Reflections on the Probability Space Induced by Moment Conditions with Implications for Bayesian Inference

A. Ronald Gallant
Penn State University

Paper: http://www.aronaldg.org/papers/reflect.pdf

Reply: http://www.aronaldg.org/papers/reply.pdf

Addendum: http://www.aronaldg.org/papers/addendum.pdf

Slides: http://www.aronaldg.org/papers/reflclr.pdf

LIML View of Bayesian Method of Moments

- **The LIML view is the most common strategy:** Use a sieve likelihood subject to (conditional) moment restrictions.

- **A good discussion of the computational issues:** Bornn, Luke, Neil Shephard, and Reza Solgi (2015) "Moment conditions and Bayesian nonparametrics," Working paper, Department of Economics, Harvard University

- **A computationally efficient implementation:** Shin, Minchul, (2015) "Bayesian GMM," Working paper, Department of Economics, University of Illinois.

- **The idea is at least 25 years old:** Gallant, A. Ronald, and George Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica* 57, 1091–1120.

- **In contrast:** I make a distributional assumption; Do not attempt complete recovery of the likelihood; Go 85 years into the past.

Motivation, Fisher (1930), 1 of 3

- Is the following correct?

- Observed variables: $x = (x_1, x_2, ..., x_n)$

  Sample mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

  Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

- If $t = \frac{\bar{x}-\theta}{s/\sqrt{n}}$ has Student's t-distribution,

- then

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \theta < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

  is a valid $(1-\alpha)\times 100\%$ credibility interval for θ.

Motivation, Fisher (1930), 2 of 3

- A line of thought leading to this construction is that an assumption of a distribution for t induces a joint distribution on the constituent random variables $(x_1, ..., x_n, \theta)$. true

- From the joint one can obtain the conditional for θ given $(x_1, ..., x_n)$. true

- And thereby make conditional probability statements on θ. false

  - Why not? Next slide.

Motivation, Fisher (1930), 3 of 3

- Although t does induce a probability space $(\mathbb{R}^n \times \Theta, \mathcal{C}, P)$,

- Where $\mathcal{C}$ is the σ-algebra of preimages of t.

- The rectangles $\mathbb{R}^n \times (a, b)$ are not in $\mathcal{C}$ and therefore cannot be assigned conditional probability.

- Hence,

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \theta < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

is **not** a valid $(1-\alpha)\times 100\%$ credibility interval for θ.

Question

- Does the probability space $(\mathcal{X} \times \Theta, \mathcal{C}, P)$ induced by a vector of moment functions $\bar{m}(x, \theta)$ permit Bayesian inference?

- Plan
  - Establish notation and state main result.
  - Look at three discrete examples to explore ideas.
  - Generalize.
  - Look at an asset pricing application.

Method of Moments, 1 of 4

- The data are $x \in \mathcal{X}$ arranged as a matrix with columns x_t.

- The parameters are $\theta \in \Theta$ regarded as random.

- Have moment functions $m(x_t, \theta)$ of dimension M with mean

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

  – There is only one θ^o for which $\mathcal{E}\bar{m}(x, \theta^o) = 0$.

  – The interesting case is when both x and θ are endogenous
    E.g., $\bar{m}(x, \theta_t) = \left(1 - \frac{1}{n} \sum_{i=1}^{n} \theta_t R_{i,t}\right)$ where $x_i = R_{i,t}$ are gross returns at time t and θ_t is the MRS at time t.

Method of Moments, 2 of 4

- Define

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

  - $W(x,\theta) = \frac{1}{n}\sum_{t=1}^{n}[m(x_t,\theta) - \bar{m}(x,\theta)]\,[m(x_t,\theta) - \bar{m}(x,\theta)]'$
  - Use HAC for W if m_t serially correlated.
  - Essential that residuals $e_t = m(x_t,\theta) - \bar{m}(x,\theta)$ be used to compute W rather than relying on $\mathcal{E}\bar{m}(x,\theta) = 0$

Method of Moments, 3 of 4

- Assert that

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

  is a likelihood and proceed directly to Bayesian inference using a prior $p^*(\theta)$.

- Amounts to a belief that $Z(x,\theta)$ is normally distributed.

  – Not essential, can assume that $Z(x,\theta)$ has a multivariate Student-t distribution or some other plausible distribution. Or, use a different $Z(x,\theta)$.

