Mostly Harmless *Bayesian* Econometrics

Dale J. Poirier

University of California, Irvine

October 4, 2018

Abstract

Think back to those early years when you took your first statistics course. You studied examples of location / scale models (i.e., the pdf is of the form $p[\sigma^{-1}(y - \mu)]$, where $\mu \in \Re$ and $\sigma \in \Re_+$). Frequentists developed simple estimators with known *ex ante* finite sampling distributions leading in many cases to "optimal" procedures for point estimation, interval estimation, hypothesis testing, and prediction. Likewise Bayesians developed corresponding *ex post* finite posterior distributions, optimal point estimates, highest posterior density (HPD) regions, Bayes factors, and predictive distributions. Conjugate priors were readily available (when sampling from the exponential family) and integrals had analytical closed-form expressions. In short, the models were simple and there were a small number of easily interpretable parameters.

Now consider your last graduate course in econometrics. The models were much more complicated, the number of variables parameters became more numerous, computational demands increased, and life moved to the magical land of *Asymptopia* where everything was consistent, normal, and efficient. If only you could return to that innocent world of your youth. Here I hope to take you on a journey to your Fountain of Youth.

The index of the widely praised book by Angrist and Pischke (2009) (hereafter AP09) contains none of the words: Bayes, Bayesian, nor likelihood. Above I have inserted "Bayesian" into their title and adopted it as the title of this monograph. I borrow those parts of AP09 that I like, and add to it an analysis incorporating these three words. The other parts of AP09 I ignore. This is not a book review of AP09. ***Instead I provide a contrasting Bayesian spin on points of agreement with AP09.***

In their preface, Angrist and Pischke (2009, p. xii) say:

> "*... that empirical work is most valuable when it uses data to answer specific causal questions*."

A distinguishing feature of their approach Angrist and Pischke (2009, pp. xii-xiii) describe as:

> "Most econometrics texts appear to take econometric models very seriously. ... We take a more forgiving and less literal-minded approach." The authors claim a "certain lack of gravitas" and " ... not much concern with asymptotic efficiency. Rather, our discussion of inference is devoted mostly to the finite-sample bugaboos that should bother practitioners."

This willingness to entertain sub-optimal efficiency is form of "econometric satisficing."

My intended audience is not econometric theorists, but rather applied econometricians, and particularly, *non-Bayesians* / frequentist. I believe there are many frequentists whose statistical practice reflects hidden beliefs about statistical matters. I do ***not*** wish to put all frequentists into a "basket of deplorables." That has been shown to not be a winning strategy. Instead my main goal is to create a "bigger statistical tent" within the econometrics community by ***interpreting non-Bayesian approaches favorably from a Bayesian standpoint***. I consider only *random sampling* and *saturated models* that reduce to simple location-scale families. I intentionally omit proofs and ignore detailing the regularity conditions under which asymptotic results follow. These can be found in the literature I cite.

Here is the brief summary of what follows. Through a lengthy list of quotations, Chapter 1 points out the wide agreement among economists that they don't care about the *literal truth* of their assumptions, be they behavioral or probabilistic. Section 1.2 argues "truth" amounts to a ***shared perception of facts***. Instead of demanding "truth" of assumptions, Section 1.3 argues a more pragmatic issue is the sensitivity of conclusions to small changes in the assumptions. Section 1.4 tackles an old issue, instrumentalism versus realism, and argues, rather than be true or realistic, models should be instruments for prediction. Together, Sections 1.5 and 1.6 argue that the rise of behavioral economics (Section 1.5) led to the so-called ***credibility revolution*** in econometrics (Sections 1.6). I wish to offer a Bayesian alternative to the widely popular AP09.

Chapter 2 discusses saturated models. Section 2.1 argues that all covariates are inherently *discrete*, and hence, there is a positive probability of *replicated* observations. I focus on simple location / scale models for clusters of observations with the same covariate values. Section 2.2 discusses an empirical example. Working with clusters within which all observations have the same covariate values leads to exact matching of covariates. Sections 2.3 and 2.4, respectfully, introduce important notation for univariate and multivariate models. Section 2.5 briefly discusses pooling of clusters, a topic of great importance in subsequent material. No whimsical assumptions are needed in this chapter other than random sampling.

Chapter 3 is a general discussion of statistical inference. This is where those whimsical distributional assumptions arise. The likelihood approach is discussed in Section 3.2 and the Bayesian approach in Section 3.3. The notion of a ***misspecified model*** (aren't they all?) is discussed in likelihood terms in Section 3.4 and in Bayesian terms in Section 3.5. A brief comparison of the asymptotics for frequentists and Bayesians is given in Section 3.6. For frequentists probability reflects a unique ***property of reality***, whereas for a Bayesians probability reflects ***personal beliefs regarding reality***.

The main purpose of Chapter 3 is to find some common ground for frequentists and Bayesians. This requires a likelihood and a prior. I argue that frequentists implicitly make assumptions about sampling properties of the data when they choose their preferred estimator. In other words, choosing an estimator presumably is made on the belief it works well for the situation at hand. So let's give the frequentist the benefit of doubt, and say this choice coincides with the maximum likelihood estimator (MLE). In Section 3.7 I argue this implies an ***implicit likelihood*** and consider examples for a variety of preferred frequentist estimators. Similarly, in Section 3.8, I argue that many other frequentist notions imply the existence of an ***implicit prior***. For example, the common remark that an estimate "has the wrong sign" certainly contains beliefs that suggest where most of the prior's mass should be. I argue a similar implicit prior exists when the preferred estimator is ***not*** uniformly convergent, then the ***belief*** that the estimator's asymptotic distribution is sufficiently accurate in the finite sample at hand reflects an ***implicit prior*** that regions of the parameter space where the convergence is weakest are assigned small positive probability. Section 3.9 provides a few conclusions.

Chapter 4 considers the case where the endogenous variables are univariate or multivariate normal. Different cases are considered: univariate known variances (Section 4.2), univariate unknown homoskedastic variances (Section 4.3), univariate unknown heteroskedastic variances (Section 4.4), multivariate known variances (Section 4.5), multivariate unknown homoskedastic variances (Section 4.6), multivariate unknown heteroskedastic variances (Section 4.7).

The argument in Chapter 2 that all data are discrete conflicts with the cases considered in Chapter 4. So Chapter 5 considers discrete endogenous variables. It turns out that the analysis is simpler in Chapter 4. Four broad cases are considered: binomial (Section 5.2), multinomial (Section 5.3), ordered multinomial (Section 5.4), and count variables (Section 5.5). Similarities to contingency table analysis are noted.

Chapter 6 addresses the issue of ***randomization***, which together with saturation is a fundamental element of the credibility revolution. After a brief history of randomization in Section 6.1, Section 6.2 discusses frequentist criticisms and qualifications of randomization. Section 6.3 proposes a reconciliation, based on Kadane and Seidenfeld (1990), of two distinct viewpoints on whether the purpose experiment is *to learn* or *to prove*.6

Finally, Chapter 7 offers a brief conclusion.