Reply to Comments on *Reflections* *

A. Ronald Gallant
Penn State University

First draft: March 1, 2015
This draft: April 10, 2015

Available at www.aronaldg.org

*Address correspondence to A. Ronald Gallant, P.O. Box 659, Chapel Hill NC 27514, USA, phone 919-428-1130; email aronldg@gmail.com.

Abstract

There are three main criticisms. The first is to misread Assumption 1 so as to think that probabilities are being put to zero after having observed the data leading to the erroneous conclusion that the proposed methods are not Bayesian. The second is due to a failure to grasp the relevance of the fifth paragraph of the Introduction leading to suggestions that alternative methods be used that are not available. The third relates to the fact that it might be advisable to make an adjustment that is analogous to a Jacobian term in some applications.

Keywords and Phrases: Moment functions, Structural Models, Bayesian inference

JEL Classification: C32, C36, E27

1 Reply to Main Issues

There are three main criticisms.

The first is due to thinking that Assumption 1 states that the probability assigned to some potential values of the data are being put to zero conditional on what particular value was actually observed. This is not true. What is being done is that labels called representers are being assigned to sets to serve as an index. No probabilities are being altered. Probabilities are determined by the probability space $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ described in the first two paragraphs of Section 3 of *Reflections* and they are never altered subsequently. This misunderstanding leads to an erroneous conclusion that the proposed methods are not Bayesian.

The second is due to a failure to grasp the relevance of the fifth paragraph of the Introduction. In that situation, to my knowledge, there are no Bayesian inference procedures for analytically intractable models that can substitute for the proposals presented in *Reflections*.

The third regards the fact that MCMC presumes that the dominating measure of a density is Lebesgue. Therefore, one cannot logically use the density defined by $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ with MCMC without a multiplicative adjustment. In the case where $Z(x, \theta)$ is one-to-one with x when θ is held fixed, the adjustment is a Jacobian term. Otherwise it relies on parameterizing a representer. Not making the adjustment can be interpreted as using a prior that differs from one's intent. Therefore, the adjustment is not needed if it does not involve θ. The effect of omitting the adjustment typically diminishes as sample size increases. One can argue that not making the adjustment is preferable to making it when $\Psi = \Phi$.

The adjustment itself is computed as follows. For $Z(x, \theta)$ of dimension M one writes the representer x^* of $C^{(\theta,z)}$ in terms of $u = (u_1, \dots, u_M)$, where u can depend on θ. E.g., $x^*(u) = (u_1, \dots, u_M, 0, \dots, 0)$. Given (x, θ) one finds all K solutions $\{\hat{u}_k\}$ to $Z(x, \theta) = Z(x^*(u), \theta)$. The adjustment is $\operatorname{adj}(x, \theta) = \sum_{k=1}^{K} |\det(\partial/\partial u')\, Z(x^*(\hat{u}_k), \theta)\,|$. Case 5 of Example 1 below is an example. There is another in the Reply to Jae-Young Kim.

I shall next expand upon these three points in the order second, first, third.

Regarding the fifth paragraph of the Introduction to *Reflections* let me elaborate by means of an example taken from Gallant and Hong (2007). As far as I know, there is no way to proceed in the Bayesian fashion for this example absent the results of *Reflections.*

Consider an asset with price P_t that pays D_t in period t. It's gross return is $R_t = (P_t + D_t)/P_{t-1}$. Standard models imply the existence of a pricing kernel $\{\theta_t\}_{t=-\infty}^{\infty}$ that satisfies the Euler equation

$$1 = \mathcal{E}_t\left(\theta_{t+1}R_{t+1}\right), \tag{1}$$

where $\mathcal{E}_t$ denotes expectation conditional on information known at time t. The goal is to infer the posterior distribution of $\{\theta_t\}_{t=1}^{n+1}$. The only knowledge of the structural model that one accepts is equation (1) and standard time series regularity conditions on $\{\theta_t, R_t\}$ and variables in the information set. Nothing more.

