Complementary Bayesian Method of Moments Strategies*

A. Ronald Gallant
Penn State University

First draft: August 7, 2016
This draft: November 5, 2016

www.aronaldg.org/papers/cb.pdf

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

Abstract

We explore practical aspects of two complementary Bayesian method of moments strategies using a macro-finance application. The first, termed here "moment constrained Bayes," uses a sieve to represent represent the density of the data. Taking the expectation of the moment conditions with respect to the sieve generates parametric restrictions on sieve parameters and introduces additional parameters from the moment conditions. An advantage of moment constrained Bayes is that it provides both an estimate of the density of the data as well as estimates of the parameters that appear in the moment conditions. The difficulty with moment constrained Bayes is computational: The parameter space is singular with respect to Lebesgue measure making Markov Chain Monte Carlo methods difficult to implement. To circumvent the computational difficulty, we use a penalty function approach, termed here the "λ-prior method," to generate draws from a close approximation to the posterior. The second Bayesian method of moments strategy, termed here "moment induced Bayes," uses a semi-pivotal Z constructed from the moment conditions and an assumed distribution Ψ for Z to infer a likelihood and thereby proceed to Bayesian inference. Moment induced Bayes provides estimates of the parameters that appear in the moment equations only. The difficulty with moment induced Bayes is that one must choose Ψ. To circumvent the difficulty, we use draws of Z from moment constrained Bayes to infer Ψ.

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

JEL Classification: C32, C36, E27

1 Introduction

We explore practical aspects of two complementary Bayesian method of moments strategies using a macro-finance application, which is

EXAMPLE 1 Let C_t denote the annual consumption endowment. Let P_{ct} denote the price of an asset that pays the consumption endowment. Let

$$R_{ct} = (P_{ct} + C_t)/P_{c,t-1} \tag{1}$$

denote the gross return on the consumption endowment. In general, if an asset S pays D_{st} per period and has price P_{st}, then its gross return is $R_{st} = (P_{st} + D_{st})/P_{s,t-1}$. Prices are real.

The Epstein-Zin (1989) and Weil (1990) variant of the Kreps-Porteus (1978) utility function is

$$U_t = \left[(1-\delta)\, C_t^{(1-1/\psi)} + \delta \left(\mathcal{E}_t U_{t+1}^{1-\gamma} \right)^{\frac{1-1/\psi}{1-\gamma}} \right]^{\frac{1}{1-1/\psi}}, \tag{2}$$

where δ is the time preference parameter, γ is the coefficient of risk aversion, and ψ is the elasticity of intertemporal substitution. Under the budget constraint $W_{t+1} = (W_t - C_t) R_{c,t+1}$, where W_t is the representative agent's wealth, the agent's intertemporal marginal rate of substitution is

$$\text{MRS}_{t,t+1} = \delta^{\beta} \left(\frac{C_{t+1}}{C_t} \right)^{-(\beta/\psi)} \left(R_{c,t+1} \right)^{(\beta-1)}. \tag{3}$$

where

$$\beta = \frac{1-\gamma}{1-1/\psi}.$$

The gross return of the asset that pays the consumption endowment satisfies the Euler equation

$$1 = \mathcal{E}_t \left(\text{MRS}_{t,t+1} R_{c,t+1} \right) \tag{4}$$

and the gross return on an asset S that pays D_{st} satisfies

$$1 = \mathcal{E}_t \left(\text{MRS}_{t,t+1} R_{s,t+1} \right). \tag{5}$$

The following data were constructed for the 86 years 1930 to 2015 as described in Subsections 2.1 and 2.2 below.

- s_t = log real gross stock return (value weighted NYSE/AMEX/NASDAQ).

- b_t = log real gross bond return (30 day T-bill return).
- c_t = log real per capita consumption growth (nondurables and services).
- w_t = log real gross wealth return, $w_t = \log(R_{ct})$ (see (1)).
- $\text{mrs}_{t-1,t}$ = log marginal rate of substitution, $\text{mrs}_{t-1,t} = \log(\text{MRS}_{t-1,t})$ (see (3)).

Let x denote an array of extent n whose columns are $x_t = (s_t, b_t, c_t, w_t)'$. The process $\{x_t\}_{t=-\infty}^{\infty}$ is assumed to be strictly stationary; i.e., the distribution of $(x_{t+1}, \ldots, x_{t+L})$ is the same as the distribution of $(x_1, \ldots, x_L)$ for any t and any L.

Given the parameters $\theta_{(2)} = (\gamma, \psi, \delta)$ and x, one can compute the pricing errors

$$\begin{aligned} e_{1,t,t-1} &= 1 - \exp(\text{mrs}_{t-1,t} + s_t) \\ e_{2,t,t-1} &= 1 - \exp(\text{mrs}_{t-1,t} + b_t) \end{aligned}$$

and form the following moment equations for the estimation of $\theta_{(2)} = (\gamma, \psi, \delta)$

$$\begin{aligned} m_1(x_t, x_{t-1}, \theta_{(2)}) &= e_{1,t,t-1} \\ m_2(x_t, x_{t-1}, \theta_{(2)}) &= e_{2,t,t-1} \\ m_3(x_t, x_{t-1}, \theta_{(2)}) &= e_{1,t,t-1} \times s_{t-1} \\ m_4(x_t, x_{t-1}, \theta_{(2)}) &= e_{1,t,t-1} \times b_{t-1} \\ m_5(x_t, x_{t-1}, \theta_{(2)}) &= e_{1,t,t-1} \times c_{t-1} \\ m_6(x_t, x_{t-1}, \theta_{(2)}) &= e_{1,t,t-1} \times w_{t-1} \\ m_7(x_t, x_{t-1}, \theta_{(2)}) &= e_{2,t,t-1} \times s_{t-1} \\ m_8(x_t, x_{t-1}, \theta_{(2)}) &= e_{2,t,t-1} \times b_{t-1} \\ m_9(x_t, x_{t-1}, \theta_{(2)}) &= e_{2,t,t-1} \times c_{t-1} \\ m_{10}(x_t, x_{t-1}, \theta_{(2)}) &= e_{2,t,t-1} \times w_{t-1} \end{aligned}$$

Let

$$\bar{m}(x, \theta_{(2)}) = \frac{1}{n} \sum_{t=2}^{n} m(x_t, x_{t-1}, \theta_{(2)}). \tag{6}$$

where

$$m(x_t, x_{t-1}, \theta_{(2)}) = \begin{pmatrix} m_1(x_t, x_{t-1}, \theta_{(2)}) \\ m_2(x_t, x_{t-1}, \theta_{(2)}) \\ \vdots \\ m_{10}(x_t, x_{t-1}, \theta_{(2)}) \end{pmatrix} \tag{7}$$

Abbreviating $m(x_t, x_{t-1}, \theta_{(2)})$ by m_t and $\bar{m}(x, \theta_{(2)})$ by $\bar{m}$, define a heteroskedatic autoregressive invariant (HAC) estimate of the variance of $\bar{m}(x, \theta_{(2)})$ by

$$W\big(x, \theta_{(2)}\big) = \sum_{\tau=-[n^{1/5}]}^{[n^{1/5}]} w\left(\frac{\tau}{[n^{1/5}]}\right) \bar{W}_\tau \tag{8}$$

where

$$w(u) = \begin{cases} 1 - 6|u|^2 + 6|u|^3 & \text{if } 0 < u < \frac{1}{2} \\ 2(1 - |u|)^3 & \text{if } \frac{1}{2} \leq u < 1 \end{cases}$$

$$\bar{W}_\tau = \begin{cases} \frac{1}{n} \sum_{t=2+\tau}^{n} (m_t - \bar{m})(m_{t-\tau} - \bar{m})' & \tau \geq 0 \\ \tilde{W}'_{n,-\tau} & \tau < 0 \end{cases} \tag{9}$$

See, e.g., Gallant (1987, p. 446, 533). Define

$$Z(x, \theta_{(2)}) = \sqrt{n} \left[W(x, \theta_{(2)})\right]^{-\frac{1}{2}} \left[\bar{m}(x, \theta_{(2)})\right]. \tag{10}$$

where $\left[W(x, \theta_{(2)})\right]^{-\frac{1}{2}}$ denotes the inverse of the Cholesky factorization of $W(x, \theta_{(2)})$. □

