TO ANSWER OR NOT TO ANSWER: THAT IS THE QUESTION

MARK FISHER

Abstract. In the first part of the paper I present a simple structural model of supply and demand, pose a policy question, and answer it in terms of the structural parameters. Then, given some data, I show how to provide a realistic assessment of the uncertainty in that answer. In the second part of the paper I present common modifications of the structural system (in particular, scaling and renormalization) that turn out to make answering the policy question more difficult or impossible.

Contents

Date: 20:45 October 26, 2016. *Filename*: `"simple supply demand"`. *Version*: April 2016.

The views expressed herein are the author's and do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. This paper is a completely reworked version of an earlier paper "Scaled impulse responses: A critical review." I thank Wally Thurman for helpful comments.

Introduction

I begin by presenting a simple structural model of supply and demand, posing a policy question, and answering it in terms of the structural parameters. The policy question involves the effect of demand shocks on equilibrium price. The answer can be obtained via impulse responses. Then, given some available data, I show how to provide a realistic assessment of the uncertainty in that answer.

The remainder of the paper presents common modifications of the structural system and shows how they make it more difficult to answer the policy question. The modifications are of two sorts. The first modification involves scaling the shocks by their standard deviations. The resulting impulse responses no longer directly address the policy question, and they have the potential adverse side-effect of providing a false sense of precision. The second modification is to renormalize the system, replacing the demand curve with an inverse demand curve. The renormalized system is no longer expressed in terms of demand shocks (either scaled on unscaled) and consequently the impulse responses (either scaled or unscaled) do not address the policy question.

In the final section of the paper I apply the likelihood preserving normalization of Waggoner and Zha (2003) and show it produces the inverse-demand-curve normalization, which (as just noted) does not permit one to answer the policy question.

Part 1. To answer

1. Simple model of supply and demand

I present a very simple structural model of supply and demand. To keep things as simple as possible I assume the supply curve is vertical and consequently equilibrium quantity is exogenous. The demand curve, however, has finite and nonzero price elasticity.

Policy question. Suppose the policy maker has a policy tool that can produce a shock to demand. The policy maker wishes to raise equilibrium price by one unit. Here is the policy question:

> How big a demand shock is required in order to raise equilibrium price by one unit?

Structural model. It is possible to use the *impulse responses* implied by a structural model to provide an answer to this question.

Here is the system of equations that characterizes the structure of this model:

$$q_t^s = u_{1t} \qquad \text{(supply equation)} \qquad (1.1a)$$

$$q_t^d = \beta\, p_t + u_{2t} \qquad \text{(demand equation)} \qquad (1.1b)$$

$$q_t^d = q_t^s = q_t, \qquad \text{(equilibrium condition)} \qquad (1.1c)$$

where q_t^d is the log of the quantity demanded, q_t^s is the log of the quantity supplied, and p_t is the log of price. Thus, β is the elasticity of demand. We require $\beta \neq 0$ in order to guarantee the equilibrium condition can be satisfied. The variables u_{1t} and u_{2t} are *structural shocks.*

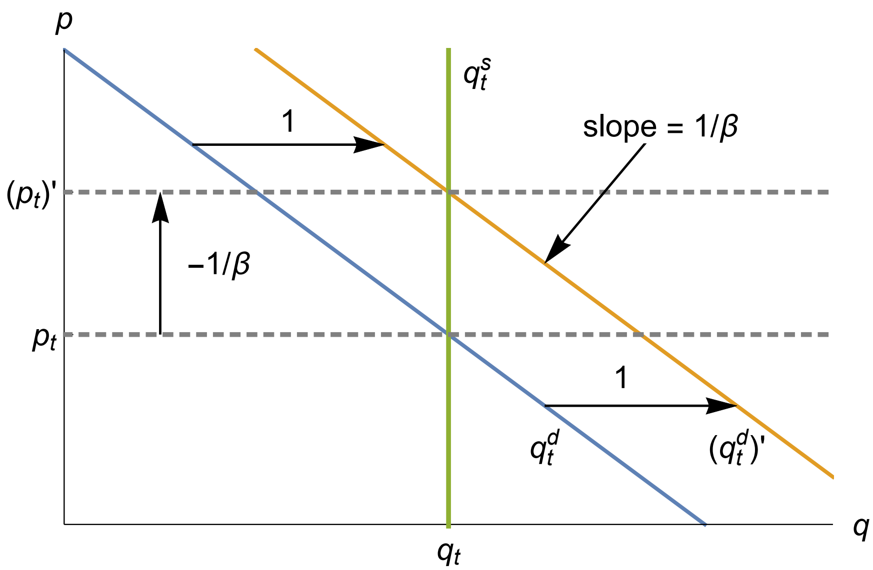

FIGURE 1. In this simple model of supply and demand, an increase in demand of one unit from q_t^d to $(q_t^d)'$ raises price by $-1/\beta$ units from p_t to $(p_t)'$, holding q_t^s fixed.

The equations are *normalized* so that u_{1t} is a supply shock and u_{2t} is a demand shock:

$$\frac{\partial q_t^s}{\partial u_{1t}} = 1 \qquad \text{and} \qquad \frac{\partial q_t^d}{\partial u_{2t}} = 1. \tag{1.2}$$

(Since quantity and price are measured in logs, a one unit change is equivalent to a one percent change.)

By imposing the equilibrium condition (thereby eliminating q_t^d and q_t^s), the structural system (1.1) can be expressed compactly as

$$B y_t = u_t, \tag{1.3}$$

where

$$B = \begin{pmatrix} 1 & 0 \\ 1 & -\beta \end{pmatrix}, \quad y_t = \begin{pmatrix} q_t \\ p_t \end{pmatrix}, \qquad \text{and} \quad u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \tag{1.4}$$

Note that each row of B has an element set to 1. The location of these elements is determined by the normalization.

Impulse responses. We can solve the system (1.3) for

$$y_t = B^{-1} u_t. \tag{1.5}$$

Impulse responses measure how the equilibrium values of quantity and price respond to the structural shocks. Consequently, the impulse responses are given by

$$B^{-1} = \begin{pmatrix} 1 & 0 \\ 1/\beta & -1/\beta \end{pmatrix}. \tag{1.6}$$

In particular, a one unit demand shock will increase equilibrium price by $-1/\beta$ units:[1]

$$\frac{dp_t}{du_{2t}} = (B^{-1})_{22} = -1/\beta. \tag{1.7}$$

See Figure 1. Consequently, it would require a demand shock of $-\beta$ units in order to increase equilibrium price by one unit.

Follow-up question. At the end of your presentation, the policy maker asks

So how big *is* $-\beta$?

The answer to this follow-up question involves statistical inference that depends on available data. In the following two sections I show how to do such inference and I present a numerical example as an illustration. The example shows that the information about β can be very non-normal and very imprecise.

2. Inference from data

The analysis in Section 1 is relatively straightforward, but making inferences from data about the sizes of the structural parameters is less so.

In this section I make distributional assumptions about the structural shocks, assumptions that involve parameters regarding the variances of the shocks. These distributional assumptions induce a distribution on the observations of equilibrium quantity and price, providing the data-generating distribution.

Conditional on a set of observations, the data-generating distribution provides a likelihood for the parameters. That likelihood is then combined with a prior distribution for the parameters in order to produce a posterior distribution. And that posterior distribution completely characterizes inferences about the structural parameters.[2,3]

In Section 3 a numerical example is presented to illustrate the issues that arise in this setting. The results are interesting and may even be surprising.

Model for the structural shocks. I assume the structural shocks u_{1t} and u_{2t} are each normally distributed with zero mean and serially independent. Moreover, I assume they are independent of each other. In particular,[4]

$$u_t \overset{\text{iid}}{\sim} \mathsf{N}(0_2, \Sigma), \tag{2.1}$$

where $0_2 = (0, 0)^\top$ and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}. \tag{2.2}$$

The *structural parameters* are $(\beta, \sigma_1^2, \sigma_2^2)$. They are the parameters that appear in either (1.3) or in (2.1). The former provides a model for the way in which equilibrium quantity and price are related to the structural shocks, while the latter provides a model for the structural shocks themselves.

---

[1]Alternatively, we can totally differentiate the second row of (1.4) to obtain $dq_t - \beta\, dp_t = du_{2t}$. Then setting $dq_t = 0$ produces $dp_t/du_{2t} = -1/\beta$.

[2]Appendix A presents a very brief introduction to Bayes' rule.

[3]Maximum likelihood estimation is considered in Appendices G and H.

[4]See Appendix B for the probability density functions for this and other distributions.

Distribution for the observations. We can use (2.1) to find the distribution of equilibrium quantity and price. Given (1.5), y_t satisfies

$$y_t \stackrel{\text{iid}}{\sim} \mathsf{N}(0_2, G), \tag{2.3}$$

where

$$G = \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix} \tag{2.4}$$

is a positive definite covariance matrix. In particular,

$$G = B^{-1}\,\Sigma\,(B^{-1})^\top = \begin{pmatrix} \sigma_1^2 & \sigma_1^2/\beta \\ \sigma_1^2/\beta & (\sigma_1^2+\sigma_2^2)/\beta^2 \end{pmatrix}. \tag{2.5}$$

We can solve (2.5) for the structural parameters in terms of the elements of G:

$$\beta = \frac{g_{11}}{g_{12}}, \quad \sigma_1^2 = g_{11}, \quad \text{and} \quad \sigma_2^2 = \frac{g_{11}\,|G|}{g_{12}^2}, \tag{2.6}$$

where $|G| = g_{11}\,g_{22} - g_{12}^2 > 0$ is the determinant of G.

Likelihood. Let $y = (y_1, \ldots, y_T)$ denote a collection of T observations. Define[5]

$$S = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} := \sum_{t=1}^{T} y_t\,y_t^\top = \begin{pmatrix} \sum_{t=1}^T y_{1t}^2 & \sum_{t=1}^T y_{1t}\,y_{2t} \\ \sum_{t=1}^T y_{1t}\,y_{2t} & \sum_{t=1}^T y_{2t}^2 \end{pmatrix} \tag{2.7}$$

and

$$\widehat{G} = \begin{pmatrix} \widehat{g}_{11} & \widehat{g}_{12} \\ \widehat{g}_{12} & \widehat{g}_{22} \end{pmatrix} := S/T. \tag{2.8}$$

The likelihood is given by

$$p(y|G) = \prod_{t=1}^{T} \mathsf{N}(y_t|0_2, G). \tag{2.9}$$

The likelihood can be expressed in terms of the structural parameters as

$$p(y|\beta,\sigma_1^2,\sigma_2^2) \propto \left\{ e^{-\frac{s_{11}}{2\,\sigma_1^2}} \left(\frac{1}{\sigma_1^2}\right)^{T/2} \right\} \left\{ e^{-\frac{s_{11} - 2\,s_{12}\,\beta + s_{22}\,\beta^2}{2\,\sigma_2^2}} \left(\frac{\beta^2}{\sigma_2^2}\right)^{T/2} \right\}. \tag{2.10}$$

Note that the likelihood depends on the data only via S and T (or equivalently $\widehat{G}$ and T). These summary statistics are sufficient for inference.

[5] "$x := y$" means x is defined as y.

Prior and posterior. Although it is the structural parameters that are of interest, it is convenient to focus on the parameters of the covariance matrix G.[6] The posterior distribution for G can be expressed as

$$p(G|y) \propto p(y|G)\, p(G), \tag{2.11}$$

where $p(G)$ is the prior distribution for G. The *conjugate prior* for G is the Inverse Wishart distribution, which has two parameters: $p(G) = \mathsf{Inv\text{-}Wishart}(G|\Psi, \nu)$, where Ψ is a positive definite matrix and $\nu > 0$ is called the degrees of freedom. With this prior, the posterior distribution for G is $p(G|y) = \mathsf{Inv\text{-}Wishart}(G|\Psi + S, \nu + T)$. Here I adopt the so-called uninformative prior, setting Ψ to the zero matrix and setting $\nu = 0$. Consequently,

$$p(G|y) = \mathsf{Inv\text{-}Wishart}(G|S, T). \tag{2.12}$$

Assuming $T > 3$, the posterior mean of G is $S/(T-3)$.[7]

At this point we could approximate the posterior distribution for the structural parameters by making draws of G from (2.12) and computing corresponding draws of $(\beta, \sigma_1^2, \sigma_2^2)$ using (2.6). Histograms of the draws, for example, would provide approximations that could be made arbitrarily accurate by increasing the number of draws of G.[8] However, it is possible to compute the posterior distribution for the structural parameters analytically and thereby obtain more accurate and in some ways more informative representations.