Method of Moments, 4 of 4

- The usual computational method is MCMC
  - Good reference: Gamerman and Lopes (2006)
  - In econometics: Chernozhukov and Hong (2003)
  - MCMC generates a (correlated) sample from the posterior. From it compute the posterior mean and standard deviation, which are the usual statistics used to report Bayesian results.
  - One can also compute the marginal likelihood from the chain (Newton and Raftery (1994)), which is used for Bayesian model comparison.

Likelihood Induced by Moment Functions
Assumptions

- Prior probability: Random variable Λ with realization θ in parameter space Θ, a subset of $\mathbb{R}^p$.

- Data: Random variable X with realization x in parameter space $\mathcal{X}$, a subset of $\mathbb{R}^K$.

- Structural model and prior $p^*(\theta)$ determine a probability space $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$
  - In simple cases the density of P^o is the likelihood times the prior.
  - We assume existence but not the ability to construct.

- A random variable $Z(x, \theta)$ defined on $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ with density $\psi(z)$ over $\mathbb{R}^M$

Likelihood Induced by Moment Functions
Main Result

- The random variable $Z(x, \theta)$ induces a probability space $(\mathcal{X} \times \Theta, \mathcal{C}, P)$ with $\mathcal{C} \subset \mathcal{C}^o$ and $P(C) = P^o(C)$ for $C \in \mathcal{C}$.
  - Note $\mathcal{C} = \{C = Z^{-1}(B) \,:\, B$ a Borel subset of $\mathbb{R}^M\}$
  - And $P(C) = \int_B \psi(z)\,dz$
- Specifying a prior augments $\mathcal{C}$ to include the rectangles $R_B = \mathcal{X} \times B$ and thereby obtain a probability space $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ that agrees with both P and P^o on $\mathcal{C}$.
  - $(\mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$ can be used for Bayesian inference
  - Complications arise if $Z(x, \theta)$ does not have some of the properties of a pivotal.

Induced Joint Density, Dice Example

Table 1. Tossing two correlated dice (X, Λ) when the probability of the difference $D = X - \Lambda$ is the primitive.

Preimage	d	$P(D = d)$	$P(D = d \,|\, \Lambda = 1)$	$P(D = d \,|\, \Lambda = 2)$
$C_{-5} = \{(1,6)\}$	-5	0	0	0
$C_{-4} = \{(1,5),(2,6)\}$	-4	0	0	0
$C_{-3} = \{(1,4),(2,5),(3,6)\}$	-3	0	0	0
$C_{-2} = \{(1,3),(2,4),(3,5),(4,6)\}$	-2	0	0	0
$C_{-1} = \{(1,2),(2,3),(3,4),(4,5),(5,6)\}$	-1	4/18	0	4/18
$C_{0} = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$	0	10/18	10/14	10/18
$C_{1} = \{(2,1),(3,2),(4,3),(5,4),(6,5)\}$	1	4/18	4/14	4/18
$C_{2} = \{(3,1),(4,2),(5,3),(6,4))\}$	2	0	0	0
$C_{3} = \{(4,1),(5,2),(6,3)\}$	3	0	0	0
$C_{4} = \{(5,1),(6,2)\}$	4	0	0	0
$C_{5} = \{(6,1)\}$	5	0	0	0

Conditional probability is $P(D = d \,|\, \Lambda = \theta) = P(C_d \cap O_\theta)/P(O_\theta)$, where O_θ is the union of the events that can occur. $Q(\Lambda = \theta) = P(O_\theta)$ is the marginal in the sense that $P(D = d) = \sum_{\theta=1}^{6} P(D = d \,|\, \Lambda = \theta) Q(\Lambda = \theta)$

Conditional Density, Dice Example, 1 of 2

- Let $\mathcal{C}$ be the smallest σ-algebra that contains the preimages in Table 1.
- Any $\mathcal{C}$-measurable f must be constant on the preimages.
- For such f the formula

$$\mathcal{E}(f \mid \Lambda = 2) = \sum_{x=1}^{6} f(x, 2) \sum_{d=-5}^{5} I_{C_d}(x, 2) P(D = d \mid \Lambda = 2) \tag{1}$$

  can be used to compute conditional expectation because f can be regarded as a function of d and the right hand side of (1) equals

$$\sum_{d=-5}^{5} f(d) P(D = d \mid \Lambda = 2).$$

Conditional Density, Dice Example, 2 of 2

- Equation (1) implies that we can view $P(D = d)$ as defining a conditional density function

$$P(X = x \,|\, \Lambda = \theta) = \sum_{d=-5}^{5} I_{C_d}(x, \theta) P(D = d \,|\, \Lambda = \theta) \tag{2}$$

that is a function of x as long as we only use it in connection with $\mathcal{C}$-measurable f.

