

## Econometrics and decision theory

Gary Chamberlain\*

*Department of Economics, Harvard University, Cambridge, MA 02138, USA*

---

### Abstract

The paper considers the role of econometrics in decision making under uncertainty. This leads to a focus on predictive distributions. The decision maker's subjective distribution is only partly specified; it belongs to a set  $\mathcal{S}$  of distributions.  $\mathcal{S}$  can also be regarded as a set of plausible data-generating processes. Criteria are needed to evaluate procedures for constructing predictive distributions. We use risk robustness and minimax regret risk relative to  $\mathcal{S}$ . To obtain procedures for constructing predictive distributions, we use Bayes procedures based on parametric models with approximate prior distributions. The priors are nested, with a first stage that incorporates qualitative information such as exchangeability, and a second stage that is quite diffuse. Special points in the parameter space, such as boundary points, can be accommodated with second-stage priors that have one or more mass points but are otherwise quite diffuse. An application of these ideas is presented, motivated by an individual's consumption decision. The problem is to construct a distribution for that individual's future earnings, based on his earnings history and on a longitudinal data set that provides earnings histories for a sample of individuals. © 2000 Elsevier Science S.A. All rights reserved.

*JEL classification:* C23; C44

*Keywords:* Expected utility; Predictive distribution; Risk robustness; Minimax regret; Bayes procedure; Longitudinal data

---

---

\* Corresponding author. Tel.: (617) 495-1869; fax: (617) 495-8570.

*E-mail address:* gary\_chamberlain@harvard.edu. (G. Chamberlain)

## 1. Portfolio choice

Consider an individual making a portfolio choice at date  $T$  involving two assets. The (gross) returns at  $t$  per unit invested at  $t - 1$  are  $y_{1t}$  and  $y_{2t}$ . The individual has observed these returns from  $t = 0$  to  $t = T$ . He has also observed the values of the variables  $y_{3t}, \dots, y_{Kt}$ , which are thought to be relevant in forecasting future returns. So the information available to him when he makes his portfolio choice is  $z \equiv \{(y_{1t}, \dots, y_{Kt})\}_{t=0}^T$ . He invests one unit, divided between an amount  $a$  in asset one and  $1 - a$  in asset two, and then holds on to the portfolio until date  $H$ . Let  $w = \{(y_{1t}, y_{2t})\}_{t=T+1}^H$  and let  $h(w, a)$  denote the value of the portfolio at  $t = H$ :

$$h(w, a) = a \prod_{t=T+1}^H y_{1t} + (1 - a) \prod_{t=T+1}^H y_{2t}.$$

How should  $a$  be chosen?

The standard approach to this problem in the microeconomics of optimal behavior under uncertainty is based on maximizing expected utility. Suppose that the investor regards  $(z, w)$  as the outcome of the random variable  $(Z, W)$  with distribution  $Q$ , and that his utility function is  $u$ . Then the problem is to choose a decision rule  $d$  that maps observations  $z$  into actions  $a$ :

$$\max_d E_Q[u(h(W, d(Z)))]. \quad (1)$$

Suppose that an econometrician is advising this individual. What role might the econometrician play? Given the utility function  $u$  and the distribution  $Q$ , all that remains is to find the optimal solution to (1). The econometrician can certainly be helpful in that task, but I am more interested in the specification of the distribution  $Q$ . In particular, given a specification for part of  $Q$ , is there useful advice on how the rest of  $Q$  might be chosen?

Section 2 sets up a framework for examining this question and makes connections with the literature. The estimation of parameters is not the primary goal, and parametric models are not introduced until Section 3. Section 4 provides an application motivated by an individual's consumption decision. The problem is to construct a distribution for that individual's future earnings, based on his earnings history and on a longitudinal data set that provides the earnings histories for a sample of other individuals.

## 2. Framework

Consider an individual making a decision under uncertainty. Various systems of axioms for rational behavior imply that he should act as if he were maximizing expected utility. See Savage (1954), Anscombe and Aumann (1963), and

Ferguson (1967). Suppose that he will observe the outcome  $z$  of a random variable  $Z$  before making his choice, and that the payoff he receives given action  $a$  depends upon a random variable  $W$ , whose outcome  $w$  is not known when he makes his choice. Let  $h(w, a)$  denote this (known) payoff function. Let  $Q$  denote the distribution of  $(Z, W)$ , and let  $u$  denote the utility function. Both  $u$  and  $Q$  are taken as given at this point; they are implied by the individual's preferences, provided the preferences satisfy the rationality axioms. We shall refer to  $Q$  as the individual's subjective (or personal) distribution. Let  $m(w, a) = u(h(w, a))$  denote the utility of the payoff if  $W = w$  and  $a$  is chosen. So the individual faces the following problem:

$$\max_{d \in \mathcal{D}} E_Q[m(W, d(Z))],$$

with

$$E_Q[m(W, d(Z))] = \int m(w, d(z)) dQ(z, w).$$

The decision rule  $d$  is a mapping from  $\mathcal{Z}$ , the range of  $Z$ , to  $\mathcal{A}$ , the set of possible actions;  $\mathcal{D}$  is the set of all such mappings. (We can avoid measurability and integrability issues by taking the range of  $(Z, W)$  to be a finite set.)

