Complementary Bayesian Method of Moments Strategies*

A. Ronald Gallant
Penn State University

First draft: August 7, 2016
This draft: February 24, 2020

Forthcoming, *Journal of Applied Econometrics*

www.aronaldg.org/papers/cb.pdf

*I should like to thank the referees; their recommendations substantially improved the MS. Conflict of interest statement: None. Address correspondence to A. Ronald Gallant, Penn State University, Department of Economics, 511 Kern Graduate Building, University Park PA 16802, USA, phone 919-428-1130; email aronldg@gmail.com.

Abstract

Methodology is proposed that addresses two problems that arise in application of the generalized method of moments representation of the likelihood in Bayesian inference: (1) a missing Jacobian term and (2) a normality assumption. The proposals are illustrated by application to the seminal application of the generalized method of moments methodology in the econometric literature: an endowment economy whose representative agent has constant relative risk aversion utility.

Keywords and Phrases: Method of moments, Bayesian inference

JEL Classification: C11, C32, C36, C58

1 Introduction

We consider Bayesian estimation of the parameters that appear in a set of over-identified, nonlinear, moment equations.

This is a well studied problem in the frequentist literature. Solutions divide into two main groups. Group (1): Estimate a nonparametric likelihood subject to the moment conditions. Empirical likelihood is an illustration of the basic idea (Owen, 1988). A example in a dynamic setting where both the likelihood and moment conditions are nonparametric is Gallant and Tauchen (1989). Group (2): Apply method of moments directly. The best known example in economic research is generalized method of moments (GMM) (Hansen, 1982). GMM is simple to implement and is more widely used in economic research than Group (1) methods.

Consideration of this problem from the Bayesian perspective follows the same division into two these two groups.

A Bayesian Group (1) approach is a difficult to implement when the moment equations are over-identified because the support of the posterior has Lebesgue measure zero. See Born, Shephard, and Solgi (2018) for a discussion of the issues, a review of the literature, and proposed remedies for likelihoods with discrete support using notions from geometric measure theory. For just identified moment equations the problem simplifies (Chamberlain and Imbens, 2003). A judicious choice of likelihood in time series applications can permit a solution (Shin, 2015). Schennach (2005) and Gallant, Hong, Leung, and Li (2019) propose approximate solutions based on asymptotics. Gallant (2020a) resolves most of the above issues by adapting the Surface Sampling Algorithm of Zappa, Holmes-Cerfon, and Goodman (2018) to the problem thereby permitting exact Bayesian analysis for any smooth likelihood estimated subject to over-identified moment conditions.

A Bayesian Group (2) approach is inspired by the simplicity of the Markov chain Monte Carlo (MCMC) computational methods proposed by Chernozhukov and Hong (2003). If $\mathcal{Q}(x,\theta) = Z'(x,\theta)Z(x,\theta)$ represents the GMM criterion function, where Z are the moment conditions normalized to unit variance, $x = [x_1, \dots, x_n]$ is a matrix containing the data, and θ is the parameter vector, the idea is to treat

$$p^*(x \mid \theta) = \phi\left[Z(x,\theta)\right] \propto \exp\left[-\frac{1}{2}\mathcal{Q}(x,\theta)\right] \tag{1}$$

as a likelihood in MCMC computations, where ϕ is the standard multivariate normal density. We term[1] equation (1), together with variants described later, variously as a method of moments representation of the likelihood or as a GMM representation of the likelihood. Examples of applications using a method of moments representation of the likelihood are Romeo (2007), Gallant and Hong (2007), Duan and Mela (2009), Yin (2009), Gallant, Giacomini, and Ragusa (2017), and Gallant and Tauchen (2017, 2018).

Using the method of moments representation of the likelihood was formally justified by Gallant (2016a). In the comments and reply between Sims (2016) and Gallant (2016b, 2016c) two problems with this approach are revealed: (1) the GMM representation of the likelihood in equation (1) is missing a Jacobian term, (2) the standard approach of assuming normality of the random variable Z as in equation (1) may be suspect. This paper addresses these two problems.

The first problem is addressed by outlining a practical procedure for constructing the Jacobian term and applying it in the seminal GMM application (Hansen and Singleton, 1982). The method suggested by Gallant, Hong, Leung, and Li (2019) can be adapted to addressing the issue of non-normality, as will be explained in Section 6.

2 Bayesian Inference from Moment Functions

2.1 Intuitive Development

Even though the parameter vector θ may be regarded as fixed, Bayesian analysis proceeds as if the data $x \in \mathcal{X}$ and parameter $\theta \in \Theta$ were random variables on $\mathcal{X} \times \Theta$ with joint density $p^o(x, \theta)$ for which the marginal for θ is the prior $p^o(\theta)$ and the conditional for x given θ is the likelihood $p^o(x \mid \theta)$. Consequently, $p^o(x, \theta) = p^o(x \mid \theta)\, p^o(\theta)$ and we are working on a probability space of the form $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$, where $\mathcal{C}^o$ denotes the Borel subsets of $\mathcal{X} \times \Theta$. The goal of Bayesian inference is to obtain the posterior $p^o(\theta \mid x) = p^o(x, \theta) \,/\! \int p^o(x, \theta)\, d\theta$. To obtain it analytically when possible. When not possible, to find a numerical method for computing $\int g(\theta)\, p^o(\theta \mid x)\, d\theta$. E.g., by Markov Chain Monte Carlo (MCMC) (Gamerman and Lopes, 2006).

[1] The term "GMM representation of the likelihood" in the present context was coined by Gallant, Giacomini, and Ragusa (2017) but as it was applied to state space models, the likelihood was a "measurement density." Gallant and Tauchen (2018) called it "method of moments representation of the likelihood." Here may be the first use of the term "GMM representation of the likelihood." It has been used in other contexts to mean maximum likelihood estimation by method of moments using the scores of the likelihood as the moment equations.

In a GMM analysis the likelihood $p^o(x \mid \theta)$ is ignored because one does not know it, or because it is too hard to compute, or because of misspecification concerns. Nonetheless, existence is assumed here.

One works, instead, with moment conditions $Z(x, \theta)$, where $\int Z(x, \theta)\, p^o(x, \theta^o)\, dx = 0$ when $\theta = \theta^o$ else nonzero.[2] The moment equations are derived from an economic agent's first order conditions, an asset pricing identity, or similar considerations.

We assume that the moment equations satisfy a semi-pivotal condition

ASSUMPTION 1 The set

$$C^{(\theta,z)} = \{x \in \mathcal{X} \,:\, Z(x, \theta) = z\}, \tag{2}$$

is not empty for any $(\theta, z) \in \Theta \times \mathcal{Z}$.

Consider shifting the data $\{x_t\}_{t=1}^n$ by $\{x_t + \tau e_i\}_{t=1}^n$, where e_i has a one in the ith element and zeros elsewhere. If, for some e_i, the shifted data is in $\mathcal{X}$ and Z is continuous and unbounded in τ, then Assumption 1 is satisfied.

Suppose that we are given $Z(x, \theta)$ defined over $\mathcal{X} \times \Theta$ and told that the random variable $z = Z(x, \theta)$ has density $\psi(z)$, which may be standard multivariate normal or some other density such as multivariate Student's t. Assuming in addition to Assumption 1 that the dimension[3] of x and z are the same, and that the mapping $Z(x, \theta)$ from x to z is one-to-one when θ is held fixed, Gallant (2016a, 2016b) proves that that the conditional density of x given θ is

$$p^*(x \mid \theta) = \left|\det[(\partial/\partial x')Z(x, \theta)]\right| \psi[Z(x, \theta)]. \tag{3}$$

We now have a likelihood; Bayesian analysis can proceed forthwith with prior $p^*(\theta) = p^o(\theta)$. Bayesian Efficient Method of Moments (EMM) is an instance where the dimension of x and z is the same and this construction is applicable (Gallant and Tauchen, 2017).

Suppose now that the situation is as in the previous paragraph but, as is usually the case, the dimension of x is larger than z. Suppose that we can find mappings $u = U(x, \theta)$ and $x = X(u, \theta)$ such that u has the same dimension as z and such that $z = Z[X(u, \theta), \theta]$.

We are now essentially in the same situation as the previous paragraph with u replacing x. However, $Z[X(u, \theta), \theta]$ will likely be sufficiently intractable that numerical methods will be required to compute $(\partial/\partial u')Z[X(u, \theta), \theta]$.

[2] Before normalization to have unit variance if not after.

[3] Meaning the number of rows times the number of columns of x.

In the above $X(u,\theta)$ does not reconstruct the data but rather maps to an $x \in \mathcal{X}$ that evaluates to the same z as the data; i.e., maps to a single point in $C^{(\theta,z)}$ given by (2). This condition is essential. One cannot arbitrarily choose some elements from $\{x_t\}_{t=1}^n$ as u. One must be able to recompute z from knowledge of u and θ; u is similar to a sufficient statistic except that, unlike a sufficient statistic, u may depend on θ.

Often it is possible to put $u = U(x,\theta)$ into the form $u = U(z,\theta)$ with inverse $z = U^{-1}(u,\theta)$ whose Jacobian can be computed analytically. In which case $z = U^{-1}(u,\theta) = Z[X(u,\theta),\theta]$ and

$$p^*(x \,|\, \theta) = \left| \det[(\partial/\partial u')U^{-1}(u,\theta)] \right| \psi[Z(x,\theta)].$$

The adjustment term becomes

$$\mathrm{adj}(x,\theta) = \left| \det[(\partial/\partial u')U^{-1}(u,\theta)] \right| \tag{4}$$

Note that, given (x,θ), the u at which the right hand side of (4) is evaluated is computed by first computing $z = Z(x,\theta)$ then computing $u = U(z,\theta)$.[4] The prior remains $p^o(\theta)$. Examples are in Gallant (2016b).

While in most instances the above suffices, a slight generalization is helpful in instances such as the the case study of Section 3. Suppose that for some invertable transformation $v = V(z)$ we can find mappings $u = U(z,\theta)$ and $x = X(u,\theta)$ such that u has the same dimension as z and such that $V(z) = Z[X(u,\theta),\theta]$. In this case, supposing that an analytic expression for $U^{-1}(u,\theta)$ is available, $z = U^{-1}(u,\theta) = V^{-1}\{Z[X(u,\theta),\theta]\}$, whence the expression for $\mathrm{adj}(x,\theta)$ is again given by (4).

If $\mathrm{adj}(x,\theta)$ does not depend on θ and MCMC is used for the computations, then the adjustment $\mathrm{adj}(u,\theta)$ is not necessary because it cancels in the MCMC ratio. For the same reason, $\mathrm{adj}(u,\theta)$ only needs to be correct to within a scale factor that does not depend on θ.

2.2 Formal Development

Here we consider a formal derivation of the function $Z(x,\theta)$ of the previous subsection and the probabilistic implications. To do so, we presume the existence of $p^o(x \,|\, \theta)$ that implies moment equations from which $Z(x,\theta)$ is constructed. We do not presume knowledge, only existence.

[4] In theory, to draw x from $p^*(x \,|\, \theta)$ for given θ, draw $z \sim \psi$, compute $u = U(z,\theta)$, map u to $C^{(\theta,z)}$ using $X(u,\theta)$, then draw x from the uniform on $C^{(\theta,z)}$ if $C^{(\theta,z)}$ is bounded or from a density over $C^{(\theta,z)}$ with large variance that is nearly flat if not bounded.

As mentioned above, in Bayesian inference θ is formally manipulated as if it were random even though one might regard it as fixed. Thus, one has a joint distribution $P^o(x, \theta)$ defined over $\mathcal{X} \times \Theta$, whose density is the product of a likelihood $p^o(x \mid \theta)$ times a prior $p^o(\theta)$.[5] The joint probability space under consideration is, therefore, $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$, where $\mathcal{C}^o$ denotes the Borel subsets of $\mathcal{X} \times \Theta$.

Recall, the data x are arranged as a matrix with columns x_t, $t = 1, 2, \ldots, n$. One sets forth moment functions $m(x_t, \theta)$ of dimension M and computes their mean

$$\bar{m}(x, \theta) = \frac{1}{n} \sum_{t=1}^{n} m(x_t, \theta). \tag{5}$$

The model implies that at the true value θ^o the unconditional expectation of the mean is zero, i.e., $\mathcal{E}\bar{m}(x, \theta^o) = 0$, and that θ^o is the only value of θ for which this is true. Put

$$Z(x, \theta) = \sqrt{n}\, [W(x, \theta)]^{-\frac{1}{2}} \left[\bar{m}(x, \theta)\right], \tag{6}$$

where

$$W(x, \theta) = \frac{1}{n} \sum_{t=1}^{n} \left[m(x_t, \theta) - \bar{m}(x, \theta)\right] \left[m(x_t, \theta) - \bar{m}(x, \theta)\right]' \tag{7}$$

and $[W(x, \theta)]^{-\frac{1}{2}}$ denotes the inverse of the Cholesky factorization of $W(x, \theta)$. If the $m(x_t, \theta)$ are serially correlated one will have to use a HAC (heteroskedatic, autoregressive invariant) variance matrix estimate of the form given by Gallant (1987, p. 446, 533) instead. In this case, residuals $e_t = m(x_t, \theta) - \bar{m}(x, \theta)$ should be used to form the estimate as in (7).

The random variable $z = Z(x, \theta)$ on $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ has some distribution $\Psi(z)$ with a support $\mathcal{Z}$. Let $\mathcal{C}$ be the smallest σ-algebra containing the preimages $C = Z^{-1}(B)$ where B ranges over the Borel subsets of $\mathcal{Z}$. Because the distribution $\Psi(z)$ of $z = Z(x, \theta)$ is determined by P^o the probability measure $P[C = Z^{-1}(B)] = \int_B d\Psi(z)$ over $(\mathcal{X} \times \Theta, \mathcal{C})$ will satisfy $P(C) = P^o(C)$ for every $C \in \mathcal{C}$.

Define $\mathcal{C}^*$ to be the smallest σ-algebra that contains all sets in $\mathcal{C}$ plus all sets of the form $R_B = (\mathcal{X} \times B)$, where B is a Borel subset of Θ. Gallant (2016a) proves that there is an extension of $(\mathcal{X} \times \Theta, \mathcal{C}, P)$ to a space $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ such that $P^o(C) = P^*(C)$ for all $C \in \mathcal{C}^*$. In particular, $P^o(C) = P(C) = P^*(C)$ for $C \in \mathcal{C}$ and $P^o(R_B) = P^*(R_B)$. The σ-algebras involved satisfy $\mathcal{C} \subset \mathcal{C}^* \subset \mathcal{C}^o$.

[5]The parameters of the likelihood may be the same as the parameter θ that enters $Z(x, \theta)$, contain elements of θ, or be distinct from θ. All three situations occur in applications and there are examples of the latter two later on. To avoid clutter and to be consistent with Gallant(2016a, 2016b, 2016c), we will assume here that the parameters are the same.

If $Z(x, \theta)$ is a semi-pivotal and has distribution Ψ with density ψ, then x has conditional density $p^*(x \mid \theta) = \text{adj}(x, \theta)\psi[Z(x, \theta)]$ defined over $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ (Gallant, 2016a, 2016b). The term $\text{adj}(x, \theta)$ is analogous to a Jacobian, as illustrated in the previous subsection. And we will illustrate further in Section 3.

2.3 Adjustment Invariance

To put the issue of invariance into perspective, note that the adjustment does two things: It makes $p^*(x \mid \theta)$ given by (3) integrate to one; It may affect the accept/reject decision in the MCMC algorithm. If one is trying to learn from data by plotting the contours of a density then it is essential that that density integrate to one. That is not our concern here. The issue here is does it affect the accept/reject decision. Especially as it takes human effort to determine $\text{adj}(x, \theta)$.

The adjustment $\text{adj}(x, \theta)$ can be regarded as a data dependent prior. In large enough samples the prior is dominated by the likelihood and therefore will not not affect the accept/reject decision. The adjustment cannot be made tighter in light of the sample size as can an ordinary prior so that the likelihood will eventually dominate. Subjectivity in the choice of prior is permitted in Bayesian inference. From this viewpoint one could go so far as to argue that omission of $\text{adj}(x, \theta)$, thinking that it is likely to be irrelevant, amounts to subjectivity in the choice of a prior.

The above thoughts are not entirely my own and thus require attribution. They come from one of the referee reports, from Sims (2016), and remarks by Chris Sims on the consequences of using contours of densities that do not integrate to one in data analysis made in the floor discussion of this paper when presented at The George Tauchen 70th Birthday Conference, Duke University, Durham, NC, November 15–16, 2019.

Although the above regarding the accept/reject decision in the MCMC algorithm is correct and identical accept/reject decisions imply invariance, it is largely beside the point. The point being not whether the draws in two MCMC chains are identical but rather are the chains drawing from the same distribution. We turn our attention to this issue.