- Simpler expressions: rewrite (2) as

$$P(X = x \,|\, \Lambda = \theta) = \frac{P(D = x - \theta)}{\sum_{x=1}^{6} P(D = x - \theta)}$$

- Similarly,

$$P(\Lambda = \theta \,|\, X = x) = \frac{P(D = x - \theta)}{\sum_{\theta=1}^{6} P(D = x - \theta)}$$

What if One Specifies a Prior?, Dice Example

- Then one knows the probabilities $P(R_\theta)$ of the rectangles

$$
\begin{aligned}
R_\theta &= \mathbb{D} \times \{\theta\} \\
\mathbb{D} &= \{1, 2, 3, 4, 5, 6\}
\end{aligned}
$$

- Let $\mathcal{C}^*$ be the smallest σ-algebra containing $\{C_d\}_{d=-5}^{5}$ and $\{R_\theta\}_{\theta=1}^{6}$

- The singleton sets $\{(x, \theta)\}$ are in $\mathcal{C}^*$ so joint probability P^* on $\mathcal{C}^*$ and conditional densities have their conventional definition

  - $P^*(X = x \mid \Lambda = \theta) = \frac{P^*(\{(x,\theta)\})}{P^*(R_\theta)}$

  - $P^*(\Lambda = \theta \mid X = x) = \frac{P^*(\{(x,\theta)\})}{P^*(R_x)}$

Indeterminacy, Dice Example, 1 of 2

For $P^*(\{(x, \theta)\})$ we have nine equations in sixteen unknowns:

$$
\begin{aligned}
\frac{4}{18} &= \sum_{i=1}^{5} P^*(\{(i, i+1)\}) \\
\frac{10}{18} &= \sum_{i=1}^{6} P^*(\{(i, i)\}) \\
\frac{4}{18} &= \sum_{i=1}^{5} P^*(\{(i+1, i)\}) \\
\frac{1}{6} &= P^*(\{(1,1)\}) + P^*(\{(2,1)\}) \\
\frac{1}{6} &= P^*(\{(1,2)\}) + P^*(\{(2,2)\}) + P^*(\{(3,2)\}) \\
\frac{1}{6} &= P^*(\{(2,3)\}) + P^*(\{(3,3)\}) + P^*(\{(4,3)\}) \\
\frac{1}{6} &= P^*(\{(3,4)\}) + P^*(\{(4,4)\}) + P^*(\{(5,4)\}) \\
\frac{1}{6} &= P^*(\{(4,5)\}) + P^*(\{(5,5)\}) + P^*(\{(6,5)\}) \\
\frac{1}{6} &= P^*(\{(5,6)\}) + P^*(\{(6,6)\})
\end{aligned}
$$

There is one linear dependency leaving eight equations in sixteen unknowns.

Indeterminacy, Dice Example, 2 of 2

- The fact that for $P^*(\{(x,\theta)\})$ we have only eight equations in sixteen unknowns is fatal.

- We have no logical basis for choosing a particular solution.

- In this instance we can set the prior $0 \leq P^*(\{(x,\theta)\}) \leq 1/6$ on eight of the probabilities, estimate them along with θ, solving for the remaining eight. Actually works well as to estimating θ but it is not an attractive general solution.

A Second Example, Mimics Fisher (1930), 1 of 2

$$
\begin{aligned}
P[Z(X,\Lambda) = z] &= \frac{1-p}{1+p} p^{|z|} \\
Z(X,\Lambda) &= X - \Lambda \\
X &\in \mathbb{N} \\
\Lambda &\in \mathbb{N} \\
\mathbb{N} &= \{0, \pm 1, \pm 2, \ldots\}
\end{aligned}
$$

- The preimages of $Z(x,\theta)$ are

$$C_z = \{(x,\theta) : x = z + \theta,\, \theta \in \mathbb{N}\} \qquad z \in \mathbb{N}$$

which lie on 45 degree lines in the (x,θ) plane.