Decompose the joint distribution  $Q$  into the marginal distribution  $Q_1$  for  $Z$  and the conditional distribution  $Q_2$  for  $W$  given  $Z$ :

$$E_Q[m(W, d(Z))] = \int \left[ \int m(w, d(z)) dQ_2(w|z) \right] dQ_1(z).$$

We shall refer to the conditional distribution  $Q_2(\cdot | z)$  as the conditional predictive distribution. Note that

$$\int m(w, d(z)) dQ_2(w|z) \leq \sup_{a \in \mathcal{A}} \int m(w, a) dQ_2(w|z),$$

which implies that

$$E_Q[m(W, d(Z))] \leq \int \left[ \sup_{a \in \mathcal{A}} \int m(w, a) dQ_2(w|z) \right] dQ_1(z).$$

We shall assume that the supremum of the inner integral is in fact obtained for some action  $a \in \mathcal{A}$ . Then the optimal decision rule  $d_Q$  satisfies

$$d_Q(z) = \operatorname{argmax}_{a \in \mathcal{A}} \int m(w, a) dQ_2(w|z).$$

The optimal action maximizes the conditional expectation of utility given the observation  $z$ . See Wald (1950, Chapter 5.1), Blackwell and Girshick (1954, Chapter 7.3), and Ferguson (1967, Chapter 1.8). Note that this argument

relies upon there being no restrictions on the form of the decision rule, i.e., no restrictions on the set  $\mathcal{D}$  of mappings from  $\mathcal{Z}$  to  $\mathcal{A}$ .

### 2.1. Frequentist evaluation

This framework has the virtue of being consistent with axioms of rational behavior. A different criterion is that the decision rule  $d$  should have good properties in repeated samples. Suppose that these repeated samples are independent draws  $\{(Z^{(j)}, W^{(j)})\}_{j=1}^J$ , each distributed according to  $Q^0$ . We shall refer to  $Q^0$  as the data-generating process (DGP). Define the risk function  $r$ :

$$r(Q, d) = - \int m(w, d(z)) dQ(z, w).$$

When evaluated at the DGP, the risk function provides a frequentist performance criterion that corresponds to (the negative of) long-run average utility: by the law of large numbers,

$$r(Q^0, d) = - \lim_{J \rightarrow \infty} \frac{1}{J} \sum_{j=1}^J m(W^{(j)}, d(Z^{(j)}))$$

with probability one (under  $Q^0$ ).

Since  $d_Q$  is chosen to minimize  $r(Q, d)$ , it follows immediately that  $d_Q$  is optimal in a frequentist sense if  $Q$  is the DGP. This optimality property is noteworthy because it is not a large-sample approximation; it is exact. The law of large numbers provides an interpretation of risk in terms of long-run average utility, as the number of repeated samples  $J$  tends to infinity. But the risk measure itself is based on an observation vector  $z$  of fixed dimension, corresponding to a fixed sample size.

The finite-sample optimality of  $d_Q$  suggests asking whether  $d_Q$  might be nearly optimal for some set of DGP's. We shall formalize this idea as follows: a decision rule  $d$  is  $\mathcal{S}$ - $\varepsilon$  risk robust if

$$\sup_{Q \in \mathcal{S}} \left[ r(Q, d) - \inf_{d' \in \mathcal{D}} r(Q, d') \right] < \varepsilon$$

for some  $\varepsilon > 0$ . The decision rule  $d_{Q^0}$  minimizes risk, and hence maximizes long-run average utility; it is not feasible, however, unless the DGP  $Q^0$  is known. A decision rule that is  $\mathcal{S}$ - $\varepsilon$  risk robust has risk within  $\varepsilon$  of this ideal, as long as the DGP is in the set  $\mathcal{S}$ . For small  $\varepsilon$ , such a rule is very attractive under the frequentist criterion. We shall refer to  $r(Q, d) - \inf_{d'} r(Q, d')$  as the *regret risk*.

The minimax bound for the set  $\mathcal{S}$  is

$$b(\mathcal{S}) \equiv \inf_{d \in \mathcal{D}} \sup_{Q \in \mathcal{S}} \left[ r(Q, d) - \inf_{d' \in \mathcal{D}} r(Q, d') \right].$$

An  $\mathcal{S}$ - $\varepsilon$  robust rule exists for  $\varepsilon > b(\mathcal{S})$ . This suggests searching for a minimax decision rule that attains the minimax bound.

Minimax arguments played a major role in Wald's (1950) development of statistical decision theory. Savage (1954, Chapters 9.7, 9.8) stressed the use of what we have called regret risk and argued that any motivation for the minimax principle depends upon the minimax bound being quite small. Robbins (1964) considered minimizing the maximum Bayes risk relative to a class of prior distributions in his development of the empirical Bayes approach to statistical decision problems.

## 2.2. Partial specification for $Q$

Our rational decision maker is characterized by a utility function  $u$  and a distribution  $Q$ . Suppose now that  $Q$  is only partly specified, and we are trying to provide guidance on specifying the rest of  $Q$ . The partial specification is that  $Q \in \mathcal{S}$ . The decision maker could be more specific, but that would be costly.

Consider using a decision rule  $d$  that is  $\mathcal{S}$ - $\varepsilon$  risk robust. Note that here  $\mathcal{S}$  is a set of distributions for  $(Z, W)$  that contains the decision maker's subjective distribution. In Section 2.1 on frequentist evaluation,  $\mathcal{S}$  was a set of distributions for  $(Z, W)$  that contained the DGP.