There are two methods that can use $m(x_t, x_{t-1}, \theta_{(2)})$ to estimate $\theta_{(2)} = (\gamma, \psi, \delta)$ within the Bayesian paradigm:

The first method, that we shall term "moment constrained Bayes," uses a likelihood $f(x \,|\, \theta_{(1)})$ to represent the density of the data, which is often a sieve. The parameter vector $\theta_{(1)}$ is determined by whatever likelihood one chooses for x; its definition will vary throughout as determined by context. Taking the expectation of the moment conditions with respect to the likelihood generates parametric restrictions

$$0 = \rho\big(\theta_{(1)}, \theta_{(2)}\big) = \int m(x_t, x_{t-1}, \theta_{(2)})\, f(x|\theta_{(1)})\, dx \tag{11}$$

Note, in passing, that for any non-random positive definite matrix $W\left(\theta_{(1)}, \theta_{(2)}\right)$, an equivalent expression for the constraint is

$$0 = \rho'\big(\theta_{(1)}, \theta_{(2)}\big)\, W^{-1}\big(\theta_{(1)}, \theta_{(2)}\big)\, \rho\big(\theta_{(1)}, \theta_{(2)}\big)\,. \tag{12}$$

The main problem with moment constrained Bayes is that under the constraint (11), the parameter space

$$\Theta = \left\{\theta{\in}\mathbb{R}^{\dim(\theta)} \,|\, \theta = \big(\theta_{(1)}, \theta_{(2)}\big)\,, 0 = \rho\big(\theta_{(1)}, \theta_{(2)}\big)\right\} \tag{13}$$

has measure zero. This makes estimation of the posterior distribution of θ subject to a prior $p(\theta)$ and constraint (11) by Markov Chain Monte Carlo (MCMC) problematic; see, e.g., Gamerman and Lopes (2006) for the practicalities of MCMC. Some proposals for dealing with the computational problem posed by moment constrained Bayes are the following.

Under the assumption that data are independent draws from a density with finite support, Bornn, Shephard, and Solgi (2016), develop MCMC methods to draw from the posterior density $f(\theta\,|\,x)$ determined by likelihood $f\big(x\,|\,\theta_{(1)}\big)$ and prior $p(\theta)$ on Θ. The main issue they address is the determination of the correct Jacobian term to account for the singularity of Θ. Shin's (2015) proposals can accommodate Markovian data. The likelihood is presumed to be a mixture of specific parametric distributions with random weights drawn from a discrete distribution. The constraint (11) becomes a constraint on the discrete distribution of the random weights. Given a prior, several MCMC samplers are developed to draw from the posterior determined by this likelihood.

Another proposal for computing moment constrained Bayes, that we term the "λ-prior method," is to use a general purpose sieve for $f\big(x\,|\,\theta_{(1)}\big)$ subject to a prior of the form

$$p_\lambda(\theta) = p(\theta) \times \exp\left[-\lambda\, n\, \rho'\big(\theta_{(1)}, \theta_{(2)}\big)\, W^{-1}\big(\theta_{(1)}, \theta_{(2)}\big)\, \rho\big(\theta_{(1)}, \theta_{(2)}\big)\right]. \tag{14}$$

Here we use the seminonparametric density (SNP) $f_{SNP}\big(x\,|\,\theta_{(1)}\big)$ proposed by Gallant and Tauchen (1989) for use in applications similar to Example 1. Its main advantage in the present context is that well tested code for estimation and simulation is available. The parameter space for likelihood $f_{SNP}\big(x\,|\,\theta_{(1)}\big)$ and prior $p_\lambda(\theta)$ is not singular so MCMC can proceed in the usual fashion. For large λ, θ draws become concentrated near the parameter space Θ thereby providing approximate draws from the posterior with likelihood $f_{SNP}\big(x\,|\,\theta_{(1)}\big)$ and prior $p(\theta)$ subject to constraint (11). For λ sufficiently large, MCMC must fail because Θ is singular. The idea is, by trial and error, to find the largest λ such that MCMC draws mix and use those draws as the approximation to the posterior on Θ. Given a proposed θ, one can easily generate a long simulation $\hat{x}$ from $f_{SNP}\big(x \,|\, \theta_{(1)}\big)$. The values for $\rho\big(\theta_{(1)}, \theta_{(2)}\big)$ and $W\big(\theta_{(1)}, \theta_{(2)}\big)$ that appear in (14) are computed by evaluating (6) and (9), respectively, with $x = \hat{x}$.

The second method, that we shall term "moment induced Bayes," assumes that $z = Z\big(x, \theta_{(2)}\big)$ given by (10) follows a distribution $\Psi(z)$ with density $\psi(z)$. From this distribution, one can infer a probability space upon which Bayesian inference can be conducted (Gallant, 2016a). The main regularity condition is the semi-pivotal condition that the set $\big\{x : Z(x, \theta_{(2)}) = z\big\}$ not be empty for any choice of $(z, \theta_{(2)})$ in the permissible parameter space $\Theta_{(2)}$ and range $\mathcal{Z}$ of Z. Often Ψ is taken to be the normal distribution Φ with density ϕ in which case one uses

$$p(x \,|\, \theta_{(2)}) = \phi(z) = (2\pi)^{-\frac{M}{2}} \exp\left\{-\frac{n}{2}\bar{m}'(x, \theta_{(2)})[W(x, \theta_{(2)})]^{-1}\bar{m}(x, \theta_{(2)})\right\}, \qquad (15)$$

as the likelihood and proceeds directly to Bayesian inference using a prior $p(\theta_{(2)})$. This is usually justified by claiming that Z is asymptotically normal. Actually, it is the density of Z not the distribution that needs to converge. Necessary and sufficient conditions that the density of Z converges to $\phi(z)$ uniformly in n and $\theta_{(2)}$, are given by Sweeting (1986).

Regardless of regularity conditions, ψ can be quite different from ϕ in finite samples. The purpose of this paper is to determine ψ for a practical application, namely Example 1, and investigate the consequences of using ϕ instead of ψ. We find that ϕ differs markedly from ψ, being concentrated at zero and having extremely long exponential tails. This departure turns out to have important consequences. Details follow.

2 Data

Collection of the data on s_t, b_t, and c_t is straightforward and discussed in the first subsection below. Construction of w_t is somewhat elaborate and discussed in the second subsection. Summary statistics are presented in the third. All data are annual for the years 1930 through 2015.

2.1 Stocks, Bonds, and Consumption

The raw data for stock returns are value weighted returns including dividends for NYSE, AMEX, and NASDAQ from the Center for Research in Security Prices data at the Wharton Research Data Services web site (http://wrds.wharton.upenn.edu).

The raw data for returns on U.S. Treasury 30 day debt are from the Center for Research in Security Prices data at the Wharton Research Data Services web site.

The raw consumption data are personal consumption expenditures on nondurables and services obtained from Table 2.3.5 at the Bureau of Economic Analysis web site (http://www.bea.gov).

Raw data are converted from nominal to real using the annual consumer price index obtained from Table 2.3.4 at the Bureau of Economic Analysis web site. Conversion of consumption to per capita is by means of the mid-year population data from Table 7.1 at the Bureau of Economic Analysis web site.

2.2 The Return to Wealth

In addition to the data described above, the construction of the return to wealth, $R_{c,t}$ requires additional data:

Raw labor income data is "compensation of employees received" from Table 2.2 at the Bureau of Economic Analysis web site.

Raw annual returns including dividends on the twenty-five Fama-French (1993) portfolios were obtained from Kenneth French's web site, http://mba.tuck.dartmouth.edu/pages /faculty/ken.french. The portfolios are the intersections of five portfolios formed on market equity and five portfolios formed on the ratio of book equity to market equity. The portfolios are for all NYSE, AMEX, and NASDAQ stocks for which equity data are not missing and book equity data are positive. The portfolios are constructed at the end of each June with breakpoints determined by the NYSE quintiles at the end of June. Complete details are at Kenneth French's web site. The advantage of the Fama-French portfolios here is that they appear to isolate and exhaust the risk factors for holding equities (Fama and French, 1992, 1993).