Reduced-form parameters. In order to compute the posterior distribution of the structural parameters $(\beta, \sigma_1^2, \sigma_2^2)$ from $p(G|y)$, first re-express the joint distribution for y_t in terms of the product of marginal and conditional distributions:[9]

$$\mathsf{N}(y_t|0_2, G) = \mathsf{N}(q_t|0, \omega_1^2)\, \mathsf{N}(p_t|\delta\, q_t, \omega_2^2), \tag{2.13}$$

where

$$\delta = \frac{g_{12}}{g_{11}}, \quad \omega_1^2 = g_{11}, \quad \text{and} \quad \omega_2^2 = \frac{|G|}{g_{11}}. \tag{2.14}$$

The parameters $(\delta, \omega_1^2, \omega_2^2)$ are sometimes referred to as the *reduced-form* parameters. Comparing (2.6) and (2.14), we see the relation between the structural parameters and the reduced-form parameters:

$$\beta = 1/\delta, \quad \sigma_1^2 = \omega_1^2, \quad \text{and} \quad \sigma_2^2 = \omega_2^2/\delta^2. \tag{2.15}$$

The properties of the Inverse Wishart distribution deliver the posterior distributions for the reduced-form parameters $(\delta, \omega_1^2, \omega_2^2)$:[10]

$$p(\delta, \omega_1^2, \omega_2^2|y) = p(\omega_1^2|y)\, p(\delta, \omega_2^2|y), \tag{2.16}$$

[6]In Appendix D, I focus on the structural parameters and arrive at essentially the same posterior distribution as that produced using the approach described in this section.

[7]I present additional material relating to the posterior for G in Appendix C.

[8]This approach is compared with the Bayesian bootstrap in Appendix H.

[9]This factorization is convenient because q_t $(= u_{1t})$ is exogenous in this system.

[10]The Jeffreys prior $p(\delta, \omega_1^2, \omega_2^2) \propto 1/(\sigma_1^2\, \sigma_2^2)$ produces a very similar posterior distribution. For example, see Zellner (1971, Chapter IX).

where

$$p(\omega_1^2|y) = \mathsf{Inv\text{-}Gamma}(\omega_1^2|\tfrac{T-1}{2}, \tfrac{T}{2}\,\widehat{g}_{11}) \tag{2.17}$$

$$p(\delta, \omega_2^2|y) = \mathsf{N}(\delta|\tfrac{\widehat{g}_{12}}{\widehat{g}_{11}}, \tfrac{\omega_2^2}{T\,\widehat{g}_{11}})\,\mathsf{Inv\text{-}Gamma}(\omega_2^2|\tfrac{T}{2}, \tfrac{T}{2}\,\tfrac{|\widehat{G}|}{\widehat{g}_{11}}). \tag{2.18}$$

In addition, the marginal posterior distribution for δ is

$$p(\delta|y) = \mathsf{Student}(\delta|\widehat{g}_{12}/\widehat{g}_{11}, |\widehat{G}|/(T\,\widehat{g}_{11}^2), T). \tag{2.19}$$

Distribution of the structural parameters. The posterior distribution for the structural parameters can be computed from the posterior distribution for the reduced-form parameters using the change-of-variables technique:

$$p(\beta, \sigma_1^2, \sigma_2^2|y) = \underbrace{p(\omega_1^2|y)\big|_{\omega_1^2=\sigma_1^2}}_{p(\sigma_1^2|y)}\underbrace{\left(p(\delta, \omega_2^2|y)\,\delta^4\right)\Big|_{\delta=1/\beta,\,\omega_2^2=\sigma_2^2/\beta^2}}_{p(\beta,\sigma_2^2|y)}. \tag{2.20}$$

We see that σ_1^2 is independent of (β, σ_2^2). The marginal distributions for β and σ_2^2 can be computed analytically from the joint distribution.

Impulse response and answer to the policy question. We can use (2.15) to express the impulse response in terms of the reduced form parameters: $dp_t/du_{2t} = -1/\beta = -\delta$. Therefore, in light of (2.19) the marginal posterior distribution for the impulse response is given by

$$-1/\beta|y \;\sim\; \mathsf{Student}(-\widehat{g}_{12}/\widehat{g}_{11}, |\widehat{G}|/(T\,\widehat{g}_{11}^2), T), \tag{2.21}$$

which simply takes the distribution for δ and changes the sign of the mean. In addition, (2.21) provides the distribution for the answer to the policy question. In particular, the distribution for $-\beta$ is computed as the reciprocal of the Student-t distribution. As such, we know that $p(-\beta|y)$ has no moments and (assuming $T > 1$) the distribution is bimodal with one positive mode and one negative mode. Therefore, before we see any data we know $-\beta$ has a bimodal posterior distribution. The locations and relative importance of the modes will be determined by the data.

Imposing a restriction on the slope of the demand curve. Economic theory suggests the restriction that the demand curve slopes downward (i.e., $\beta < 0$). In the empirical sections I will consider results both with and without the restriction. Here I provide some analytical expressions for the effect of imposing this restriction.

Referring to (2.6) and (2.15), we see that

$$\beta < 0 \iff g_{12} < 0 \iff \delta < 0. \tag{2.22}$$

Let $\mathcal{R}$ denote this restriction. It is easy to express the posterior distribution for the reduced-form parameters under the restriction:

$$p(\delta, \omega_1^2, \omega_2^2|y, \mathcal{R}) = p(\omega_1^2|y)\,p(\omega_2^2|\delta, y)\,p(\delta|y, \mathcal{R}), \tag{2.23}$$

where $p(\delta|y, \mathcal{R})$ is simply a truncated Student-t distribution. In particular,

$$p(\delta|y, \mathcal{R}) = \begin{cases} \frac{p(\delta|y)}{\alpha} & \delta < 0 \\ 0 & \text{otherwise} \end{cases}, \tag{2.24}$$

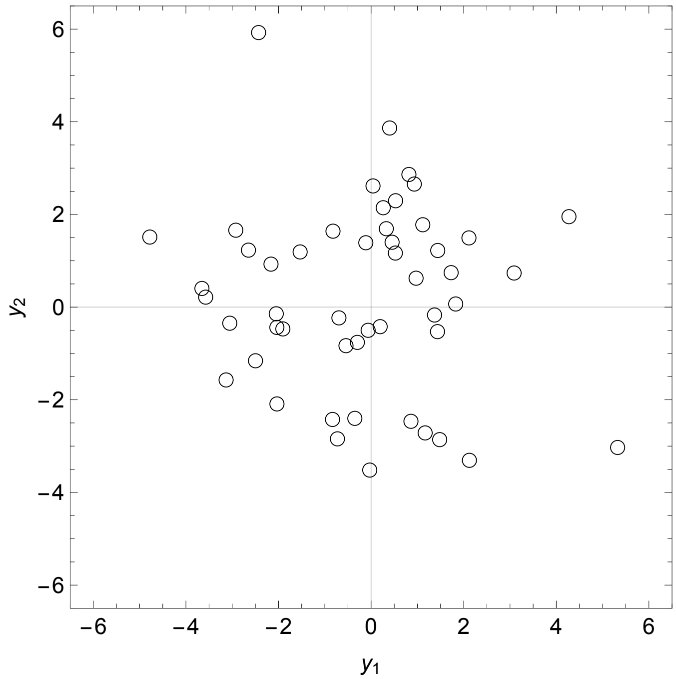

FIGURE 2. The data: $T = 50$.

where

$$\alpha = \int_{-\infty}^{0} p(\delta|y)\, d\delta. \tag{2.25}$$

The restriction does not affect the distribution for ω_1^2, but it does affect the marginal distribution for ω_2^2 (as shown in Appendix E).

The distribution for the impulse response is given by this truncated Student-t distribution:

$$-1/\beta|y, \mathcal{R} \sim \mathsf{Student}_{(0,\infty)}(-\widehat{g}_{12}/\widehat{g}_{11}, |\widehat{G}|/(T\,\widehat{g}_{11}^2), T). \tag{2.26}$$

Subject to the restriction, the distribution for the answer to the policy question, $-\beta$, is unimodal.

HPD regions. I will use highest posterior density (HPD) regions to help characterize the uncertainty in the posterior distributions. As an example consider $p(\beta|y)$, the posterior distribution for β (absent the restriction).

First define the set

$$\mathcal{H}(c) := \{\beta \in \mathbb{R} : p(\beta|y) \geq c\}, \tag{2.27}$$

where $c \geq 0$. Note that $\mathcal{H}(0) = \mathbb{R}$. Also note that for sufficiently large c, $\mathcal{H}(c) = \varnothing$. Assuming $\mathcal{H}(c)$ is not empty, then every $\beta \in \mathcal{H}(c)$ has a higher density than every $\beta \notin \mathcal{H}(c)$. Since $p(\beta|y)$ is bimodal, $\mathcal{H}(c)$ may be composed of two disjoint intervals.

Note that the probability that β is in $\mathcal{H}(c)$ is

$$\Pr[\beta \in \mathcal{H}(c)] = \int_{\mathcal{H}(c)} p(\beta|y)\, d\beta. \tag{2.28}$$

Now suppose (for example)

$$\Pr[\beta \in \mathcal{H}(c^*)] = 0.95 \tag{2.29}$$

for some c^*. Then $\mathcal{H}(c^*)$ is the 95% HPD region for $p(\beta|y)$.

3. Numerical illustration

I set the values of the structural parameters as follows:

$$\beta = -5, \quad \sigma_1^2 = 4, \quad \text{and} \quad \sigma_2^2 = 100. \tag{3.1}$$

These values imply

$$g_{11} = 4, \quad g_{12} = -4/5, \quad \text{and} \quad g_{22} = 104/25 \tag{3.2}$$

and

$$\delta = -1/5, \quad \omega_1^2 = 4, \quad \text{and} \quad \omega_2^2 = 4. \tag{3.3}$$

I generated data according to (2.3) using $T = 50$. The data are plotted in Figure 2. In addition,

$$S = \begin{pmatrix} 208.45 & -20.37 \\ -20.37 & 198.69 \end{pmatrix}. \tag{3.4}$$

Given S and T, we may discard the data without loss of information.

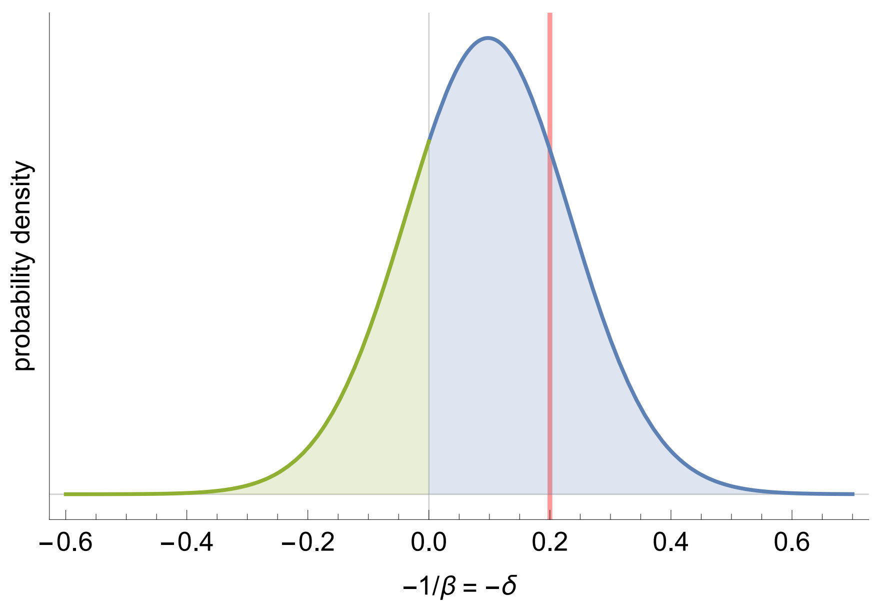

Figure 3. Posterior distribution for the impulse respond $dp_t/du_{2t} = -1/\beta = -\delta$. The probability that $-1/\beta < 0$ is about 24%. True value shown in red.

Posterior distributions. For reference, the posterior distributions for the reduced-form parameters and the joint distribution for (β, σ_2^2) are shown in Appendix F.

The distribution for the impulse response $dp_t/du_{2t} = -1/\beta = -\delta$ is a Student-t distribution, as shown in Figure 3. The probability that $-1/\beta < 0$ is about 24%. If the restriction $\beta < 0$ were imposed, then the distribution would be truncated at zero, leaving only positive support (shown in blue).

The posterior distribution for β is shown in Figure 4. The distribution for $-\beta$ is simply the distribution for β reflected across the "y" axis; it is shown in Figure 5 (with a wider range of values). The 95% highest posterior density (HPD) region[11] is composed of two intervals, each of which is quite long. The intervals are $[-87.3, -2.1]$ and $[1.4, 92.6]$. Thus, the answer to the policy question is extremely imprecise.

The posterior distribution $p(-\beta|y, \mathcal{R})$ is shown in Figure 6. The 95% HPD region is $[1.5, 59.5]$. Thus, even when the restriction $\beta < 0$ is imposed, the answer to the policy question is very imprecise. Moreover, it remains the case (due to the strong asymmetry) that the distribution for $-\beta$ is not well-described by a point-estimate and a standard error.