Let  $v(Q, a, z)$  denote the conditional expected utility under distribution  $Q$  of action  $a$  given observation  $z$ :

$$v(Q, a, z) = \int m(w, a) dQ_2(w|z).$$

Note that this is maximized by the action  $d_Q(z)$ . If the decision maker bore the cost of specifying a single subjective distribution  $Q$ , then his action upon observing  $z$  would be  $d_Q(z)$ , which would provide conditional expected utility of  $v(Q, d_Q(z), z)$ . We cannot claim that the proposed decision rule  $d$  provides conditional expected utility that is close to this optimum value, for this particular observation  $z$ . It is true that  $d$  comes within  $\varepsilon$  of the optimum when averaged over the distribution of the observation:

$$\int [v(Q, d_Q(z), z) - v(Q, d(z), z)] dQ_1(z) = r(Q, d) - r(Q, d_Q) < \varepsilon, \quad (2)$$

provided that  $Q \in \mathcal{S}$ . The decision maker may find this a persuasive argument for using  $d$ , if  $\varepsilon$  is sufficiently small and if the cost of fully specifying  $Q$  is sufficiently large. The weakness in the argument is that the observation  $z$  will be known when the decision is made, and so averaging over the distribution of the observation is problematic. Here the averaging is with respect to a subjective distribution, but the failure to condition on the observation is the basic limitation of a frequentist criterion.

To pursue this issue of conditioning on the observation, we shall say that the decision rule  $d$  is  $\mathcal{S}$ - $\varepsilon$  posterior robust at  $z$  if

$$\sup_{Q \in \mathcal{S}} [v(Q, d_Q(z), z) - v(Q, d(z), z)] < \varepsilon.$$

Note that posterior robustness depends upon the observation  $z$ . If posterior robustness holds for every observation  $z$ , then integrating over the distribution of  $Z$  as in (2) implies that  $d$  is  $\mathcal{S}$ - $\varepsilon$  risk robust. I am particularly interested in cases where the set  $\mathcal{S}$  of subjective distributions is so rich that posterior robustness is not attainable, and the decision maker is not necessarily willing to bear the cost of narrowing  $\mathcal{S}$  to obtain posterior robustness. Then risk robustness may have a useful role to play in decision making.

From a somewhat different perspective, an analysis based on a single decision rule that is  $\mathcal{S}$ - $\varepsilon$  risk robust may be of interest to a number of decision makers, provided that their subjective distributions  $Q$  are all in  $\mathcal{S}$ .

Good (1952) considers a set of subjective distributions ( $Q \in \mathcal{S}$  in our notation) and argues that a minimax solution is reasonable provided that only reasonable subjective distributions are entertained. Berger (1984) refers to risk robustness as procedure robustness, and he reviews the related literature on  $\Gamma$ -minimax and  $\Gamma$ -minimax regret criteria. I have found Morris (1983a, b) on parametric empirical Bayes methods to be particularly relevant. There are surveys of work on posterior robustness in Berger (1984, 1990, 1994) and in Wasserman (1992).

I have used the cost of specifying a subjective distribution as the motivation for considering a set  $\mathcal{S}$  of such distributions. But I have not included this cost in a formal theory of decision making. One possibility is that the individual has a fully specified distribution  $Q$ , and he can either purchase additional information before making his terminal decision, or make the decision without the additional information. Suppose that the additional information is the value  $w_1$  of  $W_1$  (the first component of  $W$ ), which, for some cost, can be known when the individual makes his terminal decision. He can calculate the action that maximizes expected utility conditional on  $Z = z$  and  $W_1 = w_1$ , and the optimized expected utility, net of the cost, corresponding to this action. Then the preposterior integral over  $w_1$ , based on the conditional distribution of  $W_1$  given  $Z = z$ , gives the expected utility from behaving optimally if he purchases  $W_1$ . This can be compared with the expected utility of behaving optimally without the purchase of  $W_1$ , in order to decide whether or not to acquire the additional information. This analysis fits into the standard framework of expected utility maximization, and it does not lead to a consideration of risk robustness (or posterior robustness).

Now suppose that the additional information comes from introspection, or from consulting with others (and drawing upon their introspection). We could still model this as acquiring, at some cost, the value of a component of  $W$ . If the individual has a joint distribution for  $(Z, W)$ , then this sequential decision

problem, in which the individual can ‘improve’ his subjective distribution, again fits into the standard framework. The idea here is that the individual always has a fully specified subjective distribution  $Q$ , although it may be hastily constructed. Instead of trying to work with a set  $\mathcal{S}$  of distributions, he faces the choice between making a terminal decision based on this  $Q$  or of acquiring additional information. This choice fits into the standard framework of expected utility maximization. It does, however, require that the individual do the preposterior analysis, which requires specifying the possible outcomes of the additional introspection, maximizing conditional on these outcomes (and on  $z$ ), and then integrating over the resulting maximized expected utilities. This calculation may itself be costly, in which case the individual may be attracted to a risk robust decision rule.

Manski (1981) adopted a different formulation for including the cost of specifying a subjective distribution. The decision maker does not have a completely specified distribution  $Q$  for  $(Z, W)$ , but only assigns probabilities to the sets in some partition of the range of  $(Z, W)$ . At some cost, he can refine the partition and assign probabilities to additional sets. Manski argues (p. 63): “It seems useless in this context to seek an optimal solution to the decision maker’s problem. As Simon (1957) properly points out, when the process of solving an idealized optimization problem is costly, the process of solving the respecified optimization problem which makes these costs explicit will generally be even more costly”.