- Given θ, for every $z \in \mathbb{N}$ there is an $x \in \mathbb{N}$ with $(x,\theta) \in C_z$ so every C_z can occur. Therefore $O_\theta = \cup_{z\in\mathbb{N}} C_z$ and $P(O_\theta) = 1$ for every $\theta \in \mathbb{N}$.

A Second Example, Mimics Fisher (1930), 2 of 2

- If $P(O_\theta) = 1$ for every $\theta \in \mathbb{N}$.

- Then
$$P(Z = z \,|\, \Lambda = \theta) = \frac{P(C_z \cap O_\theta)}{P(O_\theta)} = P(C_z) = \frac{1-p}{1+p} p^{|z|},$$
which does not depend on θ.

- Consequently,
$$P(X = x \,|\, \Lambda = \theta) = P(Z = x - \theta)$$

- Provides a rationale for choosing a solution: The conditional probability of X given Λ is the same under P^*_θ and P_θ.
$$\begin{aligned} P^*(X = x \,|\, \Lambda = \theta) &= P(Z = x - \theta) \\ P^*(X = x, \Lambda = \theta) &= P(Z = x - \theta)\, P^*(R_\theta) \end{aligned}$$

One Problem Remains

- Densities used in MCMC and other Bayesian computational methods are presumed to be with respect to Lebesgue or counting measure.

- The dominating measure for the densities constructed as in the previous slide may not be Lebesgue or counting.

- The root cause of the problem is that for given θ and z there may be more than one x that satisfies $Z(x, \theta) = z$.

- Discrete case addressed next, continuous case addressed in http://www.aronaldg.org/papers/reply.pdf

A Third Example, Sims (2015), 1 of 5

Table 2. Preimages and Probabilities for $Z(x, \theta)$

Preimage	z	$P(Z = z)$	$P(Z = z \mid \Lambda = \theta)$		
			$\theta = 1$	$\theta = 2$	$\theta = 3$
$C_1 = \{(1,1),(3,3),(4,2),(4,3)\}$	1	Ψ_1	Ψ_1	Ψ_1	Ψ_1
$C_2 = \{(1,2),(2,1),(2,3),(3,2),(4,1)\}$	2	Ψ_2	Ψ_2	Ψ_2	Ψ_2
$C_3 = \{(1,3),(2,2),(3,1)\}$	3	Ψ_3	Ψ_3	Ψ_3	Ψ_3

The sets that can occur when it is known that $\Lambda = \theta$ has occurred are those preimages C_z that contain (x, θ) for some x in the support $\mathcal{X}$ of X. Let O_θ be the union of the sets that can occur when it is known that $\Lambda = \theta$ has occurred. Conditional probability is computed as $P(Z = z \mid \Lambda = \theta) = P(C_z \cap O_\theta)/P(O_\theta)$. In this instance, O_θ is the support Θ of Λ so that $P(C_z \cap O_\theta) = \Psi_z$ and $P(O_\theta) = 1$.

A Third Example, Sims (2015), 2 of 5

Table 3. Conditional Probabilities Implied by Table 2

	$P(X = x \mid \Lambda = \theta)$		
x	$\theta = 1$	$\theta = 2$	$\theta = 3$
1	Ψ_1	Ψ_2	Ψ_3
2	Ψ_2	Ψ_3	Ψ_2
3	Ψ_3	Ψ_2	Ψ_1
4	Ψ_2	Ψ_1	Ψ_1

$P(X = x \mid \Lambda = \theta)$ is the probability of the preimage in Table 2 that contains (x, θ), which is C_1 for $P(X = 1 \mid \Lambda = 1)$, divided by the probability of the union of all sets that contain a point of the form $(\cdot, \theta)$, which is $O_1 = C_1 \cup C_2 \cup C_3$ for $P(X = 1 \mid \Lambda = 1)$. Therefore $P(X = 1 \mid \Lambda = 1) = P(C_1)/P(O_1) = \Psi_1/1$. The column probabilities do not sum to one, which is an issue that the adjustment $\text{adj}(x, \theta)$ resolves.