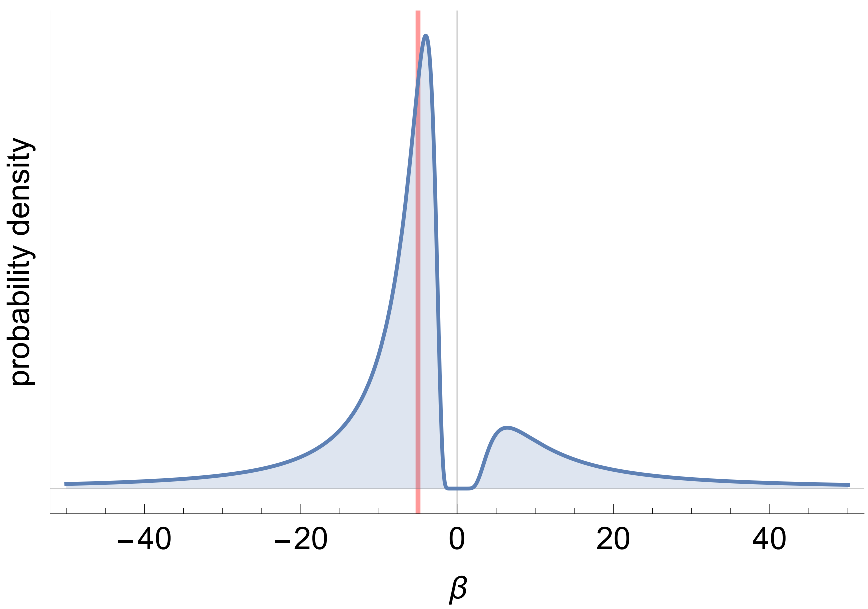

FIGURE 4. Posterior distribution for the elasticity of demand β. Approximately 90 percent of the probability is shown. The probability that $\beta > 0$ is about 24%. The true value is shown in red.

We are done. We have gone as far as we can in providing the policy maker with an answer to the policy question that provides an honest assessment of the uncertainty involved given our simple model and the available data. In this sense, we are done. The rest of the paper deals with issues related to modifying the structural system in one way or another. The modifications are commonly used and may be useful for some purposes. However, these

[11]HPD regions are described in Section 2.

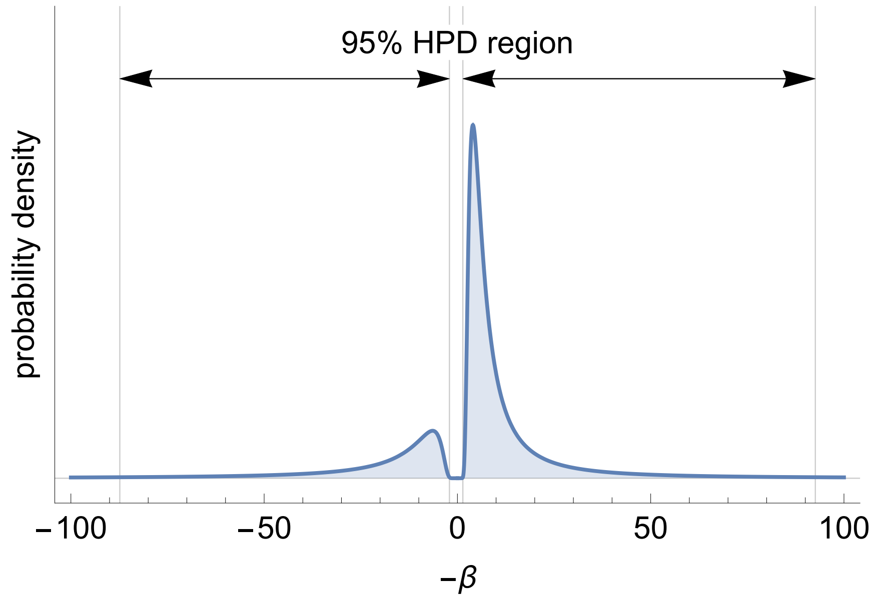

FIGURE 5. Posterior distribution for $-\beta$. The 95% HPD region is indicated.

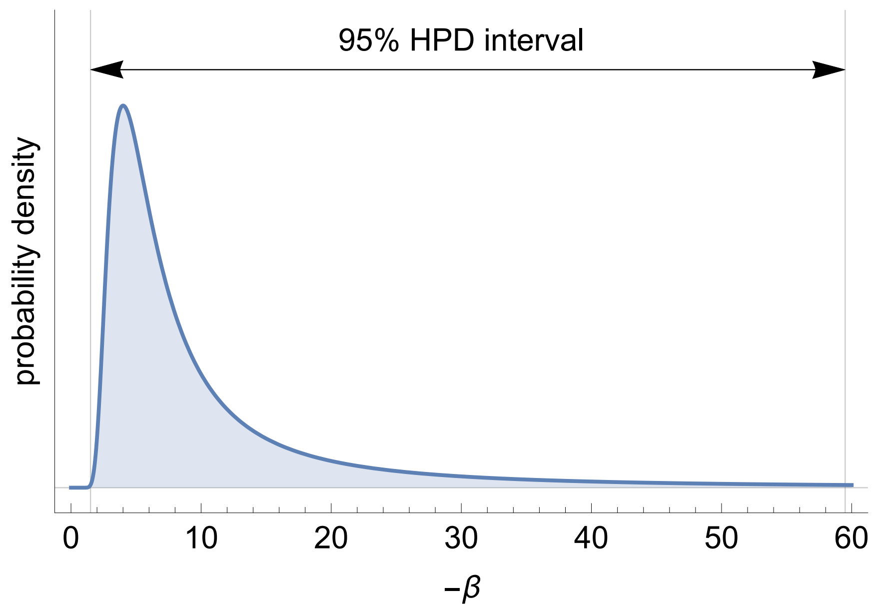

FIGURE 6. Posterior distribution of $-\beta$ given the restriction $\beta < 0$. The 95% HPD interval is $-\beta \in [1.5, 59.5]$.

modifications do not help answer the policy question I have posed. In fact, they tend to obscure the answer.

Part 2. Not to answer

4. Scaling the shocks

In Section 1, I characterized and assessed the response of equilibrium price to a *one-unit* demand shock — i.e., the magnitude of the shock being one unit. This impulse response provides the information to answer the policy question.

A popular alternative, by contrast, is to assess the response to a *one-standard-deviation* shock — i.e., the magnitude of the shock being one standard deviation of the shock's distribution. This approach is implemented by a transformation that amounts to dividing the structural shocks by their standard deviations. Such a transformation may not be as innocuous as it sounds. The transformation involves structural parameters that may be both highly uncertain and highly dependent on other structural parameters. The implicit interactions may produce strange and surprising results.

To scale the shocks, premultiply (1.3) by $\Sigma^{-1/2} = \begin{pmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{pmatrix}$ obtaining

$$C y_t = \widehat{u}_t, \tag{4.1}$$

where $C = \Sigma^{-1/2} B$ and $\widehat{u}_t = \Sigma^{-1/2} u_t$.[12] By construction

$$\widehat{u}_t \overset{\text{iid}}{\sim} \mathsf{N}(0_2, I_2) \tag{4.2}$$

where $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is the 2×2 identity matrix.

Here is the explicit representation for C:

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} 1/\sigma_1 & 0 \\ 1/\sigma_2 & -\beta/\sigma_2 \end{pmatrix}. \tag{4.3}$$

Recall that normalization of the unscaled system appeared in the matrix B as the location in each row of an element restricted to equal one. Consequently, normalization appears in C as the location in each row of an element restricted to be *positive*. In particular, $c_{11} > 0$ and $c_{21} > 0$.

Given (4.3), the impulse responses are found in

$$C^{-1} = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_1/\beta & -\sigma_2/\beta \end{pmatrix}. \tag{4.4}$$

In particular, the impulse response of p_t to the normalized shock $\widehat{u}_{2t}$ is

$$\frac{dp_t}{d\widehat{u}_{2t}} = (C^{-1})_{22} = -\sigma_2/\beta. \tag{4.5}$$

An interesting (and perhaps surprising) perspective on the relation between the scaled and unscaled impulse responses is brought out when they are expressed in terms of the reduced-form parameters. Referring to (2.15), we see the unscaled impulse response can be expressed as $-1/\beta = -\delta$ and the scaled impulse response can be expressed as

$$-\sigma_2/\beta = (-\delta/\sqrt{\delta^2})\,\omega_2 = \operatorname{sign}(-\delta)\,\omega_2. \tag{4.6}$$

[12] Note that $\sigma_1 = \sqrt{\sigma_1^2} > 0$ and $\sigma_2 = \sqrt{\sigma_2^2} > 0$ are the natural choices. Other choices are possible, but as long as one keeps track of the signs of σ_1 and σ_2 nothing of substance is affected.

Posterior distributions. The joint posterior distribution for the unscaled and scaled impulse responses is shown in Figure 7. The marginal posterior distribution for the scaled impulse response is shown in Figure 8.[13] It inherits the distinct bimodality from the joint distribution. We see that a demand shock of one standard deviation increases equilibrium price by roughly ± 2 units with odds 3-to-1 in favor of $+2$.[14]

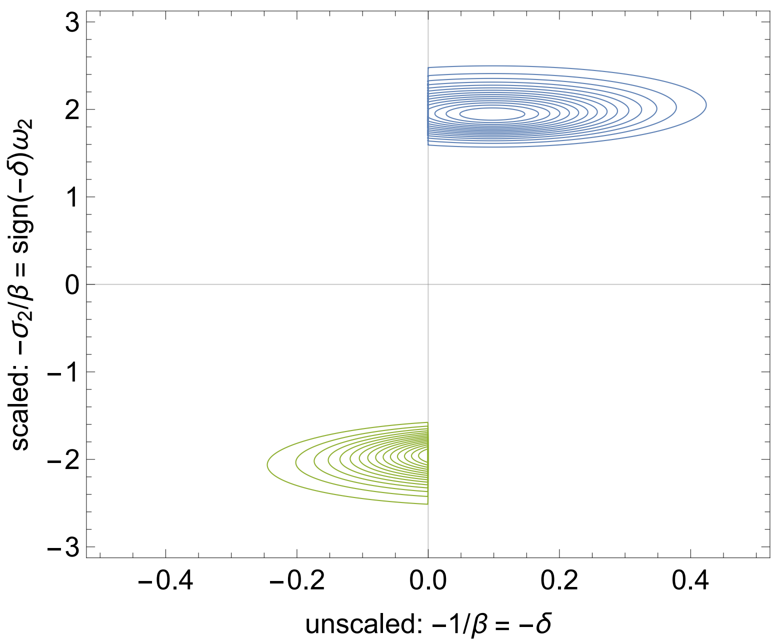

FIGURE 7. Posterior joint distribution for the unscaled $[-1/\beta = -\delta]$ and scaled $[-\sigma_2/\beta = \text{sign}(-\delta)\,\omega_2]$ impulse responses.

Before proceeding, it is interesting to examine the relation between the scaled impulse response $-\sigma_2/\beta$ and the standard deviation σ_2. The joint distribution is shown in Figure 9. The distribution for the scaled impulse response *conditional* on the standard deviation is shown in Figure 10.

Inverse scaled impulse response. The inverse of the scaled impulse response, namely $-\beta/\sigma_2$, is of more interest for our purposes at this point. It tells us the required number of standard deviations of the demand-curve shock that it takes to increase equilibrium price by one unit. The posterior distribution for $-\beta/\sigma_2$ is shown in Figure 11. Although there is uncertainty about the sign of $-\beta/\sigma_2$, there is very little uncertainty about its magnitude. We see that it takes a demand shock of roughly $\pm\frac{1}{2}$ standard deviations to increase equilibrium price by one unit.

We can eliminate the bimodality for the impulse response and its inverse by imposing the restriction $\beta < 0$. If we impose the restriction, then $dp_t/d\widehat{u}_{2t} = \omega_2$ [see (4.6)]. Note,

[13]Recall the marginal posterior distribution for the unscaled impulse response as shown in Figure 3.

[14]I show in Appendix I that the scaled impulse response is essentially (plus or minus) the standard deviation of observed prices.

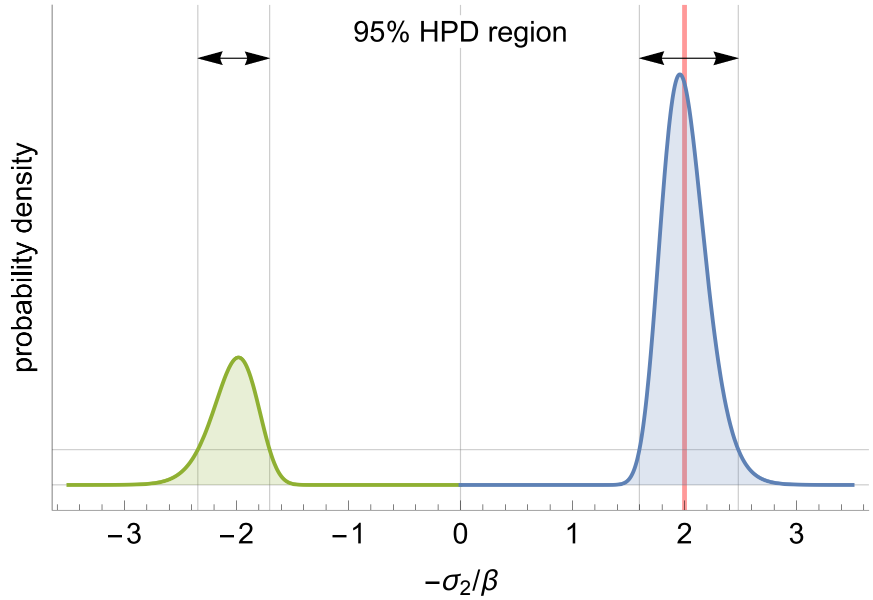

FIGURE 8. Posterior distribution for the scaled impulse response $dp_t/d\widehat{u}_{1t} = -\sigma_2/\beta$. 95% HPD region shown. True value shown in red.