Raw returns are converted from nominal to real using the annual consumer price index obtained from Table 2.3.4 at the Bureau of Economic Analysis web site. Conversion of labor income to per capita is by means of the mid-year population data from Table 7.1 at the Bureau of Economic Analysis web site.

In this subsection only, the data are real gross returns R_{st} on the Fama-French portfolios, a vector of length 25, the real gross return on the thirty day T-bill R_{bt}, real per capita consumption growth $\frac{C_t}{C_{t-1}}$, and real per capita labor income growth $\frac{L_t}{L_{t-1}}$; x_t is a vector of length 28 containing these variables and x has dimension 28 by $n = 86$. Also, in this subsection only, the definitions of θ, $m(x_t, x_{t-1}, \theta)$, $\bar{m}(x, \theta)$, $W(x, \theta)$, and $Z(x, \theta)$ will differ from those of Example 1 although they are similar entities for a similar problem.

The first step is to determine the distribution of the process $\frac{C_t}{C_{t-1}}\text{MRS}_{t-1,t}$. This is done by nonparametrically determining the ex-post realized values of $\text{MRS}_{t-1,t}$ over the period spanned by the data, multiplying by the observed $\frac{C_t}{C_{t-1}}$, and fitting an SNP conditional density function $f_{SNP}(\frac{C_t}{C_{t-1}}\text{MRS}_{t-1,t} \,|\, \frac{C_{t-1}}{C_{t-2}}\text{MRS}_{t-2,t-1}, \ldots)$ to $\frac{C_t}{C_{t-1}}\text{MRS}_{t-1,t}$. Then, we use

$$P_{ct} = \mathcal{E}_t \sum_{j=1}^{\infty} C_{t+j} M_{t,t+j} = C_0 \left(\prod_{k=1}^{t} \frac{C_k}{C_{k-1}} \right) \sum_{j=1}^{\infty} \mathcal{E}_t \prod_{k=1}^{j} \left(\frac{C_{t+k}}{C_{t+k-1}} M_{t-1+k,t+k} \right) \tag{16}$$

to express the gross return to consumption as

$$R_{ct} = \frac{P_{ct} + C_t}{P_{c,t-1}} = \frac{\frac{C_{t-1}}{C_{t-2}} \sum_{j=1}^{\infty} \mathcal{E}_t \prod_{k=1}^{j} \left(\frac{C_{t+k}}{C_{t+k-1}} M_{t+k-1,t+k} \right) + \frac{C_t}{C_{t-1}} \frac{C_{t-1}}{C_{t-2}}}{\sum_{j=1}^{\infty} \mathcal{E}_{t-1} \prod_{k=1}^{j} \left(\frac{C_{t+k-1}}{C_{t+k-2}} M_{t+k-2,t+k-1} \right)} \tag{17}$$

and use simulation from $f_{SNP}(\frac{C_t}{C_{t-1}}\text{MRS}_{t-1,t} \,|\, \frac{C_{t-1}}{C_{t-2}}\text{MRS}_{t-2,t-1}, \ldots)$ to compute the expectations in (17). It remains to discuss the computation of the ex-post realized values of $\text{MRS}_{t-1,t}$.

To compute the ex-post realized values of $\text{MRS}_{t-1,t}$ nonparametrically we follow Gallant and Hong (2007). The differences are that the data set is longer due to the passage of time and that all of the Fama-French portfolios can be used because a missing data problem has been resolved. We shall be brief; see Gallant and Hong (2007) for a more extensive exposition.

Define the instruments

$$V_t = \begin{pmatrix} R_{st} - 1 \\ R_{bt} - 1 \\ C_t/C_{t-1} - 1 \\ L_t/L_{t-1} - 1 \\ 1 \end{pmatrix},$$

where $R_{st} - 1$ denotes 1 subtracted from each element of R_{st}. Let the parameter $\theta = (\theta_1, \ldots, \theta_{86})$ denote the ex-post values of the pricing kernel. If Example 1 is misspecified, θ_t could differ from the marginal rate of substitution $\text{MRS}_{t-1,t}$ determined by (3). Denote the vector of Euler equation errors by

$$e_{t,t-1}(\theta_t) = 1 \, - \, \theta_t \begin{pmatrix} R_{s,t} \\ R_{b,t} \end{pmatrix}, \tag{18}$$

where 1 denotes a vector of 1's of length twenty-six. The moment function that determines

$$\bar{m}(x, \theta) = \frac{1}{n} \sum_{t=2}^{n} m(x_t, x_{t-1}, \theta_t) \tag{19}$$

for our estimator is

$$m(x_t, x_{t-1}, \theta_t) = V_{t-1} \otimes e_{t,t-1}(\theta), \tag{20}$$

where $\otimes$ denotes Kronecker product. The length of the vector $m(x_t, x_{t-1}, \theta_t)$ is $K = 754$ so that the number of overidentifying restrictions on $\theta_2, \ldots, \theta_{86}$ is 669. Note that θ_1 is not yet identified because θ_1 does not appear in (19); it is identified by the prior as discussed later in this subsection.

We assume that $(\theta_t R_{st}, \theta_t R_{bt})$ has a factor structure. There is one error common to all elements of $\theta_t s_t$, and twenty-six idiosyncratic errors, one for each element of $(\theta_t R_{st}, \theta_t R_{bt})$. Denote this matrix by Σ_e (or by $\Sigma_{e,t}$ if one wants to allow for heterogeneity, which makes no difference in what follows). A set of orthogonal eigen vectors U_e for Σ_e are easy to construct and can be used to diagonalize Σ_e. Similarly U_v and Σ_v for V_t.

Let $H_t(\theta) = (U_z \otimes U_e)' m(x_t, x_{t-1}\theta)$ with elements $h_{i,t}(\theta)$. Diagonalization implies that we can estimate the variance of $H_t(\theta)$ by a diagonal matrix $S_n(\theta)$ with elements

$$s_i(\theta) = \frac{1}{n} \sum_{t=2}^{n} \left(h_{t,i}(\theta) - \frac{1}{n} \sum_{t=2}^{n} h_{t,i}(\theta) \right)^2.$$

Following Gallant (2016a), a likelihood for x is, then,

$$p(x \mid \theta) \propto \exp\left\{-\frac{n}{2}\bar{m}(x,\theta)'(U_z \otimes U_e) S_n^{-1}(\theta)(U_z \otimes U_e)'\bar{m}(x,\theta)\right\}; \tag{21}$$

Next we describe the prior.

The prior for θ has the form $f_{SNP}(\theta \mid \eta)\,\phi(\eta \mid \mu, \sigma)$, where μ is the is the location of the SNP parameter η and σ is its scale. The location parameter μ is determined by fitting $f_{SNP}(\theta \mid \mu)$ to a long simulation (1665 years) of the $\{MRS_{t-1,t}\}$ in a Bansal and Yaron (2004) long-run risks economy. The values of σ are set to the values shown as the loose prior in Table 5 of Gallant and Hong (2007). There are a few additional tweaks to the prior that insure the stationarity of the process $\{\theta_t\}$ that we have glossed over, which see Gallant and Hong (2007). The prior nudges the seminonparametic density $f_{SNP}(\text{MRS}_{t-1,t} \mid \text{MRS}_{t-2,t-1}, \ldots)$ toward what would be seen in a stationary Bansal-Yaron (2004) economy. With likelihood and prior specified, MCMC draws from the posterior for θ are generated in the usual fashion. The mean of these MCMC draws is plotted in Figure 1. That is the estimated ex post realization of $\text{MRS}_{t-1,t}$. These are then multipled by $\frac{C_t}{C_{t-1}}$ and used as data to estimate the SNP density $f_{SNP}(\frac{C_t}{C_{t-1}}\text{MRS}_{t-1,t} \mid \frac{C_{t-1}}{C_{t-2}}\text{MRS}_{t-2,t-1}, \ldots)$. Simulations from this density are used to compute the expectations that appear in (17).

