharply positive, with very small probability of being zero or negative. As we shall see, only this correct method of computing the posterior gives this sharp result.

Figure 12 shows responses to money supply and demand shocks with bands computed by several methods. The uncorrected other-percentile bootstrap shows strong bias, with the 68% band on the response of prices to money demand being very wide, not including the point estimate, and nearly including the axis. The naive method produces even wider bands on this crucial response, again with the lower bound very close to the axis. The unweighted method produces somewhat narrower bands, but the 68% band still is skewed toward zero, close enough that it might suggest that the price response has at least a 2-3% chance of being zero in fact. Thus all three estimates fail to show the high precision with which this debatable response is actually estimated.

The bias-corrected bootstrap here produces bad results, as can be seen in Figure 13. Probably because of the severe bias apparent in the uncorrected bootstrap estimates in Figure 12, bias-correction as we have implemented it pushes the bootstrap distribution unreasonably far into the nonstationary region and produces strongly asymmetric, rapidly explosive intervals that at a few points even fail to include point estimates. Like those in Figure 12, though, they are skewed downward for the price-to-money-demand response and have a lower 68% limit that runs along the axis.

9. Computing Times

Forming 1000 draws from the posterior distribution for the impulse responses for our four-lag Y-M model, for example, takes less than 30 seconds using RATS code on a Pentium-90. The same is true for a bootstrap sample with a similar number of draws. For the six-variable model, the same calculation takes 2.13 minutes for the Bayesian interval and 3 minutes for the bootstrap interval. To calculate equal-tailed intervals requires storing the draws and sorting the draws for each $a_{ij}(s)$. This makes demands on sorting efficiency and on management of memory and disk storage which we have not optimized very well. Replicating these calculations 600 times to get coverage probabilities for the Y-M, for example, would take 5 hours for the other-percentile bootstrap, but our actual calculations took closer to 14 hours because of the need to sort and shuffle data on the disk. We think these times could be much reduced. Figure 11 is based on

20000 draws from the posterior, with 4936 of these discarded for negative diagonals. Figure 12 is based on 1000 bootstrap draws. The bootstrap-after-bootstrap calculations for Figure 13 used 1000 bootstrap draws in each layer of the bootstrap.

10. Conclusion

We have discussed the logical foundations of Bayesian and bootstrap inference for these models and displayed the behavior of such intervals with actual and simulated data. We hope this has led the reader to agree with us that posterior probability intervals are in principle more useful than confidence intervals and ought therefore to be the standard reporting device for these models (as in any other case where they differ from confidence intervals). Even readers who remain attached to the idea that only data, not parameters, can be given probability distributions ought to find our results useful. They confirm that unadjusted other-percentile bootstrap intervals are far worse on classical criteria than Bayesian intervals. They also show that Bayesian intervals deserve attention even on purely classical criteria. Compared to carefully constructed bootstrap intervals they are computationally more efficient, similarly asymptotically justified, and similar on the whole in the accuracy of their small-sample coverage probabilities.

References

Blanchard, O.J. and D. Quah: "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79(1989), 655-73

Box, G.E.P. and George Tiao: *Bayesian Inference in Statistical Analysis*. Addison-Wesley, 1973.

Efron, Bradley and Robert J. Tibshirani: *An Introduction to the Bootstrap*, New York and London: Chapman and Hall, 1993.

Gordon, D.B. and E.M. Leeper: "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification," *Journal of Political Economy*, 102(1994), 1228-1247.

Hall, Peter: *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, 1992.

Hildreth, Clifford: "Bayesian Statisticians and Remote Clients," *Econometrica*, 31(1963), 422-439.

Killian, Lutz: "Small-Sample Confidence Intervals for Impulse Response Functions," processed, Department of Economics, University of Pennsylvania, 1995.

Kloek, T. and H.K. van Dijk: "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica*, 46(1978), 1-19.

Koop, G.: "Aggregate Shocks and Macroeconomic Fluctuations: A Bayesian Approach," *Journal of Appllied Econometrics*, 7(1992), 395-411.

Lastrapes, W.D. and G. Selgin: "Buffer-Stock Money: Interpreting Short-Run Dynamics Using Long-Run Restrictions," *Journal of Money, Credit and Banking*, 26(1994), 34-54.

Lutkepohl, H.: "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *The Review of Economics and Statistics*, 72(1990), 53-78.

Mittnik, S. and P.A. Zadrozny: "Asymptotic Distributions of Impulse Responses, Step Responses, and Variance Decompositions of Estimated Linear Dynamic Models," *Econometrica*, 20(1993), 832-854.

Phillips, P.C.B.: "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 7(1992).

Poterba, J.M., J.J. Rotemberg and L.H. Summers: "A Tax-Based Test for Nominal Rigidities," *American Economic Review*, 76(1986), 659-675.

RATS, computer program available from Estima, 1800 Sherman Ave., Suite 612, Evanston IL 60201.

Runkle, D.E.: "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics*, **5(**1987), 437-442.

Sims, C.A.: "Are Forecasting Models Usable for Policy Analysis?" Federal Reserve Bank of Minneapolis *Quarterly Review* **10(**1986), 2-15.

Sims, C., J. Stock and M. Watson: "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58(1989), 113-144.

Sims, C.A. and H. Uhlig: "Understanding Unit Rooters: A Helicopter Tour," *Econometrica*, 59**(**1991), 1591-1599.

Table 1

Classical Performance Criteria, Cointegrated Model

Coverage Probabilities for Bootstrap Intervals					Coverage Probabilities for Bayesian Intervals				
t	1 to 1	2 to 1	1 to 2	2 to 2	*t*	1 to 1	2 to 1	1 to 2	2 to 2
1	.582	.605	.000	.553	1	.663	.658	.000	.658
2	.588	.617	.665	.653	2	.658	.660	.665	.647
3	.603	.628	.660	.667	3	.670	.670	.668	.657
4	.615	.623	.657	.662	4	.673	.678	.673	.660
6	.623	.632	.652	.657	6	.677	.678	.670	.665
8	.650	.635	.655	.653	8	.682	.683	.673	.677
12	.650	.645	.655	.652	12	.678	.680	.682	.675
16	.662	.657	.653	.650	16	.677	.675	.682	.677
24	.673	.668	.655	.652	24	.672	.680	.677	.675
32	.692	.690	.657	.652	32	.675	.677	.680	.677

Note: Monte Carlo standard error of a .68 frequency with 600 draws is .019. Bootstrap intervals have nominal coverage probability of .68. Bayesian intervals are flat-prior, equal-tail, .68 posterior probability intervals.