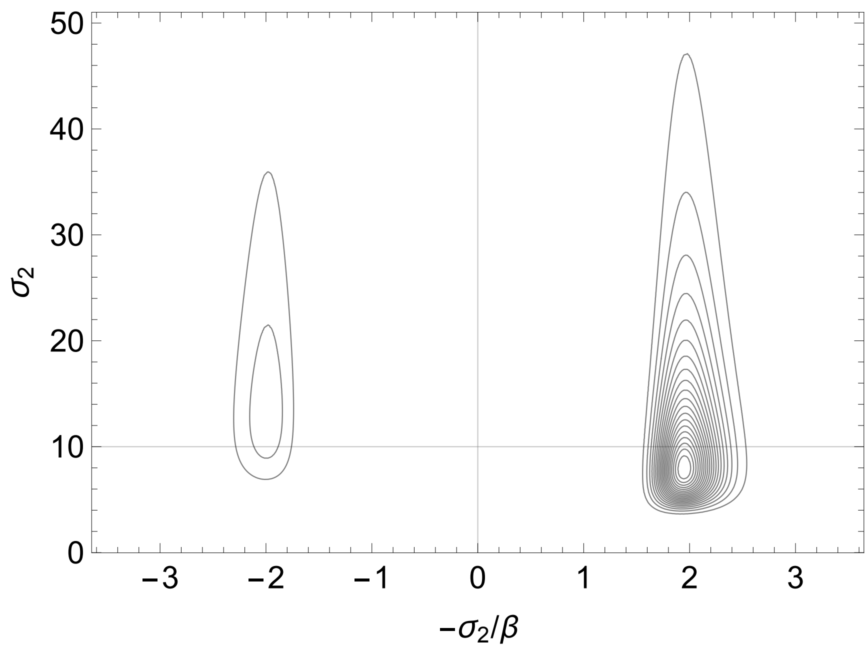

FIGURE 9. Joint posterior distribution for the scaled impulse response and the standard deviation, $p(-\sigma_2/\beta, \sigma_2|y)$. The distribution of $-\sigma_2/\beta$ conditional on $\sigma_2 = 10$ is indicated.

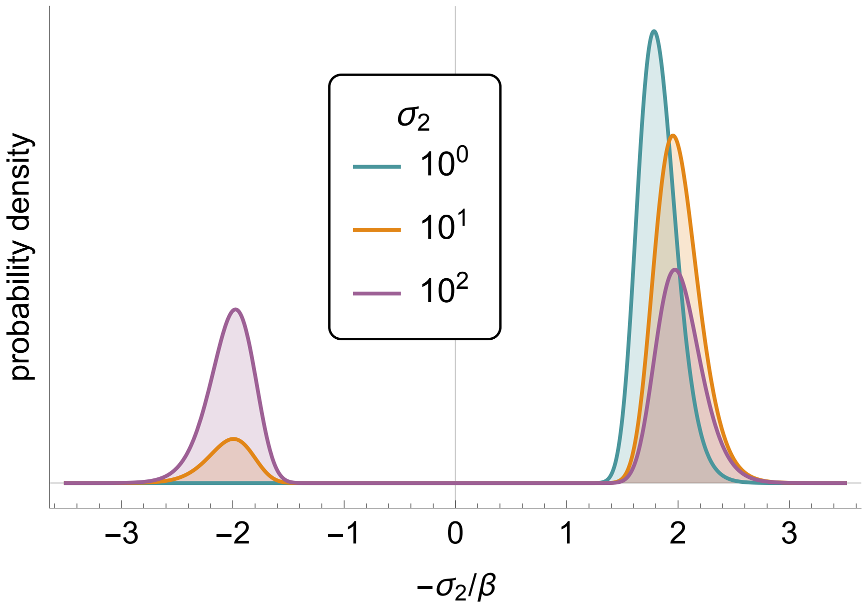

FIGURE 10. Posterior distribution for the scaled impulse response conditional on the standard deviation, $p(-\sigma_2/\beta|\sigma_2, y)$.

however, the posterior distribution of ω_2 is changed (albeit slightly) by the imposition of the restriction. (See Appendix E.) The distribution for $-\beta/\sigma_2$ subject to the restriction is simply (a rescaled version) of the distribution in Figure 11 restricted to the right of zero.

The data provide very precise inferences about the magnitude of a standardized demand shock required to raise equilibrium price by one unit. However, what is needed (to answer the policy question) is the product of (1) the required number of standard deviations and (2) the standard deviation itself: $(-\beta/\sigma_2) \times \sigma_2 = -\beta$. As we have seen, the uncertainty associated with $-\beta$ is substantial. Consequently, the distribution for $-\beta/\sigma_2$ provides a potentially misleading assessment of the uncertainty involved in the answer to the policy question. Compare Figure 11 with Figure 5. The differences in the distributions are largely accounted for by the uncertainty in σ_2.

The marginal posterior distribution for σ_2 is shown in Figure 12. It is clear from Figure 12 that the posterior uncertainty about the value for σ_2 is quite large. The 95% highest posterior density region is the interval [3.1, 179.]. There is a 4 percent change that $\sigma_2 > 200$.

5. Renormalizing the model

We now turn to the question of the effect of renormalizing the a system of structural equations. The choice of normalization is a substantive issue because the (structural) impulse responses are determined in part by the normalization.

Suppose we multiply both sides of (1.1b) by $1/\beta$ and rearrange:

$$p_t = (1/\beta)\, q_t^d + \{(-1/\beta)\, u_{2t}\}. \tag{5.1}$$

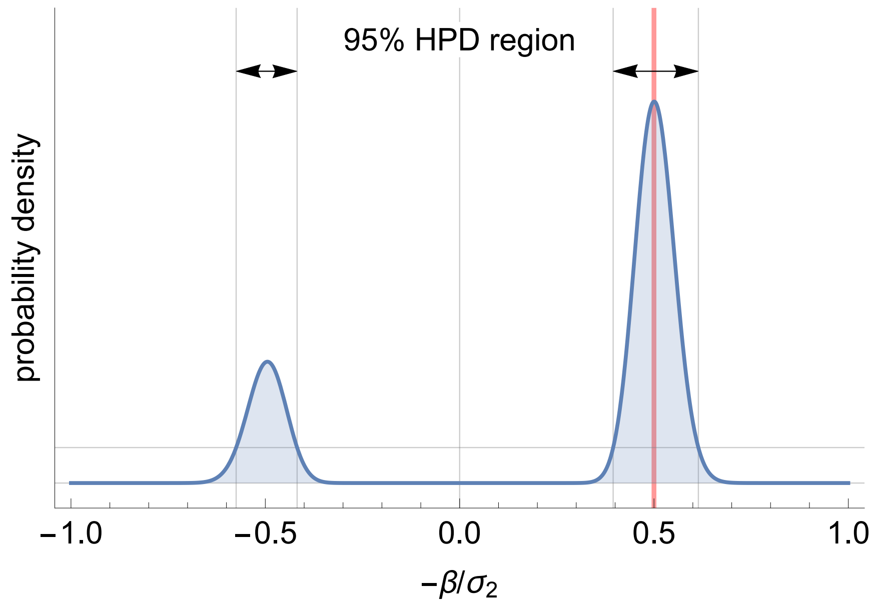

FIGURE 11. Posterior distribution for $-\beta/\sigma_2$, the number of standard deviations of the demand shock required to raise equilibrium price by one unit. True value shown in red.

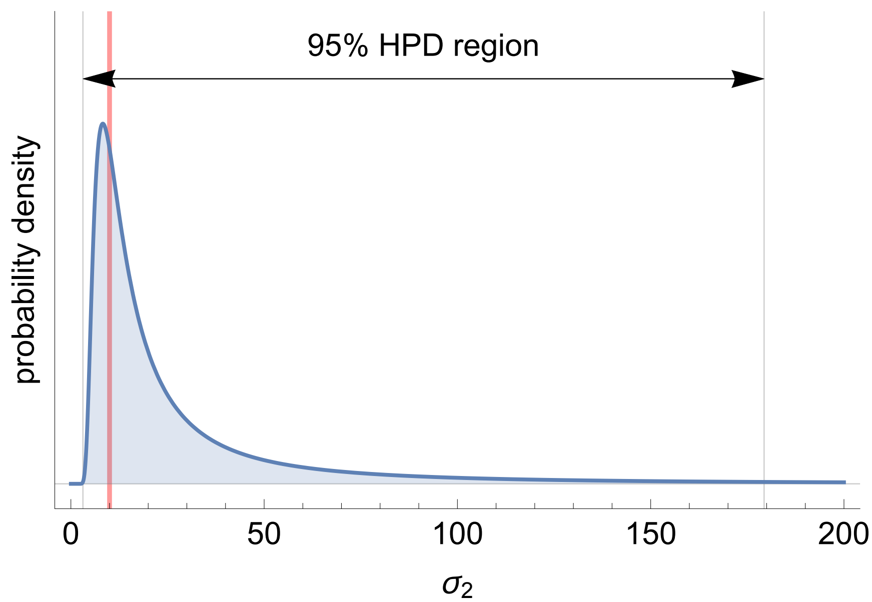

FIGURE 12. Posterior distribution for the standard deviation of a demand-curve shock σ_2. Approximately 96 percent of the probability shown. The 95% highest posterior density region is $\sigma_2 \in [3.1, 179.]$. True value shown in red.

TABLE 1. Factorizations of G. The unscaled inverse-demand factorization is the so-called triangular factorization, since Ω is diagonal and L^{-1} is lower-triangular with ones on the main diagonal as shown in (5.6). For more on the triangular factorization, see Section 4.4 in Hamilton (1994).

	Scaled	
Normalization	No	Yes
Demand	$B^{-1}\Sigma\,(B^{-1})^\top$	$C^{-1}\,I_2\,(C^{-1})^\top$
Inverse demand	$L^{-1}\Omega\,(L^{-1})^\top$	$M^{-1}\,I_2\,(M^{-1})^\top$

TABLE 2. Impulse responses in terms of the reduced-form parameters.

	Scaled	
Normalization	No	Yes
Demand	$(B^{-1})_{22} = -\delta$	$(C^{-1})_{22} = \operatorname{sign}(-\delta)\,\omega_2$
Inverse demand	$(L^{-1})_{22} = \;\;1$	$(M^{-1})_{22} = \omega_2$

This amounts to renormalizing the demand curve as an *inverse demand curve*, where the structural shock is no longer u_{2t} but rather $(-1/\beta)\,u_{2t}$. The renormalized system can be expressed as

$$q_t^s = v_{1t} \qquad \text{(supply equation)} \tag{5.2a}$$

$$p_t = \delta\, q_t^d + v_{2t} \qquad \text{(inverse demand equation)} \tag{5.2b}$$

$$q_t^d = q_t^s = q_t. \qquad \text{(equilibrium condition)} \tag{5.2c}$$

In expressing (5.2) I have used $1/\beta = \delta$ and I have given the structural shocks their own labels. The equations in (5.2) are normalized so that $v_{1t} = u_{1t}$ remains a supply-curve shock while $v_{2t} = -\delta\, u_{2t}$ is an inverse-demand-curve shock:

$$\frac{\partial q_t^s}{\partial v_{1t}} = 1 \qquad \text{and} \qquad \frac{\partial p_t}{\partial v_{2t}} = 1. \tag{5.3}$$

We can express the renormalized system as $L\, y_t = v_t$ where

$$L = \begin{pmatrix} 1 & 0 \\ -\delta & 1 \end{pmatrix}. \tag{5.4}$$

The normalization is expressed in L as the location in each row of an element set to 1. By comparing the second row of L with that of B, the renormalization is evident: $L_{22} = 1$ versus $B_{21} = 1$. With the renormalized system, we have $v_t \overset{\text{iid}}{\sim} \mathsf{N}(0_2, \Omega)$, where[15]

$$\Omega = L\, G\, L^\top = \begin{pmatrix} \omega_1^2 & 0 \\ 0 & \omega_2^2 \end{pmatrix}. \tag{5.5}$$

[15]Note $L = DB$ and $v_t = D\, u_t$, where $D = \begin{pmatrix} 1 & 0 \\ 0 & -1/\beta \end{pmatrix}$. Thus $\Omega = D\Sigma D^\top$.

The reader may recognize the renormalized system as another representation of the marginal-conditional factorization shown in (2.13).

The (unscaled) impulse responses are given by

$$L^{-1} = \begin{pmatrix} 1 & 0 \\ \delta & 1 \end{pmatrix}. \tag{5.6}$$

Notice that the impulse response of equilibrium price to a one unit inverse-demand-curve shock is identically one:

$$\frac{dp_t}{dv_{2t}} = (L^{-1})_{22} = 1. \tag{5.7}$$

The "answer" to the policy question implied by this impulse response is straightforward: Raise equilibrium price by one unit. But this answer is useless because it misses the point. The policy maker has no way to raise equilibrium price directly; the policy question is about the effect of *demand shocks* on equilibrium price.