Gallant and Hong (2007) use a likelihood $p^*(x\,|\,\theta)$ derived from moment equations that are the same as equations (36) through (43) of *Reflections* but with θ_t replacing M_t. Their $\Psi = \Phi$. The dimension of $\theta = (\theta_1, \ldots, \theta_{n+1})$ is 551 and that of m_t is $M = 810$ for their monthly data set. M is large because they use many assets rather than two as in equations (36) through (43). Computational costs are reasonable because their variance matrix is block diagonal. Their prior is obtained by analyzing simulations from a Bansal and Yaron (2004) economy. It is

$$p^*(\theta\,|\,\eta)\,p^*(\eta) = \left\{\left[\prod_{t=1}^{n} f(\theta_{t+1}|\theta_t, \ldots, \theta_1, \eta)\right] f(\theta_1|\eta)\right\}\,p(\eta),$$

where $f(\theta_{t+1}\,|\,\theta_t, \ldots, \theta_1, \eta)$ is a law of motion for $\{\theta_t\}_{t=1}^{n+1}$ and η its parameters.[1]

Their joint density

$$p^*(x, \theta, \eta) = p^*(x\,|\,\theta)\,p^*(\theta\,|\,\eta)\,p^*(\eta) \tag{2}$$

does look like a state space model, as does the example in the fifth paragraph of the Introduction of *Reflections.* The mistake is to think that they are, in fact, state space models with $\{\theta_t\}$ exogenous and latent. $\{\theta_t\}$ is not exogenous. It is endogenous so that one must go to some effort to construct a density $p^*(x\,|\,\theta)$ that can be viewed as a conditional density of x given θ.

I cannot fault anyone for this mistake. I made the same mistake myself, which mistake was forcefully and publicly pointed out to me by Lars Peter Hansen when I presented the

[1] Gallant and Hong (2007) display a plot of the posterior mean of $\{\theta_t\}_{t=1}^{n+1}$ that appears reasonable within the historical context, accept the hypothesis of recursive utility as specified in the long run risks model of Bansal and Yaron (2004), and reject the long run risks model itself.

material in Gallant and Hong (2007) at The North American Summer Meetings of the Econometric Society, Minneapolis, Minnesota, June 22 to June 25, 2006. Those remarks of Lars provoked the line of thought leading to *Reflections.*

Christian Robert is correct that if one sets models such as (2) above aside, there are many alternative Bayesian methods that one might consider and he does a good job of inventorying them. If one adds to his list those inventoried by Enrique Sentana and Dante Amengual, one has a nearly exhaustive list of alternative Bayesian methods one might consider. As to how the proposals of *Reflections* will hold up against these alternatives, my expectation is reasonably well if the examples available to date are not misleading. Method of moments has a powerful appeal in economic research and priors have some appeal as a means to deal with data limitations, so my guess is that Bayesian method of moments methods along the lines of *Reflections* will get used in applied economic research. As John Cochrane (2004) points out, most researchers find evidence based on method of moments more persuasive than evidence based on fully specified likelihoods.

Next let me discuss the confusion caused by Assumption 1 of *Reflections.* Here is the offending sentence.

> Let x^o denote the observed realization of X and let $z^o = Z(x^o, \theta)$. For $z = z^o$ we shall choose the representer of $C^{(\theta,z)}$ to be x^o so that we have $x^o = \Upsilon[Z(x^o, \theta), \theta]$ for every $\theta \in \Theta$.[2]

Chris Sims interpreted this to mean

> Within each $C^{z,\theta}$ put probability zero on all the x points except one ...

This interpretation is not correct. Actually, one is merely choosing a convenient label for the preimage $C^{(\theta,z)}$ that contains (x^o, θ). Probabilities are not being set to zero. The remarks in paragraphs three through eight of Section 1 of Chris Sims's *Comment* are therefore not relevant and the conclusion that what is proposed in *Reflections* is not Bayesian is not correct.