A Third Example, Sims (2015), 3 of 5

Table 4. The Sets $C^{(\theta,z)}$ Implied by Table 2

z	$\theta = 1$	$\theta = 2$	$\theta = 3$
1	$\{1\}$	$\{4\}$	$\{3,4\}$
2	$\{2,4\}$	$\{1,3\}$	$\{2\}$
3	$\{3\}$	$\{2\}$	$\{1\}$

$$C^{(\theta,z)} = \{x \in \mathcal{X} : Z(x,\theta) = z\} \tag{3}$$

A Third Example, Sims (2015), 4 of 5

Table 5. Dominating Measure for Table 3

	adj(x,θ)		
x	$\theta=1$	$\theta=2$	$\theta=3$
1	1	$\frac{1}{2}$	1
2	$\frac{1}{2}$	1	1
3	1	$\frac{1}{2}$	$\frac{1}{2}$
4	$\frac{1}{2}$	1	$\frac{1}{2}$

When $C^{(\theta,z)}$ given in Table 5 has more than one element, probability is split evenly among the points. E.g., $C^{(1,2)} = \{2,4\}$; therefore, adj$(2,1) =$ adj$(4,1) = \frac{1}{2}$ When $C^{(\theta,z)}$ is a singleton set, adj$(\theta,z) = 1$; therefore adj$(1,1) = 1$.

A Third Example, Sims (2015), 5 of 5

Table 3. Adjusted Conditional Probabilities

	$\text{adj}(x,\theta)P(X = x \mid \Lambda = \theta)$		
x	$\theta = 1$	$\theta = 2$	$\theta = 3$
1	Ψ_1	$\frac{1}{2}\Psi_2$	Ψ_3
2	$\frac{1}{2}\Psi_2$	Ψ_3	Ψ_2
3	Ψ_3	$\frac{1}{2}\Psi_2$	$\frac{1}{2}\Psi_1$
4	$\frac{1}{2}\Psi_2$	Ψ_1	$\frac{1}{2}\Psi_1$

Product of the entries in Tables 3 and 5.

Setup (repeated)

- Prior probability: Random variable Λ with realization θ in parameter space Θ, a subset of $\mathbb{R}^p$.

- Data: Random variable X with realization x in parameter space $\mathcal{X}$, a subset of $\mathbb{R}^K$.

- Structural model and prior $p^*(\theta)$ determine a probability space $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$
  - In simple cases the density of P^o is the likelihood times the prior.
  - We assume existence but not the ability to construct.

- A random variable $Z(x, \theta)$ defined on $(\mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$ with density $\psi(z)$ over $\mathbb{R}^M$

Abstraction, 1 of 5

- Let $\mathcal{C}$ be the smallest σ-algebra containing the preimages

$$C = \{(x,\theta) : Z(x,\theta) \in B\}$$

  where B Borel.

- Because the distribution Ψ of $Z(X,\Lambda)$ is determined by the structural model and prior, the probability distribution P induced on $(\mathcal{X} \times \Theta, \mathcal{C})$ by Ψ can be presumed to satisfy $P(C) = P^o(C)$ for every $C \in \mathcal{C}$.

- Therefore, $(\mathcal{X} \times \Theta, \mathcal{C}, P^o) = (\mathcal{X} \times \Theta, \mathcal{C}, P)$, which implies that expectations $\mathcal{E}(f)$ are computed the same on either probability space for $\mathcal{C}$-measurable f.

Abstraction, 2 of 5

Assumption

Let

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

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

Abstraction, 3 of 5

- If $C^{(\theta,z)}$ is not empty, then for every $z \in \mathcal{Z}$,

$$C^z = \{(x,\theta)\ :\ Z(x,\theta) = z\}$$

  can occur if $\Lambda = \theta$ is known to have occurred.

- The sets C^z are in $\mathcal{C}$, they are a mutually exclusive and exhaustive partitioning of the preimage $Z^{-1}(\mathcal{Z})$, and no finer partitioning of $\mathcal{X} \times \Theta$ by sets from $\mathcal{C}$ is possible.

- Therefore the conditioning set for the event $\Lambda = \theta$ is

$$O_\theta = \cup_{z \in \mathcal{Z}} C^z = Z^{-1}(\mathcal{Z}),$$

  which implies $P(O_\theta) = P^o(O_\theta) = \Psi(\mathcal{Z}) = 1$.