### 3. Parametric models

Suppose that the subjective distribution  $Q$  has a mixture form

$$Q(A \times B) = \int_{\Theta} P_{\theta}(A \times B) d\pi(\theta),$$

where  $\pi$  is a (prior) probability distribution on the parameter space  $\Theta$ . The probability distribution  $P_{\theta}$  can be decomposed into a marginal distribution  $F_{\theta}$  for  $Z$  and a conditional distribution  $G_{\theta}$  for  $W$  given  $Z$ :

$$P_{\theta}(A \times B) = \int_A G_{\theta}(B | z) dF_{\theta}(z).$$

We shall assume that  $F_{\theta}$  has density  $f(z | \theta)$  with respect to the measure  $\lambda$ :

$$F_{\theta}(A) = \int_A f(z | \theta) d\lambda(z)$$

for all  $\theta \in \Theta$ . Then we have

$$\begin{aligned} Q(A, B) &= \int_{\Theta} \left[ \int_A G_{\theta}(B|z) f(z|\theta) d\lambda(z) \right] d\pi(\theta) \\ &= \int_A \left[ \int_{\Theta} G_{\theta}(B|z) d\bar{\pi}(\theta|z) \right] q_1(z) d\lambda(z), \end{aligned}$$

where  $\bar{\pi}$  denotes the (posterior) distribution of  $\theta$  conditional on  $Z$ :

$$\bar{\pi}(C|z) = [q_1(z)]^{-1} \int_C f(z|\theta) d\pi(\theta),$$

and  $q_1(z) = \int_{\Theta} f(z|\theta) d\pi(\theta)$  is the density of  $Q_1$  with respect to  $\lambda$ . Hence, the conditional predictive distribution for  $W$  is

$$Q_2(B|z) = \int_{\Theta} G_{\theta}(B|z) d\bar{\pi}(\theta|z).$$

### 3.1. Nested prior

Suppose that the prior distribution  $\pi$  is itself a mixture:

$$\pi(C) = \int_{\Psi} \rho_{\psi}(C) d\phi(\psi).$$

A prior distribution with this nested form is sometimes referred to as a ‘hierarchical prior’. We can regard  $\phi$  as a prior distribution on  $\Psi$ , and the parameter  $\psi$  is sometimes referred to as a ‘hyperparameter’. This nested form for  $\pi$  implies an alternative mixture representation for  $Q$ :

$$Q(A \times B) = \int_{\Psi} P_{\psi}^*(A \times B) d\phi(\psi),$$

where

$$P_{\psi}^*(A \times B) = \int_{\Theta} P_{\theta}(A \times B) d\rho_{\psi}(\theta).$$

The probability distribution  $P_{\psi}^*$  can be decomposed into a marginal distribution  $F_{\psi}^*$  for  $Z$  and a conditional distribution  $G_{\psi}^*$  for  $W$  given  $Z$ :

$$P_{\psi}^*(A \times B) = \int_A G_{\psi}^*(B|z) dF_{\psi}^*(z),$$



where

$$G_{\psi}^*(B|z) = \int_{\Theta} G_{\theta}(B|z) d\bar{\rho}_{\psi}(\theta|z),$$

$$F_{\psi}^*(A) = \int_{\Theta} F_{\theta}(A) d\rho_{\psi}(\theta),$$

$\bar{\rho}$  is the posterior distribution of  $\theta$  conditional on  $\psi$ :

$$\bar{\rho}_{\psi}(C|z) = [f^*(z|\psi)]^{-1} \int_C f(z|\theta) d\rho_{\psi}(\theta),$$

and  $f^*(z|\psi)$  is the density of  $F_{\psi}^*$  with respect to the measure  $\lambda$ :

$$f^*(z|\psi) = \int_{\Theta} f(z|\theta) d\rho_{\psi}(\theta).$$

Let  $\bar{\phi}$  denote the posterior distribution of  $\psi$  conditional on  $Z$ :

$$\bar{\phi}(C|z) = \int_C f^*(z|\psi) d\phi(\psi) / \int_{\Psi} f^*(z|\psi) d\phi(\psi).$$

Then the conditional predictive distribution for  $W$  can be expressed as

$$Q_2(B|z) = \int_{\Psi} G_{\psi}^*(B|z) d\bar{\phi}(\psi|z).$$

We can refer to  $f(z|\theta)$  as the likelihood function with  $\pi$  as the prior distribution. Or we can refer to  $f^*(z|\psi)$  as the likelihood function with  $\phi$  as the prior distribution. The likelihood-prior distinction is flexible.

### 3.2. Partial specification for $Q$

Now suppose that  $Q$  is only partly specified:  $Q \in \mathcal{S}$ . One possibility is that  $\{P_{\psi}^*: \psi \in \Psi\}$  is given, but there is a set  $\Gamma$  of possible prior distributions on  $\Psi$ . Then the set of subjective distributions for the observables  $(Z, W)$  is  $\mathcal{S} = \{Q^{\phi}: \phi \in \Gamma\}$ , where  $Q^{\phi} = \int P_{\psi}^* d\phi(\psi)$ . The risk robustness criterion can be written as

$$\sup_{\phi \in \Gamma} \left[ r(Q^{\phi}, d) - \inf_{d' \in \mathcal{D}} r(Q^{\phi}, d') \right] < \varepsilon.$$

Our goal is a decision rule  $d$  that is  $\mathcal{S}$ - $\varepsilon$  risk robust. We shall focus on specifying an approximate prior  $\phi_a$  such that the corresponding Bayes rule  $d^a = d_{Q^{\phi_a}}$  is risk robust. The approximate prior need not be an element of  $\Gamma$ . One strategy is to try a uniform prior as the approximate prior and then check

robustness. If the sample observation  $z$  is sufficiently informative relative to the actual prior  $\phi$ , then the posterior distribution for  $\psi$  based on the approximate, uniform prior will be similar to the posterior distribution based on the actual prior. This corresponds to the case of stable estimation in Edwards et al. (1963): “To ignore the departures from uniformity, it suffices that your actual prior density change gently in the region favored by the data and not itself too strongly favor some other region”. If  $Q$  assigns high probability to such informative samples, then the risk  $r(Q, d^a)$  of our procedure will be close to the optimum  $r(Q, d_Q)$ . If this holds for all of the distributions  $Q$  in  $\mathcal{S}$ , then the rule  $d^a$  corresponding to the uniform prior  $\phi_a$  is risk robust.