Impulse responses to scaled shocks. We now turn to the impulse responses to scaled inverse-demand-curve shocks. Premultiply $L\, y_t = v_t$ by $\Omega^{-1/2}$ to obtain $M y_t = \widehat{v}_t$, where $M = \Omega^{-1/2}\, L$ and $\widehat{v}_t = \Omega^{-1/2}\, v_t$.[16] By construction

$$\widehat{v}_t \overset{\text{iid}}{\sim} \mathsf{N}(0_2, I_2). \tag{5.8}$$

The explicit representation for M is

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \begin{pmatrix} 1/\omega_1 & 0 \\ -\delta/\omega_2 & 1/\omega_2 \end{pmatrix}. \tag{5.9}$$

The restriction $m_{22} > 0$ reflects the inverse-demand normalization. Given (5.9), the scaled impulse responses are found in

$$M^{-1} = \begin{pmatrix} \omega_1 & 0 \\ \delta\, \omega_1 & \omega_2 \end{pmatrix}. \tag{5.10}$$

In particular, the impulse response of p_t to the normalized shock $\widehat{v}_{2t}$ is

$$\frac{dp_t}{d\widehat{v}_{2t}} = (M^{-1})_{22} = \omega_2. \tag{5.11}$$

Note that scaled impulse response is nothing more than the standard deviation of the unscaled shock. My view is that the scaled impulse response is even less helpful than the unscaled version in addressing the policy question, because the unscaled version is transparently irrelevant while the scaled version is not. Nevertheless, I show the posterior distribution of the scaled impulse response $p(\omega_2|y)$ in Figure 13 for future reference. The distribution for ω_2 is almost indistinguishable from (but not identical to [see Appendix E]) that for $-\sigma_2/\beta$ under the restriction $\beta < 0$. The interpretation, however, is quite different.

If we were to impose the restriction in both normalizations, then the two scaled impulse responses would be would have identical distributions. Nevertheless, the unscaled impulse responses would have dramatically different distributions.

[16] Note that $\omega_i = \sqrt{\omega_i^2} > 0$.

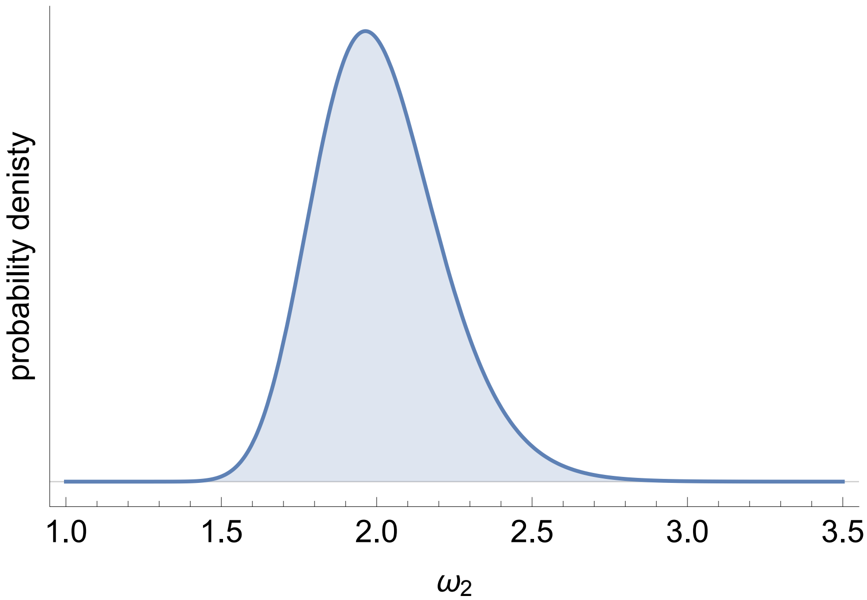

FIGURE 13. Posterior distribution for the scaled impulse response of an inverse-demand-curve shock to equilibrium price.

6. Likelihood preserving normalization

Waggoner and Zha (2003, WZ hereafter) provide a useful discussion of different normalizations.[17] Their contribution may be viewed as two-fold. First, they show that impulse responses will be affected by different normalizations. In effect, I have illustrated this point in earlier sections of this paper.

Second, as they say in the abstract to their paper, "We develop a general normalization rule that preserves the likelihood shape and maintains coherent economic interpretations for both recursive and nonrecursive models." However, their normalization rule produces the inverse-demand-curve normalization for the system in this paper and, because the impulse responses from this normalization do not answer the policy question, I therefore conclude their likelihood preserving normalization is not generally applicable.

Normalization in scaled systems. WZ discuss normalization and their rule in terms of scaled systems. Therefore, we need to examine the matrices C and M presented in (4.3) and (5.9), respectively, that characterize the scaled versions of the two normalizations of the system in this paper. Since $c_{12} = m_{12} = 0$, we can represent both C and M as

$$H = \begin{pmatrix} h_{11} & 0 \\ h_{21} & h_{22} \end{pmatrix}, \tag{6.1}$$

making the appropriate substitutions for the nonzero elements. At this level of abstraction, normalization can be understood as choosing an element from each row of H and restricting

[17] See also Hamilton et al. (2007) and Kociecki (2013).

it to be positive.[18] For the first row there is no choice: $h_{11} > 0$. For the second row, however, there is a choice: $h_{21} > 0$ versus $h_{22} > 0$. Examining C and M, the choices are seen to be $h_{21} > 0$ for the demand-curve normalization (since $c_{21} > 0$) and $h_{22} > 0$ for the inverse-demand-curve normaliztion (since $m_{22} > 0$).

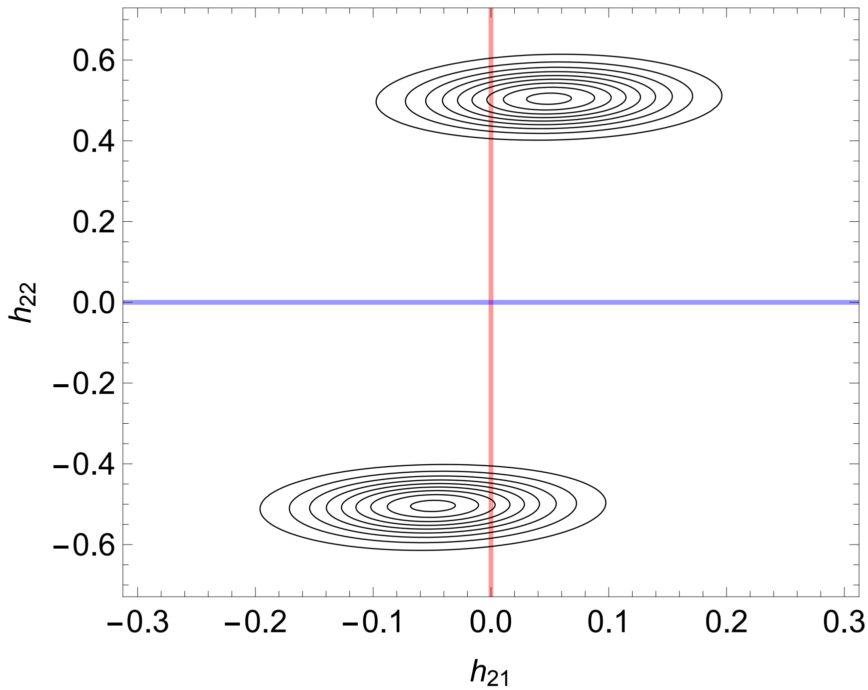

FIGURE 14. Contours of $p(y|h_{21}, h_{22})$. The area to the right of the red line is consistent with $h_{21} > 0$ and the demand-curve normalization, while the area above the blue line is consistent with $h_{22} > 0$ (i.e., the likelihood preserving normalization) and the inverse-demand-curve normalization.

Unnormalized likelihood. WZ address the effects of the choice of normalization by examining the unnormalized likelihood in terms of the elements of H. The unnormalized likelihood can be expressed as [cf. (2.10)]

$$p(y|h_{11}, h_{21}, h_{22}) \propto \left\{ e^{-\frac{s_{11}\,h_{11}^2}{2}} \left(h_{11}^2\right)^{T/2} \right\} \left\{ e^{-\frac{s_{11}\,h_{21}^2 + 2\,s_{12}\,h_{21}\,h_{22} + s_{22}\,h_{22}^2}{2}} \left(h_{22}^2\right)^{T/2} \right\}. \tag{6.2}$$

Since h_{21} and h_{22} only appear in the second factor on the right-hand side of (6.2), we restrict attention to $p(y|h_{21}, h_{22})$. This latter unnormalized likelihood is shown in Figure 14, using the values of S and T from the numerical illustration — which itself is based on

[18]If different sign choices for σ_i and ω_i had been made, then those sign choices would carry through to this discussion.

the parameter values provided by WZ for the example in their paper.[19,20] Note that the likelihood equals zero along the blue line (where $h_{22} = 0$).

Referring to Figure 14, it is evident that $h_{21} > 0$ limits the parameter space to the area to the right of the red line, while $h_{22} > 0$ limits the parameter space to the area above the blue line. If one were to impose $\beta < 0$ [or any of the equivalent restrictions in (2.22)], then the parameter space in Figure 14 would be limited to the first quadrant — regardless of the chosen normalization.

Figure 14 clearly illustrates WZ's likelihood preserving normalization for this example, namely $h_{22} > 0$. Looking at the contour plot, this choice has a strong aesthetic appeal since it "preserves" a single mode. By contrast, the alternative normalization ($h_{12} > 0$) unattractively slices through the two modes. Indeed, WZ refer to "the distortion it causes to the shape of the likelihood."[21]

To illustrate this distortion, WZ consider the distribution for $1/h_{22}$ under both normalizations. These distributions correspond to Figure 13 for $h_{22} > 0$ and Figure 8 for $h_{21} > 0$.[22] WZ first praise (their equivalent of) the unimodal distribution in Figure 13 before presenting (their equivalent of) the bimodal distribution in Figure 8 as the poster-child for the distortion.

Nevertheless, aesthetics must be overridden by subject-matter considerations: In order to answer the policy question one must adopt the demand-curve normalization, and therefore one must choose $h_{21} > 0$ and reject the likelihood preserving normalization.[23]

---

[19]See pp. 332–334 in Waggoner and Zha (2003). I show the correspondence between the two examples in Appendix J.

[20]Figure 14 corresponds to Figure 2 in Waggoner and Zha (2003). See Appendix J for a discussion of the differences in appearance between the two figures.

[21]Page 333.

[22]Figures 13 and 8 correspond to Figures 1(a) and 1(b) in Waggoner and Zha (2003). See Appendix J for a discussion of the differences in appearance between the two sets of figures.

[23]Hamilton et al. (2007, pp. 236–237) concede the inability to answer the policy question when the inverse-demand-curve normalization is used. Nevertheless, they reject the demand-curve normalization because the resulting distribution for β is poorly described by a point estimate and a standard error.

Appendix A. Bayes' rule

In this appendix I present the briefest of introductions to Bayes' rule.

Joint, marginal, and conditional distributions. Let $p(y, \theta)$ denote the (density for) the joint distribution of y and θ. Marginal distributions for y and θ can be obtained from the joint distribution by integration:

$$p(y) = \int p(y, \theta)\, d\theta \qquad \text{and} \qquad p(\theta) = \int p(y, \theta)\, dy. \tag{A.1}$$

Conditional distributions can be expressed in terms of joint and marginal distributions:

$$p(y|\theta) = \frac{p(y, \theta)}{p(\theta)} \qquad \text{and} \qquad p(\theta|y) = \frac{p(y, \theta)}{p(y)}. \tag{A.2}$$

Of course the conditional distributions integrate to one. For example,

$$\int p(y|\theta)\, dy = \int \frac{p(y, \theta)}{p(\theta)}\, dy = \frac{1}{p(\theta)} \int p(y, \theta)\, dy = \frac{p(\theta)}{p(\theta)} = 1. \tag{A.3}$$

Sometimes it is convenient to express conditional distributions as proportionalities rather than equalities:

$$p(y|\theta) \propto p(y, \theta) \qquad \text{and} \qquad p(\theta|y) \propto p(y, \theta). \tag{A.4}$$

Bayes' rule. By rearranging the two equations in (A.2), we see the joint distribution can be factored into the product of conditional and marginal distributions in either of two ways:

$$p(y, \theta) = \underbrace{p(\theta|y)\, p(y) = p(y|\theta)\, p(\theta)}_{\text{two ways}}. \tag{A.5}$$

The "two ways" can be rearranged as Bayes' rule:

$$p(\theta|y) = \frac{p(y|\theta)\, p(\theta)}{p(y)}. \tag{A.6}$$

Each of the objects in (A.6) has one or more names: $p(\theta)$ is the prior distribution for θ, $p(\theta|y)$ is the posterior distribution for θ, $p(y|\theta)$ is the likelihood for θ, and $p(y)$ is the marginal likelihood of y (sometimes known as the evidence).

Typically in (A.6), y is interpreted as data (composed of one or more observations) and θ is interpreted as one or more parameters. The data are observed and therefore fixed. Since y is fixed, $p(y)$ is a number and $p(y|\theta)$ is a function of θ. By contrast, the parameters are unobserved and therefore subject to uncertainty. That uncertainty is characterized by a probability distribution. Bayes' rule shows how to use observations to update that distribution.

Appendix B. Gallery of densities

For reference I include the densities of the distributions used in this paper. For the multivariate distributions, x and μ are vectors of length k and Σ is a $k \times k$ positive definite matrix. Scalar versions of the multivariate distributions can be obtained by setting $k = 1$.