The offending sentence should be deleted from Assumption 1 for two reasons: The notational convenience it affords is not worth the risk of the confusion it apparently can cause. It contradicts the construction of the adjustment $\text{adj}(x, \theta)$ above.

[2]If this sentence is deleted from Assumption 1, then $\Upsilon(z^o, \theta)$ has possibly some other definition and the data x^o gets mapped to its representer $x^* = \Upsilon[Z(x^o, \theta), \theta]$ instead of to itself.

The following example is taken from Chris Sims's *Comment* and will facilitate discussion of the third criticism.

EXAMPLE 1 Consider

$$(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o) \tag{3}$$

with $\mathcal{X} = (0, \infty)$, $\Theta = (1, \infty)$, $\mathcal{C}^o$ the collection of Borel sets, and

$$p^o(x, \theta) = \left(\theta\, x^{\theta-1} e^{-x^\theta}\right) p^*(\theta). \tag{4}$$

Under the transformation $z = x^\theta, u = \theta$, we have $x = z^{1/u}, \theta = u,\ \det[J(z,u)] = u^{-1} z^{1/u-1}$, and

$$\text{pdf}\,(z, u) = p^o(z^{1/u}, u)\ \det[J(z,u)] = e^{-z}\, p^*(u).$$

Integrating out u we have that $Z(x, \theta) = x^\theta$ has the exponential distribution. A moment condition for estimating θ is $m = x^\theta - 1$, which has density e^{-m-1} with support $(-1, \infty)$. But when $m = x^\theta - 1$ is substituted into e^{-m-1} the ones cancel so that $Z(x, \theta) = x^\theta$ with density $\psi(z) = e^{-z}$ that has support $(0, \infty)$ is what one actually uses. For $S(x, \theta) = \sum_{i=1}^n x_i^\theta$ computed from a random sample of size n from $p^o(x \,|\, \theta)$ we have $\psi(s) = s^{n-1} e^{-s} / \Gamma(n)$ $\square$

For $\mathcal{C}$ measurable $f(x, \theta)$, which must be of the form $f(e^{x^{-\theta}})$, we have

$$\int_1^\infty \int_0^\infty f(e^{x^{-\theta}}) \left(\theta\, x^{\theta-1} e^{-x^\theta}\right) p^*(\theta)\, dx\, d\theta = \int_1^\infty \left[\int_0^\infty f(z) \psi(z)\, dz\right] p^*(\theta)\, d\theta. \tag{5}$$

And, for the probability of a rectangle $(0, \infty) \times (c, d)$

$$\int_c^d \left[\int_0^\infty \left(\theta\, x^{\theta-1} e^{-x^\theta}\right)\, dx\right] p^*(\theta)\, d\theta = \int_c^d \left[\int_0^\infty \psi(z)\, dz\right] p^*(\theta)\, d\theta. \tag{6}$$

Equation (5) states that for $\mathcal{C}$ measurable functions, expectation with respect to the space $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$, which is the left hand side, is equal to expectation with respect to the space $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$, which is the right hand side. Similarly, equation (6) states that both spaces assign the same probability to rectangles. This, and the existence of a conditional density on $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ with respect to some dominating measure, is the main result of *Reflections*.

Now comes the point that I did not realize until I studied Chris Sims's example:

> MCMC implicitly assumes that the dominating measure on $(\mathcal{X} \times \boldsymbol{\Theta}, \mathcal{C}^*, \mathbf{P}^*)$ is Lebesgue.

Therefore, for Example 1, using $e^{-x^\theta} p^*(\theta)$ for MCMC implies that $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ has density

$$\operatorname{pdf}(\theta, x) = \frac{e^{-x^\theta} p^*(\theta)}{\int_1^\infty \int_0^\infty e^{-x^\theta} p^*(\theta)\, dx\, d\theta}.$$

with Lebesgue as the dominating measure. This differs from (4). One can take the view that the implied conditional density differs from (4), or take the view that the implied prior differs from (4), or both. Of these three views, the most useful seems to be to regard failure to make an adjustment as a distortion to one's intended prior.