Figure 1 about here

2.3 Simple Statistics

Simple statistics for these data are shown in Table 1.

Table 1 about here

3 The Distribution of Z

In this section, for Example 1, we determine a distribution Ψ for Z given by (10) using moment constrained Bayes computed with the λ-prior method. For this, we need a likelihood $f\big(x \mid \theta_{(1)}\big)$ for $x = (x_1, \ldots, x_n)$, $x_t = (s_t, b_t, c_t, w_t)'$, $n = 98$. As we shall compute the integral that appears in the expression for $\rho\big(\theta_{(1)}, \theta_{(2)}\big)$ given by (11) by averaging over a simulation from $f\big(x \mid \theta_{(1)}\big)$, the likelihood density must be easy to simulate. We compute $W\big(\theta_{(1)}, \theta_{(2)}\big)$

that appears in (12) using (8) with (9) computed from a simulation. ($[n^{\frac{1}{5}}]$ that appears in (8) is evaluated at the sample size $n = 86$, not the simulation size $N = 1000$; for Example 1, $[n^{\frac{1}{5}}] = 1$.) The SNP density can accurately approximate a stationary, Markovian process and is easy to draw from (Gallant and Tauchen, 1992); SNP is our choice.

The SNP density $f_{SNP}\big(x \,|\, \theta_{(1)}\big)$ is determined using BIC following the protocol stated in (Gallant and Tauchen, 1990) using code at `www.aronaldg.org/webfiles/snp`. That determined has transition density

$$f_{SNP}\big(x_t \,|\, x_{t-1}, \theta_{(1)}\big) = n(x_t \,|\mu_{t-1}, \Sigma_{t-1}) \tag{22}$$

where

$$\begin{aligned} \mu_{t-1} &= b_0 + Bx_{t-1} \\ \Sigma_{t-1} &= \Sigma_0 + q^2\Sigma_{t-2} + [\text{diag}(p_1, p_2, p_3, p_4)](x_{t-2} - \mu_{t-2})(x_{t-2} - \mu_{t-2})'[\text{diag}(p_1, p_2, p_3, p_4)] \end{aligned}$$

The prior is

$$p(\theta) = p\big(\theta_{(1)}, \theta_{(2)}\big) = p\big(\theta_{(1)}\big)\, p\big(\theta_{(2)}\big)\,. \tag{23}$$

The prior on $\theta_{(1)}$ is

$$p\big(\theta_{(1)}\big) = p_1\big(\theta_{(1)}\big)\, p_2\big(\theta_{(1)}\big)\,, \tag{24}$$

where $p_1\big(\theta_{(1)}\big)$ is determined from a simulation of a Bansal and Yaron (2004) economy at the annual frequency of 1665 years and $p_2\big(\theta_{(1)}\big)$ is an indicator for a support condition on $\theta_{(1)}$ that insures that simulations from (22) are stationary and mean reverting by requiring that the largest eigen values of the companion matrices of the mean μ_{t-1} and variance function Σ_{t-1} are less than one. As to $p_1\big(\theta_{(1)}\big)$, the SNP density (22) was fitted to the simulation by maximum likelihood to determine an estimate α_i and variance β_i^2 for the parameters $\theta_{(1),i}$ for the SNP density; then $p_1\big(\theta_{(1)}\big) = \prod n(\theta_{(1),i} \,|\, \alpha_i, 100 \times \beta_i^2)$.

The substantive prior on $\theta_{(2)} = (\gamma, \psi, \delta)$ is

$$p\big(\theta_{(2)}\big) = p(\gamma)p(\psi)p(\delta) \tag{25}$$

where

$$
\begin{aligned}
p(\gamma) &= I_{\{\gamma>0|\}}(\gamma)\, n[\gamma \,|\, 10.0, (100)^2] \\
p(\psi) &= I_{\{1.2<\psi<1.8\}}(\psi) \left[1+\cos\left(\pi+2\pi\frac{\psi-1.1}{2.0-1.1}\right)\right] \\
p(\delta) &= I_{\{0.981354269<\delta<0.999000999\}}(\delta) \left[1+\cos\left(\pi+2\pi\frac{\delta-0.0.981354269}{0.999000999-0.981354269}\right)\right]
\end{aligned}
$$

where 0.981354269 corresponds to 1.9 per cent per annum and 0.999000999 to 0.1 per cent. The substantive prior tightly constrains ψ and δ to the standard values of a Bansal and Yaron (2004) economy, leaving the risk aversion parameter relatively unconstrained.

If we accept (22) as an adequate representation of the likelihood, then $z = Z(x, \theta_{(2)})$ is determined by

$$
(x, \theta_{(1)}, \theta_{(2)}) \sim f_{SNP}\big(x \,|\, \theta_{(1)}\big)\; p(\theta)\, p_\lambda(\theta) \tag{26}
$$

given by (22), (23), and (14), respectively (Gallant, 2016a). The sampling procedure is to draw $\theta = \big(\theta_{(1)}, \theta_{(2)}\big)$ from $p(\theta)\, p_\lambda(\theta)$, draw x from $f_{SNP}\big(x \,|\, \theta_{(1)}\big)$, and set $z = Z(x, \theta_{(2)})$. We consider λ that are powers of 10. The largest order λ for which draws of θ from $p(\theta)\, p_\lambda(\theta)$ will mix is $\lambda = 10^2$. After transients died off, a θ-chain of length $R = 200000$ was retained. For every tenth θ in the chain a sample x from $f_{SNP}\big(x \,|\, \theta_{(1)}\big)$ of size $n = 86$ was generated and $z = Z\big(x, \theta_{(2)}\big)$ evaluated to form a z-chain of length 20000.

Marginal histograms from the z-chain are shown in Figure 2. What is immediately apparent from Figure 2 is that a using a density $\psi(z)$ determined from the z-chain instead of $\phi(z)$ in (15) to implement moment induced Bayes by MCMC will have the effect of reducing the number of $\theta_{(2)}$ draws from the $\theta_{(2)}$-chain that have large values of z.

(Figure 2 about here)

(Figure 3 about here)

The moment constrained estimates of the substantive parameters $\theta_{(2)} = (\gamma, \psi, \delta)$ obtained via the λ-prior method are of some interest. These are presented in Table 2. The chain for the prior for $\lambda = 10^3$ will not mix, so these values are missing.

(Table 2 about here)

We now proceed to the determination of ψ from the z-chain. Our intention is not to match the distribution exactly but rather come reasonably close with a distribution that has elliptical contours with mean and mode of zero. These seem to be reasonable a priori considerations. A mean of zero is dictated by the scientific theory. A mode of zero and elliptical contours are motivated by a desire to treat euler equation errors symmetrically.

Figure 4 plots the left-most quantiles of the z-chain against the corresponding quantiles of a 2 d.f. t-distribution; the details of the construction are in the figure legend. Figure 5 plots the left-most quantiles of the z-chain against the corresponding quantiles of a 30 d.f. t-distribution. Comparing the two figures, it would seem that a conclusion that the left tails of z are exponential is warranted. One reaches the same conclusion regarding the right tails from similar plots (not shown).

Figure 2 suggests a classic "witches hat" distribution that can represented as a mixture of two multivariate normal distributions with zero mean and variance-covariance matrices that are proportional, viz.,

$$\psi(z) = p\, n_M(z \,|\, 0, \alpha\Sigma) + (1-p)\, n_M(z \,|\, 0, \Sigma), \tag{27}$$

where $n_M(z \,|\, 0, \Sigma) = (2\pi)^{-\frac{M}{2}} [\det(\Sigma)]^{-\frac{1}{2}} \exp(-\frac{1}{2} z' \Sigma^{-1} z)$. The result of a fit of this density to the z-chain yields visual results that are qualitatively similar to Figures 2 and 4 whereas the equivalent of Figure 3 differs dramatically, having the classic normal symmetric shape about zero rather then the decentered, asymmetric shapes seen it Figure 3.