The only sets to which $p^*(x \mid \theta)$ given by (3) can assign probability are sets in C^*. Consider a set of the form

$$A \times \{\theta^o\} = \{(x, \theta^o) : |Z(x, \theta^o) - z^o| < \delta\}$$

for small $\delta > 0$, arbitrary $z^o \in \mathcal{Z}$, and arbitrary $\theta^o \in \Theta$. Now $A \times \{\theta^o\}$ is in C^* and

$$P^*(A \,|\, \theta = \theta^o) = \int_{\{z\,:\,z=Z(x,\theta^o),\, x\in A\}} \psi(z)\, dz.$$

Apply the change of variables[6] $u = U(z, \theta^o)$, $z = U^{-1}(u, \theta^o)$, $dz = (\partial/\partial u')U^{-1}(u, \theta^o)\, du$, whence

$$P^*(A \,|\, \theta = \theta^o) = \int_{\{u\,:\,u=U[Z(x,\theta^o),\theta^o],\, x\in A\}} |(\partial/\partial u')U^{-1}(u, \theta^o)|\, \psi[U^{-1}(u, \theta^o)]\, du.$$

Because $P^*(A \,|\, \theta = \theta^o)/\delta$ will be computed the same for any choice of $u = U(z, \theta)$, MCMC computations using the adjustment described in Subsection 2.1 will be drawing from the same distribution.

2.4 Constructing the Adjustment Term

We next describe a method for constructing the mappings $u = U(z, \theta)$ and $x = X(u, \theta)$ such that u has the same dimension as z and $z = U^{-1}(u, \theta) = Z[X(u, \theta), \theta]$. Recall that the data x are a matrix with n columns; i.e., $x = [x_1, \ldots, x_n]$.

Let k be the largest column index such that $\dim[x_k, \ldots, x_n] \geq \dim(z)$. Find a function $u = U(z, \theta)$ with inverse $z = U^{-1}(u, \theta)$ and functions $X_k(u, \theta), \ldots, X_n(u, \theta)$ that map u to $[x_k, \ldots, x_n]$ such that

$$m[X_k(u, \theta), \theta] + \ldots + m[X_n(u, \theta), \theta] = z. \tag{8}$$

Next find functions $X_1(u, \theta), \ldots, X_{k-1}(u, \theta)$ such that

$$\begin{aligned} \sum_{t=1}^{k-1} m[X_t(u, \theta), \theta] &= 0 \\ n\, W[X(u, \theta), \theta] &= I \end{aligned}$$

Equivalently, find $X_{(1)}(u, \theta)$ such that

$$Z[X(u, \theta), \theta] = z \tag{9}$$

where

$$\begin{aligned} X(u, \theta) &= \left[X_{(1)}(u, \theta), X_{(2)}(u, \theta)\right] \\ X_{(1)}(u, \theta) &= [X_1(u, \theta), \ldots, X_{k-1}(u, \theta)] \\ X_{(2)}(u, \theta) &= [X_k(u, \theta), \ldots, X_n(u, \theta)]\,. \end{aligned}$$

[6] If, instead, one is using the first construction of Subsection 2.1, then use the change of variables $u = U(x, \theta^o)$, $z = Z[X(u, \theta^o), \theta^o]$, $dz = (\partial/\partial u')Z[X(u, \theta^o), \theta^o]\, du$.

As $X_{(1)}(u,\theta)$ is never actually used, it need only exist. Existence is plausible because one is solving a system of M equations in $(k-1)\dim x_t$ unknowns for which the semi-pivotal condition suggests that a solution ought to exist.

Existence can be checked numerically if there is doubt. One method for checking is to use a simple iterative approach using random orthogonal matrices O of dimension s. A method for constructing O is to assign random numbers to the elements of a square matrix and apply Gram-Schmidt. One transforms successive blocks $x_{(i)}$ of length s from $\text{vec}[x_1,\ldots,x_{k-1}]$ by drawing random orthogonal O and random v from the uniform over $(0,\tau)$ and replacing $x_{(i)}$ by $x^*_{(i)} = x_{(i)} + vO$. Denote $x = X(u,\theta)$ by x^* after replacement. One accepts transforms for which $\| z - Z(x^*,\theta) \|$ decreases then moves on to the next block of $\text{vec}[x_1,\ldots,x_{k-1}]$, wrapping from end to beginning as necessary. One continues thus until $\| z - Z(x^*,\theta) \|$ is smaller than a reasonable tolerance. Both τ and the dimension s of O are tuning parameters. This is repeated for random draws θ from the prior until one is satisfied that $X(u,\theta)$ exists.

2.5 Asymptotic Normality

A verification that the limiting density of Z is normal under mild regularity conditions is in the Online Appendix (Gallant, 2020b).

3 CRRA Utility

Our examples are asset pricing applications under constant relative risk aversion (CRRA) utility. In this section we describe the stochastic discount factor corresponding to CRRA utility, describe the standard moment conditions $\bar{m}(x,\theta)$ used to estimate the parameters $\theta = (\beta,\gamma)$ entering the CRRA utility function, and derive the adjustment term for the normalized moment conditions $Z(x,\theta)$.

Let lsr_t denote gross geometric stock returns observed at time t; i.e., $\text{lsr}_t = \log(P_t + D_t) - \log(P_{t-1})$, where P_t is the stock price and D_t is the dividend. Let lcg_t denote log consumption growth observed at time t; i.e., $\text{lcg}_t = \log(C_t) - \log(C_{t-1})$. In a Lucas (1978) exchange economy $D_t = C_t$. Data is a matrix x with columns

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \text{lsr}_t \\ \text{lcg}_t \end{pmatrix} \tag{10}$$

The stochastic discount factor (SDF) for CRRA utility is the marginal rate of substitution $M_{t,t-1} = \beta \left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}$ where β is the representative agent's discount factor, γ is a risk aversion parameter, and the elasticity of intertemporal substitution (EIS) is $1/\gamma$. The SDF discounts gross returns to unity so

$$\mathcal{E}_{t-1}\left[1 - \exp(\log\beta - \gamma \mathrm{lcg}_t + \mathrm{lsr}_t)\right] = 0,$$

where $\mathcal{E}_{t-1}$ is conditional expectation given $C_{t-1}, C_{t-2}, \ldots$. We shall refer to the term in brackets as the Euler equation error.

Standard moment conditions for estimating $\theta = (\beta, \gamma)$ are

$$m(x_t, \theta) = \begin{pmatrix} 1 \\ \mathrm{lsr}_{t-1} \\ \mathrm{lcg}_{t-1} \end{pmatrix} \left[1 - \exp(\log\beta - \gamma \mathrm{lcg}_t + \mathrm{lsr}_t)\right], \tag{11}$$

for $t = T_0, \ldots, n$.[7] By the law of iterated expectations, the $m(x_t, \theta)$ are uncorrelated under $p^o(x, \theta)$ so that

$$\bar{m}(x,\theta) = \frac{1}{n}\sum_{t=1}^{n} m(x_t,\theta) \tag{12}$$

$$W(x,\theta) = \frac{1}{n}\sum_{t=1}^{n} \left[m(x_t,\theta) - \bar{m}(x,\theta)\right]\left[m(x_t,\theta) - \bar{m}(x,\theta)\right]' \tag{13}$$

$$Z(x,\theta) = \sqrt{n}\left[W(x,\theta)\right]^{-\frac{1}{2}}\left[\bar{m}(x,\theta)\right]. \tag{14}$$

To derive the adjustment factor $\mathrm{adj}(x,\theta)$ consider the last two moment equations. Our right hand side choices and solution is below followed by a discussion of the logic involved. For some $c > 0$, e.g., $c = 4$ if the distribution $\Psi(z)$ of Z is well approximated by the normal, larger if $\Psi(z)$ is fat tailed, consider

$$m(x_{n-1},\theta) = \begin{pmatrix} 1 \\ \mathrm{lsr}_{n-2} \\ \mathrm{lcg}_{n-2} \end{pmatrix} \Big[1 - \exp(\log\beta - \gamma \mathrm{lcg}_{n-1} + \mathrm{lsr}_{n-1})\Big] = \begin{pmatrix} 1 - e \\ z_2 \\ z_3 \end{pmatrix}$$

$$m(x_n,\theta) = \begin{pmatrix} 1 \\ \mathrm{lsr}_{n-1} \\ \mathrm{lcg}_{n-1} \end{pmatrix} \left[1 - \exp(\log\beta - \gamma \mathrm{lcg}_n + \mathrm{lsr}_n)\right] = \begin{pmatrix} \tanh\left(\frac{1}{c}z_1\right) \\ (1-\log\beta)\tanh\left(\frac{1}{c}z_1\right) \\ 0 \end{pmatrix}$$

[7] $T_0 > 1$ allows for lags in $m(x_t,\theta)$ and in $W(x,\theta)$ when W is HAC. Thus, x_t in $m(x_t,\theta)$ is to be interpreted as x_t itself and whatever additional lags may be needed.

A solution is

$$\begin{aligned}
\mathrm{lsr}_{n-2} &= \frac{z_2}{1-e} \\
\mathrm{lcg}_{n-2} &= \frac{z_3}{1-e} \\
\mathrm{lsr}_{n-1} &= 1 - \log\beta \\
\mathrm{lcg}_{n-1} &= 0 \\
\mathrm{lsr}_{n} &= \log\left[1 - \tanh\left(\frac{1}{c} z_1\right)\right] - \log\beta \\
\mathrm{lcg}_{n} &= 0
\end{aligned}$$

The logic behind the choices and solution is as follows. Starting with $m(x_n, \theta)$, we wish to assign z_1 to the right hand side of $m_1(x_n, \theta)$, which is the Euler equation error. The difficulty here is that the Euler equation error is bounded above by one so z_1 must be transformed. Of the obvious choices, $z_1/(1 + |z_1|)$ is not differentiable and $z_1/(1 + z_1^2)$ is not one-to-one. The one used, $\tanh\left(\frac{1}{c} z_1\right)$, seems natural to one who is familiar with the various representations of the logistic distribution. Otherwise it may seem bizarre. But any tractable distribution function could be used here. With this choice, setting $\mathrm{lcg}_n = 0$ determines lsr_n. This, in turn determines both lsr_{n-1} and the right hand side of $m_2(x_n, \theta)$. Setting $\mathrm{lcg}_{n-1} = 0$, hence $m_3(x_n, \theta) = 0$, is a convenient choice for lcg_{n-1}. With lcg_{n-1} and lsr_{n-1} determined, $m_1(x_{n-1}, \theta) = 1 - e$ thus allowing us to set $m_2(x_{n-1}, \theta) = z_2$ and $m_3(x_{n-1}, \theta) = z_3$, which determines lcg_{n-2} and lsr_{n-2}.

The mapping $u = U(z, \theta)$ is, thus,

$$\begin{aligned}
u_1 &= \log\left[1 - \tanh\left(\frac{1}{c} z_1\right)\right] - \log\beta \qquad (15) \\
u_2 &= \frac{z_2}{1-e} \\
u_3 &= \frac{z_3}{1-e}
\end{aligned}$$

Its inverse $z = U^{-1}(u, \theta)$ is

$$\begin{aligned}
z_1 &= c\left\{\operatorname{arctanh}\left[1 - \exp(u_1 + \log\beta)\right]\right\} \qquad (16) \\
z_2 &= (1-e)u_2 \\
z_3 &= (1-e)u_3.
\end{aligned}$$

The Jacobian [8] $(\partial/\partial u')U^{-1}(u,\theta)$ has elements

$$\begin{aligned}
(\partial/\partial u_1)U_1^{-1}(u,\theta) &= -c\,\frac{\exp(u_1+\log\beta)}{1-[1-\exp(u_1+\log\beta)]^2} \\
&= -c\,\frac{1-\tanh\left(\frac{1}{c}z_1\right)}{1-\left[\tanh\left(\frac{1}{c}z_1\right)\right]^2} \\
(\partial/\partial u_2)U_2^{-1}(u,\theta) &= (1-e) \\
(\partial/\partial u_3)U_3^{-1}(u,\theta) &= (1-e) \\
(\partial/\partial u_j)U_i^{-1}(u,\theta) &= 0 \quad \text{if } i \neq j
\end{aligned}$$

Setting

$$\begin{aligned}
X_{n-2}(u,\theta) &= \begin{pmatrix} u_2 \\ u_3 \end{pmatrix} \\
X_{n-1}(u,\theta) &= \begin{pmatrix} 1-\log\beta \\ 0 \end{pmatrix} \\
X_n(u,\theta) &= \begin{pmatrix} u_1 \\ 0 \end{pmatrix},
\end{aligned}$$

the mapping $x = X(u,\theta)$ implied by the above is

$$\begin{aligned}
X_{(1)}(u,\theta) &= [X_1(u,\theta),\ldots,X_{n-3}(u,\theta)] \\
X_{(2)}(u,\theta) &= [X_{n-2}(u,\theta), X_{n-1}(u,\theta), X_n(u,\theta)] \\
X(u,\theta) &= \left[X_{(1)}(u,\theta), X_{(2)}(u,\theta)\right]
\end{aligned}$$

where $X_{(1)}(u,\theta)$ is determined such that

$$\begin{aligned}
\sum_{t=T_0}^{n-3} m[X_t(u,\theta),\theta] &= 0 \\
n\,W[X(u,\theta),\theta] &= I
\end{aligned}$$

or, equivalently, where $X_{(1)}(u,\theta)$ is determined such that

$$Z[X(u,\theta),\theta] = V(z),$$

[8] On a machine $\tanh\left(\frac{1}{c}z_1\right)$ can equal ± 1. For the event $+1$, $(\partial/\partial u_1)Z_1(u,\theta) = -\frac{c}{2}$ by l'Hospital's rule. To guard against the event -1, the constant c can be chosen larger than 4 in (17). In this paper, when the event -1 occurs, the proposed MCMC draw is rejected.

where

$$v = V(z) = \begin{pmatrix} 1 - e + \tanh\left(\frac{1}{c}z_1\right) \\ z_2 + (1 - \log\beta)\tanh\left(\frac{1}{c}z_1\right) \\ z_3 \end{pmatrix}$$

with inverse

$$z = V^{-1}(v) = \begin{pmatrix} c\left[\operatorname{arctanh}(v_1 - 1 + e)\right] \\ v_2 - (1 - \log\beta)(v_1 - 1 + e) \\ v_3 \end{pmatrix}.$$

The adjustment is

$$\operatorname{adj}(x, \theta) = c\,(1 - e)^2 \left| \frac{1 - \tanh\left(\frac{1}{c}z_1\right)}{1 - \left[\tanh\left(\frac{1}{c}z_1\right)\right]^2} \right|. \tag{17}$$

4 An Exchange Economy

In this section we consider the adjustment factor in an instance where assuming normality for Z is justified.

Following Sargent and Stachurski (2018), we consider the following specification of a pure exchange economy of Lucas (1978)

$$\begin{aligned} \text{Log Endowment:}\quad & y_t = \alpha y_{t-1} + e_t \\ \text{Consumption:}\quad & C_t = Y_t = e^{y_t} \\ \text{Random shock:}\quad & e_t \sim N(0, \sigma^2) \\ \text{Utility function :}\quad & \mathcal{E}_0\left(\sum_{t=0}^{\infty} \beta^t \, \frac{(C_t)^{1-\gamma} - 1}{1 - \gamma} \right) \end{aligned}$$

$N(\mu, \sigma^2)$ denotes the normal distribution function with mean μ and variance σ^2 and $n(\mu, \sigma^2)$ the normal density function. The time increment is one year. Lower case is the natural logarithm of upper case quantities, i.e., $c_t = \log(C_t)$, $y_t = \log(Y_t)$. Shocks are independent. The stationary distribution of y_t is $N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$.

The stochastic discount factor is $M_{t+1,t} = \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$, which implies that the price at time t of a claim on the endowment is

$$\begin{aligned} P(y_t) &= \mathcal{E}_t\left\{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}[Y_{t+1}+P(y_{t+1})]\right\} \qquad (18) \\ &= \beta Y_t^{\gamma}\mathcal{E}_t\left[Y_{t+1}^{1-\gamma}+Y_{t+1}^{-\gamma}P(y_{t+1})\right] \\ &= \beta Y_t^{\gamma}\mathcal{E}_t\left(Y_{t+1}^{1-\gamma}\right)+\beta Y_t^{\gamma}\mathcal{E}_t\left[Y_{t+1}^{-\gamma}P(y_{t+1})\right] \\ &= \beta Y_t^{\gamma}e^{\alpha(1-\gamma)y_t+\frac{1}{2}(1-\gamma)^2\sigma^2}+\beta Y_t^{\gamma}\mathcal{E}_t\left[Y_{t+1}^{-\gamma}P(y_{t+1})\right]. \end{aligned}$$

$P(\cdot)$ is defined as the solution of the Euler condition (18). To compute $P(\cdot)$ we use the fixed point algorithm set forth in Sargent and Stachurski (2018). The changes we make are to use a natural cubic spline (Schumaker, 2015) to represent $P(\cdot)$ instead of a grid approximation and Gaussian quadrature (Golub and Welsch, 1969) instead of Monte Carlo integration to compute $\mathcal{E}_t\left[Y_{t+1}^{-\gamma}P(y_{t+1})\right]$. We make the first change to reduce the number of grid points required and provide a differentiable approximation with differentiable inverse; we make the second for improved numerical accuracy. Rather than $P(y)$ we usually work with $p(y) = \log P(y)$ and its inverse $y = q(p)$. Once one has computed $P(y_i)$ at some set of not necessarily equally spaced points y_i it is a trivial matter to compute $p(y)$ and $q(p)$ from these same points.