Table 2

Bayesian Performance Criteria, Cointegrated Model

Sample 1						Sample 2				
t	1 to 1	2 to 1	1 to 2	2 to 2		*t*	1 to 1	2 to 1	1 to 2	2 to 2
1	.634	.624	.000	.594		1	.613	.604	.000	.632
2	.629	.614	.621	.657		2	.622	.598	.692	.671
3	.648	.634	.629	.657		3	.639	.616	.704	.683
4	.673	.651	.640	.667		4	.670	.645	.707	.712
6	.686	.694	.645	.683		6	.710	.685	.717	.732
8	.694	.701	.652	.683		8	.723	.716	.726	.736
12	.715	.712	.667	.672		12	.750	.746	.744	.748
16	.725	.726	.677	.675		16	.757	.752	.765	.758
24	.717	.719	.680	.680		24	.787	.782	.768	.764
32	.700	.707	.677	.677		32	.781	.776	.774	.771
Sample 3						**Sample 4**				
t	1 to 1	2 to 1	1 to 2	2 to 2		*t*	1 to 1	2 to 1	1 to 2	2 to 2
1	.602	.648	.000	.720		1	.576	.608	.000	.700
2	.622	.645	.658	.650		2	.636	.637	.623	.663
3	.659	.655	.660	.658		3	.649	.663	.609	.673
4	.676	.689	.666	.675		4	.665	.673	.614	.678
6	.702	.708	.677	.696		6	.676	.693	.606	.634
8	.701	.724	.682	.699		8	.663	.678	.606	.626
12	.729	.733	.693	.698		12	.660	.669	.612	.615
16	.749	.737	.703	.702		16	.669	.667	.618	.617
24	.764	.762	.703	.708		24	.684	.682	.614	.618
32	.780	.773	.696	.708		32	.698	.694	.614	.615

Note: Monte Carlo standard error of a .68 frequency with 1000 draws is .015. Bootstrap intervals have nominal coverage probability of .68. The samples are random draws with initial *y*'s both 1, $\Omega = \begin{bmatrix} 1.16 & 1 \\ 1 & 1 \end{bmatrix}$, sample size 100. The four randomly drawn estimated coefficient matrices were

	Truth		1		2		3		4	
A'	0.9	0.1	1.060	.329	1.180	.355	1.174	.364	1.268	.536
	0.1	0.9	-.058	.674	-.172	.654	-.165	.646	-.263	.473
c'	1	1	1.056	.987	.737	.769	.700	.718	1.006	.888

Table 3

Cointegrated Model

Ratios of Interval Lengths				
Bayesian over Bootstrap				
t	1 to 1	2 to 1	1 to 2	2 to 2
1	1.044	1.040	1.000	1.029
2	1.027	1.034	1.024	1.028
3	1.014	1.023	.996	1.009
4	.999	1.010	.973	.986
6	.968	.980	.933	.946
8	.941	.954	.901	.914
12	.900	.914	.853	.864
16	.874	.885	.819	.829
24	.847	.855	.776	.784
32	.836	.843	.750	.755

Note: Monte Carlo standard errors of these ratios vary, but all are below .03.

Table 4

Classical Performance Criteria, Random Walks with Drift

Bootstrap Interval Coverage Probabilities					Bayesian Interval Coverage Probabilities				
t	1 to 1	2 to 1	1 to 2	2 to 2	*t*	1 to 1	2 to 1	1 to 2	2 to 2
1	.542	.688	.000	.510	1	.652	.640	.000	.630
2	.460	.617	.535	.447	2	.530	.528	.295	.480
3	.437	.575	.535	.427	3	.442	.433	.295	.387
4	.453	.530	.535	.435	4	.377	.373	.295	.330
6	.462	.503	.535	.442	6	.320	.320	.295	.295
8	.470	.507	.535	.467	8	.287	.305	.295	.280
12	.482	.517	.535	.485	12	.282	.285	.295	.282
16	.485	.517	.535	.497	16	.288	.288	.295	.280
24	.500	.517	.535	.510	24	.278	.285	.295	.283
32	.505	.527	.535	.520	32	.292	.283	.295	.295

Note: Monte Carlo standard error of a .68 frequency with 600 draws is .019. Bootstrap intervals have nominal coverage probability of .68. Bayesian intervals are flat-prior, equal-tail, .68 posterior probability intervals.

Table 5

Bayesian Performance Criteria, Random Walks with Drift

Posterior Probability of Bootstrap Interval				
t	**1 to 1**	**2 to 1**	**1 to 2**	**2 to 2**
1	.616	.618	.000	.569
2	.677	.645	.568	.707
3	.706	.640	.563	.737
4	.713	.626	.567	.722
6	.679	.628	.574	.721
8	.660	.643	.581	.718
12	.672	.686	.632	.715
16	.688	.715	.664	.768
24	.741	.804	.696	.823
32	.769	.851	.734	.860
	A'	.965	.061	
		.036	.942	
	c'	1.170	.751	

Note: Monte Carlo standard error of a .68 frequency with 1000 draws is .015. Bootstrap intervals have nominal coverage probability of .68. The samples are random draws with initial *y*'s both 1.

Table 6

Random Walks with Drift

Ratios of Interval Lengths				
Bayesian over Bootstrap				
t	**1 to 1**	**2 to 1**	**1 to 2**	**2 to 2**
1	1.039	1.026	1.000	1.027
2	.968	.963	.934	.951
3	.916	.912	.905	.898
4	.878	.873	.878	.859
6	.819	.811	.830	.803
8	.771	.762	.787	.758
12	.688	.680	.707	.679
16	.611	.603	.632	.606
24	.461	.450	.484	.461
32	.313	.304	.336	.319

Note: Monte Carlo standard errors of these ratios vary, but all are below .06.