(Figure 4 about here)

(Figure 5 about here)

4 Moment Induced Bayesian Estimates

Conceptually a likelihood $f\big(x \,|\, \theta_{(1)}\big)$, a prior $p(\theta) = p\big(\theta_{(1)}\big)\; p\big(\theta_{(2)}\big)$, and a restriction $0 = \int m(x_t, x_{t-1}, \theta_{(2)})\, f(x|\theta_{(1)})\, dx$ determine a joint probability space $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$. The prior is the marginal distribution of θ on this probability space. Marginalizing gives a space $(\mathcal{X} \times \Theta_{(2)}, \mathcal{C}^o_{(2)}, P^o_{(2)})$ with prior $p\big(\theta_{(2)}\big)$ as the marginal distribution of $\theta_{(2)}$. Both the joint and marginal probability spaces determine the same same density $\psi(z)$ for the random variable $Z(x, \theta_{(2)})$ given by (10).

Conversely, an assumption that $Z(x, \theta_{(2)})$ has density $\psi(z)$ together with an assumption that $\theta_{(2)}$ has density $p(\theta_{(2)})$ induces a probability space $(\mathcal{X} \times \Theta_{(2)}, \mathcal{C}^*_{(2)}, P^*_{(2)})$. The details of its construction are in Gallant (2016a). The two probability spaces are equivalent in the sense that $P^o_{(2)}(C) = P^*_{(2)}(C)$ for every $C \in \mathcal{C}^*_{(2)}$. Therefore the probability space $(\mathcal{X} \times \Theta_{(2)}, \mathcal{C}^*_{(2)}, P^*_{(2)})$ can be used as a substitute for $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ for the purpose of Bayesian inference regarding $\theta_{(2)}$. The semi-pivotal condition that the set $\{x : Z(x, \theta_{(2)}) = z\}$ not be empty for any choice of $(z, \theta_{(2)})$ in the parameter space $\Theta_{(2)}$ and range $\mathcal{Z}$ of Z implies that the likelihood $f(x \,|\, \theta_{(2)})$ on $(\mathcal{X} \times \Theta_{(2)}, \mathcal{C}^*, P^*)$ is $f(x \,|\, \theta_{(2)}) = \psi[Z(x, \theta_{(2)})]$. Therefore, one can use the likelihood $f(x \,|\, \theta_{(2)})$ and prior $p(\theta_{(2)})$ to generate a MCMC chain for $\theta_{(2)}$. See the discussion of this claim in Gallant (2016b, 2016c). Briefly, there is a missing adjustment term that has the effect of making the effective prior different from $p(\theta_{(2)})$. The adjustment depends only on $Z(x, \theta_{(2)})$ and has no effect on the determination of $\psi(z)$, which is the focus of this paper. The adjustment term is usually disregarded in applications as we do here. Indeed, Gallant (2016b) argues that it should be disregarded.

To summarize, the moment induced Bayes method assumes that $z = Z(x, \theta_{(2)})$ given by (10) follows a distribution $\Psi(z)$ with density $\psi(z)$. One uses

$$p(x \,|\, \theta_{(2)}) = \psi(z) \tag{28}$$

as the likelihood and proceeds directly to Bayesian inference using a prior $p(\theta_{(2)})$.

For Example 1 and prior $p(\theta_{(2)}) = p(\gamma)p(\psi)p(\delta)$ given by (25) we constructed $\psi(z)$ as described in Section 3 and used it to compute the posterior distribution of $\theta_{(2)}$ under the likelihood (28); the exercise was repeated with ϕ replacing ψ. The results are shown in Table 3.

(Table 3 about here)

The results shown in Table 3 are unsatisfactory because the prior is so influential that it makes little difference whether one uses Ψ or Φ to implement moment induced Bayes. We shall need to consider another example where the data are more influential.

In an earlier version of this paper, Ψ was determined to closely mimic Figures 2 and 3. This did have a dramatic influence on estimates that we attribute to the asymmetric, decentered Ψ that obtained. A decentered Ψ contradicts the assumptions of the moment induced Bayesian estimation strategy. Imposing symmetry is less defensible although intuition strongly suggests that symmetry is desirable.

Oddly enough, it is not volatility in $W\big(x, \theta_{(2)}\big)$ given by (8) that causes $z = Z(x, \theta_{(2)}$ to be volatile as one might expect but rather volatility in $\mathrm{MRS}_{t,t+1}$ given by (3) that causes $\bar{m}(x, \theta_{(2)})$ given by (6) to be volatile; the marginal rate of substitution under Epstein-Zin-Weil utility is an extremely unstable function.

5 Conclusion

We explored the practical aspects of two complementary Bayesian method of moments strategies using a macro-finance application. The first, moment constrained Bayes, uses a sieve to represent represent the density of the data. Taking the expectation of the moment conditions with respect to the sieve generates parametric restrictions on sieve parameters and introduces additional parameters from the moment conditions. An advantage of moment constrained Bayes is that it provides both an estimate of the density of the data as well as estimates of the parameters that appear in the moment conditions. The difficulty with moment constrained Bayes is computational: The parameter space is singular with respect to Lebesgue measure making Markov Chain Monte Carlo methods difficult to implement. To circumvent the computational difficulty, we used a penalty function approach, the λ-prior method, to generate draws from a close approximation to the posterior. The second Bayesian method of moments strategy, moment induced Bayes, uses a semi-pivotal Z constructed from the moment conditions and an assumed distribution Ψ for Z to infer a likelihood and thereby proceed to Bayesian inference. Moment induced Bayes provides estimates of the parameters that appear in the moment equations only. The difficulty with moment induced Bayes is that one must choose Ψ. To circumvent the difficulty, we use draws of Z from moment constrained Bayes to infer Ψ.

6 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.

Bornn, Luke, Neil Shephard, and Reza Solgi (2016), "Moment Conditions and Bayesian Nonparametrics," Working paper, Department of Economics, Harvard. http://scholar.harvard.edu/fil

Coppejans, Mark, and A. Ronald Gallant (2002), "Cross-Validated SNP Density Estimates," *Journal of Econometrics* 110, 27–65.

Epstein, L. G., and S. Zin. (1989). "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica 57,* 937–969.

Fama, E., and K. French. (1992). "The Cross-Section of Expected Stock Returns." *Journal of Finance 59,* 427–465.

Fama, E., and K. French. (1993). "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics 33,* 3–56.

Gallant, A. R. (1987), *Nonlinear Statistical Models,* New York: Wiley.

Gallant, A. Ronald (2016a), "Reflections on the Probability Space Induced by Moment Conditions with Implications for Bayesian Inference," *Journal of Financial Econometrics* 14, 227–228.

Gallant, A. Ronald (2016b), "Reply to Comment on Reflections," *Journal of Financial Econometrics* 14, 284–294.

Gallant, A. Ronald (2016c), "Addendum to Reply to Reflections," Working paper, Department of Economics, Penn State University, http://www.aronaldg.org/papers/addendum.pdf

Gallant, A. Ronald, and Han Hong (2007), "A Statistical Inquiry into the Plausibility of Recursive Utility," *Journal of Financial Econometrics* 5, 523–559.

Gallant, A. R., and G. Tauchen. (1989). "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications." *Econometrica 57,* 1091–1120.

Gallant, A. Ronald, and George Tauchen (1990), "SNP: A Program for Nonparametric Time Series Analysis, Version 9.1, User's Guide," www.aronaldg.org/webfiles/snp.

Gallant, A. Ronald, and George Tauchen (1992), "A Nonparametric Approach to Nonlinear Time Series Analysis: Estimation and Simulation," in Brillinger, David, Peter Caines, John Geweke, Emanuel Parzen, Murray Rosenblatt, and Murad S. Taqqu eds. (1992), *New Directions in Time Series Analysis, Part II.* Springer–Verlag, New York, 71–92.

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

Kreps, D. M., and E. L. Porteus. (1978). "Temporal Resolution of Uncertainty and Dynamic Choice." *Econometrica 46,* 185–200.

Shin, Minchuyl (2015), "Bayesian GMM," Working paper, Department of Economics, University of Illinois. http://www.econ.uiuc.edu/ mincshin/BGMM_ver05

Sweeting, T. (1986), "On a Converse to Scheffe's Theorem," *The Annals of Statistics* 14, 1252–1256.