The multivariate normal distribution:

$$\mathsf{N}(x|\mu, \Sigma) = (2\,\pi)^{-\frac{k}{2}}\, |\Sigma|^{-\frac{1}{2}}\, e^{-\frac{1}{2}\,(x-\mu)^\top \Sigma^{-1}(x-\mu)}. \tag{B.1}$$

The multivariate Student distribution ($\nu > 0$):

$$\mathsf{Student}(x|\mu,\Sigma,\nu) = \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})}\,(\nu\,\pi)^{-\frac{k}{2}}\,|\Sigma|^{-\frac{1}{2}}\left(1+\frac{1}{\nu}\,(x-\mu)^\top\Sigma^{-1}(x-\mu)\right)^{-\frac{1}{2}\,(\nu+k)}. \tag{B.2}$$

If $z \sim \mathsf{Student}(\mu,\sigma^2,\nu)$, then $x = 1/z \sim \mathsf{Inv\text{-}Student}(\mu,\sigma^2,\nu)$ where

$$\mathsf{Inv\text{-}Student}(x|\mu,\sigma^2,\nu) = \mathsf{Student}(1/x|\mu,\sigma^2,\nu)/x^2. \tag{B.3}$$

The Inverse Gamma distribution (x is a scalar and $a, b > 0$):

$$\mathsf{Inv\text{-}Gamma}(x|a,b) = \frac{b^a\,e^{-\frac{b}{x}}}{\Gamma(a)\,x^{a+1}}, \tag{B.4}$$

where $\Gamma(\,\cdot\,)$ is the gamma function.

The Inverse Wishart distribution (X and Ψ are $k \times k$ positive definite matrices):

$$\mathsf{Inv\text{-}Wishart}(X|\Psi,\nu) = \frac{|\Psi|^{\frac{\nu}{2}}}{2^{\frac{\nu\,k}{2}}\,\Gamma_k(\frac{\nu}{2})}\,|X|^{-\frac{1}{2}\,(\nu+k+1)}\,e^{-\frac{1}{2}\,\mathrm{tr}(\Psi X^{-1})}, \tag{B.5}$$

where $\mathrm{tr}(\,\cdot\,)$ is the trace function and $\Gamma_k(\,\cdot\,)$ is the multivariate gamma function.

Appendix C. Additional material related to $p(G|y)$

Predictive distribution. Given the observed data, uncertainty about the parameters is embodied in the posterior distribution, $p(G|y)$. Taking account of that parameter uncertainty, one can compute a probability distribution for new observations. This *predictive distribution* is

$$\begin{aligned} p(y_{T+1}|y) &= \int p(y_{T+1}|G)\,p(G|y)\,dG \\ &= \int \mathsf{N}(y_{T+1}|0,G)\,\mathsf{Inv\text{-}Wishart}(G|S,T)\,dG \\ &= \mathsf{Student}(y_{T+1}|0_2,\widehat{G},T-1). \end{aligned} \tag{C.1}$$

Prior versus likelihood. Suppose we divide our data into two groups: a "training set" composed of the the first T_0 observations from which we develop proper prior and a "holdout set" of the remaining $T_1 = T - T_0$ observations that characterize the likelihood. Let $y^0 = (y_1,\ldots,y_{T_0})$, $y^1 = (y_{T_0+1},\ldots,y_T)$, $S_0 = \sum_{t=1}^{T_0} y_t\,y_t^\top$, and $S_1 = \sum_{t=T_0+1}^{T} y_t\,y_t^\top$. Then $T_0 + T_1 = T$, $S_0 + S_1 = S$, and $y^0 \cup y^1 = y$.

Let the prior based on the training set be

$$p(G|y^0) = \mathsf{Inv\text{-}Wishart}(G|S_0,T_0). \tag{C.2}$$

Then the posterior distribution for G is unaffected by the choice of T_0:

$$p(y^1|G)\,p(G|y^0) \propto \mathsf{Inv\text{-}Wishart}(G|S,T) = p(G|y). \tag{C.3}$$

Moreover, since the posterior distribution is unaffected by the choice of T_0, the predictive distribution is unaffected as well.

Marginal likelihood. By contrast, the marginal likelihood of the hold-out data y^1 *is* affected by the choice of T_0. The marginal likelihood is given by

$$p(y^1|y^0) = \int p(y^1|G)\, p(G|y^0)\, dG. \tag{C.4}$$

We can obtain some insight by expressing the marginal likelihood in terms of a rearrangement of Bayes' rule:

$$p(y^1|y^0) = \frac{p(y^1|G)\, p(G|y^0)}{p(G|y)}. \tag{C.5}$$

Since the left-hand side of (C.5) is independent of G, we may evaluate the right-hand side by choosing any convenient value for G. Given a fixed value for G, changes in T_0 will leave the denominator of the right-hand side unchanged. However, changes in T_0 *will* affect the numerator because observations are transferred from one factor to the other, thereby producing changes in the product (as one may verify).

Appendix D. An explicit prior for the structural parameters

Here I put a prior directly on the structural parameters $(\beta, \sigma_1^2, \sigma_2^2)$. In particular, let $p(\beta, \sigma_1^2, \sigma_2^2) = p(\beta)/(\sigma_1^2\, \sigma_2^2)$ where $p(\beta)$ is specified below. Recall the likelihood $p(y|\beta, \sigma_1^2, \sigma_2^2)$ is given by (2.10). Then the marginal likelihood for β is

$$p(y|\beta) = \iint \frac{p(y|\beta, \sigma_1^2, \sigma_2^2)}{\sigma_1^2\, \sigma_2^2}\, d\sigma_1^2\, d\sigma_2^2 \propto \left(\frac{\beta^2}{s_{11} - 2\, s_{12}\, \beta + s_{22}\, \beta^2} \right)^{T/2}. \tag{D.1}$$

There are two features of note. First, this likelihood equals zero at $\beta = 0$: $p(y|\beta)|_{\beta=0} = 0$. Second, this likelihood does not go to zero as β gets large in magnitude: $\lim_{\beta\to\infty} p(y|\beta) = \lim_{\beta\to-\infty} p(y|\beta) > 0$. Owing to this latter feature, this likelihood is not normalizable and thus a proper prior for β is required.

Recall that the structural model has no solution if $\beta = 0$, so it is natural (but not necessary) to adopt a prior for which the density goes to zero as β goes to zero. One such prior (that allows for both positive and negative β) is obtained by letting the reciprocal of β have a normal distribution centered at zero:

$$1/\beta \sim \mathsf{N}(0, s^2), \tag{D.2}$$

for some $s > 0$. In this case, the density for β is

$$p(\beta) = \begin{cases} \frac{e^{-\frac{1}{2\, s^2\, \beta^2}}}{\sqrt{2\pi}\, s\, \beta^2} & \beta \neq 0 \\ 0 & \beta = 0 \end{cases}. \tag{D.3}$$

This distribution is symmetric around zero, bimodal [with modes at $\pm 1/(\sqrt{2}\, s)$], and has no finite moments.

It turns out that $p(\beta|y)$ is not sensitive to the free parameter s if $s \gg 1$. In fact, the posterior distribution for β given $s \gg 1$ is essentially the same as that produced using the Inverse Wishart distribution as described in the main text. [See Figure 4.] The reason for the similarity is that the implicit prior (in the main text) for $\delta = 1/\beta$ is a approximately $\mathsf{N}(0, s^2)$ with $s \gg 1$.

Alternative prior for β. Suppose intead $p(\beta) = \mathsf{N}(\beta|0, s^2)$. The posterior distribution is not available analytically. A large number of finite moments can be computed numerically. The length of highest posterior density regions for β are quite sensitive to the value of s; they can be shortened or lengthened dramatically. The posterior distribution converges pointwise to zero as $s \to \infty$.

Appendix E. Marginal distribution of ω_2^2 given $\delta < 0$

In this appendix I show how the marginal posterior distribution for ω_2^2 is modified by the imposing the restriction $\delta < 0$.

The marginal distribution for ω_2^2 is modified via the dependence between δ and ω_2^2 [see (2.18)]. It is convenient to express this dependence as

$$p(\omega_2^2|\delta, y) = \frac{p(\delta|\omega_2^2, y)\, p(\omega_2^2|y)}{p(\delta|y)}. \tag{E.1}$$

It follows that[24]

$$p(\omega_2^2|y, \mathcal{R}) = \int p(\omega_2^2|\delta, y)\, p(\delta|y, \mathcal{R})\, d\delta = p(\omega_2^2|y)\, \alpha^{-1} \int_{-\infty}^{0} p(\delta|\omega_2^2, y)\, d\delta. \tag{E.2}$$

Therefore,

$$\frac{p(\omega_2^2|y, \mathcal{R})}{p(\omega_2^2|y)} = \alpha^{-1} \int_{-\infty}^{0} p(\delta|\omega_2^2, y)\, d\delta. \tag{E.3}$$

In the numerical example, the effect on the marginal distribution for ω_2^2 is not large but nevertheless the systematic shift in the densities is evident in Figure 15.

Appendix F. Additional posterior distributions

Reduced-form parameters. The joint posterior distribution for (δ, ω_2^2) is shown in Figure 16. Although δ and ω_2^2 are not independent, the dependence between them is not strong. The marginal posterior distributions for (ω_1^2, ω_2^2) and δ are shown in Figures 17 and 18 respectively. Note that the probability that $\delta > 0$ is about 24 percent.

Structural parameters. The joint distribution for (β, σ_2^2) is shown in Figure 19. It has two "branches." The two branches are the consequence of the bimodality of β mentioned in Section 2. Within each of the two branches β and σ_2^2 are highly dependent.

Appendix G. Maximum likelihood estimation

Given the likelihood for the structural parameters (2.10), define the log-likelihood

$$L(\theta) := \log\big(p(y|\beta, \sigma_1^2, \sigma_2^2)\big), \tag{G.1}$$

where $\theta = (\beta, \sigma_1^2, \sigma_2^2)$. The maximum likelihood values $\widehat{\theta}$ are given by the (unique) solution to $\nabla L(\theta) = 0$. They turn out to be the "plug-in" values using $\widehat{G} = S/T$ [compare with (2.6)]:

$$\widehat{\beta} = \frac{\widehat{g}_{11}}{\widehat{g}_{12}}, \quad \widehat{\sigma}_1^2 = \widehat{g}_{11}, \quad \text{and} \quad \widehat{\sigma}_2^2 = \frac{\widehat{g}_{11}\, |\widehat{G}|}{\widehat{g}_{12}^2}. \tag{G.2}$$

[24] Refer to (2.23)–(2.25) for the notation used in (E.2).

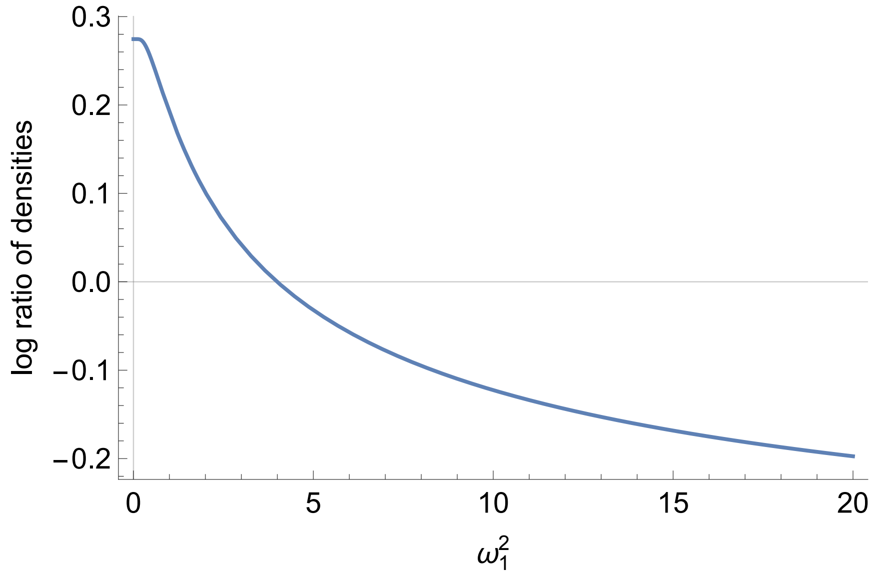

FIGURE 15. $\log\big(p(\omega_2^2|y,\mathcal{R})/p(\omega_2^2|y)\big)$, where $p(\omega_2^2|y,\mathcal{R})$ denotes the posterior distribution for ω_2^2 given the restriction $\delta < 0$.

For the numerical illustration, the maximum likelihood values are

$$\widehat{\beta} = -10.2, \quad \widehat{\sigma}_1^2 = 4.2, \quad \text{and} \quad \widehat{\sigma}_2^2 = 411.9. \tag{G.3}$$

The asymptotic covariance matrix can be approximated by $(-\nabla^2 L(\theta)|_{\theta=\widehat{\theta}})^{-1}$. Therefore, the asymptotic distribution for $\widehat{\beta}$ can be expressed as

$$\widehat{\beta} \overset{a}{\sim} \mathsf{N}(\beta, \widehat{\sigma}_\beta^2), \tag{G.4}$$

where the asymptotic variance for $\widehat{\beta}$ is

$$\widehat{\sigma}_\beta^2 = \frac{\widehat{g}_{11}^2\,|\widehat{G}|}{\widehat{g}_{12}^4\,T} = \frac{\widehat{g}_{11}}{\widehat{g}_{12}^2}\left(\frac{\widehat{\sigma}_2^2}{T}\right). \tag{G.5}$$

The asymptotic distribution for $\widehat{\beta}$ is unimodal and allows for the possibility that a confidence interval may cross zero. The distribution for $\widehat{\beta}$ ignores the global aspects of the likelihood (as they do not matter asymptotically). The asymptotic distribution for $\widehat{\beta}$ is shown in Figure 20.