Table 7

Classical Performance Criteria, GNP-M1 Model

Bootstrap Interval Coverage Probabilities					Bayesian Interval Coverage Probabilities			
Y to Y	M to Y	Y to M	M to M	*t*	Y to Y	M to Y	Y to M	M to M
.432	.687	.000	.360	1	.595	.668	.000	.595
.540	.678	.717	.502	2	.618	.672	.685	.543
.603	.692	.685	.592	3	.577	.687	.643	.557
.625	.670	.688	.622	4	.593	.690	.613	.518
.657	.702	.687	.653	6	.585	.677	.618	.523
.650	.713	.685	.655	8	.560	.670	.610	.510
.682	.717	.697	.675	12	.555	.642	.627	.518
.703	.732	.707	.687	16	.557	.623	.608	.513
.712	.730	.742	.712	24	.570	.642	.617	.517
.725	.742	.748	.750	32	.602	.632	.610	.520

Note: Monte Carlo standard error of a .68 frequency with 600 draws is .019. Bootstrap intervals have nominal coverage probability of .68. Bayesian intervals are flat-prior, equal-tail, .68 posterior probability intervals.

Table 8

Bayesian Performance Criteria, GNP-M1 Model

Posterior Probability of Bootstrap Interval				
t	Y to Y	M to Y	Y to M	M to M
1	.613	.685	.000	.809
2	.723	.722	.680	.799
3	.716	.697	.683	.780
4	.742	.713	.678	.760
6	.729	.751	.732	.725
8	.726	.769	.765	.702
12	.778	.790	.818	.671
16	.788	.822	.841	.675
24	.801	.831	.868	.716
32	.809	.832	.883	.776

Note: Monte Carlo standard error of a .68 frequency with 1000 draws is .015. Bootstrap intervals have nominal coverage probability of .68. The samples are random draws with initial *y*'s both 1.

Table 9

GNP-M1 Model

Ratios of Mean Lengths: Bayesian/Bootstrap				
t	Y to Y	M to Y	Y to M	M to M
1	1.022	1.016	1.000	1.020
2	.995	.988	1.015	.983
3	.967	.961	.982	.949
4	.938	.929	.945	.918
6	.895	.880	.892	.872
8	.841	.829	.836	.828
12	.759	.762	.750	.756
16	.716	.730	.708	.721
24	.652	.691	.644	.677
32	.603	.661	.599	.647

Table 10

Classical Performance Criteria, Blanchard-Quah Model

Bootstrap Interval Coverage Probabilities					Bayesian Interval Coverage Probabilities				
t	Y to S	U to S	Y to D	U to D	*t*	Y to S	U to S	Y to D	U to D
1	.657	.650	.050	.632	1	.653	.662	.273	.672
2	.648	.653	.108	.425	2	.658	.655	.270	.618
3	.647	.642	.225	.295	3	.650	.668	.372	.462
4	.657	.642	.337	.297	4	.658	.660	.445	.448
6	.678	.647	.550	.452	6	.642	.637	.615	.505
8	.670	.672	.593	.525	8	.647	.623	.652	.600
12	.687	.705	.630	.635	12	.690	.705	.658	.632
16	.728	.775	.645	.707	16	.752	.792	.693	.725
24	.747	.845	.783	.780	24	.768	.860	.845	.853
32	.737	.910	.928	.865	32	.760	.923	.952	.920

Note: Monte Carlo standard error of a .68 frequency with 600 draws is .019. Bootstrap intervals have nominal coverage probability of .68. Bayesian intervals are flat-prior, equal-tail, .68 posterior probability intervals.

Table 11

Bayesian Performance Criteria, Blanchard-Quah Model

Posterior Probabilities of Bootstrap Intervals				
t	Y to S	U to S	Y to D	U to D
1	.710	.696	.452	.644
2	.706	.715	.583	.616
3	.701	.717	.655	.611
4	.688	.712	.700	.661
6	.693	.696	.712	.673
8	.703	.698	.705	.731
12	.687	.718	.700	.692
16	.693	.713	.689	.684
24	.668	.714	.670	.676
32	.655	.721	.707	.718

Note: Monte Carlo standard error of a .68 frequency with 1000 draws is .015. Bootstrap intervals have nominal coverage probability of .68. The samples are random draws with initial *y*'s both 1.

Table 12

Blanchard-Quah Model

Ratios of Mean Lengths: Bayesian/Bootstrap				
t	Y to S	U to S	Y to D	U to D
1	1.009	1.004	.996	1.008
2	1.005	1.001	.985	1.006
3	.997	.999	.979	.993
4	.990	.992	.979	.984
6	.972	.981	.997	.984
8	.962	.975	.991	.994
12	.972	.985	.995	.999
16	.979	.983	1.005	1.016
24	.976	.942	.984	1.001
32	.963	.886	.958	.969

Note: Monte Carlo standard errors of these figures vary, but none are over .06.