Weil, P. (1990). "Nonexpected Utility in Macroeconomics." *The Quarterly Journal of Economics 105,* 29–42.

Table 1. Simple Statistics for the Data

Series	Mean	Standard Deviation	Skewness	Excess Kurtosis
s_t	0.08609	0.19501	-1.16717	2.14135
b_t	0.03211	0.02880	0.85028	0.25196
c_t	0.01996	0.02198	-1.40754	5.06603
w_t	0.01617	0.29628	0.40303	1.17462

The data are annual for the years 1930 through 2015; the sample size is $n = 86$. s_t is log real gross stock return. b_t is log real gross bond return. c_t is log real per capita consumption growth. w_t is log real gross wealth return.

Table 2. Moment Constrained Estimates

	Estimates			Prior		
Parameter	Mean	Mode	Std. Dev.	Mean	Mode	Std. Dev.
			$\lambda = 10^1$			
γ	56.838	49.494	27.830	53.353	61.89	35.835
ψ	1.5019	1.4244	0.1121	1.5140	1.5247	0.1061
δ	0.9902	0.9867	0.0032	0.9902	0.9927	0.0033
			$\lambda = 10^2$			
γ	60.158	104.97	29.811	55.767	24.270	37.490
ψ	1.5104	1.6201	0.1055	1.5162	1.3945	0.1084
δ	0.9902	0.9908	0.0032	0.9902	0.9876	0.0032
			$\lambda = 10^3$			
γ	65.909	75.632	28.393	NA	NA	NA
ψ	1.5001	1.4800	0.1090	NA	NA	NA
δ	0.9902	0.9868	0.0032	NA	NA	NA

The estimation method is moment constrained Bayes using the lambda prior computational method with λ as shown as described in Section 1. Data are real, annual, per capital consumption for the years 1930–2015 and real, annual stock, bond, and wealth returns for the same years from BEA (2016) and CRSP (2016) that are used to form the moment functions (6) through (6). The prior is $p(\theta_{(1)})p(\gamma,\psi,\delta)p_\lambda(\theta)$ where $p(\theta_{(1)})$ is given by (24), $p_\lambda(\theta)$ is given by (14), and $p(\gamma,\psi,\delta)$ is an independence prior with γ normal with mean 10 and standard deviation 100 constrained to have positive support, ψ with a cosine density with support (1.2,1.8); and δ a cosine density with support 0.1% to 1.9% per annum; see (25). The columns labeled mean, mode, and standard deviation are the mean and standard deviations of an MCMC chain (Gamerman and Lopes (2006) of length 200,000 collected past the point where transients have dissipated. The proposal is move-one-at-a-time random walk. The prior chain for $\lambda = 10^3$ will not mix.

Table 3. Moment Induced Estimates

	Z Normal			Z Empirical			Z Prior		
Parameter	Mean	Mode	Std. Dev.	Mean	Mode	Std. Dev.	Mean	Mode	Std. Dev.
γ	97.439	72.765	50.216	117.98	14.777	69.183	85.166	9.0768	61.906
ψ	1.5006	1.4937	0.1057	1.4956	1.4867	0.1053	1.5011	1.4997	0.1088
δ	0.9902	0.9901	0.0032	0.9903	0.9898	0.0032	0.9902	0.9901	0.0032

The estimation method is moment induced Bayes as described in Section 4. Data are real, annual, per capital consumption for the years 1930–2016 and real, annual stock, bond, and wealth returns for the same years from BEA (2016) and CRSP (2016) that are used to form the moment functions (6) through (6). In the columns labeled Z Normal the distribution of Z given by (10) is presumed to be Normal. In the columns labeled Z Empirical the distribution of Z given by (10) is presumed to be the distribution (27) derived in Section 3. The prior is an independence prior with γ normal with mean 10 and standard deviation 100 constrained to have positive support, ψ with a cosine density with support (1.2,1.8); and δ a cosine density with support 0.1% to 1.9% per annum; see (25). The columns labeled mean, mode, and standard deviation are the mean and standard deviations of an MCMC chain (Gamerman and Lopes (2006) of length 200,000 collected past the point where transients have dissipated. The proposal is move-one-at-a-time random walk.

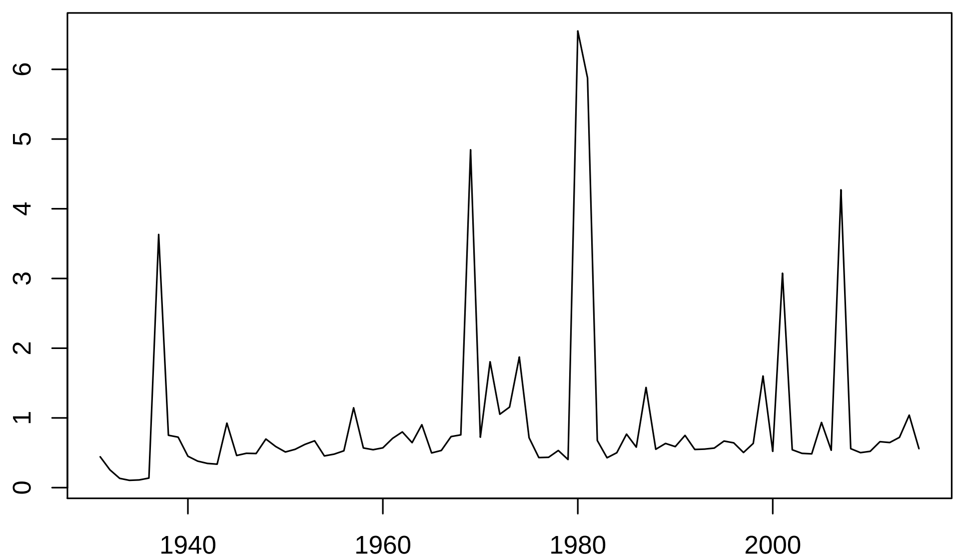

Figure 1. The Posterior Mean of the Marginal Rate of Substitution. Plotted is the posterior mean of $\theta_2 = \mathrm{MRS}_{1931,1932}$ through $\theta_{86} = \mathrm{MRS}_{2014,2015}$.