The geometric return on the endowment is $r_{dt} = \log(Y_t + P_t) - \log(P_{t-1})$. The geometric return on an asset that pays one unit of the endowment one year hence with certainty is $r_{ft} = -\log\beta - (1-\alpha)\gamma y_t - \frac{1}{2}\gamma^2\sigma^2$. Following convention, call r_{dt} the geometric stock return, r_{ft} the geometric risk free rate, and $P(y_t)$ the stock price. For a given value of $\rho = (\alpha, \sigma, \beta, \gamma)$, simulated values of x_t given by (10) are

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} r_{dt} \\ y_t \end{pmatrix}. \qquad (19)$$

The prior is shown in Table A1 of the Online Appendix (Gallant, 2020b); it is a normal independence prior with mean $= (0.95, 0.02, 0.95, 12.5)$ and sdev $= (0.01, 0.01, 0.01, 2.0)$. Support conditions and effective support as determined by a simulation of size 10,000 are shown in the table; only $0.01 < \sigma$ binds. The mean of the prior was calibrated to get an excess return and standard deviations on the geometric stock return and geometric risk free rate as close (subjectively) to US values[9] over the last 100 years as an exchange economy

[9]See Table A1 legend for the values.

with CRRA utility will permit. The fixed point iterations in (18) do not converge for all choices of ρ; on this see Sargent and Stachurski (2018). The standard deviations of the prior are close to the upper limit of convergence. Withal, the effective support of the prior is reasonable. This is the tight prior. A loose prior is also used in estimation; it is the same but with standard deviations multiplied by ten.

We next examine whether a normality assumption is warranted for this economy and the relative magnitude of the adjustment term.

Normality is addressed in Figure A1 of the Online Appendix. It is a quantile-quantile plot of $z = Z(x, \theta)$ given by (14) against the quantiles of a standard normal and a Student's t-distribution on three degrees freedom. The z draws are computed from $x = [x_1, \ldots, x_n]$ and (β, γ) obtained by drawing $(\alpha, \sigma, \beta, \gamma)$ from the prior described in Table A1 then drawing x from $p(x_t \,|\, x_{t-1}, \alpha, \sigma, \beta, \gamma)$ described above. Complete details are in the table legend. The desiderata is how closely the plot follows a straight degree line. The components Z_1, Z_2, and Z_3 of Z are effectively uncorrelated by construction so that the joint distribution of Z can be reasonably considered to be the product of the marginals shown in Figure A1. The plot suggests that the assumption that Z follows the normal density is reasonable at $n = 100$ and that assuming a density with fat tails is not reasonable. Plots for $n = 50$ and $n = 1000$ (not shown in the Online Appendix) are nearly identical to Figure A1.

Table 1 about here.

Table 2 about here.

The relative magnitude of the adjustment term is addressed in Table 1 here and in Tables A3 and A4 of the Online Appendix. The adjustment entries in lower panels of Tables A3 through A4 are about the same as in Table 1 but likelihood entries of the upper panels are larger due to the increased sample sizes. In these tables are shown that instance in $N = 1000$ draws from the prior for which the adjustment has the maximal effect over the values of β and γ shown in the table for sample sizes $n = 50, 100, 1000$. In each legend are the values of the ρ corresponding to that draw and the range of MCMC draws for the loose prior, which indicates which values of β and γ shown in the table are relevant. Complete details are in the legend of Table 1, For $n = 50$ the smallest value of $|\log_{10}(\text{likelihood})| - |\log_{10}(\text{adjustment})|$ over the two panels of Tables 1 is 0.6185 which would suggest that the adjustment term can affect the first significant digit of $\log p^*(x, \theta)$ and thereby affect the MCMC accept/reject condition. Table 2 presents MCMC results.[10] Panels 1 and 3 are relevant to $n = 50$. In panel 1, the prior dominates; little can be said. Freed from the necessity of successful simulation, one can use a loser prior as in panel 3. In this case the adjustment shifts estimates but not significantly relative to standard deviations. Standard deviations are larger. See the table legend for complete details.

For $n = 100$ the smallest value of $|\log_{10}(\text{likelihood})| - |\log_{10}(\text{adjustment})|$ over the two panels of Tables A3 is 0.7131. As now the likelihood is larger relative to the adjustment than for $n = 50$ one expects the adjustment to be less important as born out in the second and fourth panels of Table 2.

For $n = 1000$ the situation is interesting. The smallest value of $|\log_{10}(\text{likelihood})| - |\log_{10}(\text{adjustment})|$ over the two panels of Tables A4 is 3.0316, which suggests that the adjustment cannot affect the MCMC accept/reject condition. There is one value where the adjustment is outside IEEE double float limits and thus would entail an automatic rejection of a proposed MCMC draw at that θ value. But the likelihood is so small at that θ that it would almost certainly be rejected if proposed and, indeed, the entire MCMC chain stays well away from that value. As seen from the third and fifth panels of Table 2, the adjustment has a modest affect for the case $n = 1000$.

This same dependence on sample size of the effect of the adjustment can be seen by comparing lines 3 and 5 of Gallant (2016b).

5 Discounted Corporate Profits

In this section we consider correction for non-normality. To do so we define a parameter vector ρ that has $\theta = (\beta, \gamma)$ of CRRA utility as a subvector, determine a prior for ρ, and describe how a draw x is computed from a draw from that prior. From the pair (x, θ) thus drawn, draws from $z = Z(x, \theta)$ can be computed.

The case study in this section is similar to the exchange economy of Section 4 but for

[10] In Table 2 and throughout the mode is computed as that parameter in an MCMC chain with the largest posterior, which is only an approximation to the true mode. It's advantage over other measures of central tendency, such as the mean, is that the parameter value has actually occurred in the chain and therefore satisfies support conditions and therefore can be used in subsequent calculations.

which distribution of Z violates a normality assumption. The case study relies on a model free extraction of the SDF from returns on the Fama and French (1993) portfolios and the 30-day T-bill for the years 1930 through 2015 subject to a yield curve prior using the methods proposed by Gallant and Hong (2007). The yield curve prior is

$$p(\theta) = \prod_{t=1930}^{2015} \phi[(Y_{1,t} - 0.00896)/0.01]\phi[(Y_{30,t} - 2.0)/0.01)] \tag{20}$$

where ϕ is the standard normal density function. Complete details regarding this model free SDF extraction and data sources are in Gallant and Tauchen (2018). The extracted SDF is considered to be observed data hereafter.

Consider the trivariate series

$$y_t = \begin{pmatrix} \log(SDF_{t-1,t}) \\ \log(GDP_t) - \log(CP_t) \\ \log(CP_t) - \log(CP_{t-1}) \end{pmatrix} = \begin{pmatrix} sdf_{t-1,t} \\ gdp_t - cp_t \\ \Delta cp_{t-1,t} \end{pmatrix} \tag{21}$$

where CP_t denotes annual corporate profits in year t, GDP_t denotes gross domestic product, and $SDF_{t-1,t}$ denotes the SDF.

Let $sdf_{0,t} = \sum_{s=1}^{t} sdf_{t-1,t}$, $SDF_{0,t} = \prod_{s=1}^{t} SDF_{t-1,t} = \exp(sdf_{0,t})$. The time zero present value of the cash flow CP_t is

$$PV_{0,t}(CP) = \mathcal{E}(CP_t\, SDF_{0,t} \,|\, \mathcal{F}_0) = \mathcal{E}\left[\exp\left(\sum_{s=1}^{t} \Delta cp_{s-1,s} + \sum_{s=1}^{t} sdf_{s-1,s}\right) \Bigg| \mathcal{F}_0\right] \tag{22}$$

where $\mathcal{F}_t$ denotes the time t information set. The time zero discounted value of the sum of corporate profits through time t is the sum $DCP_{0,t} = \sum_{s=1}^{t} PV_{0,s}(CP)$. For a risk free payoff of one dollar at time t, the time zero present value is

$$PV_{0,t}(1) = \mathcal{E}(SDF_{0,t} \,|\, \mathcal{F}_0). \tag{23}$$

The corresponding yield is

$$YLD_t = -\log[PV_{0,t}(1))]/t. \tag{24}$$

Assume that (21) follows the one-lag vector autoregression (VAR)

$$y_t = b_0 + By_{t-1} + e_t, \tag{25}$$

with initial condition y_0 where the e_t are independent, trivariate normal with mean zero and variance $S = RR'$, where R is the Cholesky factor of S. Analytic expressions for (22) under

(25) are given in Gallant and Tauchen (2018) as well as data sources for CP and GDP. Note that (22) will depend on y_0. When we need to call attention to this fact we write $DCP_{0,t} \,|\, y_0$, $YLD_t \,|\, y_0$.

Using observed data for (21) as the y_t and given values for $(b_0, B, R, \beta, \gamma)$, where β and γ are the subjective discount factor and the risk aversion parameter of CRRA utility, respectively, geometric returns to corporate profits can be computed as

$$lsr_t = \log[(DCP_{0,30} \,|\, y_0 = y_t) + CP_t] - \log(DCP_{0,30} \,|\, y_0 = y_{t-1}) \tag{26}$$

and log consumption growth can be computed as

$$lcg_t = [\log(\beta) - sdf_{t-1,t}]/\gamma. \tag{27}$$

These values for lsr_t and lcg_t, arranged as in (10), are data x for computing $Z(x, \theta)$ given by (14) with $\theta = (\beta, \gamma)$.

We next derive a prior for $(b_0, B, R, \beta, \gamma)$. From CP and GDP data for the years 1960 through 2015, the VAR (25) was estimated subject to a prior comprised of the product of $\phi[\mathcal{E}(lcg) - 0.04]$, $\phi[\mathrm{Var}(lcg) - 0.01]$, and equation (20),[11] where $\mathcal{E}(\cdot)$ and $\mathrm{Var}(\cdot)$ are with respect to the stationary distribution of (25), and support conditions $0.8 \leq \beta \leq 0.99$ and $0 \leq \gamma \leq 100$. The mode[12] and variance matrix of the posterior of the parameter vector

$$\rho = (b_0, \mathrm{vec}(B), \mathrm{vech}(R), \beta, \gamma) \tag{28}$$

are used to define a prior for for ρ; i.e, the mode becomes the mean of the prior and the variance matrix becomes the variance matrix of the prior. The mean and standard deviations of the prior are displayed in Table A6 of the Online Appendix (Gallant, 2020b); its correlation matrix is shown in Table A7 of the Online Appendix; it is sparse.

Figure 1 is a quantile-quantile plot of $z = Z(x, \theta)$ drawn from $\Psi(z)$[13] for lsr and lcg computed as described in this section and prior described in Tables A6 and A7 against the quantiles of a standard normal and a Student's t-distribution on three degrees freedom. Complete details are in the table legend. The desiderata is how closely the solid plot follows a straight degree line. Fat, t-like tails are indicated by the dotted line being straight. The

[11] With 1930 replaced by 1960 in (20); the yields entering (20) are computed using (24).

[12] Recall that the mode is computed as that parameter in an MCMC chain with the largest posterior.

[13] One draws ρ from the prior defined by Tables A6 and Table A7, computes $x_t = (lsr_t, lcg_t)$ given that ρ using (26), (27), and the observed data y_t, then computes $Z(x, \theta)$ for the subvector $\theta = (\beta, \rho)$ of ρ.

components Z_1, Z_2, and Z_3 of Z are effectively uncorrelated by construction so that the joint distribution of Z can be reasonably considered to be the product of the marginals shown in Figure 1. Figure 1 suggests that presuming that Z_2 and Z_3 follow the normal density is reasonable for $n = 50$. However, Z_1 is decidedly not normal with fat t-like tails and right skew.

Figure 1 about here.

Non-normality can be corrected by letting Z_1 follow

$$\psi(z_1) = \left| \frac{d}{du}\Phi^{-1}[F_{Z_1}(z_1)] \frac{d}{dz_1} F_{Z_1}(z_1) \right| \phi \left\{ \Phi^{-1}[F_{Z_1}(z_1)] \right\}, \tag{29}$$

where Φ is the distribution function of the standard normal and F_{z_1} is a natural cubic spline interpolation of the empirical distribution of z_1, As seen from Figure A3 of the Online Appendix, plots of the quantiles of $\tilde{z} = (\Phi^{-1}[F_{z_1}(z_1)], z_2, z_3)$ against normal quantiles are essentially straight lines.

Table 3 about here.

Table 3 here and Tables A9 and A10 of the Online Appendix are evaluations of the likelihood $p^*(x \,|, \theta) = \psi[Z(x, \theta)]$ with Z given by (14) and ψ the standard normal density for the panel labeled "Log Likelihood" and ψ given by (29) for the panel labeled "Log Transformed Likelihood" evaluated over a grid with $0.80 \leq \beta \leq 0.99$, $0.05 \leq \gamma \leq 100$; complete details are in the legend of Table 3. Table 3 and Tables A9 and A10 suggest, firstly, that identification is poor. Secondly, they suggest that both normal and transformed ψ will determine approximately the same posterior location parameters for $\theta = (\beta, \gamma)$ but that transformed ψ will have larger posterior standard deviations.

Table 4 about here.

The predictions from the analysis of Table 3 and Tables A9 and A10 are imperfectly borne out in Table 4. A difficulty we face with this case study is that the fat tails of the distribution of Z_1 affect the MCMC draws: draws for γ mostly cluster about the mode but there are a number of draws well to the right, some near 100. This can be partly mitigated by focusing on the mode and interquartile range (IQR) of the draws rather than the mean and standard deviation.

6 Complementary Methods

An alternative implementation of Bayesian method of moments uses an an explicit likelihood $f(x \mid \rho)$ that has a non-parametric interpretation such as a sieve. This approach was termed Group (1) in the Introduction. The elements of ρ do not contain elements of the parameter θ of the unconditional moment conditions $\bar{m}(x, \theta)$ defined by (5). Taking the expectation of the moment conditions with respect to the likelihood generates parametric restrictions

$$0 = \bar{g}(\rho, \theta) = \int \bar{m}(x, \theta)\, f(x|\rho)\, dx \tag{30}$$

. The difficulty with the Group (1) approach is that the parameter space

$$\{(\rho, \theta) \in \mathcal{R} \times \Theta \mid 0 = \bar{g}(\rho, \theta)\} \tag{31}$$

has Lebesgue measure zero. This makes estimation of the posterior distribution of (ρ, θ) subject to a joint prior $p(\rho, \theta)$ and the constraint (30) by Markov Chain Monte Carlo (MCMC) problematic. Here we shall presume that the joint prior has the form $p(\rho, \theta) = p(\rho)p^o(\theta)$. Relevant statistical methods — Bornn, Shephard, and Solgi (2018), Shin (2015), Gallant, Hong, Leung, and Li (2019), Schennach (2005), and Gallant (2020a) — were discussed earlier.

The method proposed by Gallant, Hong, Leung, and Li (2019) is best suited to our purpose here because we are not going to estimate θ but rather to choose a θ and force $f(x \mid \rho)$ to accommodate. The method is simple. One uses a general purpose sieve for $f(x \mid \rho)$ subject to a prior of the form

$$p_\lambda(\rho, \theta) \propto p^o(\theta) \times p(\rho) \times \exp\left[-\lambda \frac{n}{2}\, \bar{g}'(\rho, \theta)\bar{g}(\rho, \theta)\right]. \tag{32}$$

The larger is λ the more $f(x \mid \rho)$ is forced to approximately satisfy $\bar{g}'(\rho, \theta) = 0$. To have a name, call it the λ-prior method.

To implement the λ-prior method we use the seminonparametric density (SNP) $f_{SNP}(x \mid \rho)$ proposed by Gallant and Tauchen (1989). Its main advantage in the present context is that well tested code for estimation and simulation is available. The parameter space for likelihood $f_{SNP}(x \mid \rho)$ and prior $p_\lambda(\rho, \theta)$ does not have measure zero so MCMC can proceed in the usual fashion. For large λ, ρ draws become concentrated near the parameter space (31) thereby providing approximate draws from the posterior with likelihood $f_{SNP}(x \mid \rho)$ and prior $p(\rho)p^o(\theta)$ subject to constraint (30). For λ sufficiently large, MCMC must fail because the effective parameter space collapses to (31). 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.

The plan is to use the λ-prior method to infer a likelihood $f(x \mid \rho)$ and a prior $p(\rho)$ from knowledge of $\bar{m}(x, \theta)$, $p^o(\theta)$, and the observed data that can be used to construct diagnostics for the normality of $Z(x, \theta)$ similar to Figure 1. One proceeds as follows.

A specification $f_{SNP}(x \mid \rho)$ is chosen by fitting to the data using the BIC model selection criterion (Schwarz, 1978). One can experiment with alternative specifications to see if they impact results.

Figure 2 about here.

Next, discretize the prior $p^o(\theta)$ using a quadrature rule (Golub and Welsch, 1969). A quadrature rule approximates $p^o(\theta)$ by a discrete density that assign probability p_i^o to points θ_i^o for $i = 1, ..., K$. If the quadrature rule puts some points outside the support of θ, these points are discarded and the p_i^o are renomalized to sum to one.

For each support point θ_i^o, apply the λ-prior method to $f_{SNP}(x \mid \rho)$, successively, with

$$p_\lambda^i(\rho, \theta) \propto p^o(\theta) \times \exp\left[-\lambda \, \frac{n}{2} \, \bar{g}'(\rho, \theta_i^o) \bar{g}(\rho, \theta_i^o)\right] \tag{33}$$

for $i = 1, ..., K$. Let ρ_i^o be the mode of the posterior density $p_\lambda^i(\rho \mid x) \propto f_{SNP}(x \mid \rho) \, p_\lambda^i(\rho, \theta)$.