Figure 1

Example 2 p.d.f. for $m = 0$ and $m = 1$

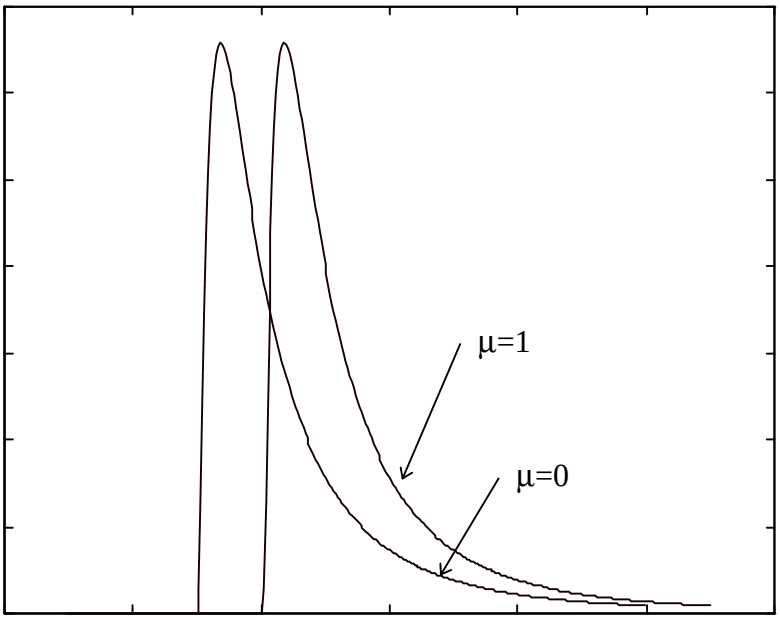


Figure 2

Example 2 Likelihood, $X = 1$

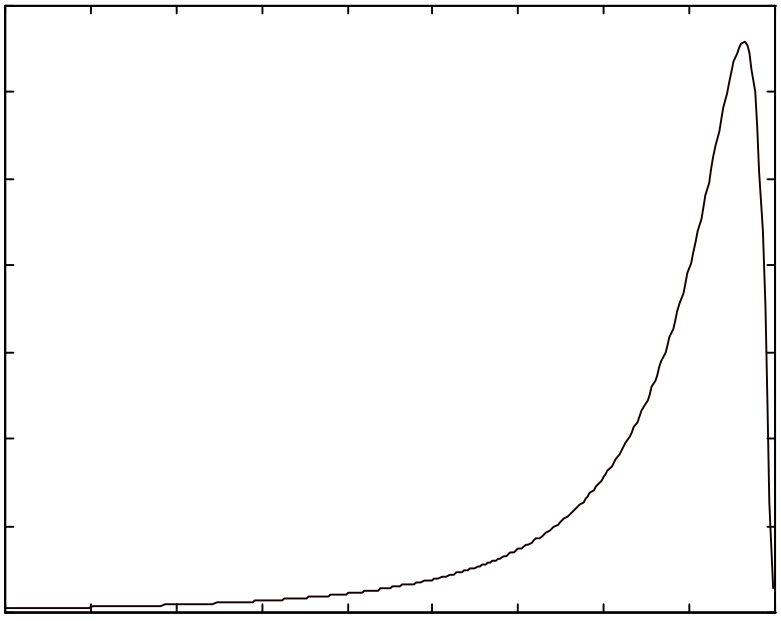


Figure 3
Example 3 p.d.f. for μ=1 and μ=1/*e*

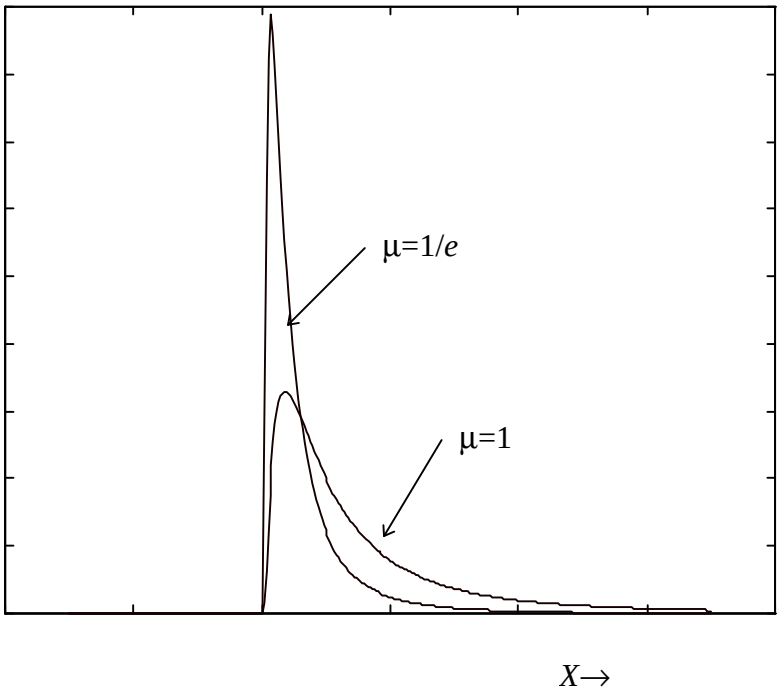


Figure 4
Example 3 Likelihood for X=1

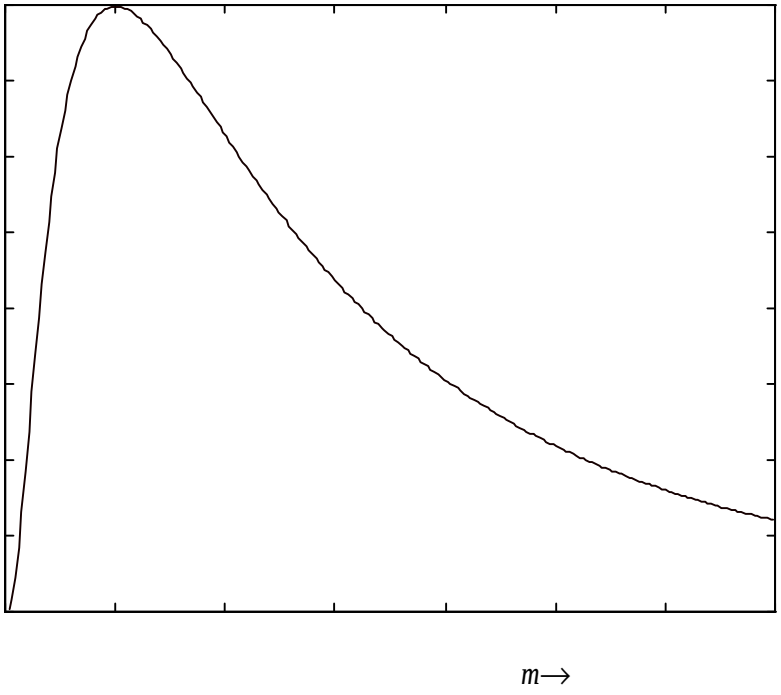
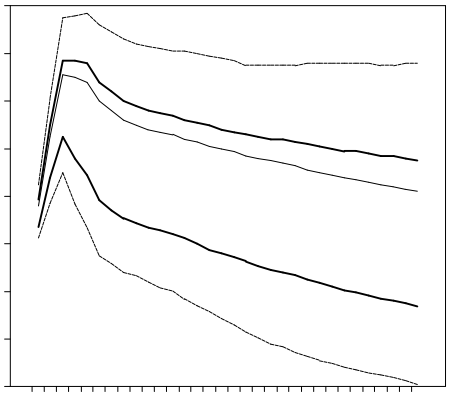