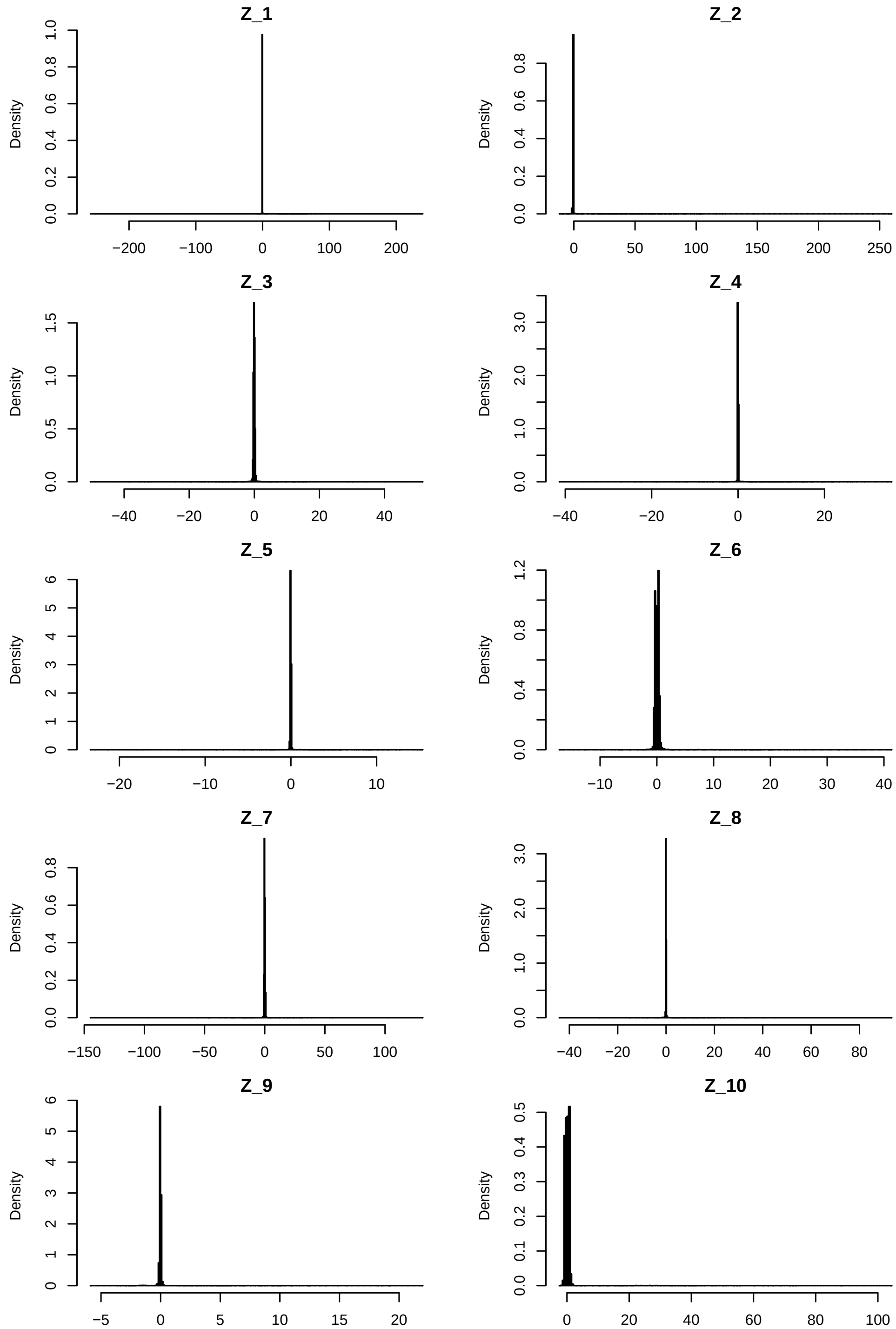

Figure 2. The Distribution of Z. For Example 1, draws from the prior distribution $p(\theta)p_\lambda(\theta)$ for $\theta = \big(\theta_{(1)}, \theta_{(2)}\big)$ of the moment constrained Bayes estimator were generated method with $\lambda = 10^2$. For each θ draw, a sample x of size $n = 86$ for (s_t, b_t, c_t, w_t) was generated from $f_{SNP}\big(x \,|\, \theta_{(1)}\big)$ and used to compute $z = Z\left(x, \theta_{(2)}\right)$. Shown in Figure 2 are the histograms of the coordinates $z_1, \ldots, z_{10}$ of z computed from every tenth draw of $N = 200,000$ total draws for θ.

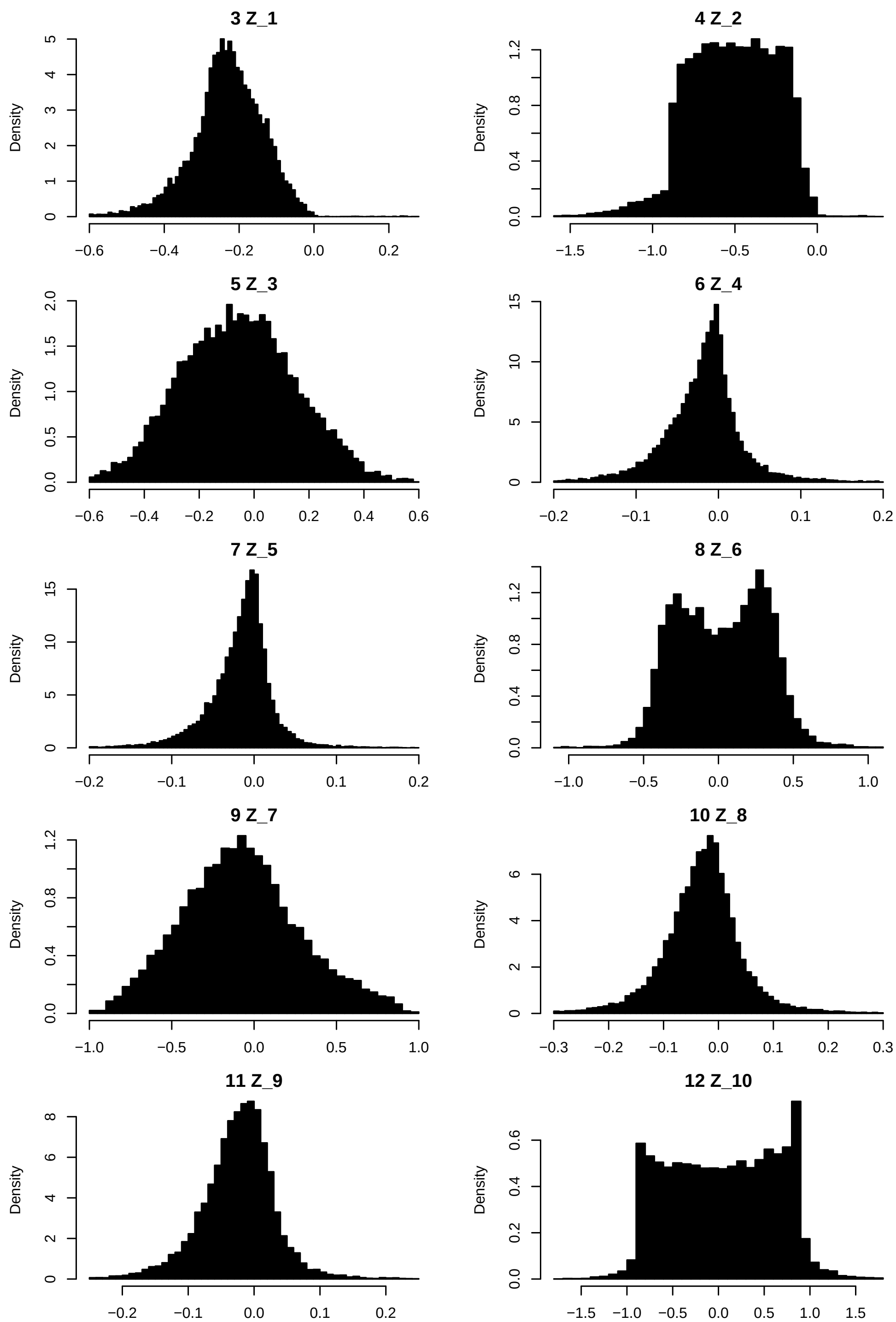

Figure 3. The Trimmed Distribution of Z. For Example 1, draws from the prior distribution $p(\theta)p_\lambda(\theta)$ for $\theta = \big(\theta_{(1)}, \theta_{(2)}\big)$ of the moment constrained Bayes estimator were generated method with $\lambda = 10^2$. For each θ draw, a sample x of size $n = 86$ for (s_t, b_t, c_t, w_t) was generated from $f_{SNP}\big(x \,|\, \theta_{(1)}\big)$ and used to compute $z = Z\left(x, \theta_{(2)}\right)$. Shown in Figure 2 are trimmed histograms of the coordinates $z_1, \ldots, z_{10}$ of z computed from every tenth draw of $N = 200,000$ total draws for θ.