Special points in the parameter space, such as boundary points, can be accommodated with an approximate prior that has one or more mass points but is otherwise quite diffuse. There is an example of this in Section 4.

### 3.3. Loss function

We have focused on problems where the optimal action requires a predictive distribution. The estimation of parameters is not the primary goal; the role of the parametric model is to aid in the construction of the conditional predictive distribution for  $W$ .

We can, however, define a loss function with the parameter as one of its arguments, and then express the risk function as expected loss. This loss function  $L: \Psi \times \mathcal{Z} \times \mathcal{A} \rightarrow \mathcal{R}$  is defined as follows:

$$L(\psi, z, a) = - \int m(w, a) dG_{\psi}^*(w | z). \quad (3)$$

Note that  $L$  depends upon the observation  $z$  as well as on the parameter  $\psi$  and the action  $a$ ; this is necessary in order to include prediction problems. Then with  $Q^\phi = \int P_{\psi}^* d\phi(\psi)$ , we have

$$r(P_{\psi}^*, d) = \int L(\psi, z, d(z)) f^*(z | \psi) d\lambda(z)$$

$$r(Q^\phi, d) = \int r(P_{\psi}^*, d) d\phi(\psi).$$

The optimal decision rule can be expressed as

$$d_{Q^\phi}(z) = \operatorname{argmin}_{a \in \mathcal{A}} \int L(\psi, z, a) d\bar{\phi}(\psi | z).$$

So the optimal action is chosen to minimize posterior expected loss;  $d_{Q^\phi}$  is the Bayes rule with respect to the prior distribution  $\phi$ . See Wald (1950, Chapter 5.1), Blackwell and Girshick (1954, Chapter 7.3), and Ferguson (1967, Chapter 1.8).

The risk function  $r(P_{\psi}^*, d)$  is sometimes referred to as a classical risk function, with the stress on its being a function of the parameter  $\psi$ , and  $r(Q^{\phi}, d)$  is referred to as the Bayes risk, with the stress on using the prior  $\phi$  to integrate over the parameter space. I prefer to regard them both as functions of the distribution of the observables  $(Z, W)$  (and of the decision rule  $d$ ). That  $Q^{\phi}$  has a mixture form is not a fundamental difference; the mixing distribution  $\phi$  could assign unit mass to some particular point  $\psi$ .

The finite-sample optimality of Bayes decision rules played an important role in Wald's (1950) development of statistical decision theory. With  $\mathcal{S} = \{P_{\psi}^*: \psi \in \Psi\}$ , a decision rule  $d \in \mathcal{D}$  is *admissible* if there is no other rule  $d' \in \mathcal{D}$  with  $r(P_{\psi}^*, d') \leq r(P_{\psi}^*, d)$  for all  $\psi \in \Psi$  with strict inequality for some  $\psi \in \Psi$ . Wald's complete class theorem establishes that any admissible decision rule can be obtained as a Bayes rule for some prior distribution (or sequence of prior distributions).

If the estimation of parameters is a primary goal, then we can begin with the specification of a loss function and use it to examine the risk properties of an estimator. For example, we could consider estimating some scalar component  $\psi_k$  with a (piecewise) linear loss function:

$$L(\psi, z, a) = \begin{cases} c_1 |\psi_k - a| & \text{if } a \leq \psi_k, \\ c_2 |\psi_k - a| & \text{otherwise,} \end{cases} \quad (4)$$

with  $c_1, c_2 > 0$ . Then the posterior expected loss is minimized by setting the estimate  $d_{Q^{\phi}}(z)$  equal to the  $c_1/(c_1 + c_2)$  quantile of the posterior distribution of  $\psi_k$  (as in Koenker and Bassett (1978, 1982), in a different context).

The linear loss function is also useful in prediction problems, if we do not want to commit to a specific utility function that is based on the economics of the problem. Let  $w_k$  be some scalar component of  $w$ ,

$$-m(w, a) = \begin{cases} c_1 |w_k - a| & \text{if } a \leq w_k, \\ c_2 |w_k - a| & \text{otherwise,} \end{cases} \quad (5)$$

and define  $L(\psi, z, a)$  as in (3), where now the dependence on  $z$  is needed. Then the posterior expected loss is minimized by setting the forecast  $d_{Q^{\phi}}(z)$  equal to the  $c_1/(c_1 + c_2)$  quantile of the conditional predictive distribution of  $W_k$  (as in Koenker and Bassett (1978, 1982), in a similar context).

### 3.4. Coverage probability

As in Section 3.2, suppose that the parametric family  $\{P_{\psi}^*: \psi \in \Psi\}$  is given, and there is a set  $\Gamma$  of prior distributions. Consider an approximate prior distribution  $\phi_a$ , with corresponding subjective distribution  $Q^{\phi_a}$  for  $(Z, W)$ . The linear loss function in (4) implies that the optimal estimate of  $\psi_k$  is the  $c_1/(c_1 + c_2)$

quantile of the posterior distribution. If the posterior distribution of  $\psi_k$  is continuous, then

$$\bar{\phi}_a(\{\psi: \psi_k \leq d^a(z)\} | z) = c_1/(c_1 + c_2),$$

where  $d^a = d_{Q^{\phi_a}}$ . If  $Q = \int P_{\psi}^* d\phi(\psi)$  is the DGP, then the corresponding frequentist coverage probability for  $d^a$  is

$$\begin{aligned} \text{cover}(Q, d^a) &= \int F_{\psi}^*(\{z: \psi_k \leq d^a(z)\}) d\phi(\psi) \\ &= \int \bar{\phi}(\{\psi: \psi_k \leq d^a(z)\} | z) dQ_1(z), \end{aligned}$$

where  $Q_1 = \int F_{\psi}^* d\phi(\psi)$ .