For Z, because z is one-to-one with x when θ is fixed, the adjustment is the Jacobian term $\theta\, x^{\theta-1}$. If one uses $\left(\theta\, x^{\theta-1}\right) e^{-x^\theta}\, p^*(\theta)$ for MCMC one gets $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*) = (\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$.

For S, write the representer of $C^{(\theta,s)}$ as $x^*(u) = (u, 0, \ldots, 0)$. The solution to $S(x,\theta) = S(x^*(u),\theta)$ is $u = s^{1/\theta}$. Then the adjustment is $\operatorname{adj}(x,\theta) = (d/du) S(x^*(u))\big|_{u=s^{1/\theta}} = \theta\, u^{\theta-1}\big|_{u=s^{1/\theta}} = \theta\, s^{1-1/\theta} = \theta \left(\sum_{i=1}^n x_i^\theta\right)^{1-1/\theta}$.

To gain a feel for how much these considerations matter, I simulated a random sample from $p^o(x \,|\, \theta) = \left(\theta\, x^{\theta-1}\right) e^{-x^\theta}$ with $\theta = 5$. For the prior I used $p^*(\theta) \propto I_{(1,\infty)}(\theta)\, n(\theta \,|\, 5, 5^2)$,[3] and considered five cases:

1. Bayes using Example 1, i.e., MCMC using

   $$\log \operatorname{pdf}(x \,|\, \theta) = n \log \theta + (\theta - 1) \sum \log x_i - \sum x_i^\theta$$

   with prior $p^*(\theta)$.

2. $Z(x,\theta) = \sum {x_i}^\theta - n$ implemented with MCMC using

   $$\log \operatorname{pdf}(x \,|\, \theta) = (n-1) \log\Big(\sum {x_i}^\theta\Big) - \sum {x_i}^\theta - \log \Gamma(n)$$

   with prior $p^*(\theta)$.[4]

3. Method of *Reflections* Section 2 with $m(x_i,\theta) = {x_i}^\theta - 1$, $\Psi = \Phi$, and prior $p^*(\theta)$.

4. Method of *Reflections* Section 2 with $m(x_i,\theta) = \theta^{-1} - \log x_i - (\log x_i) x_i^\theta$, which is the score of Example 1, $\Psi = \Phi$, and prior $p^*(\theta)$.

[3] These choices were haphazard and I did not consider any others. Should one wish to try others themselves, the code is at `http://www.aronaldg.org/webfiles/mle`.

[4] With the $\log \Gamma$ term omitted because it is a normalization term that does not affect MCMC.

Table 1. Summary of MCMC Draws for Example 1

	$n=2$		$n=20$		$n=200$		$n=2000$	
Case	Mean	SDev.	Mean	SDev.	Mean	SDev.	Mean	SDev.
1	4.0473	1.8675	5.2498	0.8260	5.2982	0.2865	5.0725	0.08509
2	5.4622	3.0674	3.5376	1.5615	4.8828	1.4957	5.0299	0.27804
3	1.9934	0.7954	5.1121	3.1283	6.0428	1.2143	5.0967	0.27663
4	6.2694	3.3036	6.7336	2.6303	5.3563	0.3084	5.0800	0.08819
5	5.5203	2.7735	4.8472	1.4248	5.6847	0.8992	5.0709	0.26997

The total number of MCMC draws was 100,000. Means and standard errors were computed from every 25th draw. All chains were started at $\theta = 5$. Trace plots indicated that transients died out within 10 draws. The code is at `http://www.aronaldg.org/webfiles/mle`.

5. $Z(x,\theta) = \sum {x_i}^{\theta} - n$ with adjustment, i.e., implemented with MCMC using

$$\log \text{pdf}\,(x \mid \theta) = \log\theta + (1 - 1/\theta)\log\left(\sum x_i^{\theta}\right) + (n-1)\log\left(\sum {x_i}^{\theta}\right) - \sum {x_i}^{\theta} - \log\Gamma(n)$$

with prior $p^*(\theta)$.