Abstraction, 4 of 5

- Let $\mathcal{C}^*$ be the smallest σ-algebra that contains all sets in $\mathcal{C}$ plus all rectangles of the form $R_B = (\mathbb{R}^K \times B) \cap (\mathcal{X} \times \Theta)$, where B is a Borel subset of $\mathbb{R}^p$.

- Define a measure P^* on $\mathcal{C}^*$ by the densities

$$p^*(x \,|\, \Lambda = \theta) = \mathrm{adj}(x, \theta)\, \psi[Z(x, \theta)] \tag{4}$$

$$p^*(x, \theta) = p^*(x \,|\, \Lambda = \theta)\, p^*(\theta).$$

  - May or may not want to include this term, see Reply.

- For given θ and $\mathcal{C}$-measurable f, which must be a function of the form $f[Z(x, \theta)]$,

$$\int f[Z(x, \theta)]\, p^*(x \,|\, \Lambda = \theta)\, dx = \int_{\mathcal{Z}} f(z)\, \psi(z)\, dz, \tag{5}$$

Abstraction, 5 of 5

- Using (4) and (5) one can verify that

$$(\mathcal{X} \times \Theta, \mathcal{C}, P_\theta^o) = (\mathcal{X} \times \Theta, \mathcal{C}, P_\theta) = (\mathcal{X} \times \Theta, \mathcal{C}, P_\theta^*).$$

$$(\mathcal{X} \times \Theta, \mathcal{C}^*, P_\theta^o) = (\mathcal{X} \times \Theta, \mathcal{C}^*, P_\theta^*).$$

- For any $\mathcal{C}$-measurable f, $\mathcal{E}(f)$ will be computed the same under any of these three probability measures: P_θ^o, P_θ, or P_θ^*.

- Only $\mathcal{C}$-measurable f arise in applications.

- The probability space $(\mathcal{X} \times \Theta, \mathcal{C}^*, P_\theta^*)$ induced by the moment functions $Z(x, \theta)$ and the prior p^* is the one that is relevant for Bayesian inference.

Commentary

This is the joint density used in MCMC

$$p^*(x,\theta) = \text{adj}(x,\theta)\,\psi[Z(x,\theta)]p^*(\theta)$$

- Sample size large or $\text{adj}(x,\theta)$ does not depend on θ, then setting $\text{adj}(x,\theta) = 1$ is OK.

- Group terms thusly $\psi[Z(x,\theta)]\{\text{adj}(x,\theta)\,p^*(\theta)\}$ then omitting $\text{adj}(x,\theta)$ can be viewed as using a data dependent prior.

- Group terms thusly $\{\text{adj}(x,\theta)\,\psi[Z(x,\theta)]\}p^*(\theta)$ then can be viewed as a particular choice of likelihood subject to the restriction that $Z \sim \Psi(z)$.

- My own view is that one should set $\text{adj}(x,\theta) = 1$ regardless.

Habit Persistence Asset Pricing, 1 of 3

Driving Processes

$$\text{Consumption: } c_t - c_{t-1} = g + v_t$$

$$\text{Dividends: } d_t - d_{t-1} = g + w_t$$

$$\text{Random shocks: } \begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim \text{NID}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\sigma_w \\ \rho\sigma\sigma_w & \sigma_w^2 \end{pmatrix}\right]$$

The time increment is one year. Lower case denotes logarithms of upper case quantities; i.e. $c_t = \log(C_t)$, $d_t = \log(D_t)$. From Campbell and Cochrane (1999).

Habit Persistence Asset Pricing, 2 of 3

Utility function

$$\mathcal{E}_0 \left(\sum_{t=0}^{\infty} \delta^t \frac{(S_t C_t)^{1-\gamma} - 1}{1-\gamma} \right),$$

Habit persistence

$$\text{Surplus ratio: } s_t - \bar{s} = \phi(s_{t-1} - \bar{s}) + \lambda(s_{t-1}) v_t$$

$$\text{Sensitivity function: } \lambda(s) = \begin{cases} \frac{1}{\bar{S}}\sqrt{1 - 2(s - \bar{s})} - 1 & s_t \leq s_{\max} \\ 0 & s_t > s_{\max} \end{cases}$$

$\mathcal{E}_t$ is conditional expectation with respect to $S_t, S_{t-1}, \ldots$. Lower case denotes logarithms of upper case quantities: $s_t = \log(S_t)$. $\bar{S}$ and $s_{\max}$ can be computed from model parameters $\theta = (g, \sigma, \rho, \sigma_w, \phi, \delta, \gamma)$ as $\bar{S} = \sigma\sqrt{\frac{\gamma}{1-\phi}}$ and $s_{\max} = \bar{s} + \frac{1}{2}(1 - \bar{S}^2)$. From Campbell and Cochrane (1999).