Likewise, the linear loss function in (5) implies that the optimal forecast of  $W_k$  is the  $c_1/(c_1 + c_2)$  quantile of the conditional predictive distribution. If this distribution is continuous, then

$$Q_2^{\phi_a}(\{w: w_k \leq d^a(z)\} | z) = c_1/(c_1 + c_2).$$

The corresponding frequentist coverage probability for  $d^a$  is

$$\begin{aligned} \text{cover}(Q, d^a) &= Q(\{(z, w): w_k \leq d^a(z)\}) \\ &= \int Q_2(\{w: w_k \leq d^a(z)\} | z) dQ_1(z). \end{aligned}$$

It follows, as in Pratt (1965), that the posterior probability and the coverage probability are equal for  $d^a$  if  $Q^{\phi_a}$  is the DGP:

$$\text{cover}(Q^{\phi_a}, d^a) = c_1/(c_1 + c_2).$$

This suggests asking whether the posterior probability and the coverage probability are nearly equal for some set of DGPs. See Morris (1983b). A corresponding robustness measure is

$$\sup_{Q \in \mathcal{S}} |\text{cover}(Q, d^a) - \text{cover}(Q, d_Q)|.$$

I prefer to not focus on this robustness measure because it is not derived from an explicit loss function. Nevertheless, a substantial difference between posterior probability and coverage probability suggests a lack of robustness. Furthermore, a useful check on our numerical work can be based on this result that posterior probability equals coverage probability for  $d_Q$  when  $Q$  is the DGP.

#### 4. Application: Dynamic models for longitudinal data

Consider an individual trying to forecast his future earnings, in order to guide savings and other decisions. We shall focus on how he might combine his personal earnings history with the data on the earnings trajectories of other individuals.

At the beginning of each period, the individual has some amount of financial wealth. He receives labor earnings, and financial wealth plus labor earnings gives cash on hand. He chooses to consume some of this and invest the rest. The return on this investment gives financial wealth at the beginning of the next period, and the process repeats. Labor income in future periods is uncertain. A decision rule specifies consumption at each date as a function of cash on hand at that date and of variables (whose values are known at that date) that are used in forming conditional distributions for future earnings. Such a rule leads to a distribution for the consumption stream, and the individual uses expected utility preferences to rank the distributions corresponding to different rules. The objective is to choose a decision rule that maximizes expected utility.

Recent work on this problem includes Skinner (1988), Caballero (1990), Deaton (1991), Hubbard et al. (1994, 1995), and Carroll (1997). These papers adopt specifications for preferences and for the conditional distribution of future earnings. They use analytical and numerical methods to solve for optimal decision rules, and then summarize properties of the optimal paths for consumption and for the stock of financial wealth.

We shall work with a simple version of this problem in order to illustrate our framework. The decision maker, denoted  $i = 1$ , has access to the earnings histories for himself and  $N - 1$  other individuals over  $T + 1$  periods:  $z = \{y_i^T\}_{i=1}^N$ , where  $y_i^T = \{y_{it}\}_{t=0}^T$  and  $y_{it}$  is the log of earnings for individual  $i$  in period  $t$ . Let  $w$  denote the decision maker's future earnings:  $w = \{y_{1t}\}_{t=T+1}^H$ . He regards  $(z, w)$  as the realization of the random variable  $(Z, W)$ , which has subjective distribution  $Q$ .

The decision maker has specified the following parametric model,  $\{P_\theta: \theta \in \Theta\}$ , for log earnings:

$$Y_{it} = \gamma Y_{i,t-1} + \alpha_i + U_{it}, \quad (6)$$

$$U_{it} | \{Y_{i0} = y_{i0}\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \quad (i = 1, \dots, N; t = 1, \dots, H), \quad (7)$$

with  $\theta = (\gamma, \alpha_1, \dots, \alpha_N, \log \sigma^2)$ . He has the following nested prior,  $\{\rho_\psi: \psi \in \Psi\}$ :

$$\alpha_i | \{Y_{i0} = y_{i0}\}_{i=1}^N \stackrel{\text{ind}}{\sim} \mathcal{N}(\tau_1 + \tau_2 y_{i0}, \sigma_v^2), \quad (8)$$

with  $\psi = (\gamma, \tau_1, \tau_2, \log \sigma^2, \lambda)$  and  $\lambda \equiv \sigma_v^2 / \sigma^2$ . Our question is whether some approximate prior  $\phi_a$  on  $\Psi$  can be recommended as being suitably robust.

Models of this sort, with extensions to allow for serial correlation in  $U_{it}$ , have been used by Hause (1977), Lillard and Willis (1978), MaCurdy (1982), Abowd and Card (1989), and others. An issue, however, is the role of the normality assumptions. In particular, MaCurdy (1982) maximized a (quasi) likelihood function that was based on normality assumptions, but he based his parameter inferences on large-sample approximations that do not require specifying a particular parametric family of distributions. Abowd and Card (1989) used a minimum-distance method that, under large-sample approximations, can provide more efficient estimates of the parameters than quasi-maximum likelihood, when the normality assumption is false.