Approximate draws z from the distribution $\Psi(z)$ of $Z(x, \theta)$ defined by (6) are obtained by drawing i with probability p_i^o, $i = 1, ..., K$, drawing x from $f_{SNP}(x \mid \rho_i^o)$, and putting $z = Z(x, \theta_i^o)$. One repeats until one has a sample of z as large as desired.

The method just proposed cannot recover the likelihood $p^o(x \mid \rho)$ that actually generates the data because the data x is only observed at a single value of the parameter pair (ρ, θ). Thus, under smoothness conditions for quasi-maximum likelihood applied to sieves, the best one can hope for is a local approximation to $p^o(x \mid \rho)$, presuming that one of the θ_i^o is actually that which generated the data. Therefore, all the proposed diagnostic can do is produce a class of models local to the correct model. Passing or failing the diagnostic implies that there is a class of models compatible with the observed data for which Z is or is not normally distributed. Stated differently, one constructs a model that can generate the observed data. For that model Z is or is not normally distributed.

Figure 2 here, for the discounted cash flow example, and A4 of the Online Appendix (Gallant, 2020b), for the exchange economy example, show the quantiles of $\Psi(z)$ plotted against the quantiles of the standard normal. The quantiles of $\Psi(z)$ are obtained from z draws obtained as just described. Figures 2 and A4 can be compared to Figures 1 and A1, respectively. That these figures are constructed for different sample sizes n does not materially affect their appearance for $n \leq 500$. The comparison suggests that the approximation to $\Psi(z)$ obtained via λ-prior is a reasonable indicator to whether or not $\Phi(z)$ can be used as an approximation in an application.

7 Conclusion

This paper has addressed two issues that arise in the practical application of the GMM representation of the likelihood in Bayesian inference: (1) a missing Jacobian term and (2) a normality assumption.

Practicable methods for addressing these two problems in an application are proposed and illustrated by application to the seminal application of the GMM methodology: an endowment economy whose representative agent has constant relative risk aversion utility.

While the main contribution of the paper is the proposed methodology, the illustration does provide some interesting anecdotal evidence that is consistent with the anecdotal evidence in Gallant (2016b): violation of (1) or (2) has less serious consequences in applications than one might expect for moderately large sample sizes.

8 References

Bornn, Luke, Neil Shephard, and Reza Solgi (2018), "Moment Conditions and Bayesian Nonparametrics," Journal of the Royal Statistical Society, Series B 81(1), 5–43.

Chamberlain, Gary and Guido Imbens (2003), "Nonparametric applications of Bayesian inference," *Journal of Business and Economic Statistics* 21, 12–18.

Chernozhukov, Victor, and Han Hong (2003), "An MCMC Approach to Classical Estimation," *Journal of Econometrics* 115(2), 293–346.

Duan, Jason A., and Carl F. Mela (2009), "The Role of Spatial Demand on Outlet Location and Pricing," *Journal of Marketing Research* XLVI 260-278

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, 229–247.

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

Gallant, A. Ronald (2016c), "Addendum to Reply to Reflections," Manuscript, http://www.aronaldg.org/papers/addendum.pdf.

Gallant, A. Ronald (2020a), "Nonparametric Bayes Subject to Overidentified Moment Conditions," Manuscript, http://www.aronaldg.org/papers/npb.pdf

Gallant, A. Ronald (2020b), "Online Appendix for Complementary Bayesian Method of Moments Strategies," Manuscript, http://www.aronaldg.org/papers/cb_online.pdf

Gallant, A. Ronald, Raffaella Giacomini, Giuseppe Ragusa (2017), "Bayesian Estimation of State Space Models Using Moment Conditions," *Journal of Econometrics* 201, 198–211.

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. Ronald, Han Hong, Michael P. Leung, and Jessie Li (2019), "Constrained and Penalized Estimation using MCMC," Manuscript, http://www.aronaldg.org/papers/ce.pdf.

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

Gallant, A. Ronald, and George Tauchen (2017), "SNP: A Program for Nonparametric Time Series Analysis, Version 9.1 Users Guide," Manuscript, www.aronaldg.org/webfiles/snp

Gallant, A. Ronald, and George Tauchen (2017), “Exact Bayesian Moment Based Inference for the Distribution of the Small-Time Movements of an Ito Semimartingale,” *Journal of Econometrics* 201, 140–155.

Gallant, A. Ronald, and George Tauchen (2018), “Cash Flows Discounted Using a Model Free SDF Extracted under a Yield Curve Prior,” Manuscript, http://www.aronaldg.org/papers/dcf.pdf.

Gamerman, Dani and Hedibert F. Lopes (2006), “Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd ed.),” Boca Raton, FL: Chapman and Hall.

Golub, Gene H., and John H. Welsch (1969), ”Calculation of Gaussian Quadrature Rules,” *Mathematics of Computation* 23, 221–230.

Hansen, L. P. (1982), “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica* 50, 1029–1054.

Hansen, L. P., and Singleton, K. J. (1982), “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models,” *Econometrica* 50 1269–1286.

Lucas, Robert E. (1978), “Asset Prices in an Exchange Economy,” *Econometrica* 46, 1429–1445.

Owen, Art B. (1988), “Empirical likelihood ratio confidence intervals for a single functional,” *Biometrika* 75(2), 237–249.

Romeo, Charles (2007), “A Gibbs Sampler for Mixed Logit Analysis of Differentiated Product Markets Using Aggregate Data,” *Computational Economics* 29(1), 33–68.

Sargent, Thomas J., and John Stachurski (2018), “QuantEcon.lectures-python3 PDF, Ch. 7.6 Asset Pricing II: The Lucas Asset Pricing Model,” Manuscript, https://lectures.quantecon.org.

Schennach, Susanne M. (2005), “Bayesian Exponentially Tilted Empirical Likelihood,” *Biometrika* 92, 31–46.

Schumaker, Larry L. (2015), *Spline Functions: Computational Methods,* Society for Industrial and Applied Mathematics, Philadelphia, PA

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics* 6, 461–464.

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

Sims, Christopher A. (2016), "Comment on: Reflections on the Probability Space Induced by Moment Conditions with Implications for Bayesian Inference," *Journal of Financial Econometrics* 14, 272–277,

Yin, Guosheng (2009), "Bayesian Generalized Method of Moments," *Bayesian Analysis* 4, 191–208.

Zappa, Emilio, Miranda Holmes-Cerfon,and Jonathan Goodman (2018), "Monte Carlo on Manifolds: Sampling Densities and Integrating Functions," *Communications on Pure and Applied Mathematics* 71, 2609–2647.

Table 1. Exchange Economy Likelihood and Adjustment, $n = 50$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-95.45	-4295.71	-70.45	-14.64	-6.03	-4.13	-3.86	-4.04	-4.29	-4.49	-4.62
10.45	-69.96	-3823.10	-51.70	-11.02	-5.10	-3.96	-3.95	-4.19	-4.45	-4.63	-4.73
20.40	-49.85	-3262.43	-36.57	-8.26	-4.47	-3.93	-4.09	-4.37	-4.61	-4.77	-4.85
30.35	-34.66	-2619.06	-24.70	-6.27	-4.12	-4.00	-4.27	-4.56	-4.78	-4.91	-4.96
40.30	-23.73	-1919.42	-15.77	-4.96	-4.00	-4.16	-4.49	-4.77	-4.96	-5.06	-5.07
50.25	-16.17	-1219.97	-9.53	-4.24	-4.08	-4.40	-4.74	-4.99	-5.14	-5.20	-5.19
60.20	-10.99	-607.12	-5.71	-4.03	-4.31	-4.70	-5.01	-5.21	-5.32	-5.34	-5.30
70.15	-7.45	-179.24	-4.06	-4.25	-4.69	-5.04	-5.30	-5.45	-5.51	-5.49	-5.41
80.10	-5.38	-13.38	-4.26	-4.81	-5.17	-5.43	-5.60	-5.69	-5.69	-5.63	-5.52
90.05	-4.97	-62.53	-5.87	-5.65	-5.74	-5.85	-5.92	-5.93	-5.88	-5.77	-5.63
100.00	-6.41	-361.68	-8.48	-6.69	-6.37	-6.29	-6.24	-6.17	-6.06	-5.91	-5.73

Log Adjustment

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	1.78	1.78	1.78	1.86	2.04	2.24	2.44	2.60	2.74	2.84	2.92
10.45	1.78	1.78	1.78	1.90	2.10	2.31	2.50	2.66	2.78	2.88	2.95
20.40	1.78	1.78	1.79	1.95	2.18	2.39	2.56	2.71	2.82	2.91	2.97
30.35	1.79	1.78	1.81	2.02	2.26	2.46	2.63	2.76	2.87	2.94	3.00
40.30	1.82	1.78	1.85	2.11	2.35	2.54	2.69	2.81	2.91	2.98	3.02
50.25	1.87	1.78	1.92	2.22	2.45	2.62	2.76	2.87	2.95	3.01	3.05
60.20	1.96	1.78	2.05	2.36	2.55	2.70	2.82	2.92	2.99	3.04	3.07
70.15	2.13	1.78	2.25	2.51	2.67	2.79	2.89	2.97	3.03	3.07	3.09
80.10	2.40	1.92	2.53	2.67	2.78	2.87	2.95	3.01	3.06	3.10	3.12
90.05	2.78	6.78	2.90	2.85	2.90	2.96	3.01	3.06	3.10	3.12	3.14
100.00	3.24	13.16	3.33	3.05	3.02	3.04	3.08	3.11	3.13	3.15	3.16

For sample size $n = 50$, ρ from the prior shown in Table A1 of the Online Appendix (Gallant, 2020b) was sampled with $N = 1000$ repetitions. For each repetition, x with columns given by (19) was simulated from the exchange economy of Section 4. For $\theta = (\beta, \gamma)$ as shown, the likelihood is $p^*(x \,|, \theta) = \phi[Z(x, \theta)]$ with Z given by (14) and ϕ the standard normal density. The adjustment is given by (17). Shown are values for that ρ among the $N = 1000$ repetitions for which the difference between the maximum and minimum adjustment was largest. That value is $\rho = (0.9415, 0.01103, 0.9431, 17.89)$. Under the loose prior, the range of the MCMC draws reported in Table A5 were $0.8006 \leq \beta \leq 0.9899$, $0.07502 \leq \gamma \leq 84.46$ with adjustment and $0.8007 \leq \beta \leq 0.9898$, $0.01369 \leq \gamma \leq 82.19$ without.

Table 2. Exchange Economy Estimates

	No Adjustment			Adjustment		
Parameter	Mean	Mode	Sdev	Mean	Mode	Sdev
			$n = 50$, tight prior			
β	0.9530	0.9550	0.00513	0.9540	0.9559	0.00510
γ	13.650	12.365	1.4849	13.603	12.233	1.4800
			$n = 100$, tight prior			
β	0.9595	0.9603	0.00422	0.9604	0.9611	0.00428
γ	13.603	12.233	1.4800	13.994	12.994	1.3948
			$n = 1000$, tight prior			
β	0.9312	0.9315	0.00242	0.9442	0.9445	0.00169
γ	14.995	14.673	0.8594	14.900	14.539	0.7894
			$n = 50$, loose prior			
β	0.9177	0.9561	0.03884	0.9228	0.9569	0.03904
γ	31.145	12.025	12.893	31.725	11.939	14.016
			$n = 100$, loose prior			
β	0.9416	0.9614	0.03017	0.9449	0.9623	0.03022
γ	33.272	13.301	13.402	34.613	13.189	14.213
			$n = 1000$, loose prior			
β	0.9280	0.9296	0.00493	0.9433	0.9441	0.00282
γ	16.539	15.414	2.0841	15.870	14.912	1.8355

The x with columns given by (19) are a simulation from the exchange economy of Section 4 with with ρ set to the values shown in Tables 1 and Tables A3 and A4 of the Online Appendix (Gallant, 2020b). Without adjustment the likelihood is $p^*(x\,|,\theta) = \phi[Z(x,\theta)]$ with Z given by (14) and ϕ denoting the standard normal density. With adjustment the likelihood is $p^*(u\,|,\theta) = \text{adj}(u,\theta)\phi[Z(u,\theta)]$ with u defined by (15), Z defined by (16), and adj defined by (17). $\theta = (\beta,\gamma)$, a subvector of ρ. The tight prior for θ is the last two rows shown in Table A1. The loose prior is the same with standard deviations multiplied by ten. Estimates are from an MCMC chain (Gamerman and Lopes, 2006) of length $N = 200000$ collected past the point where transients have died out. The proposal is move-one-at-a-time random walk. Mean and standard deviation are computed with a stride of 100; mode is that over the entire chain.

Table 3. Discounted Corporate Profits Likelihood, $n = 50$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-237.32	-14.67	-5.56	-5.96	-6.29	-6.35	-6.30	-3.17	-3.17	-3.17	-3.17
10.45	-202.61	-14.00	-5.54	-5.96	-6.29	-6.34	-6.28	-3.17	-3.17	-3.17	-3.17
20.40	-171.87	-13.39	-5.52	-5.96	-6.28	-6.32	-6.27	-3.17	-3.17	-3.17	-3.17
30.35	-144.75	-12.82	-5.51	-5.96	-6.27	-6.31	-6.25	-3.17	-3.17	-3.17	-3.17
40.30	-120.93	-12.29	-5.49	-5.96	-6.27	-6.30	-6.24	-3.17	-3.17	-3.17	-3.17
50.25	-100.10	-11.80	-5.48	-5.96	-6.26	-6.29	-6.23	-3.17	-3.17	-3.17	-3.17
60.20	-82.00	-11.34	-5.46	-5.96	-6.25	-6.28	-6.21	-3.17	-3.17	-3.17	-3.17
70.15	-66.39	-10.92	-5.45	-5.96	-6.24	-6.27	-6.20	-3.17	-3.17	-3.17	-3.17
80.10	-53.02	-10.52	-5.44	-5.96	-6.24	-6.26	-6.18	-3.17	-3.17	-3.17	-3.17
90.05	-41.69	-10.15	-5.43	-5.96	-6.23	-6.24	-6.17	-3.17	-3.17	-3.17	-3.17
100.00	-32.19	-9.81	-5.42	-5.96	-6.22	-6.23	-6.15	-3.17	-3.17	-3.17	-3.17

Log Transformed Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-14.67	-7.39	-5.77	-5.96	-6.32	-6.37	-6.31	-3.21	-3.21	-3.21	-3.21
10.45	-14.18	-7.26	-5.74	-5.97	-6.31	-6.36	-6.30	-3.21	-3.21	-3.21	-3.21
20.40	-13.14	-7.15	-5.70	-5.97	-6.31	-6.35	-6.29	-3.21	-3.21	-3.21	-3.21
30.35	-13.24	-7.08	-5.66	-5.97	-6.30	-6.34	-6.27	-3.21	-3.21	-3.21	-3.21
40.30	-12.14	-7.03	-5.62	-5.97	-6.29	-6.32	-6.26	-3.21	-3.21	-3.21	-3.21
50.25	-11.09	-7.01	-5.59	-5.97	-6.28	-6.31	-6.24	-3.21	-3.21	-3.21	-3.21
60.20	-9.52	-6.99	-5.56	-5.97	-6.28	-6.30	-6.23	-3.21	-3.21	-3.21	-3.21
70.15	-9.94	-6.96	-5.53	-5.97	-6.27	-6.29	-6.22	-3.21	-3.21	-3.21	-3.21
80.10	-7.76	-6.92	-5.50	-5.97	-6.26	-6.28	-6.20	-3.21	-3.21	-3.21	-3.21
90.05	-7.89	-6.86	-5.48	-5.97	-6.25	-6.26	-6.19	-3.21	-3.21	-3.21	-3.21
100.00	-8.30	-6.79	-5.46	-5.97	-6.25	-6.25	-6.17	-3.21	-3.21	-3.21	-3.21

For sample size $n = 50$, ρ from the prior shown in Table A6 was sampled with $N = 1000$ repetitions. For each repetition, x with columns given by (10) was computed from simulated cash flows as described in Section 5. For $\theta = (\beta, \gamma)$ as shown, the likelihood is $p^*(x\,|, \theta) = \psi[Z(x, \theta)]$ with Z given by (14) and ψ the standard normal density for the panel labeled "Log Likelihood" and (29) for the panel labeled "Log Transformed Likelihood"; Shown are values for that ρ among the $N = 1000$ repetitions for which the absolute value of the difference between the likelihoods was largest. For this draw $\theta = (\beta, \gamma) = (0.8991, 6.359)$. The range of the MCMC draws reported in Table 4 were $0.8008 \leq \beta \leq 0.9900$, $0.05273 \leq \gamma \leq 99.68$ with adjustment and $0.8049 \leq \beta \leq 0.9900$, $0.40184 \leq \gamma \leq 94.57$ without.