Habit Persistence Asset Pricing, 2 of 3

Marginal Rate of Substitution

$$M_t = \delta \left(\frac{S_t C_t}{S_{t-1} C_{t-1}} \right)^{-\gamma}.$$

Euler Equations

$$\begin{aligned} 1 &= \mathcal{E}_{t-1}\left[M_t (P_{d,t} + D_t)/P_{d,t-1} \right] \\ 1 &= \mathcal{E}_{t-1}\left[M_t (P_{f,t} + I_t)/P_{f,t-1} \right] \end{aligned}$$

- $(P_{d,t} + D_t)/P_{d,t-1}$ is gross real return on a stock

- $(P_{f,t} + I_t)/P_{f,t-1}$ the same for a bond.

The Prior

The prior is

$$\pi(\theta) = \prod_{i=1}^{7} \mathrm{N}\left[\theta_i \,\middle|\, \theta_i^*, \left(\frac{\tau_i \theta_i^*}{1.96}\right)^2\right]$$

- θ_i^* are Campbell and Cochrane's calibrated values

$$\begin{aligned} \theta^* &= (g, \sigma, \sigma_w, \rho, \phi, \delta, \gamma) \\ &= (0.0189, 0.015, 0.122, 0.2, 0.87, 0.89, 2.00) \end{aligned}$$

- For, e.g., $\tau_i = 0.1$ the prior states that the marginal probability that θ_i is within 10% of θ_i^* is 95%.

- Support conditions: $-0.5 < g < 0.5$, $\sigma > 0$, $\sigma_w > 0$, $-1 < \rho < 1$, $-1 < \phi < 1$, $0.7 < \delta < 1.05$, and $1 < \gamma < 20$.

Footnotes

- Problem: ϕ and δ are not identified.
  Solution: Set τ_i for ϕ and δ to smaller of overall τ or 0.1

- Problem: Habit model breaks when confronted with the Great Depression or Great Recession.
  Solution: Set τ_i for g and σ to smaller of overall τ or 0.3

- Problem: $1 < \gamma$ truncates posterior when overall $\tau > 0.5$
  Solution: None.

- Comment: All the above problems disappear if data from Great Depression or Great Recession are excluded.

Table 3. Data Characteristics

Variable	Mean	Std. Dev.
log consumption growth	0.02183	0.01256
log dividend growth	0.02117	0.1479
ρ	0.2399	
log income growth	0.02175	0.01925
geometric stock return	0.04355	0.1736
geometric bond return	0.02044	0.02969

Data are real, annual, per capital consumption and income for the years 1950–2013 and real, annual stock and bond returns for the same years from BEA (2013) and CRSP (2013). ρ is the correlation between log consumption growth and log dividend growth.

Moment Functions

$$
\begin{aligned}
m_{1,t} &= c_t - c_{t-1} - g \\
m_{2,t} &= \sigma^2 - (c_t - c_{t-1} - g)^2 \\
m_{3,t} &= \sigma_w^2 - (d_t - d_{t-1} - g)^2 \\
m_{4,t} &= \rho - (c_t - c_{t-1} - g)(d_t - d_{t-1} - g)/(\sigma\sigma_w) \\
m_{5,t} &= 1.0 - M_t(P_{d,t} + D_t)/P_{d,t-1} \\
m_{6,t} &= 1.0 - M_t(P_{f,t} + I_t)/P_{f,t-1}) \\
m_{7,t} &= r_{d,t-1} m_{5,t} \\
m_{8,t} &= r_{f,t-1} m_{5,t} \\
m_{9,t} &= (\ell_{t-1} - \ell_{t-2}) m_{5,t} \\
m_{10,t} &= r_{d,t-1} m_{6,t} \\
m_{11,t} &= r_{f,t-1} m_{6,t} \\
m_{12,t} &= (\ell_{t-1} - \ell_{t-2}) m_{6,t}
\end{aligned}
$$

ℓ_t is the log of labor income at time t.