.

Table 1 indicates that, as is usually the case, priors become swamped by data eventually so that the consequence of omission of an adjustment diminishes with sample size. This is consistent with interpreting no adjustment as causing the actual prior[5] to differ from $p^*(\theta)$. Does presence or absence of the adjustment affect whether or not method of moments with a prior can be regarded as a Bayesian procedure? No.[6] It is only that the prior is not exactly what one specifies. Operationally this does not matter much because method of moments is

[5] Which may be data dependent. And there may be integrability issues. See Chris Sims's *Comment.* With respect to integrability, using a prior with compact support is, in general, good practice with MCMC and usually eliminates this consideration. However, even then, one should check to be sure posterior mass is not accumulating at a boundary.

[6] If one accepts the use of data dependent priors. I do, others do not. With respect to observational data, one's views represent accumulated experience so that one can argue that all priors for observational data sets that span more than one time period are data dependent.

well understood, especially if $\Phi = \Psi$ when it becomes the Chernozhukov and Hong (2003) estimator, in which case the effect of imposing a prior is easily anticipated. What is relevant is that one can legitimately use the Bayesian inference apparatus with method of moments regardless of whether one makes an adjustment or not.

Reflections has shown that moment conditions allow a probability space $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ to be deduced from a probability space $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ that was derived from a structural economic model and a prior. And that these two probability spaces have enough in common that they can serve as substitutes. This part is pure mathematics and I believe that the logic is correct. If one can actually infer Ψ from $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$, there again should be no logical problem. Information is lost because $\mathcal{C}^* \subset \mathcal{C}^o$. I like to think of this as similar to the information loss that occurs when one divides the range of a continuous variable into intervals and uses a discrete distribution to assign probability to each interval. Both the continuous and discrete distributions assign the same probability to each interval but the discrete distribution cannot assign probability to subintervals.

Difficulties begin when one chooses moment conditions and asserts a distribution Ψ for Z, then uses $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ for Bayesian inference with little regard for $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ other than for suggestions as to what Z and Ψ ought to be. The Gallant and Hong (2007) example above is an illustration of this approach. After doing so, is one now obliged to demonstrate the existence of a fully specified structural model to justify the approach?

One can try to circumvent the problem by claiming that asymptotic normality justifies the choice $\Psi = \Phi$, allowing one to omit a demonstration of existence. This approach is fraught with peril[7] as discussed in the *Comments*, especially that of Chris Sims. What is not clear to me is to what extent the various counter examples apply in the particular case where the object of interest is expressed explicitly as a function of the data in a reasonably intelligent fashion and MCMC is the computational device. The counter example to Kwan in Chris Sims's *Comment* is a case in point. It requires the use of an expression that presupposes knowledge of the true value of the parameter if one regards θ as fixed or requires use of a moment condition that does not have unconditional expectation zero if one regards θ as random. In fairness, the paragraphs in Chris Sims's *Comment* following the counter example

[7] Case 3 of Table 1 for $n = 2$ is an illustration of the consequence of extreme violation of normality.

do blunt its impact and perhaps even make the same point that I'm trying to make here.

A more extreme way to deal with the question of does there exist a $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ that implies $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ is simply to set $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ equal to $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$.[8] This actually seems legitimate to me. One is insisting that $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ is the only information one is willing to use to address the inference problem at hand.

My personal view is that choosing $\Psi = \Phi$ and omitting the adjustment is the best approach. At least one understands what one is doing: Using method of moments in the style of Chernozhukov and Hong (2003) but pushing results toward one's beliefs with a prior. The most compelling reason for doing so is that the data are sparse and that doing so is far more scientifically credible and transparent than calibration.

I shall next respond to each *Comment* individually. A failure to react to a point made in a *Comment* that has already been discussed in this section means that I do not see how reacting to it is helpful, usually because I agree with it.

2 Reply to John Geweke

I have no problems at all with John Geweke's *Comment*. He has identified several concerns and problems that some of the proposals in *Reflections* entail and I think them well justified.