Table 4. Discounted Cash Flow Estimates

	No Adjustment			Adjustment		
Parameter	Mean	Mode	IQR	Mean	Mode	IQR
			$n = 50$			
β	0.9328	0.9556	0.05590	0.9328	0.9559	0.05582
γ	39.595	29.902	17.422	42.082	31.859	18.922
			$n = 100$			
β	0.9311	0.9541	0.05653	0.9311	0.9538	0.05666
γ	30.151	25.574	8.6111	31.737	26.849	9.3107
			$n = 1000$			
β	0.9310	0.9556	0.05631	0.9305	0.9528	0.05669
γ	40.288	28.940	4.3800	37.489	25.680	5.7771
			$n = 50$, transformed Z			
β	0.9327	0.9558	0.05562	0.9327	0.9574	0.05606
γ	35.500	29.562	18.321	39.249	38.540	18.720
			$n = 100$, transformed Z			
β	0.9315	0.9539	0.05652	0.9308	0.9533	0.05677
γ	26.797	25.657	8.4857	28.833	29.564	8.3368
			$n = 1000$, transformed Z			
β	0.9327	0.9604	0.05570	0.9322	0.9567	0.05606
γ	32.624	27.330	3.0234	35.546	28.259	5.0233

The data x are a simulation of log corporate profit returns and log consumption growth as described in the legend of Table A6 with ρ set to the column labeled Mean in Table A6 of the Online Appendix (Gallant, 2020b); specifically, $\theta = (\beta, \gamma) = (0.9532, 24.5030)$ for the subvector θ of ρ. Without adjustment the likelihood is $p^*(x\,|, \theta) = \phi[Z(x, \theta)]$ with Z given by (14) and ϕ denoting the standard normal density. With adjustment the likelihood is $p^*(u\,|, \theta) = \text{adj}(u, \theta)\phi[Z(u, \theta)]$ with u defined by (15), Z defined by (16), and adj defined by (17). Transformed Z are computed as described in the legend to Figure A3. Estimates are from an MCMC chain (Gamerman and Lopes, 2006) of length $N = 800000$ collected past the point where transients have died out. The proposal is move-one-at-a-time random walk. Mean and interquartile range (IQR), and mode are computed with a stride of one.

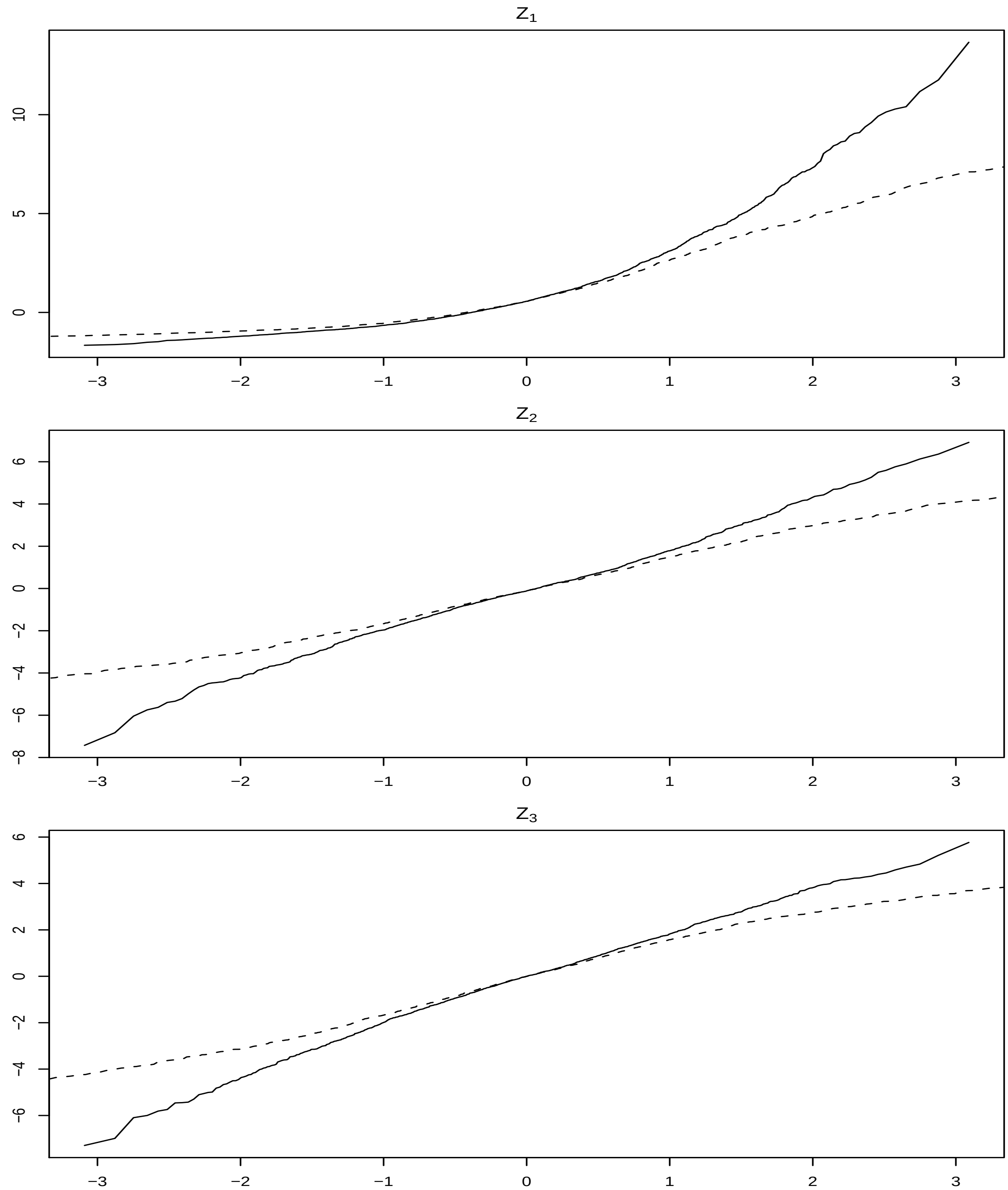

Figure 1. Discounted Cash Flow Raw Q-Q Plots. For the discounted cash flows setup of Section 5, $N = 2000$ values for $\rho = (b_0, B, R, \beta, \gamma)$ were drawn from the prior shown in Table A6 of the Online Appendix (Gallant, 2020b). For each ρ, data $x = \{x_t\}_{t=1}^{n}$ were computed for $n = 50$ as described in the legend to Table A6 of the Online Appendix (2020b). For each x and subvector $\theta = (\beta, \gamma)$ of ρ the random variable $z = Z(x, \theta)$ was computed as described in the legend to Table A6. From the $N = 2000$ values of z thus computed, quantiles at probabilities 0.001 through 0.999 at increments of 0.001 were computed for each of the elements z_1, z_2, and z_3 of z. Plotted are the z quantiles (vertical axis) against the normal quantiles (solid line) and the t-quantiles (dotted line).

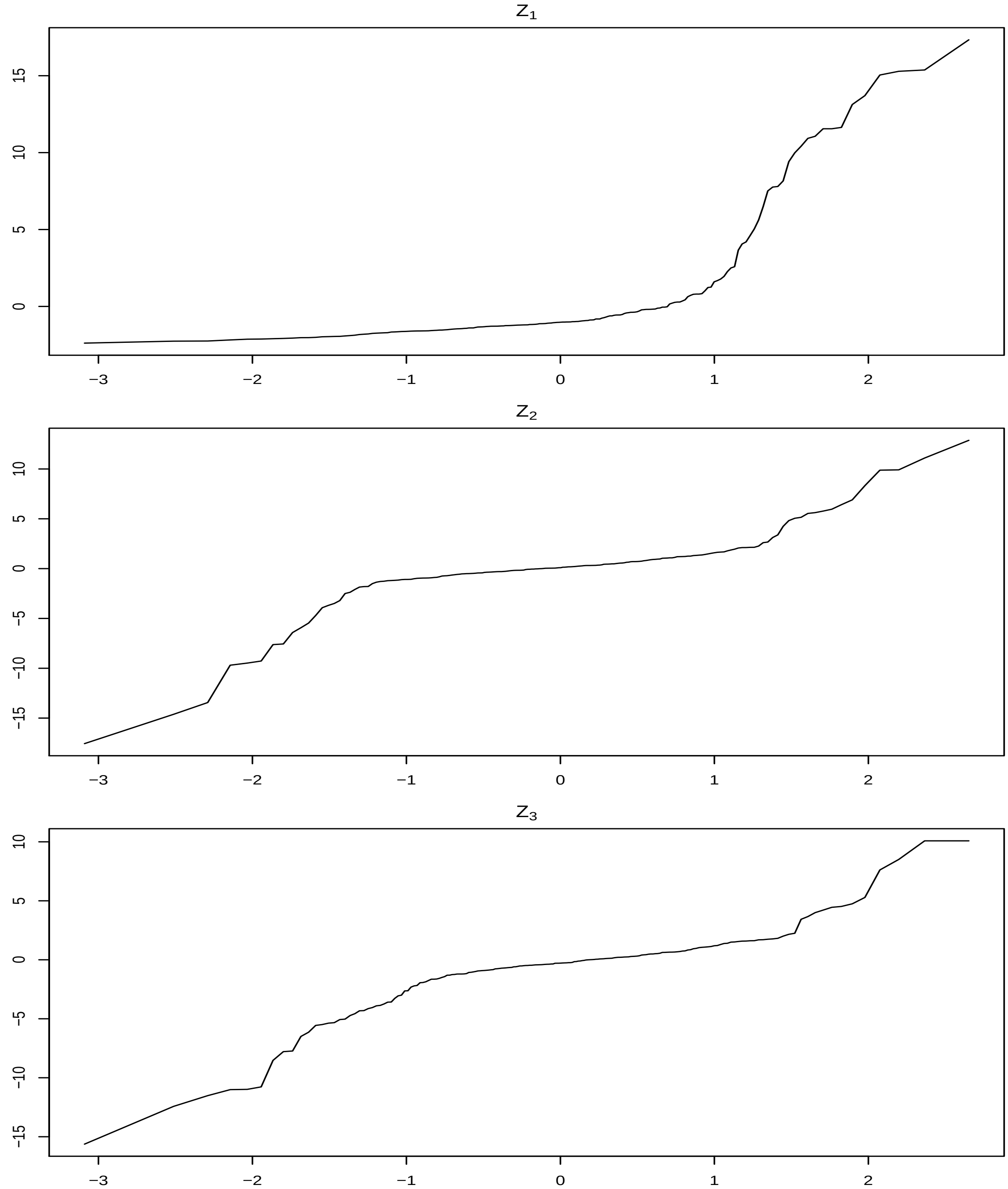

Figure 2. Complementary Q-Q Plots for Discounted Cash Flows. The prior $p^o(\theta)$ given in the legend of Table A6 of the Online Appendix (Gallant, 2020b) for the discounted cash flows setup of Section 5 was discretized by a quadrature rule to obtain a discrete prior $(\theta_i^o, p_i^o), i = 1, \ldots, 16$. Using the λ-prior method described in Section 6 with $p_\lambda^i(\rho, \theta)$ defined by (33), the discrete prior $(\rho_i^o, p_i^o), i = 1, \ldots, 16$, was determined for the likelihood $f_{SNP}(x, \rho)$. In the notation of Gallant and Tauchen (2017), the BIC determined SNP specification has parameters $L_u = 1$, and $K_z = 4$, with all other parameters zero. Sample size is $n = 500$; $\lambda = 10^{1.25} \approx 1.25$. $N = 1000$ values for p were drawn from the discrete prior (ρ_i^o, p_i^o). For each ρ drawn, data x of length $n = 500$ were drawn from $f_{SNP}(x, \rho)$. For each x, the random variable $z = Z(x, \theta)$ defined by (14) was computed. From the $N = 1000$ values of z thus computed, quantiles at probabilities 0.001 through 0.999 at increments of 0.005 were computed for each of the elements z_1, z_2, and z_3 of z. Similarly for the normal. Plotted are the z quantiles against the normal quantiles.

Online Appendix for Complementary Bayesian Method of Moments Strategies by A. Ronald Gallant

www.aronaldg.org/papers/cb_online.pdf

Asymptotic Normality

The probability space on which $Z(x, \theta)$ is defined has density $p^o(x \,|\, \theta)p^o(\theta)$, which is a mixture. To draw, one draws θ from $p^o(\theta)$ and then x from the conditional density $p^o(x \,|\, \theta)$. Fix such a θ draw. For data x generated according to $p^o(x \,|\, \theta)$ from that draw, define

$$F_n(z|\theta) = \int I[Z(x,\theta) \le z]\, p^o(x|\theta)dx.$$

Asymptotic normality for that θ is defined as

$$\lim_{n\to\infty} F_n(z \,|\, \theta) = \Phi(z) \tag{34}$$

at every continuity point of the right hand side, which is everywhere for the normal $\Phi(z)$. Conditions that imply asymptotic normality for cross-sectional data are given by Theorem 9 of Gallant(1987, p. 211) and for time series data by Theorem 10 Gallant(1987, p. 581).

To emphasize the dependence on sample size, write $\Psi_n(z)$ for the distribution of Z, and note that

$$\Psi_n(z) = \int F_n(z|\theta)\, p^o(\theta)\, d\theta. \tag{35}$$

The dominated convergence theorem (Royden and Fitzpatrick, 2010, p. 88), (34) implies

$$\lim_{n\to\infty} \Psi_n(z) = \int \lim_{n\to\infty} F_n(z|\theta)\, p^o(\theta)\, d\theta = \Phi(z).$$

Thus, asymptotic normality for any draw from $p^o(\theta)$ is enough to imply asymptotic normality of $Z(x, \theta)$ on the probability space $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$; asymptotic normality uniform over Θ is not required.[14]

Denote the density that corresponds to Ψ_n by ψ_n. To obtain $\lim_{n\to\infty} \psi_n(z) = \phi(z)$, $\psi_n(z)$ must be bounded and asymptotically equicontinuous (Sweeting, 1986). Uniformity is not required.

[14] I am indebted to Shengbo Zhu for pointing this out to me.

Additional References

Royden, H. L., and P. M. Fitzpatrick (2010), *Real Analysis, Fourth Edition*, Saddle River, NJ: Prentice-Hall.

Sweeting, Trevor J. (1986), “On a converse to Scheffe’s theorem,” *The Annals of Statistics* 14, 1252–1256.

Table A1. Exchange Economy Prior

	Normal Factor		Support Factor		Range in Simulation	
Parameter	Mean	Sdev	Left	Right	Left	Right
α	0.95	0.01	-0.99	0.99	0.914	0.988
σ	0.2	0.01	0.01	100.0	0.010	0.058
β	0.95	0.01	0.8	0.99	0.908	0.987
γ	12.5	2.0	0.0	100.0	4.719	20.10

For parameter $\rho = (\alpha, \sigma, \beta, \gamma)$ the prior has the form $\prod_{i=1}^{4} n(\rho_i, \text{Mean}_i, \text{Sdev}_i) I(\text{Left}_i < \rho_i < \text{Right}_i)$, for the values shown in the columns labeled Normal Factor and Support Factor, respectively. The effective support of the prior as determined from a simulation of size 10,000 is shown in the columns labeled Range in Simulation. At $\rho = (0.95, 0.02, 0.95, 12.5)$, in a simulation of size 200,000, the log endowment has mean 0.00000 and standard deviation 0.02024, the geometric stock return has mean 0.04490 and standard deviation 0.14332, the geometric risk free rate has mean 0.02019 and standard deviation 0.04032.

Table A2. Exchange Economy Likelihood and Adjustment, $n = 50$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-95.45	-4295.71	-70.45	-14.64	-6.03	-4.13	-3.86	-4.04	-4.29	-4.49	-4.62
10.45	-69.96	-3823.10	-51.70	-11.02	-5.10	-3.96	-3.95	-4.19	-4.45	-4.63	-4.73
20.40	-49.85	-3262.43	-36.57	-8.26	-4.47	-3.93	-4.09	-4.37	-4.61	-4.77	-4.85
30.35	-34.66	-2619.06	-24.70	-6.27	-4.12	-4.00	-4.27	-4.56	-4.78	-4.91	-4.96
40.30	-23.73	-1919.42	-15.77	-4.96	-4.00	-4.16	-4.49	-4.77	-4.96	-5.06	-5.07
50.25	-16.17	-1219.97	-9.53	-4.24	-4.08	-4.40	-4.74	-4.99	-5.14	-5.20	-5.19
60.20	-10.99	-607.12	-5.71	-4.03	-4.31	-4.70	-5.01	-5.21	-5.32	-5.34	-5.30
70.15	-7.45	-179.24	-4.06	-4.25	-4.69	-5.04	-5.30	-5.45	-5.51	-5.49	-5.41
80.10	-5.38	-13.38	-4.26	-4.81	-5.17	-5.43	-5.60	-5.69	-5.69	-5.63	-5.52
90.05	-4.97	-62.53	-5.87	-5.65	-5.74	-5.85	-5.92	-5.93	-5.88	-5.77	-5.63
100.00	-6.41	-361.68	-8.48	-6.69	-6.37	-6.29	-6.24	-6.17	-6.06	-5.91	-5.73

Log Adjustment

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	1.78	1.78	1.78	1.86	2.04	2.24	2.44	2.60	2.74	2.84	2.92
10.45	1.78	1.78	1.78	1.90	2.10	2.31	2.50	2.66	2.78	2.88	2.95
20.40	1.78	1.78	1.79	1.95	2.18	2.39	2.56	2.71	2.82	2.91	2.97
30.35	1.79	1.78	1.81	2.02	2.26	2.46	2.63	2.76	2.87	2.94	3.00
40.30	1.82	1.78	1.85	2.11	2.35	2.54	2.69	2.81	2.91	2.98	3.02
50.25	1.87	1.78	1.92	2.22	2.45	2.62	2.76	2.87	2.95	3.01	3.05
60.20	1.96	1.78	2.05	2.36	2.55	2.70	2.82	2.92	2.99	3.04	3.07
70.15	2.13	1.78	2.25	2.51	2.67	2.79	2.89	2.97	3.03	3.07	3.09
80.10	2.40	1.92	2.53	2.67	2.78	2.87	2.95	3.01	3.06	3.10	3.12
90.05	2.78	6.78	2.90	2.85	2.90	2.96	3.01	3.06	3.10	3.12	3.14
100.00	3.24	13.16	3.33	3.05	3.02	3.04	3.08	3.11	3.13	3.15	3.16

For sample size $n = 50$, ρ from the prior shown in Table A1 was sampled with $N = 1000$ repetitions. For each repetition, x with columns given by (19) was simulated from the exchange economy of Section 4. For $\theta = (\beta, \gamma)$ as shown, the likelihood is $p^*(x \,|, \theta) = \phi[Z(x, \theta)]$ with Z given by (14) and ϕ the standard normal density. The adjustment is given by (17). Shown are values for that ρ among the $N = 1000$ repetitions for which the difference between the maximum and minimum adjustment was largest. That value is $\rho = (0.9415, 0.01103, 0.9431, 17.89)$. Under the loose prior, the range of the MCMC draws reported in Table A5 were $0.8006 \leq \beta \leq 0.9899$, $0.07502 \leq \gamma \leq 84.46$ with adjustment and $0.8007 \leq \beta \leq 0.9898$, $0.01369 \leq \gamma \leq 82.19$ without.