The sampling distribution can be computed analytically. Note that

$$\widehat{G}|G,T \sim \mathsf{Wishart}(G, T-1). \tag{G.6}$$

Consequently,

$$\widehat{\delta} = \frac{\widehat{g}_{12}}{\widehat{g}_{11}} \sim \mathsf{Student}(g_{12}/g_{11}, |G|/((T-1)\,g_{11}^2), T-1). \tag{G.7}$$

It follows that the sampling distribution for $\widehat{\beta}$ is the reciprocal of a Student-t distribution and is thus bimodal. See Figure 21. The probability that $\widehat{\beta}$ is positive is about 8 percent.

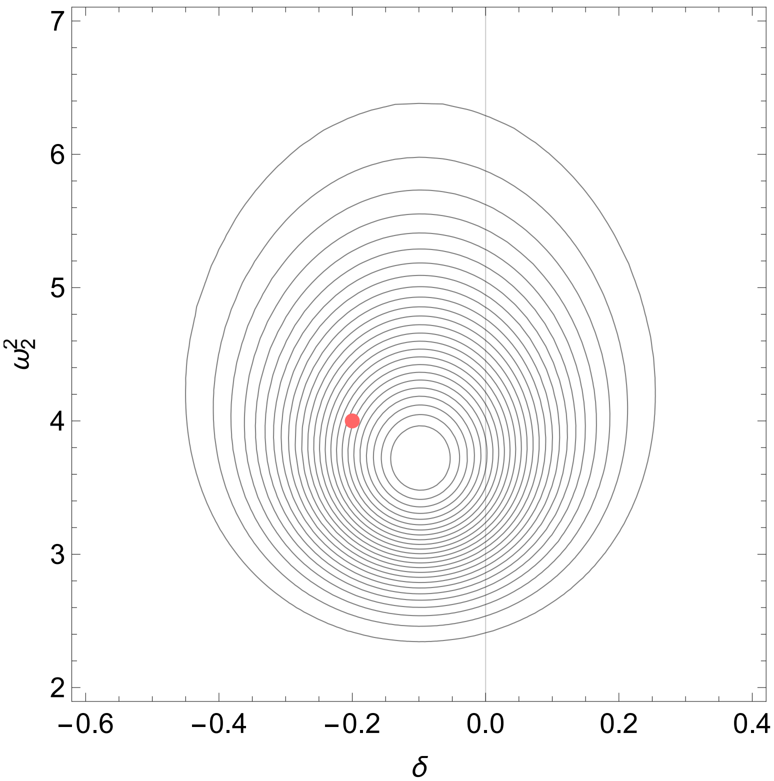

FIGURE 16. Contours of the posterior distribution for (δ, ω_2^2). More than 99 percent of the probability is in the region shown. True value shown in red.

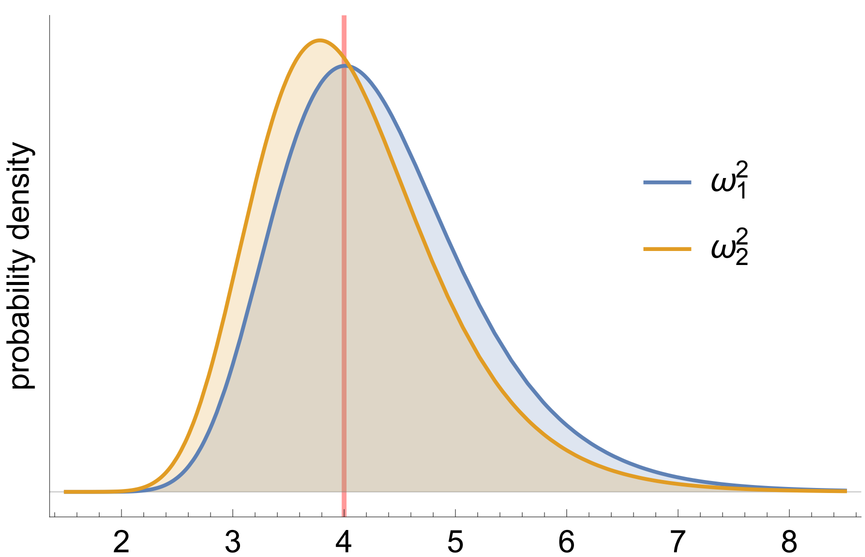

FIGURE 17. Posterior distributions of ω_1^2 and ω_2^2. True value shown in red.

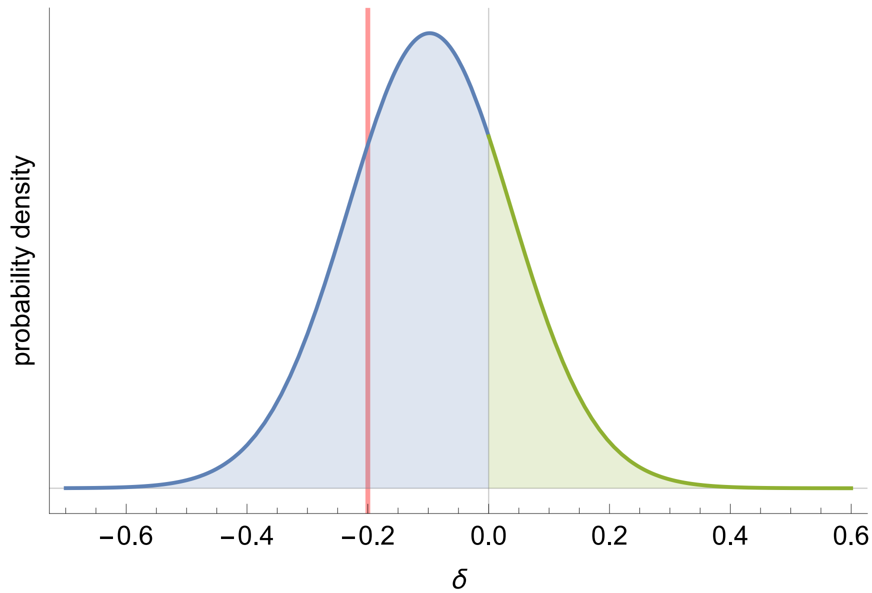

FIGURE 18. Posterior distribution of δ. (True value shown in red.) Probability of $\delta > 0$ is about 24 percent.

The is somewhat less that the 23 percent we have seen for the posterior probability that β is positive.

In general, of course, we will not know the true values and will not be able to compute the sampling distribution this way. Instead, we can appeal to the bootstrap. Appendix H shows (among other things) that the bootstrapped distribution for $\widehat{\beta}$ has much in common with the sampling distribution.

Appendix H. The bootstrap

There are two types of bootstrap. The *frequentist bootstrap* produces a sampling distribution for an estimator such as the maximum likelihood estimator $\widehat{\beta}$ [see (G.2)]. As such, the bootstrap provides a distribution that may be more consistent with the global features of the likelihood than the asymptotic distribution for $\widehat{\beta}$. By contrast, the *Bayesian bootstrap* produces a posterior distribution for an unknown parameter such as β. As such, the bootstrap provides an alternative posterior distribution that can be compared with what is presented in the main text as a robustness check. The two bootstrap distributions (the sampling distribution for $\widehat{\beta}$ and the posterior distribution for β) are quite similar, although their interpretations and uses are quite different.[25]

Each of the two bootstrap distributions depends on the data $y = (y_1, \ldots, y_T)$ and a set of random weights $w = (w_1, \ldots, w_T)$. The weights are nonnegative and sum to one. The probability distribution for the weights is different for the two types of bootstrap. Let

[25] The frequentist bootstrap was introduced by Efron (1979) and the Bayesian bootstrap was introduced by Rubin (1981). Lancaster (2009) provides an interesting comparison.

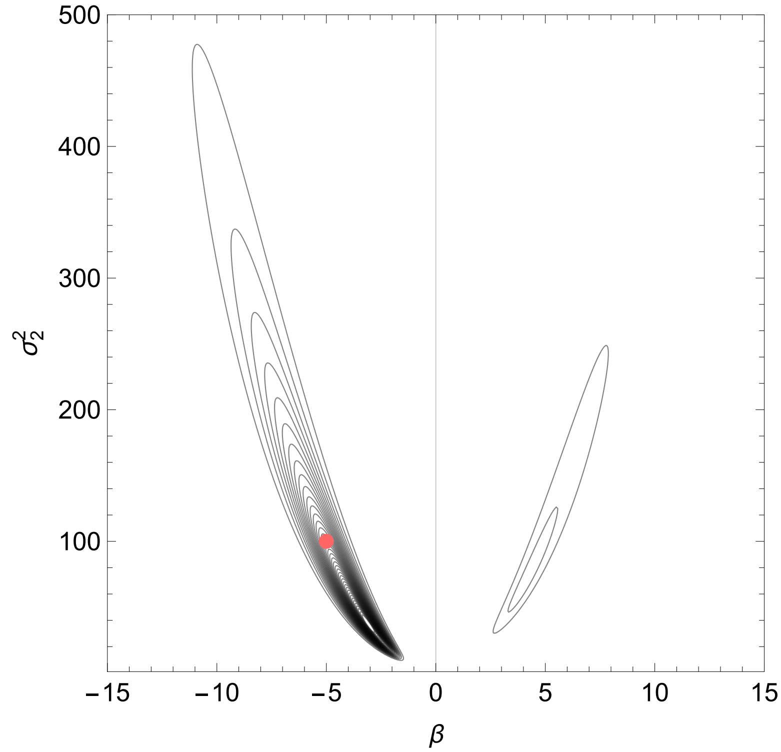

FIGURE 19. Contours for the posterior distribution for (β, σ_2^2). Approximately 61 percent of the probability is in the region shown. True value shown in red.

$\underline{w} = (1/T, \ldots, 1/T)$ so that $T\underline{w} = (1, \ldots, 1)$. For the frequentist bootstrap w is drawn from a normalized multinomial distribution with equal probabilities,

$$Tw \sim \mathsf{Multinom}(T, \underline{w}). \tag{H.1}$$

For the Bayesian bootstrap w is drawn from a flat Dirichlet distribution,

$$w \sim \mathsf{Dirichlet}(T\underline{w}). \tag{H.2}$$

The two distributions for w have the same mean, $\underline{w}$, and similar variances and covariances. However, the interpretations of the distributions for w is conceptually quite different. For the frequentist bootstrap, the weights represent the selection of observations in the process of resampling the data with replacement. For the Bayesian bootstrap, the weights are unoberved parameters and (H.2) is the posterior distribution of those parameters.[26]

[26]The Bayesian bootstrap is sometimes referred to a smoothed version of the frequentist bootstrap because the Dirichlet distribution is continuous and the normalized multinomial distribution is discrete. I prefer to think of the frequentist bootstrap and a “quick and dirty” approximation to the Bayesian bootstrap.

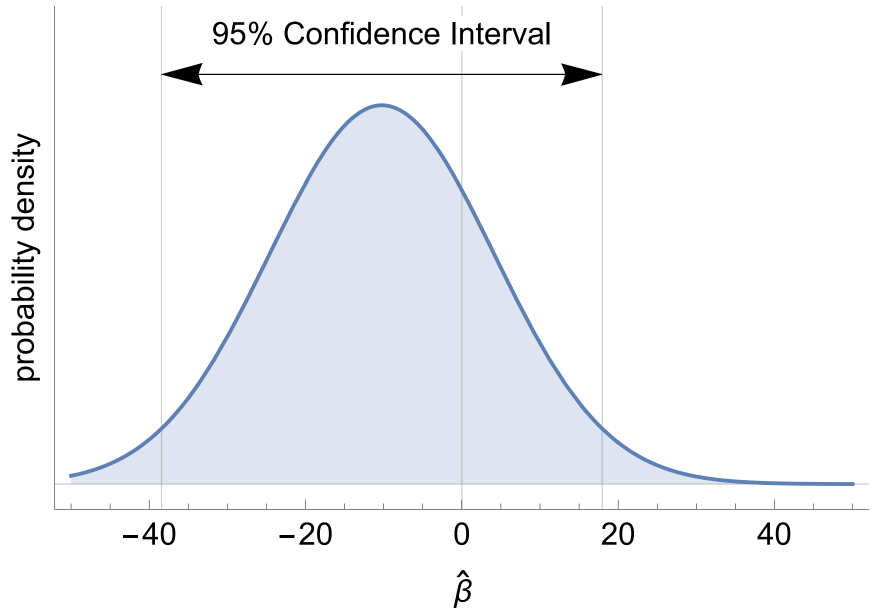

FIGURE 20. Asymptotic distribution for $\widehat{\beta}$ [see (G.4)].