The parameter estimates alone, however, do not provide a distribution for future earnings, and such a distribution is necessary for the consumption problem. Deaton (1991) combines estimates from MaCurdy (1982) with normality assumptions to generate distributions for future earnings. Hubbard et al. (1994) use minimum-distance methods in their parameter estimation, but make normality assumptions to provide distributions for future earnings. There has been recent work using more general models to provide conditional predictive distributions for earnings. Geweke and Keane (1996) use three-point mixtures of normal distributions. Chamberlain and Hirano (1997) allow for individual-specific persistence in earnings volatility. Hirano (1998) works with Dirichlet process mixtures of normal distributions. In future work, I would like to apply this paper's framework to these more general models. Also it would be good to pay particular attention to the risk of very low earnings, as stressed in Carroll (1997).

#### 4.1. Parameter estimation

We shall begin by examining the estimation of  $\gamma$ , using the linear loss function in (4). Then we shall see how our conclusions are affected when we focus instead on the predictive distribution of  $W$ . Our choice of DGPs is based on the sample from the Panel Study of Income Dynamics used in Chamberlain and Hirano (1997). The DGP  $Q$  assigns a gamma distribution to  $h = 1/\sigma^2$  with shape parameter equal to 5.0. The 0.1 and 0.9 quantiles for  $\sigma$  are 0.24 and 0.43. Conditional on  $\sigma^2$ , the components of  $(\gamma, \tau_1, \tau_2)$  are independent normals with variances proportional to  $\sigma^2$ . We consider three specifications for the mean of  $\gamma$ : 0.2, 0.5, 0.8. The mean of  $\tau_1$  is 0, and the mean of  $\tau_2$  is 0.25. The standard deviations for  $(\gamma, \tau_1, \tau_2)$  in the (unconditional)  $t$ -distribution are 0.22. We consider five specifications for  $\lambda$ : uniform on (0, 0.1), (0.1, 0.3), (0.3, 1), (1, 5), (5, 25). The DGP is completed by setting the distribution of  $y_{i0}$  to be normal with mean 0 and standard deviation 0.45. So there are 15 DGPs in  $\mathcal{S}$ .

##### 4.1.1. Uniform prior

Our first approximate prior is uniform: the parameter space  $\Psi = \mathcal{R}^4 \times [0, \infty)$ , and the density of  $\phi_a$  (with respect to Lebesgue measure on  $\Psi$ ) is constant. The

Table 1  
 $\max_{Q \in \mathcal{Q}} [r(Q, d^a) - r(Q, d_Q)]$

Uniform prior $N$	$T$		
	2	4	10
100	0.11	0.03	0.005
1000	0.03	0.004	0.000

loss function is

$$L(\psi, z, a) = |\gamma - a|,$$

so the optimal decision rule (estimator)  $d^a$  is the median of the posterior distribution for  $\gamma$  that is implied by  $\phi_a$ . The risk of  $d^a$  under the DGP  $Q$  is  $r(Q, d^a)$ , and we shall compare this with the lower bound,  $r(Q, d_Q)$ , which would be attained if the decision maker knew the DGP and used the median of the corresponding posterior distribution of  $\gamma$ . The risks are approximated numerically by taking 5000 independent samples from each of the DGPs, and calculating the average loss. We shall consider sample sizes of  $N = 100, 1000$ , and  $T = 2, 4, 10$ .

The risk comparison for the uniform prior is given in Table 1.

The uniform prior appears to be quite robust when  $N = 100, T = 10$  or  $N = 1000$  and  $T = 4$  or 10. There is little to be gained then from considering a more informative prior. With  $N = 100$  and  $T = 2$ , however, the risk difference of 0.11 is substantial. For each of the three values for  $E(\gamma)$ , the maximal regret risk occurs for the DGP with  $\lambda$  uniform on  $(0, 0.1)$ . This suggests that there are particular difficulties associated with the  $\lambda = 0$  boundary of the parameter space. There are also large discrepancies between posterior probability and coverage probability in this case. For example, the DGP with  $E(\gamma) = 0.5$  and  $\lambda$  uniform on  $(0, 0.1)$ , which we shall denote by  $Q_6$ , gives the coverage probabilities in Table 2. When  $N = 100$  and  $T = 2$ , the coverage probability for the posterior median is far from 0.5; it is only 0.12.

#### 4.1.2. Point-mass prior

In response to this poor performance for small values of  $\lambda$ , we shall consider a second approximate prior. It differs from the first one only in the distribution for  $\lambda$ . This distribution assigns point mass of 0.10 to  $\lambda = 0$ , and probability of 0.90 to a gamma distribution with mean 12.5 and standard deviation 11.9. (The shape parameter is 1.1) We shall refer to prior-1 as uniform and to prior-2 as point mass. The risk comparison for the point mass prior is given in Table 3.