Table A3. Exchange Economy Likelihood and Adjustment, $n = 100$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-211.99	-16640.04	-170.40	-39.47	-15.89	-8.78	-6.06	-4.92	-4.50	-4.45	-4.61
10.45	-155.11	-13398.32	-126.74	-29.62	-12.51	-7.39	-5.46	-4.70	-4.47	-4.53	-4.75
20.40	-109.32	-10306.79	-91.12	-21.81	-9.87	-6.33	-5.03	-4.58	-4.51	-4.66	-4.92
30.35	-73.48	-7474.01	-62.69	-15.78	-7.88	-5.56	-4.76	-4.55	-4.61	-4.82	-5.12
40.30	-46.57	-5002.77	-40.72	-11.29	-6.43	-5.05	-4.63	-4.61	-4.76	-5.02	-5.33
50.25	-27.68	-2978.95	-24.59	-8.12	-5.46	-4.76	-4.64	-4.74	-4.97	-5.25	-5.56
60.20	-15.73	-1462.56	-13.67	-6.07	-4.91	-4.69	-4.76	-4.95	-5.21	-5.50	-5.80
70.15	-9.34	-484.22	-7.26	-4.97	-4.72	-4.80	-4.98	-5.22	-5.49	-5.78	-6.06
80.10	-7.08	-58.66	-4.61	-4.70	-4.88	-5.08	-5.30	-5.55	-5.81	-6.07	-6.33
90.05	-8.10	-51.30	-5.04	-5.16	-5.33	-5.52	-5.72	-5.93	-6.16	-6.39	-6.60
100.00	-12.42	-508.94	-8.15	-6.29	-6.07	-6.10	-6.21	-6.36	-6.54	-6.72	-6.89

Log Adjustment

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	1.78	1.78	1.78	1.79	1.86	2.01	2.21	2.42	2.63	2.81	2.95
10.45	1.78	1.78	1.78	1.80	1.90	2.08	2.29	2.50	2.70	2.87	3.00
20.40	1.78	1.78	1.78	1.82	1.96	2.15	2.37	2.59	2.78	2.93	3.06
30.35	1.78	1.78	1.78	1.86	2.03	2.24	2.47	2.67	2.85	2.99	3.11
40.30	1.79	1.78	1.79	1.92	2.12	2.35	2.56	2.76	2.92	3.06	3.16
50.25	1.81	1.78	1.81	2.00	2.23	2.46	2.66	2.85	3.00	3.12	3.20
60.20	1.87	1.78	1.87	2.13	2.37	2.58	2.77	2.93	3.07	3.17	3.25
70.15	2.05	1.78	2.01	2.30	2.52	2.71	2.88	3.02	3.14	3.23	3.30
80.10	2.43	1.78	2.30	2.51	2.68	2.84	2.98	3.11	3.21	3.29	3.34
90.05	3.04	6.38	2.75	2.77	2.87	2.98	3.09	3.20	3.28	3.34	3.38
100.00	3.84	15.45	3.37	3.06	3.06	3.12	3.20	3.28	3.35	3.39	3.42

As for Table A2 but with sample size $n = 100$, $\rho = (0.9410, 0.01000, 0.9542, 15.00)$, and range $0.0.8105 \leq \beta \leq 0.9896$, $0.004808 \leq \gamma \leq 85.11$ with adjustment and $0.8002 \leq \beta \leq 0.9899$, $0.9028 \leq \gamma \leq 78.52$ without.

Table A4. Exchange Economy Likelihood and Adjustment, $n = 1000$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-1726.42	-113176.30	-1334.75	-282.69	-92.93	-36.92	-18.53	-14.47	-16.73	-21.69	-27.40
10.45	-1216.58	-80780.43	-949.73	-196.16	-63.81	-26.35	-15.65	-15.17	-19.10	-24.74	-30.57
20.40	-816.54	-54890.15	-643.10	-129.38	-42.29	-19.39	-14.68	-16.94	-22.07	-28.11	-33.89
30.35	-514.72	-34764.91	-406.01	-79.84	-27.40	-15.59	-15.40	-19.64	-25.57	-31.75	-37.34
40.30	-299.62	-19758.34	-230.73	-45.29	-18.29	-14.58	-17.62	-23.18	-29.51	-35.63	-40.88
50.25	-158.29	-9316.57	-110.38	-23.74	-14.24	-16.02	-21.14	-27.43	-33.85	-39.70	-44.51
60.20	-76.30	-2996.26	-38.55	-13.50	-14.65	-19.63	-25.83	-32.32	-38.52	-43.93	-48.20
70.15	-40.56	-419.18	-9.06	-13.14	-19.00	-25.18	-31.54	-37.76	-43.47	-48.29	-51.93
80.10	-42.62	-151.26	-16.16	-21.46	-26.85	-32.44	-38.16	-43.69	-48.68	-52.77	-55.70
90.05	-79.04	-2276.10	-54.96	-37.47	-37.83	-41.23	-45.57	-50.03	-54.09	-57.33	-59.49
100.00	-149.07	-7531.50	-121.60	-60.35	-51.59	-51.38	-53.69	-56.74	-59.67	-61.96	-63.29

Log Adjustment

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	1.78	1.78	1.78	1.78	1.78	1.81	2.06	2.71	3.57	4.37	5.00
10.45	1.78	1.78	1.78	1.78	1.78	1.87	2.28	3.08	3.95	4.69	5.26
20.40	1.78	1.78	1.78	1.78	1.80	1.99	2.61	3.49	4.32	5.00	5.51
30.35	1.78	1.78	1.78	1.78	1.85	2.23	3.03	3.92	4.69	5.30	5.75
40.30	1.78	1.78	1.78	1.79	1.99	2.62	3.51	4.35	5.05	5.59	5.99
50.25	1.78	1.78	1.78	1.85	2.31	3.15	4.02	4.78	5.40	5.87	6.21
60.20	1.78	1.78	1.79	2.12	2.89	3.76	4.55	5.21	5.74	6.14	6.42
70.15	2.00	1.78	2.08	2.86	3.67	4.41	5.07	5.62	6.06	6.40	6.63
80.10	3.82	9.54	3.82	4.05	4.54	5.08	5.58	6.02	6.38	6.65	6.83
90.05	6.84	30.26	6.53	5.40	5.44	5.73	6.08	6.41	6.68	6.89	7.02
100.00	9.87	inf	9.25	6.77	6.32	6.37	6.56	6.78	6.98	7.12	7.20

As for Table A2 but with sample size $n = 1000$, $\rho = (0.9377, 0.01009, 0.9424, 17.84)$, and range $0.8909 \leq \beta \leq 0.9496$, $12.92 \leq \gamma \leq 43.86$ with adjustment and $0.8539 \leq \beta \leq 0.9383$, $12.86 \leq \gamma \leq 40.35$ without.

Table A5. Exchange Economy Estimates

	No Adjustment			Adjustment		
Parameter	Mean	Mode	Sdev	Mean	Mode	Sdev
			$n = 50$, tight prior			
β	0.9530	0.9550	0.00513	0.9540	0.9559	0.00510
γ	13.650	12.365	1.4849	13.603	12.233	1.4800
			$n = 100$, tight prior			
β	0.9595	0.9603	0.00422	0.9604	0.9611	0.00428
γ	13.603	12.233	1.4800	13.994	12.994	1.3948
			$n = 1000$, tight prior			
β	0.9312	0.9315	0.00242	0.9442	0.9445	0.00169
γ	14.995	14.673	0.8594	14.900	14.539	0.7894
			$n = 50$, loose prior			
β	0.9177	0.9561	0.03884	0.9228	0.9569	0.03904
γ	31.145	12.025	12.893	31.725	11.939	14.016
			$n = 100$, loose prior			
β	0.9416	0.9614	0.03017	0.9449	0.9623	0.03022
γ	33.272	13.301	13.402	34.613	13.189	14.213
			$n = 1000$, loose prior			
β	0.9280	0.9296	0.00493	0.9433	0.9441	0.00282
γ	16.539	15.414	2.0841	15.870	14.912	1.8355

The x with columns given by (19) are a simulation from the exchange economy of Section 4 with with ρ set to the values shown in Tables A2 through A4. Without adjustment the likelihood is $p^*(x\,|,\theta) = \phi[Z(x,\theta)]$ with Z given by (14) and ϕ denoting the standard normal density. With adjustment the likelihood is $p^*(u\,|,\theta) = \text{adj}(u,\theta)\phi[Z(u,\theta)]$ with u defined by (15), Z defined by (16), and adj defined by (17). $\theta = (\beta,\gamma)$, a subvector of ρ. The tight prior for θ is the last two rows shown in Table A1. The loose prior is the same with standard deviations multiplied by ten. Estimates are from an MCMC chain (Gamerman and Lopes, 2006) of length $N = 200000$ collected past the point where transients have died out. The proposal is move-one-at-a-time random walk. Mean and standard deviation are computed with a stride of 100; mode is that over the entire chain.

Table A6. Discounted Cash Flow Prior

	Normal Factor		Support Factor		Range in Simulation	
Parameter	Mean	Sdev	Left	Right	Left	Right
$b_0(1)$	-1.6040	0.0143	$-\infty$	∞	-1.6506	-1.5578
$b_0(2)$	0.8663	0.0234	$-\infty$	∞	0.7835	0.9388
$b_0(3)$	-0.9020	0.0376	$-\infty$	∞	-1.0379	-0.7732
$B(1,1)$	-0.0003	0.0006	$-\infty$	∞	-0.0021	0.0016
$B(2,1)$	0.0186	0.0093	$-\infty$	∞	-0.0128	0.0501
$B(3,1)$	-0.0226	0.0103	$-\infty$	∞	-0.0558	0.0158
$B(1,2)$	-0.0020	0.0005	$-\infty$	∞	-0.0036	-0.0002
$B(2,2)$	0.6730	0.0095	$-\infty$	∞	0.6450	0.7083
$B(3,2)$	0.3457	0.0126	$-\infty$	∞	0.2979	0.3921
$B(1,3)$	-0.0017	0.0061	$-\infty$	∞	-0.0210	0.0183
$B(2,3)$	-0.1782	0.0941	$-\infty$	∞	-0.4941	0.1746
$B(3,3)$	0.3144	0.1018	$-\infty$	∞	-0.0225	0.6647
$R(1,1)$	1.7636	0.0083	$-\infty$	∞	1.7313	1.7941
$R(1,2)$	-0.1093	0.0402	$-\infty$	∞	-0.2404	0.0450
$R(2,2)$	0.0314	0.0027	$-\infty$	∞	0.0214	0.0422
$R(1,3)$	-0.2343	0.0336	$-\infty$	∞	-0.3386	-0.1280
$R(2,3)$	-0.1373	0.0121	$-\infty$	∞	-0.1867	-0.0929
$R(3,3)$	0.1575	0.0128	$-\infty$	∞	0.1141	0.2076
β	0.9532	0.0547	0.8	0.99	0.8003	0.9899
γ	24.5030	28.2425	0	100	0.0479	99.5641

The prior to determine the distribution of Z (e.g., Figure A2), is over $\rho = (b_0, B, R, \beta, \gamma)$. The prior is multivariate normal with means and standard errors as shown here and correlations as shown in Table A7 subject to support conditions shown here. The VAR $y_t = b_0 + By_{t-1} + Ru$ with u independent normal determines the the distribution of $y = \log$ (marginal rate of substitution, GDP/corporate profits, corporate profits). For each $\tilde{\rho}$ drawn from the prior, a realization $\{y_t\}_{t=1}^n$ of the VAR with initial condition y_0 for 2015 is computed. From each realization, the return on corporate profits conditional on y_t is computed analytically from the drawn $(\tilde{b}_0, \tilde{B}, \tilde{R})$ and the marginal rate of substitution is converted to consumption growth using CRRA utility with drawn $(\tilde{\beta}, \tilde{\gamma})$. The resulting bivariate geometric returns to corporate profits and log consumption growth series is used to compute the Z corresponding to $\tilde{\rho}$. In a simulation of size 5,000 with ρ set to the prior mean shown here, log consumption growth has mean 0.0020 and standard deviation 0.1441, the geometric return on corporate profits has mean 0.06310 and standard deviation 0.07255. To estimate $\theta = (\beta, \gamma)$ from bivariate returns and consumption growth data using the method described in Section 2 the prior is bivariate normal with means and standard deviations as shown here for (β, γ), zero correlation, and support conditions $0.8 \leq \beta \leq 0.99$ and $0 \leq \gamma \leq 100$.

Table A7. Discounted Cash Flow Prior Correlations

	$b_0(1)$	$b_0(2)$	$b_0(3)$	$B(1,1)$	$B(2,1)$	$B(3,1)$	$B(1,2)$	$B(2,2)$	$B(3,2)$	$B(1,3)$	$B(2,3)$	$B(3,3)$	$R(1,1)$	$R(1,2)$	$R(2,2)$	$R(1,3)$	$R(2,3)$	$R(3,3)$	β	γ
$b_0(1)$	$b_0(1)$																			
$b_0(2)$		$b_0(2)$																		
$b_0(3)$		−∘	$b_0(3)$																	
$B(1,1)$				$B(1,1)$																
$B(2,1)$		+∘	−∘		$B(2,1)$															
$B(3,1)$		−∘	+∘		−●	$B(3,1)$														
$B(1,2)$				−•			$B(1,2)$													
$B(2,2)$	−∘	−•						$B(2,2)$												
$B(3,2)$			−●					−∘	$B(3,2)$											
$B(1,3)$							+∘			$B(1,3)$										
$B(2,3)$			−∘		+∘	−∘					$B(2,3)$									
$B(3,3)$			+∘		−∘	+∘					−•	$B(3,3)$								
$R(1,1)$	−•												$R(1,1)$							
$R(1,2)$		+∘						−∘	+∘		+∘	−∘	+•	$R(1,2)$						
$R(2,2)$															$R(2,2)$					
$R(1,3)$	+∘															$R(1,3)$				
$R(2,3)$			−∘		+∘	−∘					+∘	−∘					$R(2,3)$			
$R(3,3)$			+∘		−∘	+∘					−∘	+∘					−●	$R(3,3)$		
β																			β	
γ																				γ

The prior over $\rho = (b_0, B, R, \beta, \gamma)$ that is used to determine the distribution of Z is multivariate normal with means and standard errors as shown in Table A6 and correlation matrix as shown here. A blank indicates a correlation of less than 0.25 in absolute value. A ∘ indicates a correlation with absolute value between 0.25 and 0.5 with sign as shown. Similarly a • for between 0.5 and 0.75 and ● for larger than 0.75.