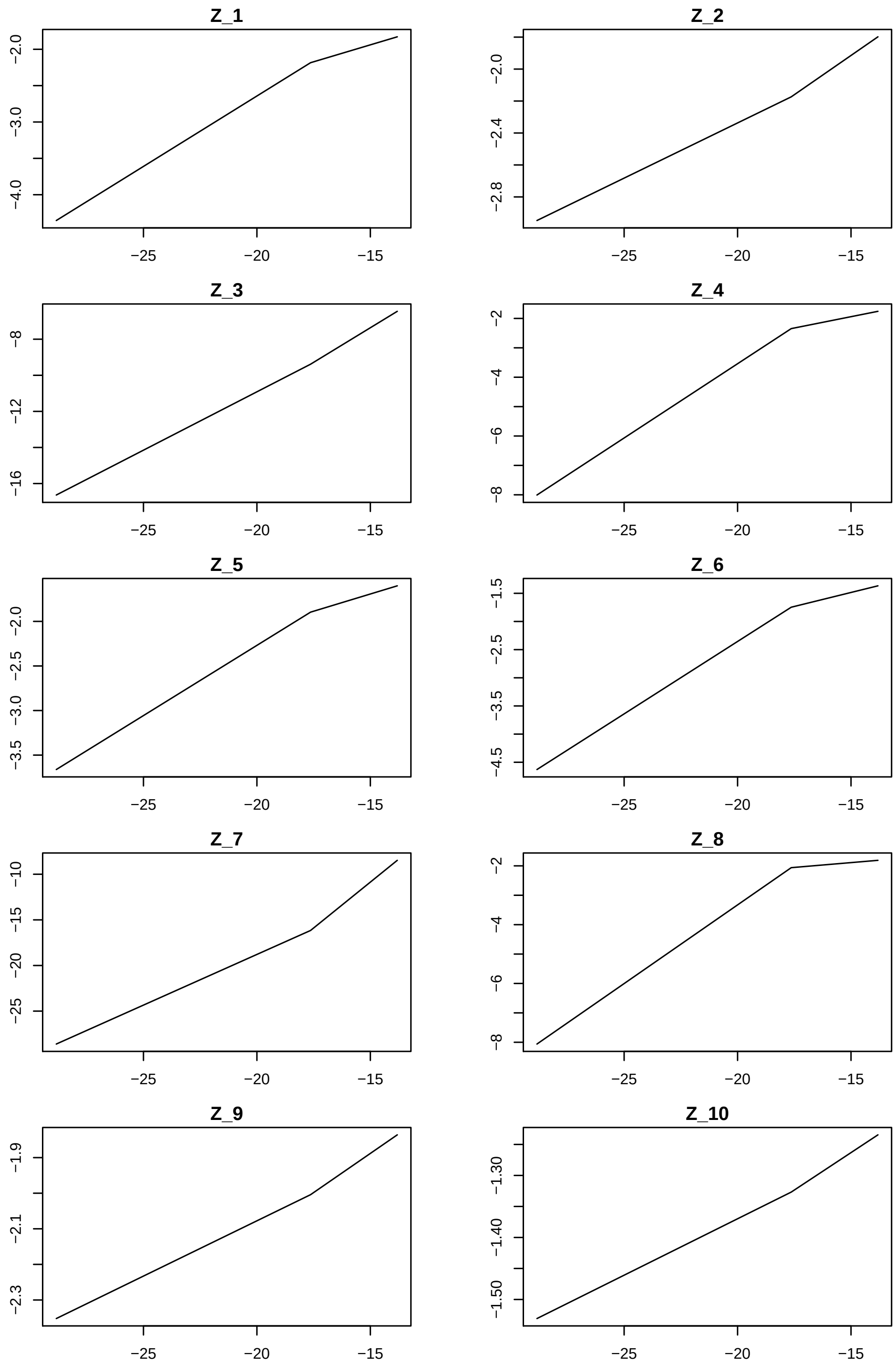

Figure 4. Left Q-Q Plots, Z vs. 2 d.f. t. For Example 1, $N = 20,000$ draws from $z = Z\left(x, \theta_{(2)}\right)$ were generated as described in the legend of Figure 2. For each z_i, quantiles at probabilities 2/2000 through 60/2000 with increment 1/2000 were computed from these draws. Quantiles for the t-distribution with 2 degrees freedom were computed for probabilities 2/2000 through 60/2000 with increment 1/2000. Shown in Figure 4 are plots of the z quantiles against the t quantiles with the t quantiles on the horizontal axis.

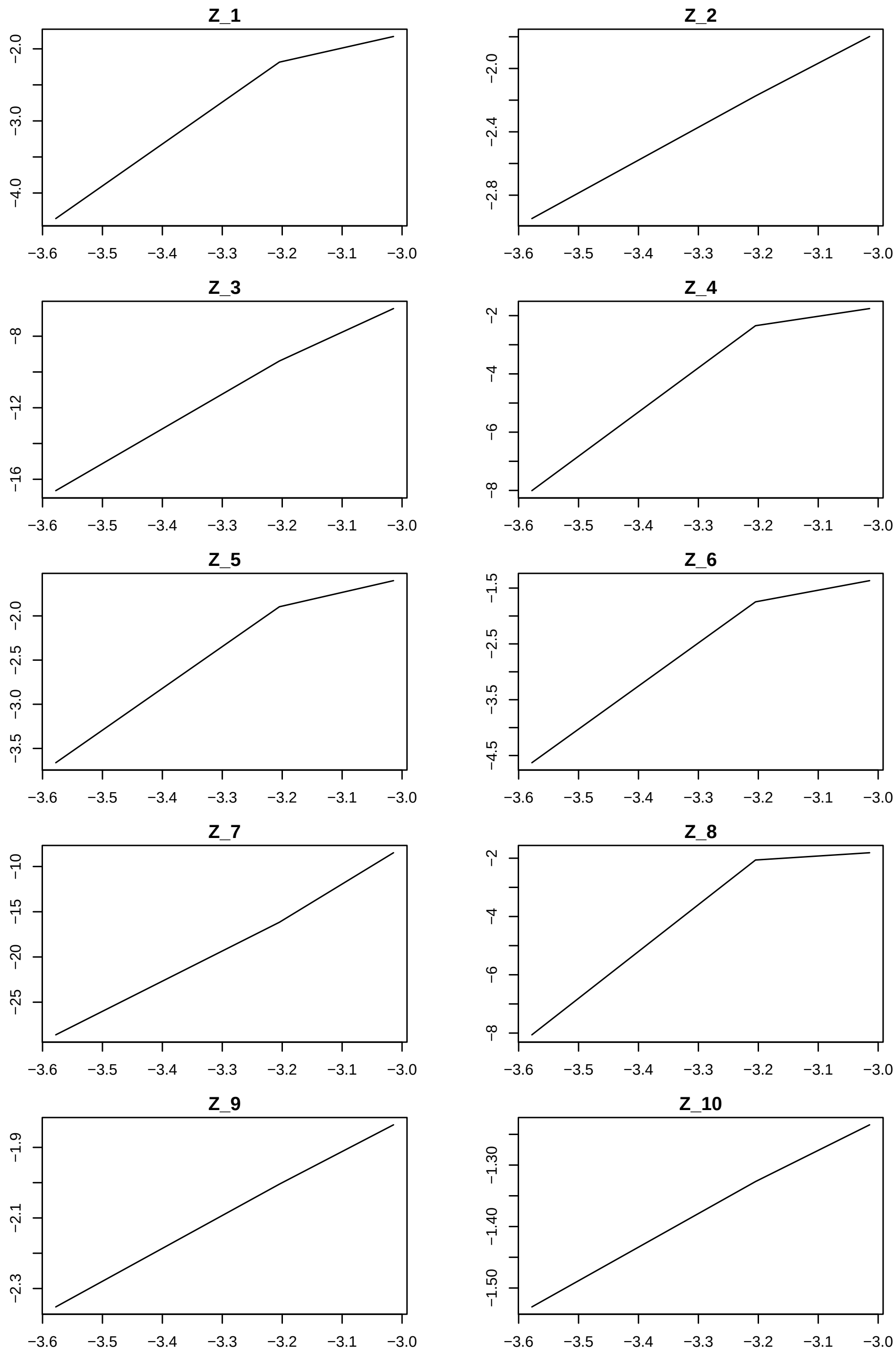

Figure 5. Left Q-Q Plots, Z vs. 30 d.f. t. For Example 1, $N = 20,000$ draws from $z = Z\left(x, \theta_{(2)}\right)$ were generated as described in the legend of Figure 2. For each z_i, quantiles at probabilities 2/2000 through 60/2000 with increment 1/2000 were computed from these draws. Quantiles for the t-distribution with 30 degrees freedom were computed for probabilities 2/2000 through 60/2000 with increment 1/2000. Shown in Figure 4 are plots of the z quantiles against the t quantiles with the t quantiles on the horizontal axis.