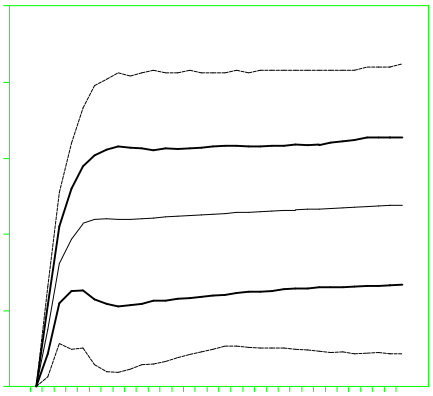
Figure 5

68% and 98% Other-Percentile Bootstrap Bands, Y-M Model

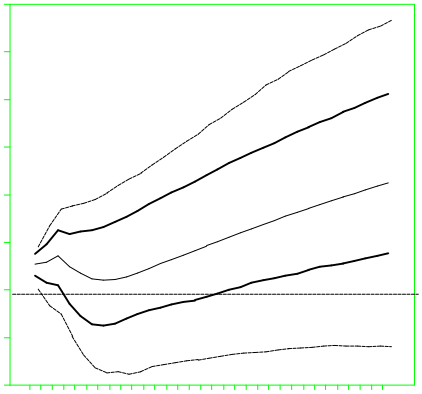
Responses to Y *Responses to M*

Y





M



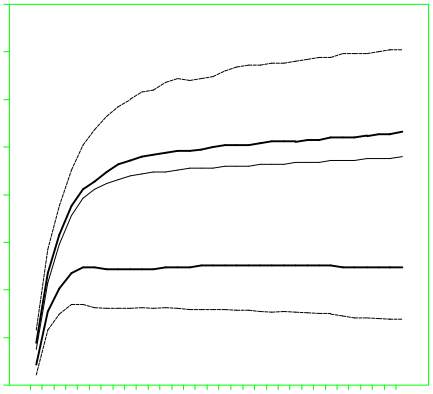


Figure 6

.68 and .95 Posterior Probability Bands, Y-M Model

Responses to Y *Responses to M*