Regarding his Example 1, his model (1) implies that there can be no Ψ with support $\mathbb{R}$ which rules out the normal, as stated in John Geweke's *Comment*. Knowing θ and z rules out certain x which was a feature of the example of Section 3.1 of *Reflections*. Therefore specification of Ψ and Z partially determines a prior. Nonetheless, John Geweke's Example 1 does make some interesting points, especially the potential sensitivity of conclusions to which moments are chosen.

With respect to John Geweke's remarks on Section 4 of *Reflections*, see also the *Comment* of Oliver Linton and Wu Ruochen.

[8]Logical consistency would require that $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ is $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ with the adjustment applied.

3 Reply to Enrique Sentana and Dante Amengual

The *Comment* by Enrique Sentana and Dante Amengual is a very nice survey of the literature. I have three remarks. The first is to note that the motivating problem discussed at the beginning of Section 1 of *Reply* rules out many comparisons. The second is that the behavior of method of moments estimators depends more on skill at selecting moments than anything else with the best choice usually being the estimating equations of maximum likelihood. Thus, comparison of procedures is hard to do in an objective fashion. Comparison of regularity conditions and scope of applicability, as Enrique Sentana and Dante Amengual advocate, is, of course, exempt from the second remark. The third remark is that my guess is that most applications of the notions in *Reflections* will occur in time series settings. This too limits the scope of comparisons.

4 Reply to Oliver Linton and Wu Ruochen

Oliver Linton and Wu Ruochen provide a very nice extension to the analysis of the habit model presented in Section 4 of *Reflections*. In this connection, see also John Geweke's discussion of Section 4. Of most interest in Oliver Linton and Wu Ruochen's discussion is the role that the prior plays and the lack of stability of habit model parameter estimates, especially when periods of more volatile consumption enter the data. They maintained the tight priors on the parameters g, σ, ϕ, and δ in their analysis that were imposed in Section 4 of *Reflections*. Had they relaxed these the instability would have been far worse. The data and code for Section 4 are at `http://www.aronaldg.org/webfiles/mle` should one want to experiment themselves.

5 Reply to Chris Sims

Before reading please read Section 1, Reply to Main Points.

The logic of the paper is that one assumes a presample $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ and derives the space $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ used for inference, not the other way around. When it comes to applications one may cut corners by starting with $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ without worrying much about $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$. But using a risky method in practice does not negate the logic of the paper.

The statement "Within each $C^{(z,\theta)}$ put probability zero on all the x points except one ..." in paragraph three of Section 1 of *Comment* by Chris Sims is not correct. Within $C^{(z,\theta)}$ there may be many x for which $Z(x,\theta) = z$, all that is being done is choosing one to index $C^{(z,\theta)}$. A convenient property of the index x^* is that $Z(x^*,\theta) = z$. Probabilities are determined by the probability space $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ described in the first two paragraphs of Section 3 of *Reflections* and they are never altered subsequently. The remarks in paragraphs three through eight of Section 1 of Chris Sims's *Comment*, which follow from the misconception that probabilities are set to zero after observing the data, are therefore not relevant and therefore the conclusion that what is proposed in *Reflections* is not Bayesian is incorrect.

When Chris Sims is discussing his continuous example, which begins in paragraph nine of Section 1 of his *Comment*, he makes this statement:

> Is there then a joint distribution for θ and x in which $p(x \mid \theta) = \psi(x \mid \theta)$ and the unconditional distribution of x^θ is exponential? No.

But his analysis answers this question:

> Is there then a joint distribution for θ and x in which $p(x \mid \theta) = \psi(x \mid \theta)$ and the unconditional distribution of x^θ is exponential **when all densities are with respect to Lebesgue measure**?