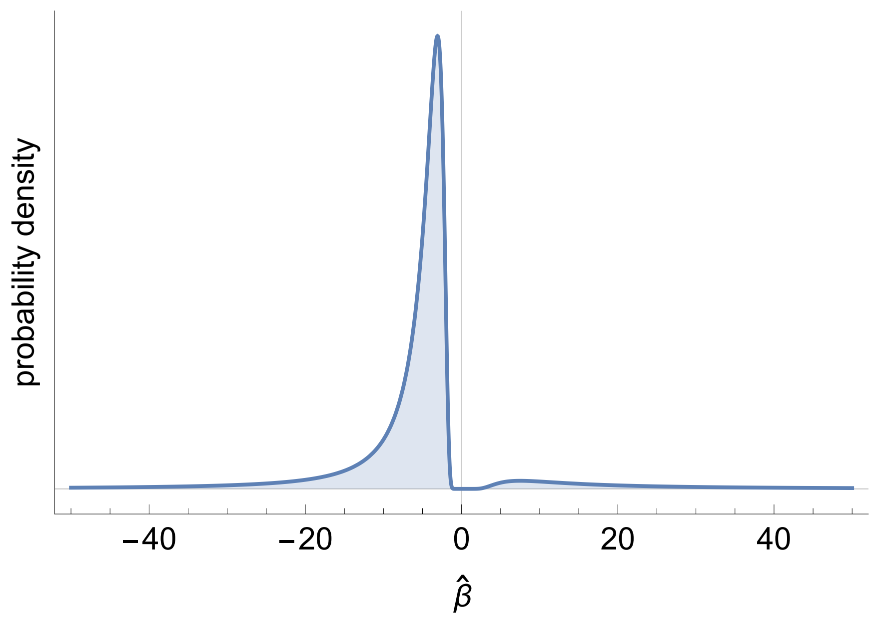

FIGURE 21. Sampling distribution for $\widehat{\beta}$ [see (G.4)]. Approximately 96% of the probability is shown. Approximately 8% of the distribution is positive.

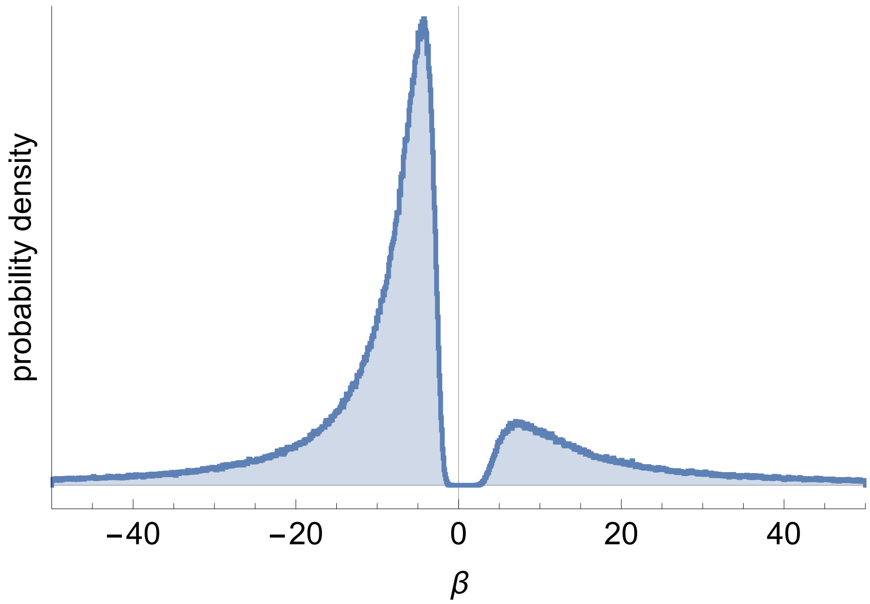

FIGURE 22. Posterior distribution for β computed using 10^6 draws from the Bayesian bootstrap. Approximately 90 percent of the probability shown. Compare with Figure 4.

Both types of bootstrap combine the data and the weights in the same way. In particular, let

$$\widetilde{G} = \sum_{t=1}^{T} w_t \, (y_t \, y_t^\top) = \begin{pmatrix} \sum_{t=1}^{T} w_t \, y_{1t}^2 & \sum_{t=1}^{T} w_t \, y_{1t} \, y_{2t} \\ \sum_{t=1}^{T} w_t \, y_{1t} \, y_{2t} & \sum_{t=1}^{T} w_t \, y_{2t}^2 \end{pmatrix}. \tag{H.3}$$

Given the similarities in the distributions for w between the two types of bootstrap, the distributions for $\widetilde{G}$ are similar as well. For example, $E[\widetilde{G}] = \widehat{G}$ $(= S/T)$ for both distributions.

The interpretation of $\widetilde{G}$ is different between to the two types of bootstrap. For the frequentist bootstrap, it is an estimator for which the sampling distribution is determined by the distribution for w (in conjunction with y, which represents the population distribution for y_t and is therefore fixed); for the Bayesian bootstrap it is a functional of the posterior distribution for w (where the "value" associated with category t is given by $y_t \, y_t^\top$).

Given $\widetilde{G}$, define[27]

$$\widetilde{\beta} = \frac{\widetilde{g}_{11}}{\widetilde{g}_{12}}, \quad \widetilde{\sigma}_1^2 = \widetilde{g}_{11}, \quad \text{and} \quad \widetilde{\sigma}_2^2 = \frac{\widetilde{g}_{11} \, |\widetilde{G}|}{\widetilde{g}_{12}^2}. \tag{H.4}$$

The similarity in the two distributions for $\widetilde{G}$ carries over the the distributions for the structural parameters via (H.4).

[27] In passing, note that $E[\widetilde{\beta}] \neq \widehat{g}_{11}/\widehat{g}_{12}$ owing to Jensen's inequality.

The bootstrap distributions are computed as follows. First make R draws $\{w^{(r)}\}_{r=1}^{R}$ from the appropriate distribution for w. Next, for each $w^{(r)}$ compute $\widetilde{G}^{(r)}$ using (H.3). Finally, compute any function of $\widetilde{G}^{(r)}$ desired, such as $\widetilde{\beta}^{(r)}$ using (H.4).

From the Bayesian perspective, the density of the posterior distribution for β may be approximated by a histogram of $\{\widetilde{\beta}^{(r)}\}_{r=1}^{R}$.[28] See Figure 22. Note the striking similarity between this density and that shown in Figure 4. This is driven by the fact that the posterior distribution for G given by the Bayesian bootstrap is very similar to that given by the Inverse Wishart distribution.

From the frequentist perspective, the bootstrap distribution for $\widetilde{\beta}$ provides (an approximation to) the sampling distribution for the maximum likelihood estimator $\widehat{\beta}$. The frequentist bootstrap distribution for w produces an almost identical distribution for $\widetilde{\beta}$ as shown in Figure 22. Thus the bootstrap approximation to the sampling distribution for $\widehat{\beta}$ given $T = 50$ appears to have much in common with the actual sampling distribution as shown in Figure 21.

Appendix I. Source of the precision of the scaled impulse response

The source of the the precision of the scaled impulse response $(-\sigma_2/\beta)$ and its inverse can be found in the likelihood for (β, σ). The likelihood is also the source of the imprecision of β and of σ_2. Figure 23 shows the likelihood for (β, σ_2).[29] The red lines show where $\sigma_2/\beta = \pm 2$. The lines appear to follow the ridges of the likelihood.

The ridges can be characterized as

$$\sigma_2/\beta \approx \pm\sqrt{\widehat{g}_{22}}. \tag{I.1}$$

Recall $\widehat{g}_{22}$ is an element of $\widehat{G} = S/T$. It is the maximum likelihood estimate of g_{22}, the variance of p_t. Thus the slopes of the ridges depend on the standard deviation of observed prices. Given the example, $\sqrt{\widehat{g}_{22}} = 1.99$. (Compare this with $\sqrt{g_{22}} = 2.04$.) Approximation (I.1) provides a convenient caricature of the likelihood.

Approximation (I.1) can be derived in two steps. First, compute the two local maxima for β conditional on σ_2:

$$\beta^* = \frac{\widehat{g}_{12}}{2\,\widehat{g}_{22}} \pm \sqrt{\left(\frac{\widehat{g}_{12}}{2\,\widehat{g}_{22}}\right)^2 + \frac{\sigma_2^2}{\widehat{g}_{22}}}. \tag{I.2}$$

Second, take the limit

$$\lim_{\sigma_2 \to \infty} \beta^*/\sigma_2 = \pm\sqrt{1/\widehat{g}_{22}}. \tag{I.3}$$

This result can be expressed as (I.1).

[28] See the paragraph that follows (2.12) about drawing from $p(G|y)$.

[29] See the second factor on the right-hand side of (2.10) for the analytical expression. Previously we treated this factor as the likelihood for (β, σ_2^2).

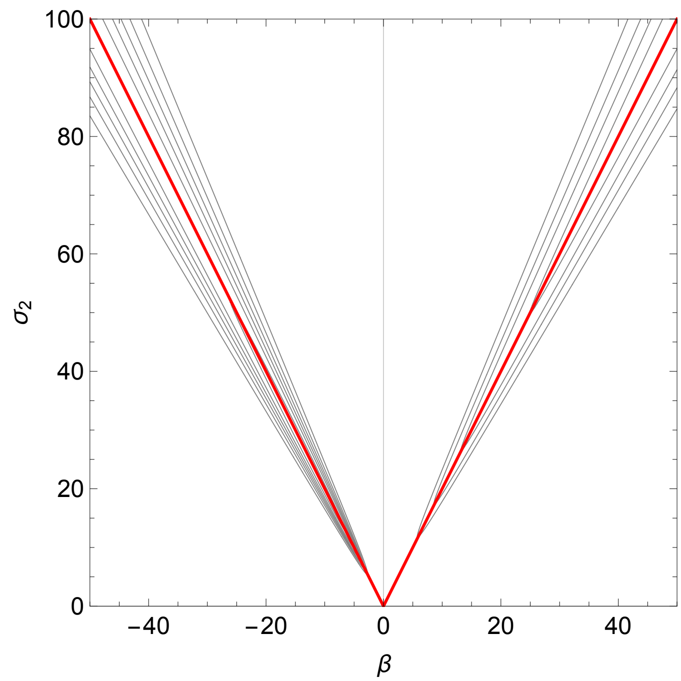

FIGURE 23. Contours of the likelihood for (β, σ_2). The red lines show where $\sigma_2/\beta = \pm 2$. The lines appear to follow the ridges of the likelihood.

Appendix J. Correspondence between examples

In this appendix I show the correspondence between the example in Waggoner and Zha (2003, WZ hereafter) and the example in this paper.[30]

The crucial feature of WZ's example is their matrix A which plays the role of my matrix H [see (6.1)]:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \tag{J.1}$$

The correspondence between the two matrices is complicated by the fact that I reversed the order of the variables and transposed the equations relative to the way WZ expressed the system. Assuming $a_{12} = 0$, the relation between H and A is given by[31]

$$\begin{pmatrix} h_{11} & 0 \\ h_{21} & h_{22} \end{pmatrix} = \begin{pmatrix} a_{22} & 0 \\ a_{21} & a_{11} \end{pmatrix}. \tag{J.2}$$

Consequently, in their notation the normalizations $a_{21} > 0$ and $a_{11} > 0$ correspond to $h_{21} > 0$ and $h_{22} > 0$, respectively.

[30]The example in WZ is dynamic and includes a lag (characterized by by their matrix A_1). To simplify the example, I have ignored the lagged term because it has no effect on the issues under consideration.

[31]$H = PA^\top P$, where $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is a permutation matrix.

In order to generate the data for their example, WZ provide the following numerical values:

$$A = \begin{pmatrix} 0.5 & 0 \\ 0.1 & 0.5 \end{pmatrix}. \tag{J.3}$$

These are the values I have used for my numerical illustration. In addition, WZ set $T = 50$ as I do. However, because the datasets are randomly generated my data do not exactly match theirs. Nevertheless, the features of the likelihoods computed from the data are very similar, suggesting that neither of our datasets is atypical.

In addition to data differences, there is another reason why the corresponding plots for $1/h_{11}$ would not be identical. In particular, I am displaying posterior distributions which involve a prior for ω_2^2 proportional to $1/\omega_2^2$, while WZ compute the distribution directly from the likelihood itself.

Finally, WZ's Figure 2, which corresponds to my Figure 14, requires a more detailed discussion. I believe the axes in their Figure 2 are mislabeled and the labels should be swapped. Differences in appearance between WZ's Figure 2 and my Figure 14 are attributable to the following three things (in addition to data differences): Their plot is transposed relative to mine, the domain of their plot is substantially larger than that of mine, and they display contours for the log of the likelihood whereas I display contours for the likelihood itself.

References

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics 7*(1), 1–26.

Hamilton, J. D. (1994). *Times series analsis.* Princeton, NJ: Princeton University Press.

Hamilton, J. D., D. F. Waggoner, and T. Zha (2007). Normalization in econometrics. *Econometric Reviews 26*(2–4), 221–252.

Kociecki, A. (2013). Towards understanding the normalization in structural VAR models. Working paper No. 47645, Munich Personal RePEc Archive.

Lancaster, T. (2009, May). Explaining bootstraps and robustness. Paper may be found at `http://www.rcfea.org/papers/2009/bayesian/Lancaster.pdf` as of February 2016.

Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics 9*, 130–134.

Waggoner, D. F. and T. Zha (2003). Likelihood preserving normalization in multiple equation models. *Journal of Econometrics 114*, 329–347.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* John Wiley & Sons, Inc.

Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street N.E., Atlanta, GA 30309–4470

E-mail address: `mark.fisher@atl.frb.org`

URL: `http://www.markfisher.net`