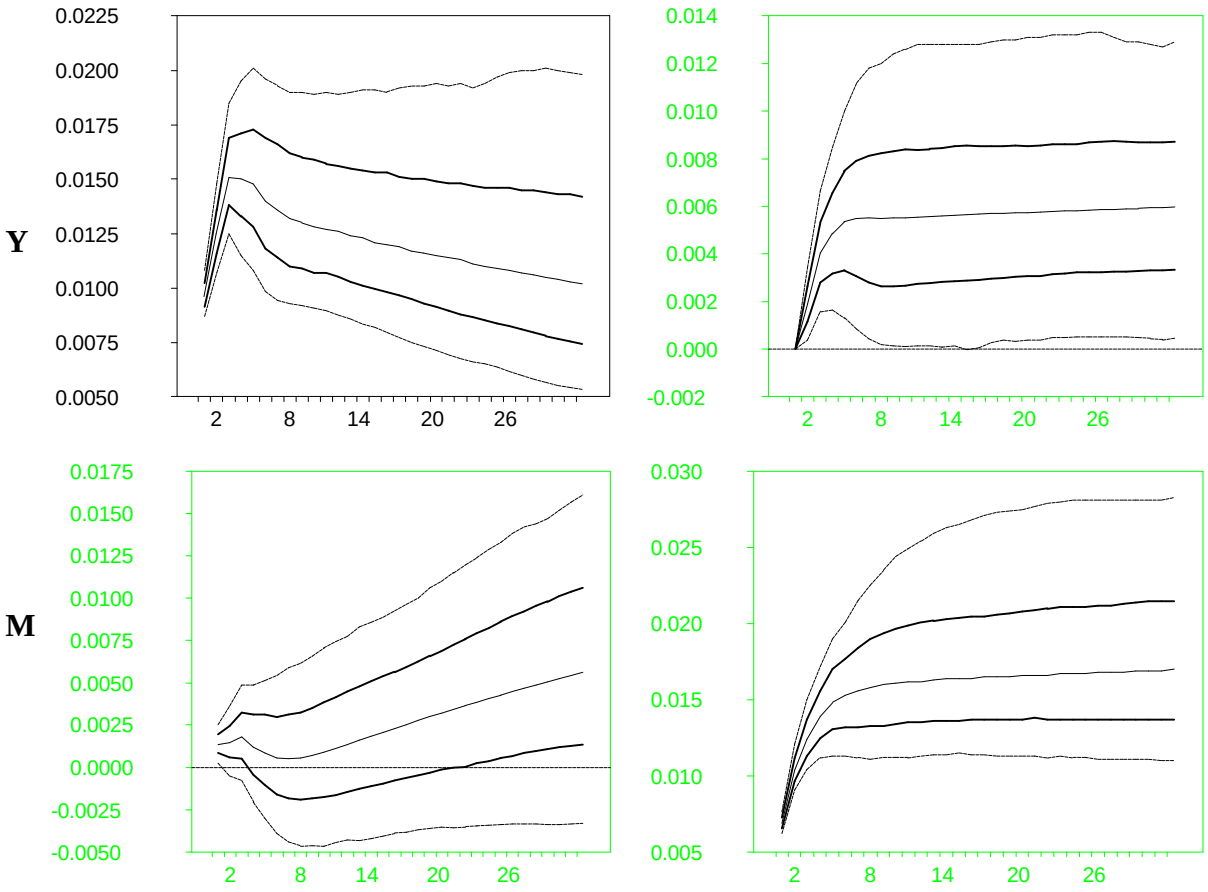


Figure 7

68% and 95% Bias-Corrected Bootstrap Bands, Y-M Model

Responses to Y *Responses to M*

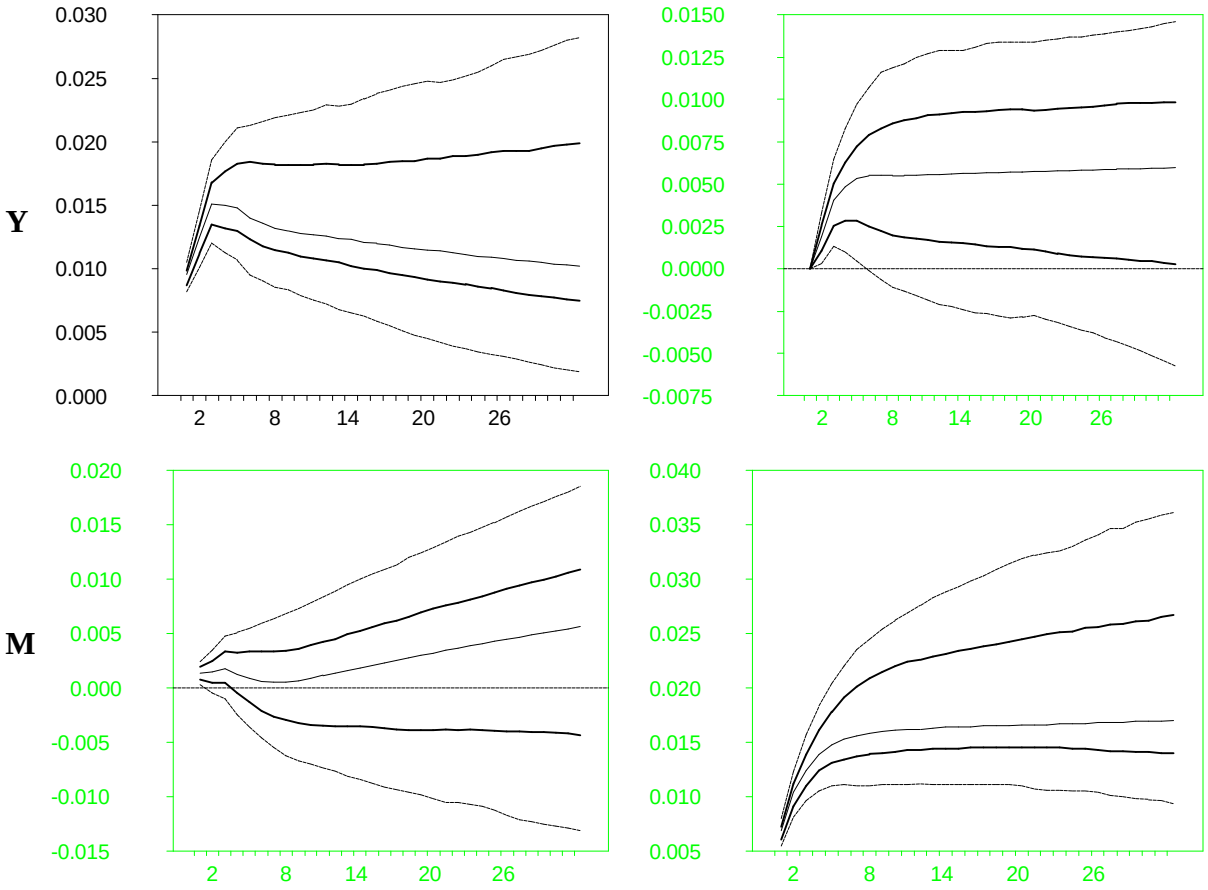


Figure 8

68% and 95% Other-Percentile Bootstrap Bands, B-Q Model

Responses to Supply *Responses to Demand*