Table A8. Discounted Corporate Profits Likelihood, $n = 50$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-237.32	-14.67	-5.56	-5.96	-6.29	-6.35	-6.30	-3.17	-3.17	-3.17	-3.17
10.45	-202.61	-14.00	-5.54	-5.96	-6.29	-6.34	-6.28	-3.17	-3.17	-3.17	-3.17
20.40	-171.87	-13.39	-5.52	-5.96	-6.28	-6.32	-6.27	-3.17	-3.17	-3.17	-3.17
30.35	-144.75	-12.82	-5.51	-5.96	-6.27	-6.31	-6.25	-3.17	-3.17	-3.17	-3.17
40.30	-120.93	-12.29	-5.49	-5.96	-6.27	-6.30	-6.24	-3.17	-3.17	-3.17	-3.17
50.25	-100.10	-11.80	-5.48	-5.96	-6.26	-6.29	-6.23	-3.17	-3.17	-3.17	-3.17
60.20	-82.00	-11.34	-5.46	-5.96	-6.25	-6.28	-6.21	-3.17	-3.17	-3.17	-3.17
70.15	-66.39	-10.92	-5.45	-5.96	-6.24	-6.27	-6.20	-3.17	-3.17	-3.17	-3.17
80.10	-53.02	-10.52	-5.44	-5.96	-6.24	-6.26	-6.18	-3.17	-3.17	-3.17	-3.17
90.05	-41.69	-10.15	-5.43	-5.96	-6.23	-6.24	-6.17	-3.17	-3.17	-3.17	-3.17
100.00	-32.19	-9.81	-5.42	-5.96	-6.22	-6.23	-6.15	-3.17	-3.17	-3.17	-3.17

Log Transformed Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-14.67	-7.39	-5.77	-5.96	-6.32	-6.37	-6.31	-3.21	-3.21	-3.21	-3.21
10.45	-14.18	-7.26	-5.74	-5.97	-6.31	-6.36	-6.30	-3.21	-3.21	-3.21	-3.21
20.40	-13.14	-7.15	-5.70	-5.97	-6.31	-6.35	-6.29	-3.21	-3.21	-3.21	-3.21
30.35	-13.24	-7.08	-5.66	-5.97	-6.30	-6.34	-6.27	-3.21	-3.21	-3.21	-3.21
40.30	-12.14	-7.03	-5.62	-5.97	-6.29	-6.32	-6.26	-3.21	-3.21	-3.21	-3.21
50.25	-11.09	-7.01	-5.59	-5.97	-6.28	-6.31	-6.24	-3.21	-3.21	-3.21	-3.21
60.20	-9.52	-6.99	-5.56	-5.97	-6.28	-6.30	-6.23	-3.21	-3.21	-3.21	-3.21
70.15	-9.94	-6.96	-5.53	-5.97	-6.27	-6.29	-6.22	-3.21	-3.21	-3.21	-3.21
80.10	-7.76	-6.92	-5.50	-5.97	-6.26	-6.28	-6.20	-3.21	-3.21	-3.21	-3.21
90.05	-7.89	-6.86	-5.48	-5.97	-6.25	-6.26	-6.19	-3.21	-3.21	-3.21	-3.21
100.00	-8.30	-6.79	-5.46	-5.97	-6.25	-6.25	-6.17	-3.21	-3.21	-3.21	-3.21

For sample size $n = 50$, ρ from the prior shown in Table A6 was sampled with $N = 1000$ repetitions. For each repetition, x with columns given by (10) was computed from simulated cash flows as described in Section 5. For $\theta = (\beta, \gamma)$ as shown, the likelihood is $p^*(x \,|, \theta) = \psi[Z(x, \theta)]$ with Z given by (14) and ψ the standard normal density for the panel labeled "Log Likelihood" and (29) for the panel labeled "Log Transformed Likelihood"; Shown are values for that ρ among the $N = 1000$ repetitions for which the absolute value of the difference between the likelihoods was largest. For this draw $\theta = (\beta, \gamma) = (0.8991, 6.359)$. The range of the MCMC draws reported in Table A11 were $0.8008 \le \beta \le 0.9900$, $0.05273 \le \gamma \le 99.68$ with adjustment and $0.8049 \le \beta \le 0.9900$, $0.40184 \le \gamma \le 94.57$ without.

Table A9. Discounted Corporate Profits Likelihood, $n = 100$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-335.85	-4.20	-4.03	-4.52	-4.54	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
10.45	-279.67	-4.07	-4.05	-4.52	-4.54	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
20.40	-230.82	-3.96	-4.06	-4.53	-4.54	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
30.35	-188.51	-3.86	-4.07	-4.53	-4.55	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
40.30	-152.08	-3.77	-4.08	-4.53	-4.55	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
50.25	-120.91	-3.69	-4.09	-4.53	-4.55	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
60.20	-94.47	-3.62	-4.11	-4.54	-4.55	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
70.15	-72.27	-3.55	-4.12	-4.54	-4.55	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
80.10	-53.88	-3.49	-4.13	-4.54	-4.55	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
90.05	-38.91	-3.44	-4.14	-4.54	-4.55	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17
100.00	-27.02	-3.40	-4.15	-4.54	-4.55	-4.50	-4.45	-4.41	-3.17	-3.17	-3.17

Log Transformed Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-9.47	-4.49	-4.02	-4.79	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
10.45	-8.82	-4.42	-4.03	-4.79	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
20.40	-8.31	-4.35	-4.04	-4.79	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
30.35	-7.64	-4.26	-4.04	-4.79	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
40.30	-8.37	-4.17	-4.05	-4.79	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
50.25	-9.19	-4.07	-4.06	-4.80	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
60.20	-7.95	-3.96	-4.07	-4.80	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
70.15	-6.63	-3.85	-4.07	-4.80	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
80.10	-6.27	-3.75	-4.08	-4.80	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
90.05	-7.73	-3.65	-4.09	-4.80	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33
100.00	-7.49	-3.55	-4.10	-4.80	-4.89	-4.84	-4.79	-4.75	-3.33	-3.33	-3.33

As for Table A8 but with sample size $n = 100$, $\theta = (\beta, \gamma) = (0.9133, 9.427)$. $0.8036 \leq \beta \leq 0.9900$, $0.4313 \leq \gamma \leq 94.32$ with adjustment and $0.8028 \leq \beta \leq 0.9900$, $0.2902 \leq \gamma \leq 94.53$ without.

Table A10. Discounted Corporate Profits Likelihood, $n = 1000$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-3370.20	-651.29	-90.64	-5.13	-5.45	-6.22	-5.87	-5.47	-5.17	-4.97	-4.83
10.45	-2654.69	-589.82	-82.32	-4.87	-5.54	-6.25	-5.88	-5.47	-5.17	-4.97	-4.83
20.40	-2053.04	-534.00	-74.76	-4.66	-5.63	-6.27	-5.88	-5.47	-5.17	-4.97	-4.83
30.35	-1553.27	-483.26	-67.87	-4.49	-5.71	-6.30	-5.89	-5.47	-5.17	-4.97	-4.83
40.30	-1144.87	-437.10	-61.61	-4.35	-5.80	-6.32	-5.89	-5.47	-5.17	-4.97	-4.83
50.25	-818.50	-395.08	-55.90	-4.25	-5.88	-6.34	-5.90	-5.47	-5.17	-4.97	-4.83
60.20	-565.54	-356.81	-50.70	-4.18	-5.96	-6.37	-5.90	-5.47	-5.18	-4.97	-4.83
70.15	-377.53	-321.93	-45.97	-4.13	-6.04	-6.39	-5.91	-5.47	-5.18	-4.97	-4.83
80.10	-245.61	-290.13	-41.65	-4.11	-6.12	-6.41	-5.91	-5.47	-5.18	-4.97	-4.83
90.05	-159.99	-261.14	-37.73	-4.10	-6.19	-6.43	-5.92	-5.48	-5.18	-4.97	-4.83
100.00	-110.14	-234.71	-34.15	-4.12	-6.27	-6.45	-5.92	-5.48	-5.18	-4.97	-4.83

Log Transformed Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-299.60	-29.99	-39.02	-5.28	-8.55	-8.46	-7.62	-7.87	-7.29	-5.87	-7.12
10.45	-206.38	-30.87	-32.52	-5.00	-10.00	-8.30	-7.60	-7.88	-7.29	-5.87	-7.12
20.40	-131.38	-30.91	-30.32	-4.97	-8.32	-8.17	-7.59	-7.89	-7.30	-5.87	-7.12
30.35	-74.79	-29.45	-28.53	-4.67	-7.81	-8.05	-7.58	-7.90	-7.30	-5.87	-7.12
40.30	-36.72	-28.59	-26.99	-4.34	-7.54	-7.95	-7.57	-7.90	-7.31	-5.87	-7.12
50.25	-15.79	-27.87	-25.65	-4.33	-7.37	-7.86	-7.56	-7.91	-7.31	-5.87	-7.12
60.20	-11.52	-27.18	-24.49	-4.30	-7.11	-7.78	-7.55	-7.92	-7.31	-5.87	-7.12
70.15	-16.05	-26.50	-23.54	-4.19	-7.20	-7.70	-7.54	-7.93	-7.32	-5.87	-7.12
80.10	-31.39	-25.82	-22.87	-4.02	-8.62	-7.64	-7.53	-7.93	-7.32	-5.87	-7.12
90.05	-52.84	-25.13	-21.83	-3.92	-8.38	-7.57	-7.53	-7.94	-7.32	-5.87	-7.12
100.00	-72.31	-24.43	-19.40	-4.19	-7.98	-7.52	-7.52	-7.94	-7.32	-5.87	-7.12

As for Table A8 but with sample size $n = 1000$, $\theta = (\beta, \gamma) = (0.8832, 25.24)$. $0.8012 \leq \beta \leq 0.9900$, $4.635 \leq \gamma \leq 99.94$ with adjustment and $0.8046 \leq \beta \leq 0.9900$, $7.306 \leq \gamma \leq 99.27$ without.

Table A11. Discounted Cash Flow Estimates

	No Adjustment			Adjustment		
Parameter	Mean	Mode	IQR	Mean	Mode	IQR
			$n = 50$			
β	0.9328	0.9556	0.05590	0.9328	0.9559	0.05582
γ	39.595	29.902	17.422	42.082	31.859	18.922
			$n = 100$			
β	0.9311	0.9541	0.05653	0.9311	0.9538	0.05666
γ	30.151	25.574	8.6111	31.737	26.849	9.3107
			$n = 1000$			
β	0.9310	0.9556	0.05631	0.9305	0.9528	0.05669
γ	40.288	28.940	4.3800	37.489	25.680	5.7771
			$n = 50$, transformed Z			
β	0.9327	0.9558	0.05562	0.9327	0.9574	0.05606
γ	35.500	29.562	18.321	39.249	38.540	18.720
			$n = 100$, transformed Z			
β	0.9315	0.9539	0.05652	0.9308	0.9533	0.05677
γ	26.797	25.657	8.4857	28.833	29.564	8.3368
			$n = 1000$, transformed Z			
β	0.9327	0.9604	0.05570	0.9322	0.9567	0.05606
γ	32.624	27.330	3.0234	35.546	28.259	5.0233

The data x are a simulation of log corporate profit returns and log consumption growth as described in the legend of Table A6 with ρ set to the column labeled Mean in Table A6; specifically, $\theta = (\beta, \gamma) = (0.9532, 24.5030)$ for the subvector θ of ρ. Without adjustment the likelihood is $p^*(x \,|, \theta) = \phi[Z(x, \theta)]$ with Z given by (14) and ϕ denoting the standard normal density. With adjustment the likelihood is $p^*(u \,|, \theta) = \text{adj}(u, \theta)\phi[Z(u, \theta)]$ with u defined by (15), Z defined by (16), and adj defined by (17). Transformed Z are computed as described in the legend to Figure A3. Estimates are from an MCMC chain (Gamerman and Lopes, 2006) of length $N = 800000$ collected past the point where transients have died out. The proposal is move-one-at-a-time random walk. Mean and interquartile range (IQR), and mode are computed with a stride of one.

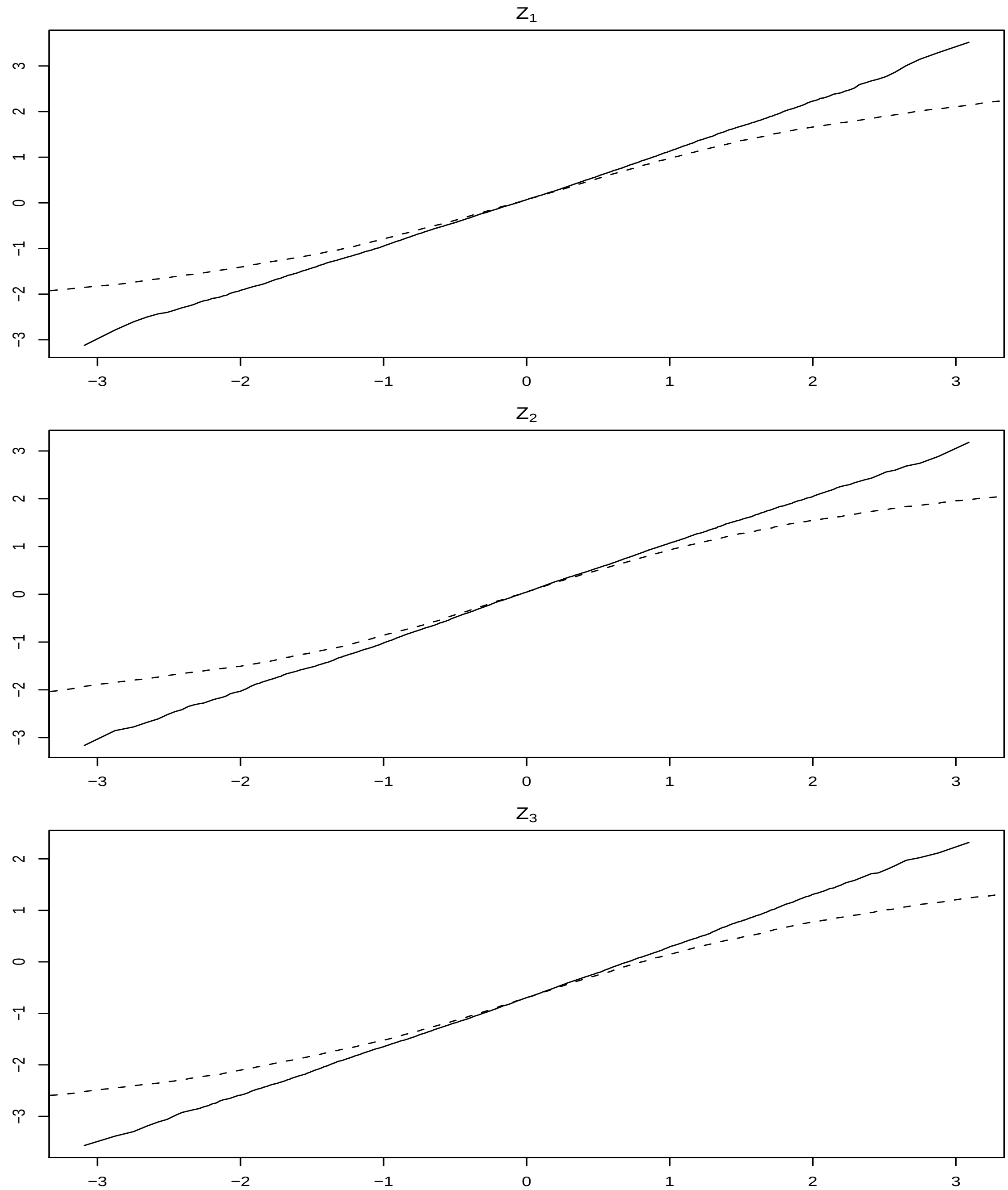

Figure A1. Exchange Economy Q-Q Plots. For the exchange economy of Section 4, $N = 10000$ values for $\rho = (\alpha, \sigma, \beta, \gamma)$ were drawn from the prior shown in Table A1. For each ρ, data x with columns given by (19) were computed for $t = 1, \ldots, n$, $n = 100$. For each x and subvector $\theta = (\beta, \gamma)$ of ρ the random variable $z = Z(x, \theta)$ defined by (14) was computed. From the $N = 10000$ values of z thus computed, quantiles at probabilities 0.001 through 0.999 at increments of 0.001 were computed for each of the elements z_1, z_2, and z_3 of z. Similarly for the normal and Student's t-distribution on 3 degrees freedom. Plotted are the z quantiles (vertical axis) against the normal quantiles (solid line) and the t-quantiles (dotted line).