As to what's going wrong with Chris Sims's analysis, it appears to me that he is using the change of variable formula in the wrong direction. What he is doing is starting with a probability space $(\mathcal{Z}, \mathcal{B}, P_Z)$, setting forth a random variable $X(z)$, and deducing a space $(\mathcal{X}, \mathcal{C}, P_X)$ using the change of variable formula $P_X(C) = P_Z\left[X^{-1}(C)\right]$. Moreover, he is assigning conditional probability to sets of the form $(a < x < b)$. This cannot be done: line segments parallel to an axis are not preimages of $Z(x,\theta) = x^\theta$. *Reflections* is using the change of variable formula in the opposite direction. We know $(\mathcal{Z}, \mathcal{B}, P_Z)$ and are trying to infer $P_{(X,\theta)}$ using $Z(x,\theta)$. The only sets C to which $P_{(X,\theta)}$ can assign probability are of the form $C = Z^{-1}(B)$. The probability we assign to C is $P_{(X,\theta)}(C) = P_Z\left[Z^{-1}(B)\right]$. Moreover, when we condition on θ, we cannot enlarge the σ-algebra to include line segments parallel to an axis. We can only assign conditional probability to sets of the form $C = Z^{-1}(B)$. All this relates to the probability space $(\mathcal{X} \times \Theta, \mathcal{C}, P)$. We then add a prior to get the space $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$. At that point we can, if desired, re-express $p^*(x \mid \theta)p^*(\theta)$ so that Lebesgue is the dominating measure. Case 5 of Example 1 above, is an illustration. There is another in the Reply to Jae-Young Kim.

Setting aside the erroneous conclusion that the methods proposed in *Reflections* are not Bayesian due to misreading Assumption 1, Chris Sims's other main points are correct despite not following the logic of *Reflections* in his mathematics: One should be aware that claiming that asymptotic normality of Z justifies setting $\Psi = \Phi$ carries risks that one might try to reduce in an application if one can, e.g., by simulation or checking regularity conditions. And one should either make an adjustment to $p^*(x \mid \theta)p^*(\theta)$ analogous to a Jacobian term as in Case 5 of Reply to Main Points to account for the fact that MCMC presumes that the dominating measure is Lebesgue or be aware that the actual prior one is using may not be the stated prior. The exceptions to the second point are that one can omit the adjustment without consequence if it does not depend on θ, that the need for the adjustment usually diminishes as sample size increases, and that one may prefer not to use it for the sake of transparency.

6 Reply to Christian Robert

Before reading, please read Section 1, Reply to Main Points.

Christian Robert's *Comment* can be characterized as a collection of many isolated remarks, some similar to those of Chris Sims, and some of which are opinions. Some opinions I share, some do not, but everyone is entitled to opinions, and arguing about opinions here will not accomplish much. I'll restrict attention to those remarks that are more in the nature of fact, are not similar to those made by Chris Sims, and for which I think responding will be helpful.

I think that Christian Robert does not appreciate the relevance of the fifth paragraph of the Introduction of *Reflections* and that this lack of appreciation flavors many of his remarks. Or he does, and it really bothers him to get the notion of what is a parameter and what is data entangled in such a fashion.

Regarding Christian Robert's Abstract, *Reflections* is actually not about the construction of priors, excepting when they get constructed inadvertently via violation of Assumption 1 as discussed in Sections 3.1 and 3.3 of *Reflections*. In the end, Assumption 1 gets imposed and the prior is as specified in the second paragraph of Section 3. *Reflections* is, in the main, a paper about the construction of likelihoods. Christian Robert's conclusion about priors in his Abstract regards a peripheral aspect of the paper.

On rereading Section 3 of *Reflections* I admit that there is some ambiguity as to what is intended. What is intended is that one has a prior but that when trying to construct a likelihood using moment conditions one can get oneself in trouble by simultaneously partially specifying a prior that can interfere with imposing the intended prior. Upon discovering this fact in Sections 3.1, 3.3, and 3.6, one gets rid the problem with Assumption 1.

Regarding "I would suggest examining the range of prior$\times$likelihood pairs that agree with this partial property when using the regular Borel σ-algebra," I would argue that that was done in Section 3.3 of *Reflections* and that it did not seem to accomplish anything useful. It rather suggested to me that one should eliminate this case by imposing Assumption 1.