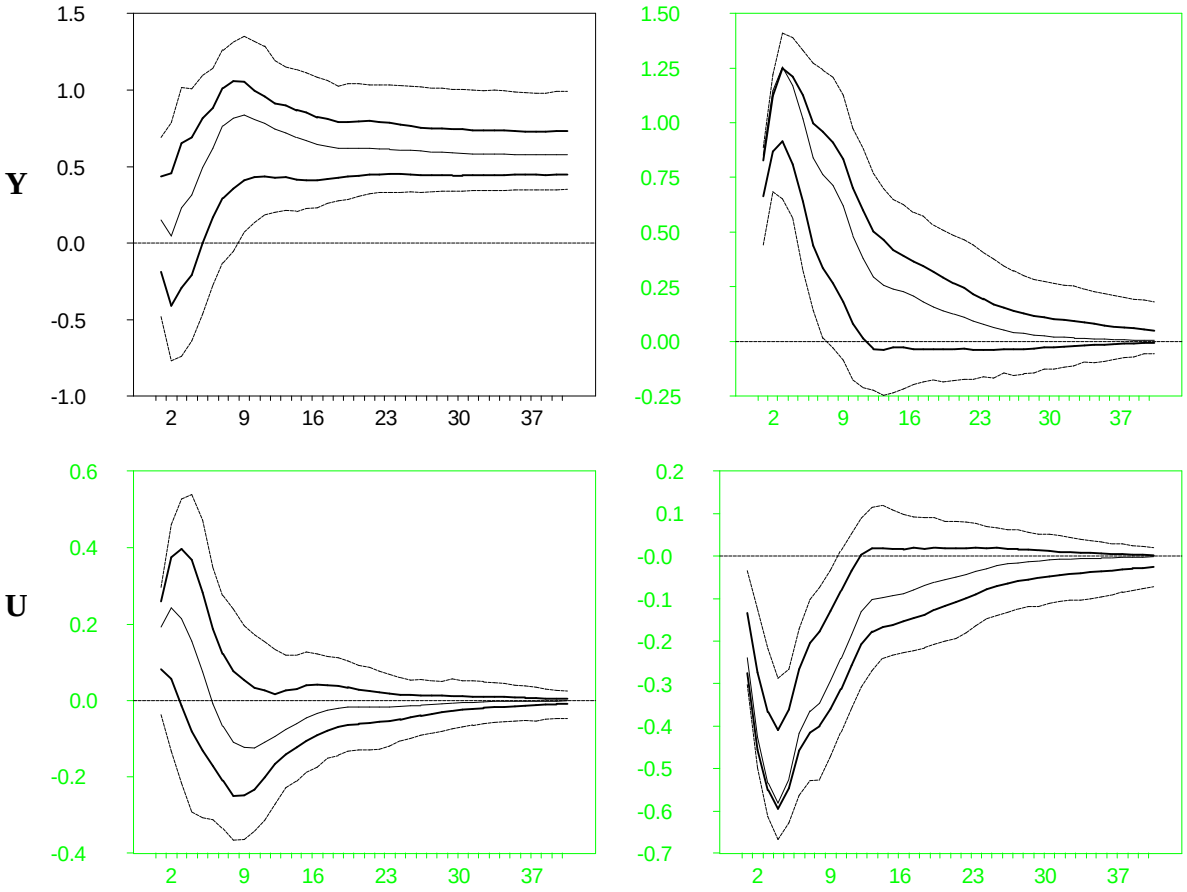


Figure 9

.68 and .95 Posterior Probability Bands, B-Q Model

Responses to Supply *Responses to Demand*

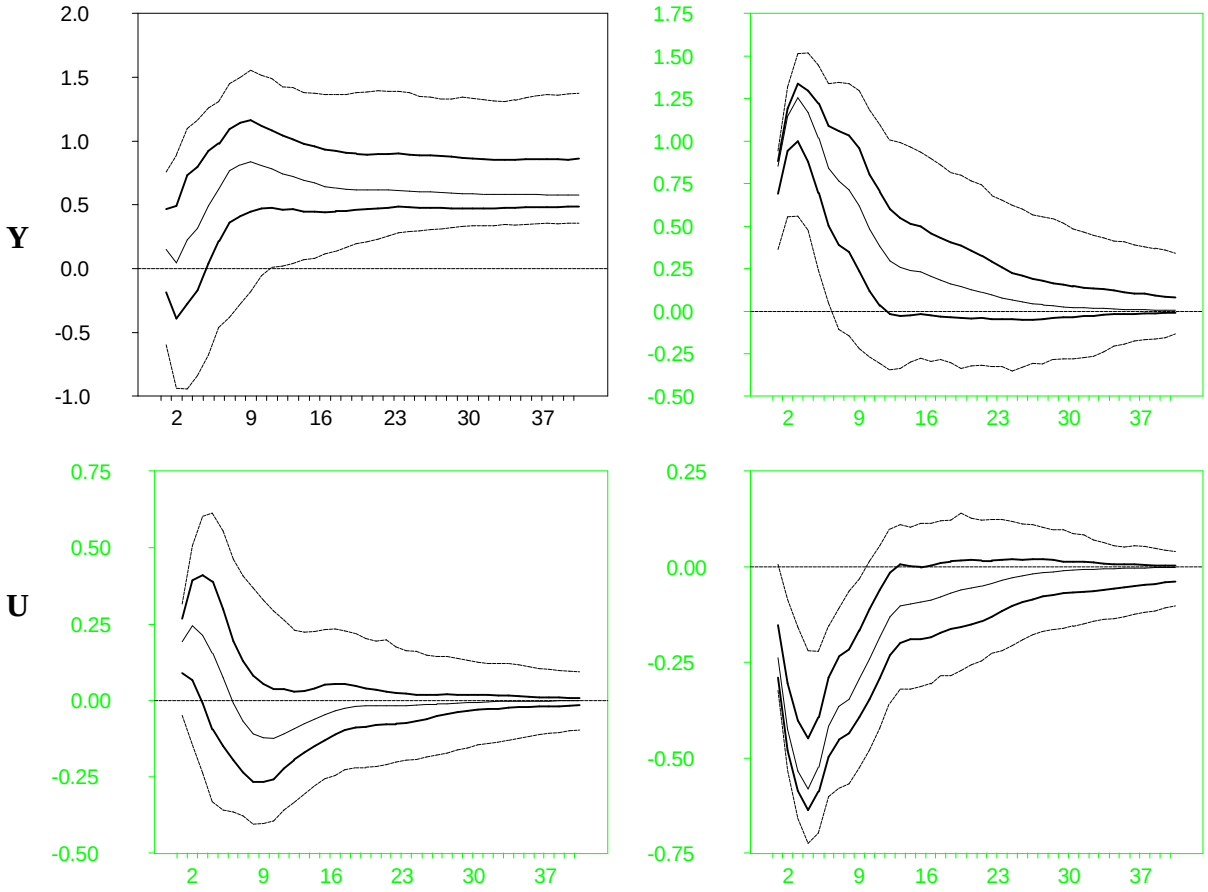


Figure 10

68% and 95% Bias-Corrected Bootstrap Bands, B-Q Model

Responses to Supply *Responses to Demand*