Table 4. Parameter Estimates for the Habit Model

	Prior Scale									
	$\tau = 0.01$		$\tau = 0.1$		$\tau = 0.5$		$\tau = 1$		$\tau = 2$	
Parameter	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
g	0.01896	0.0001040	0.02086	0.0006600	0.02196	0.001098	0.02198	0.001109	0.02205	0.001175
σ	0.01486	7.509e-05	0.01204	0.0004407	0.01093	0.0006138	0.01096	0.0006441	0.01091	0.0006507
σ_w	0.1121	0.0005665	0.1152	0.005595	0.1260	0.01581	0.1296	0.01781	0.1301	0.01884
ρ	0.2000	0.0009874	0.2002	0.01023	0.2033	0.04839	0.2065	0.08464	0.2189	0.1249
ϕ	0.8676	0.004260	0.8187	0.03066	0.8337	0.03649	0.8329	0.03685	0.8339	0.03520
δ	0.8886	0.004502	0.8742	0.02799	0.8898	0.03248	0.8873	0.03449	0.8799	0.03442
γ	2.0001	0.0150	1.9979	0.1038	2.0536	0.4894	2.3679	0.8108	3.0291	1.2303
Model Prob.	0		0.0036		0.4023		0.3345		0.2597	

Data are real, annual, per capital consumption and income for the years 1950–2013 and real, annual stock and bond returns for the same years from BEA (2013) and CRSP (2013) that are used to form the moment functions (32) through (43) with years prior to 1950 used for lags. The likelihood given by (45) is an assertion that the average of these moment functions over the data is normally distributed with variance given by a one lag HAC weighting matrix with Parzen weights (Gallant, 1987, p. 446). The prior is given by (31) with scale τ as shown in the table. It is an independence prior that states that the marginal probability is 95% that a parameter is within $\tau \times 100\,\%$ of Campbell and Cochrane's (1999) calibrated values with the exceptions of ϕ and δ which are as shown for the first two panels and 0.1 for last three panels and g and σ which are as shown for the first two panels and 0.3 for the last three panels. The columns labeled mean and standard deviation are the mean and standard deviations of an MCMC chain (Gamerman and Lopes (2006), Chernozukov and Hong, 2003) of length 100,000 collected past the point where transients have dissipated. The proposal is move-one-at-a-time random walk. Posterior model probabilities are computed using the Newton and Raftery (1994) $\hat{p}^4$ method for computing the marginal likelihood from an MCMC chain when assigning equal prior probability to each model. The software and data for this example are at `http://www.aronaldg.org/webfiles/mle`.

Comparison

- $\theta = (g, \sigma, \sigma_w, \rho, \phi, \delta, \gamma)$
- Campbell and Cochrane (1999)

$$\theta = (0.0189, 0.015, 0.122, 0.20, 0.87, 0.89, 2.00)$$

- This paper

$$\theta = (0.0220, 0.011, 0.126, 0.20, 0.83, 0.89, 2.05)$$

- Frequentist, ϕ and δ fixed

$$\theta = (0.0214, 0.011, 0.149, 0.09, 0.87, 0.89, 1.06)$$

- Aldrich and Gallant (2011) for years 1930-2008

$$\theta = (0.0200, 0.017, 0.111, 0.19, 0.86, 0.89, 1.96)$$

- Remarks:
  - No efficiency loss: Coefficients of variation of this paper and Aldrich and Gallant comparable.
  - Frequentist estimates of ρ and γ are anomalous. Aldrich and Gallant comparable.

Takeaway

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

(or something similar) may be used as a likelihood for Bayesian inference where

- $\bar{m}(x,\theta) = \frac{1}{n}\sum_{t=1}^{n} m(x_t,\theta)$
- $W(x,\theta) = \frac{1}{n}\sum_{t=1}^{n} [m(x_t,\theta) - \bar{m}(x,\theta)]\,[m(x_t,\theta) - \bar{m}(x,\theta)]'$
  - Use HAC for W if m_t serially correlated.
- Provided $Z(x,\theta) = \sqrt{n}\,[W(x,\theta)]^{-\frac{1}{2}}\,[\bar{m}(x,\theta)]$ (or something similar) satisfies an easily verified support condition.