Regarding Newton and Raftery (1994) and its potential for disaster, my own experience is that all methods for computing the marginal likelihood from an MCMC chain when the normalizing constants of densities are not known are disasters. The best of the lot is supposed to be method f_5 from Gamerman and Lopes (2006, section 7.2.1). But it doubles the computational cost by requiring draws from the prior, has an iterative phase that does not always converge, and gives results that appear to me to be no less ridiculous than Newton and Raftery's method in applications. At least with Newton and Raftery one understands what one is doing: comparing harmonic means.

In my view, Zellner (1997) has no relation at all to *Reflections* other than that the words Bayesian and moments appear in their titles. I also do not understand how whether or not Gallant and McCulloch (2009) is an ABC method relates to *Reflections*.

All this said, Christian Robert's *Comment* contains many useful observations and suggestions as to how the developments in *Reflections* could have been done differently and maybe better.

7 Reply to Wei Wei and Asger Lunde

This is a nice application. It never would have occurred to me to proceed in this fashion. I would have used a simulation estimator for this sort of problem out of habit; e.g., Gallant and Tauchen (2015). A simulation estimator requires numerical approximation whereas Wei Wei and Asger Lunde's approach does not. Very nice!

8 Reply to Jae-Young Kim

I believe that Jae-Young Kim's analysis of Student's t is correct: The Bayes credibility interval is indistinguishable from the frequentist confidence interval if one interprets $p^*(\theta) = c$ to mean a proper prior that is nearly flat everywhere.[9] This because the representer[10] $x^*(s) = (\mu/s, 0, \ldots, 0)$ with $s = \frac{1}{n}(1-t)$ solves[11] $T(x^*(s), \mu) = t$ and $(\partial/\partial s)T(x^*(s), \mu)$ is a constant so that no adjustment is needed.

Most of Jae-Young Kim's other comments have been touched on above. I'll respond to those where more seems needed. As to what are "first principles," how about knowledge of Lindley (1985) and Gallant (1997)? The values shown in Table 1 come from the last four unnumbered equations in Section 3.3. I.e., after die Λ is thrown die X can be equal to, move up one from, or down one from the realization θ of Λ with equal probability, except that X is equal to θ with probability one when θ is 1 or 6. The die Λ is fair. The absence of rectangles means that one can not assign probability to the interval $a < \theta < b$, which is the usual operation with a posterior distribution.

9 Additional References

Bansal, R., and A. Yaron. (2004). "Risks For the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance 59,* 1481–1509.

Cochrane, John H. (2004), *Asset Pricing (Revised Edition)*, Princeton University Press,

[9] E.g., a normal with large variance.

[10] I have not yet settled in my own mind whether or not results are sensitive to the choice of the parameterization of a representer. I think not, but am not certain. If they are, it seems sensible to find a parameterization for which no adjustment is necessary, as here.

[11] Use l'Hospital's rule when $t = 1$.

Ch. 16, esp. pp. 292 & 298.

Gallant, A. Ronald (1997), *An Introduction to Econometric Theory: Measure Theoretic Probability and Statistics with Applications to Economics*, Princeton University Press

Gallant, A. Ronald, and George Tauchen (2015), "Efficient Indirect Bayesian Estimation: With Application to the Dominant Generalized Blumenthal-Getoor Index of an Ito Semimartingale," Working paper, Department of Economics, Penn State University, http://www.aronaldg.org/papers/hfemm.pdf

Gamerman, D., and Lopes, H. F., (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd Edition)*, Chapman & Hall.

Lindley, Dennis (1985), *Making Decisions, Second Edition*, Wiley.

Zellner, Arnold (1997), "The Bayesian method of moments," In *Applying Maximum Entropy to Econometric Problems, Advances in Econometrics, Volume 12.* T. B. Fomby and R. Hill, eds. Emerald Group Publishing Limited, Bingley, 85 – 105.