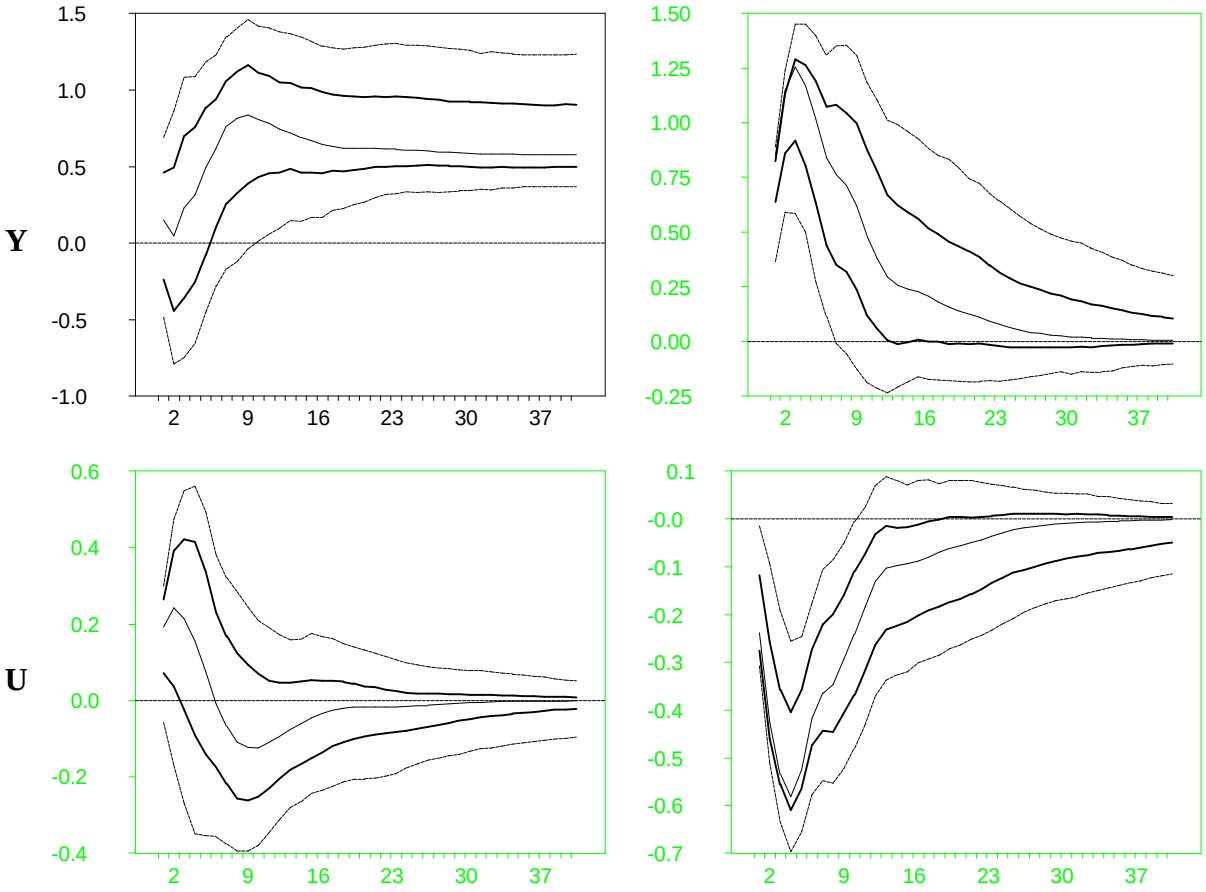


Figure 11

Six-Variable Model, Impulse Responses with .68 and .95 Probability Bands

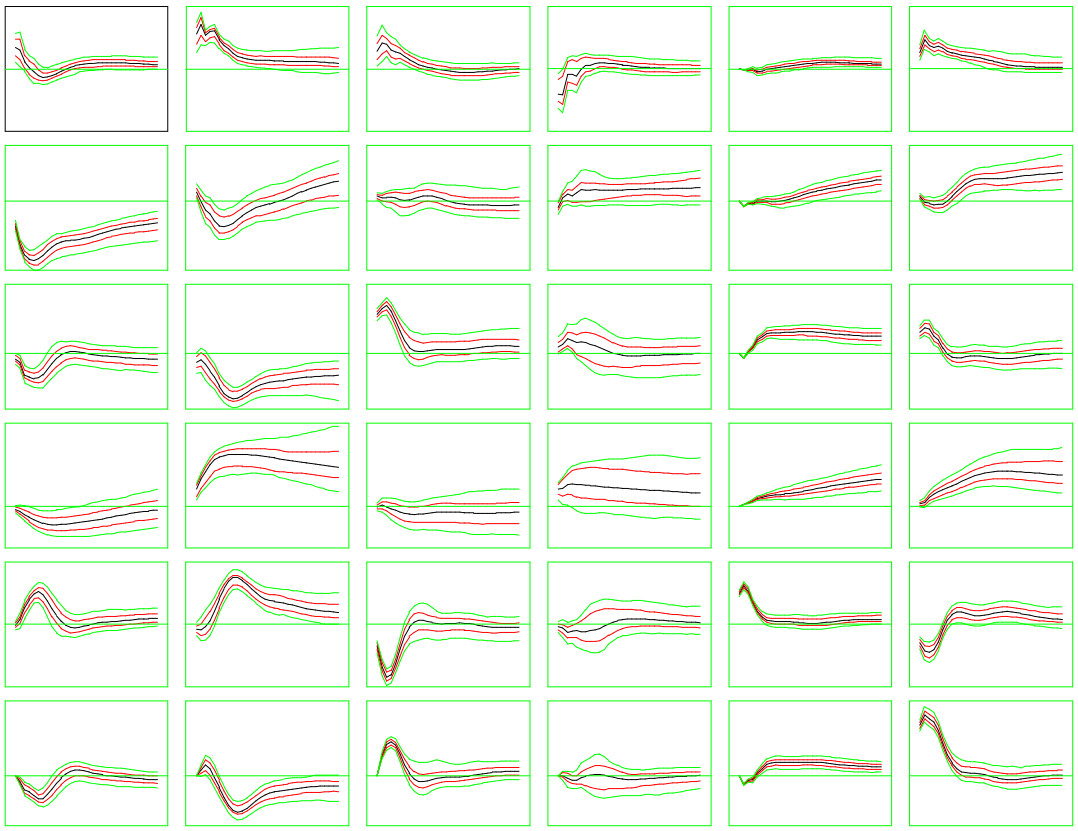


Figure 12

Six-Variable Model, Responses to Money Supply and Demand with .68 Error Bands

Figure 13

Six-Variable Model, Bias-Corrected Bootstrap 68% Confidence Bands