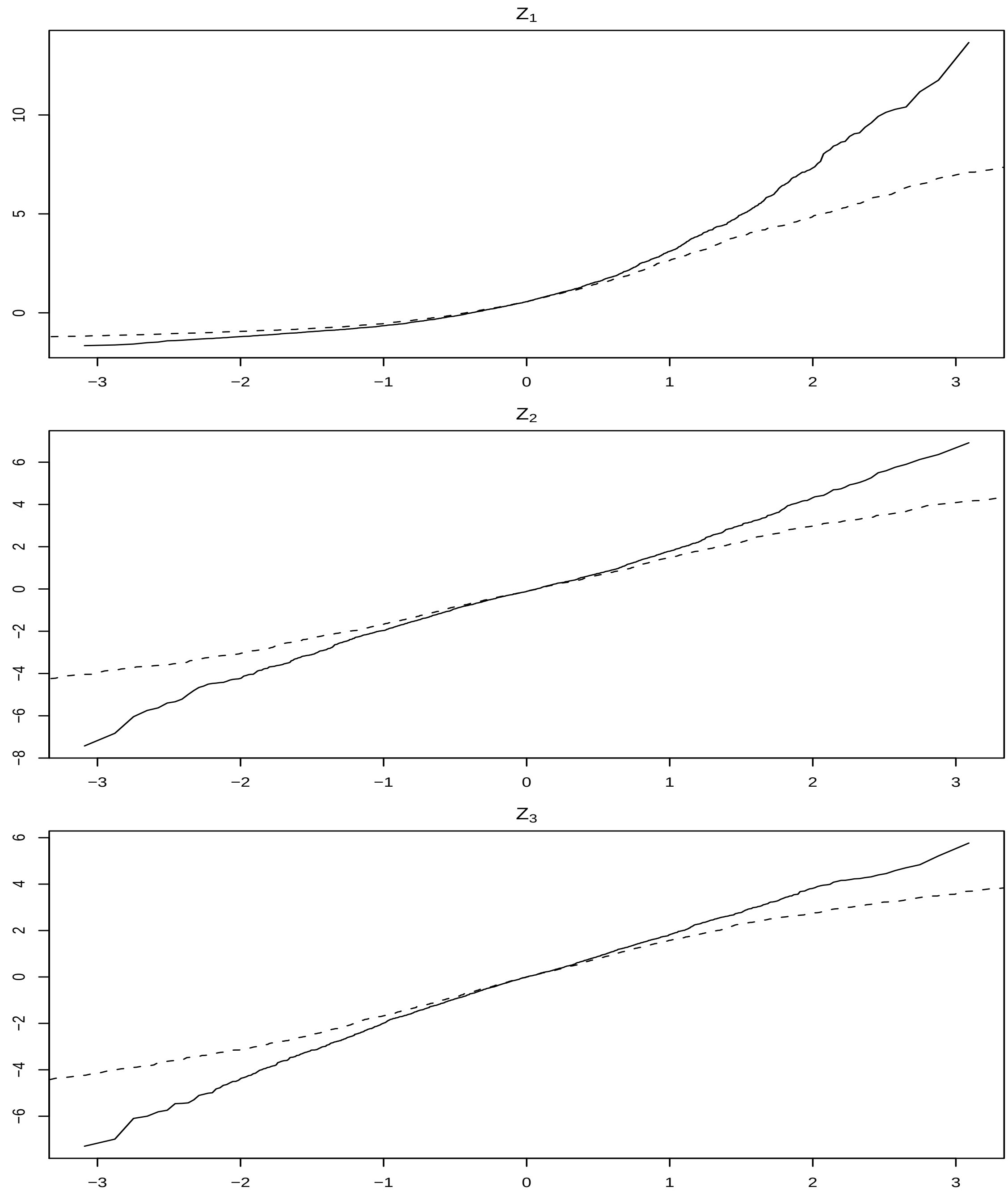

Figure A2. Discounted Cash Flow Raw Q-Q Plots. For the discounted cash flows setup of Section 5, $N = 2000$ values for $\rho = (b_0, B, R, \beta, \gamma)$ were drawn from the prior shown in Table A6. For each ρ, data $x = \{x_t\}_{t=1}^{n}$ were computed for $n = 50$ as described in the legend to Table A6. For each x and subvector $\theta = (\beta, \gamma)$ of ρ the random variable $z = Z(x, \theta)$ was computed as described in the legend to Table A6. From the $N = 2000$ values of z thus computed, quantiles at probabilities 0.001 through 0.999 at increments of 0.001 were computed for each of the elements z_1, z_2, and z_3 of z. Plotted are the z quantiles (vertical axis) against the normal quantiles (solid line) and the t-quantiles (dotted line).

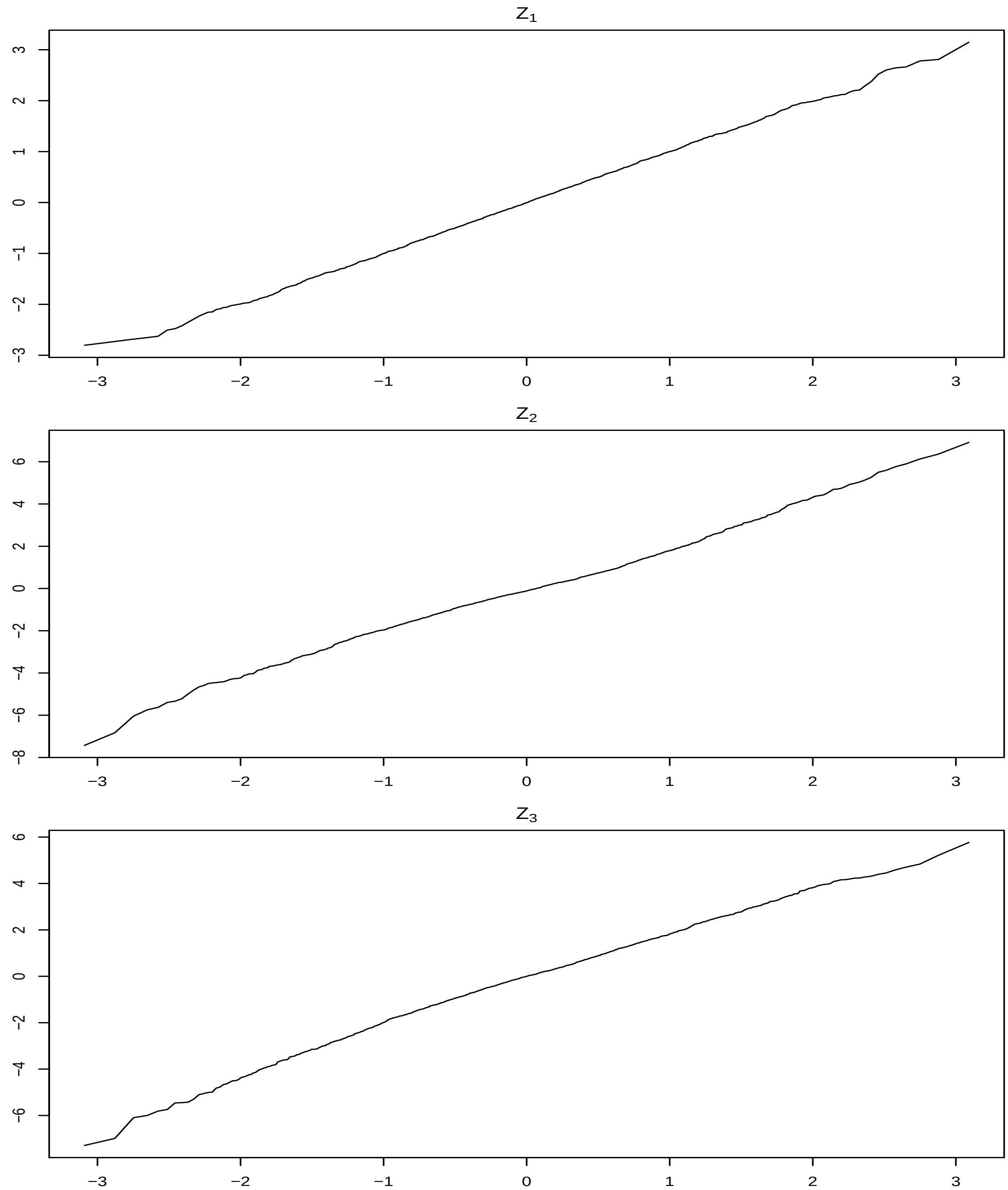

Figure A3. Discounted Cash Flow Transformed Q-Q Plots. For the discounted cash flows setup of Section 5 and draws from the prior shown in Table A6, the random variable $z = Z(x, \theta)$ was computed as described in the legend for Figure A2. Each $z = (z_1, z_2, z_3)$ was transformed to $\tilde{z} = \big(\Phi^{-1}[F_{z_1}(z_1)], z_2, z_3\big)$, where Φ is the distribution function of the standard normal and F_{z_1} is a natural cubic spline interpolation of the empirical distribution of z_1. The correlations among the $\tilde{z}$ are -0.0451, 0.0088, 0.0015 in the order (1,2), (1,3), (2,3). From the $N = 2000$ values of $\tilde{z}$ thus computed, quantiles at probabilities 0.001 through 0.999 at increments of 0.001 were computed for each of the elements of $\tilde{z}$. Plotted are the $\tilde{z}$ quantiles (vertical axis) against the normal quantiles.

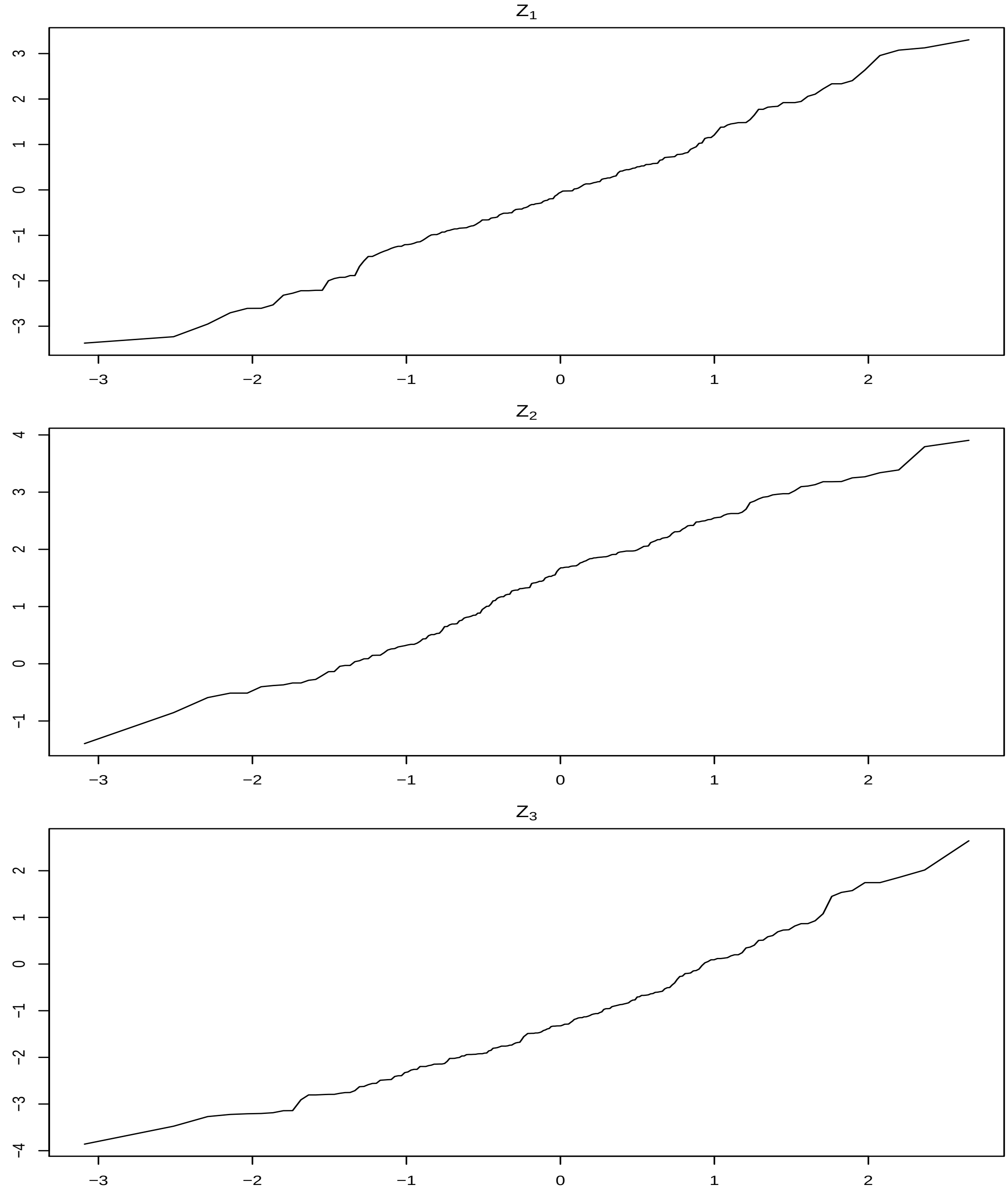

Figure A4. Complementary Q-Q Plots for Exchange Economy. The prior $p^o(\theta)$ given in the legend of Table A1 for the exchange economy of Section 4 was discretized by a quadrature rule to obtain a discrete prior $(\theta_i^o, p_i^o), i = 1, \ldots, 16$. Using the λ-prior method described in Section 6 with $p_\lambda^i(\rho, \theta)$ defined by (33), the discrete prior $(\rho_i^o, p_i^o), i = 1, \ldots, 16$, was determined for the likelihood $f_{SNP}(x, \rho)$. In the notation of Gallant and Tauchen (2017), the BIC determined SNP specification has parameters $L_u = 1,$, $K_z = 4$, and $I_z = 3$, with all other parameters zero. Sample size is $n = 500$; $\lambda = 10^{2.5} \approx 316$. $N = 1000$ values for p were drawn from the discrete prior (ρ_i^o, p_i^o). For each ρ drawn, data x of length $n = 500$ were drawn from $f_{SNP}(x, \rho)$. For each x, the random variable $z = Z(x, \theta)$ defined by (14) was computed. From the $N = 1000$ values of z thus computed, quantiles at probabilities 0.001 through 0.999 at increments of 0.005 were computed for each of the elements z_1, z_2, and z_3 of z. Similarly for the normal. Plotted are the z quantiles against the normal quantiles.

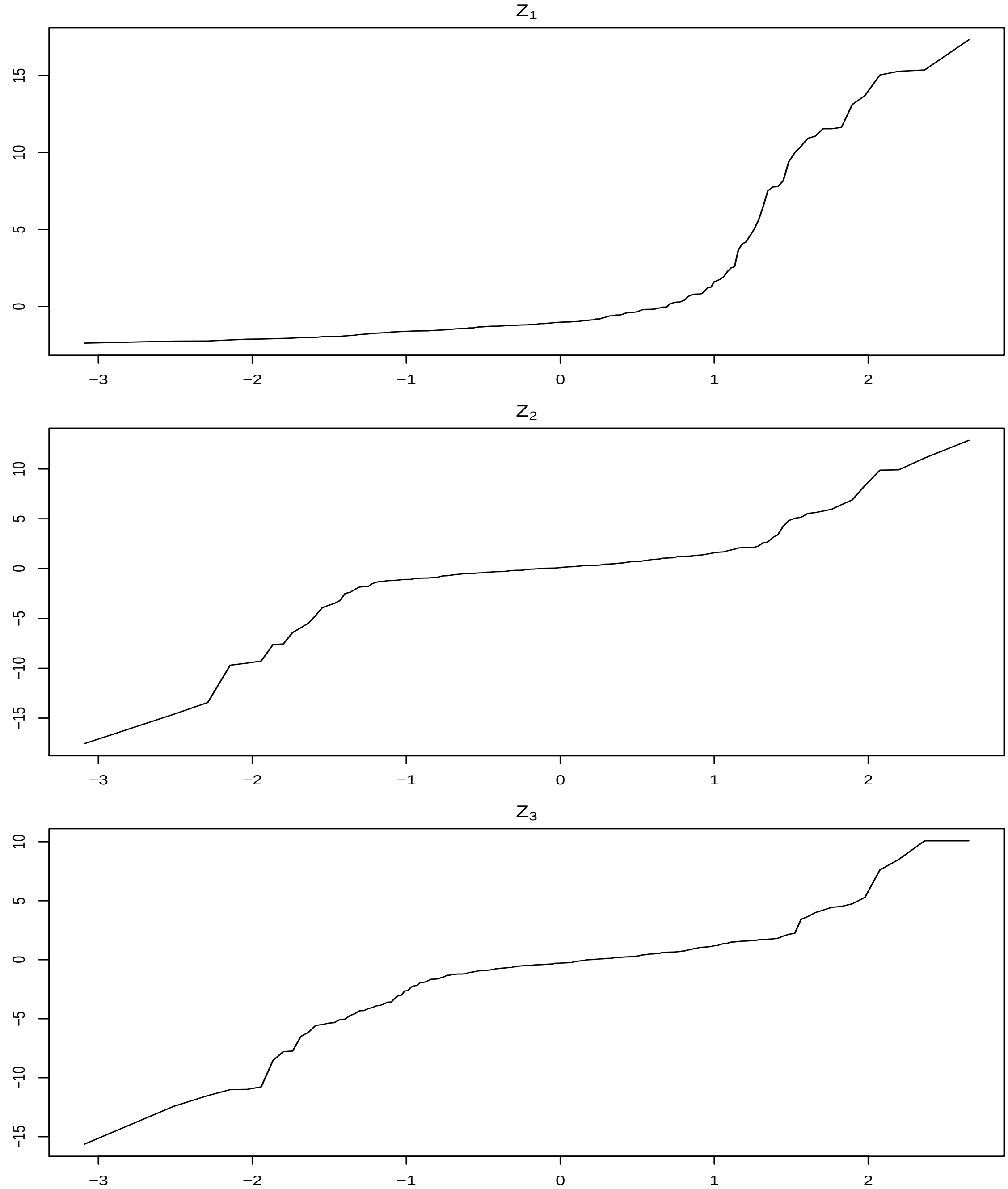

Figure A5. Complementary Q-Q Plots for Discounted Cash Flows. The prior $p^o(\theta)$ given in the legend of Table A6 for the discounted cash flows setup of Section 5 was discretized by a quadrature rule to obtain a discrete prior $(\theta_i^o, p_i^o), i = 1, \ldots, 16$. Using the λ-prior method described in Section 6 with $p_\lambda^i(\rho, \theta)$ defined by (33), the discrete prior $(\rho_i^o, p_i^o), i = 1, \ldots, 16$, was determined for the likelihood $f_{SNP}(x, \rho)$. In the notation of Gallant and Tauchen (2017), the BIC determined SNP specification has parameters $L_u = 1$, and $K_z = 4$, with all other parameters zero. Sample size is $n = 500$; $\lambda = 10^{1.25} \approx 1.25$. $N = 1000$ values for p were drawn from the discrete prior (ρ_i^o, p_i^o). For each ρ drawn, data x of length $n = 500$ were drawn from $f_{SNP}(x, \rho)$. For each x, the random variable $z = Z(x, \theta)$ defined by (14) was computed. From the $N = 1000$ values of z thus computed, quantiles at probabilities 0.001 through 0.999 at increments of 0.005 were computed for each of the elements z_1, z_2, and z_3 of z. Similarly for the normal. Plotted are the z quantiles against the normal quantiles.