Table 2  
Cover ( $Q_6, d^a$ )

Uniform prior $N$	$T$		
	2	4	10
100	0.12	0.20	0.35
1000	0.23	0.38	0.46

Table 3  
 $\max_{Q \in \mathcal{S}} [r(Q, d^a) - r(Q, d_Q)]$

Point-mass prior $N$	$T$		
	2	4	10
100	0.09	0.04	0.01
1000	0.05	0.01	0.002

When  $N = 100$  and  $T = 2$ , the Bayes rule for the point-mass prior does much better for the DGPs with  $\lambda$  uniform on  $(0, 0.1)$ , with maximal regret risk of 0.04 instead of 0.11. The coverage probability is also much improved: the posterior median under the point mass prior has coverage probability (under  $Q_6$ ) of 0.43 instead of 0.12. But there are tradeoffs. The maximal regret risk for the point-mass prior is 0.09, which occurs for the DGP with  $E(\gamma) = 0.8$  and  $\lambda$  uniform on  $(0.3, 1)$ . The minimax criterion favors the point mass prior, but not by very much. At the other sample sizes, the minimax criterion favors the uniform prior. These results also hold separately for each of the three specifications for  $E(\gamma)$ , with  $\mathcal{S}$  consisting of the five DGPs corresponding to the specifications for  $\lambda$ .

#### 4.1.3. 0.05 and 0.95 quantiles

Now consider the loss function in (4) with  $c_1 = 1$  and  $c_2 = 19$ . The loss for underestimating  $\gamma$  is  $|\gamma - a|$ , but the loss for overestimating  $\gamma$  is  $19 \cdot |\gamma - a|$ . So the Bayes rule sets the estimate  $a$  equal to the 0.05 quantile of the posterior distribution of  $\gamma$ . The minimax regret criterion again favors the point-mass prior when  $N = 100$  and  $T = 2$ , and favors the uniform prior at the other sample sizes. These results also hold when  $c_1 = 19$  and  $c_2 = 1$ , with the Bayes rule equal to the 0.95 quantile of the posterior distribution of  $\gamma$ .

#### 4.1.4. Conditional prior

So far we have considered three estimators: based on the uniform prior, the point-mass prior, and (for the regret calculations) the DGP prior. Numerical



quadrature has been used to calculate these estimators; Appendix A provides some detail on the calculations. I have also considered the following computationally simpler estimator, which does not require quadrature; it is developed in Appendix A. The (marginal) posterior density for  $\lambda$  is obtained in closed form, using the uniform prior. The posterior mode,  $\hat{\lambda}_{\text{mode}}$ , is calculated numerically. The posterior distribution of  $\gamma$  conditional on  $\lambda$  is a  $t$ -distribution, and it is evaluated at  $\hat{\lambda}_{\text{mode}}$ . Then the  $c_1/(c_1 + c_2)$  quantile of this posterior distribution is used as the decision rule for the linear loss function in (4). This estimator does poorly at the 0.05 and 0.95 quantiles, presumably because it does not allow for any uncertainty regarding  $\lambda$ . When  $N = 100$  and  $T = 2$ , the maximal regret risks are 0.27 and 0.71 for the loss functions with  $c_1/(c_1 + c_2) = 0.05$  and 0.95. The corresponding results for the uniform prior (not conditioning on  $\lambda$ ) are 0.21 and 0.36, and the point-mass prior gives 0.14 and 0.22.

#### 4.1.5. Fixed effects

We can examine the role of the nested prior in (8) by considering a uniform prior on the original parameter space  $\Theta = \mathcal{R}^{N+2}$ . In particular, the prior density for  $(\alpha_1, \dots, \alpha_N)$  is constant on  $\mathcal{R}^N$ , which corresponds to setting  $\lambda = \infty$ . Then the posterior mean for  $\gamma$  can be obtained from a least-squares regression that includes  $N$  individual-specific dummy variables. (The posterior mean and median coincide; see the appendix.) This is known as the fixed-effects ( $d^{\text{fe}}$ ) or within-group estimator. It is inconsistent as  $N \rightarrow \infty$  with  $T$  fixed in our autoregression model. The inconsistency is particularly interesting because this is a Bayes estimator. We would obtain very similar results using a proper prior, say uniform on  $(-10^{10}, 10^{10})^{N+2}$ . If we use this prior to construct a DGP, the (slightly modified) fixed-effects estimator will be optimal in terms of mean square error (or mean absolute error) for that DGP. It will, however, be extremely nonrobust for the (plausible) family of DGPs that we have been considering.

The risk comparisons for the fixed-effects estimator, using mean absolute error, are shown in Table 4.

$\mathcal{S}$  contains the same 15 DGPs that were used before. The maximal regret risk when  $N = 100$  and  $T = 2$  is 0.60. Using the nested (uniform) prior on  $\Psi$ , the

Table 4  
 $\max_{Q \in \mathcal{S}} [r(Q, d^{\text{fe}}) - r(Q, d_Q)]$

Fixed effects $N$	$T$		
	2	4	10
100	0.60	0.31	0.12
1000	0.64	0.33	0.13

maximal regret risk is 0.11. In fact, the fixed-effects estimator is dominated, with higher risk at each of the DGPs in  $\mathcal{S}$ .

Since the fixed-effects estimator is a Bayes estimator, a modified version using a proper prior will be admissible, and so we will not have risk dominance over a sufficiently wide set of DGPs. Nevertheless, risk comparisons can provide a critique of the fixed-effects estimator that does not rely upon large-sample arguments. To illustrate, consider adding to  $\mathcal{S}$  the DGP with  $E(\gamma) = 0.5$  and  $\lambda$  uniform on (1000,1001). The mean absolute error for estimating  $\gamma$  is about 0.004 under this DGP (when  $N = 100$  and  $T = 2$ ) for each of the following estimators: the optimal (infeasible) estimator that uses the DGP prior; the uniform, nested-prior estimator; and the fixed-effects estimator. Suppose that the risk is in fact slightly less for the fixed-effects estimator than for the nested-prior estimator, so we no longer have dominance. Then the critique of the fixed-effects estimator could be as follows: (i) the DGP for which the fixed-effects estimator does well is not plausible; (ii) even if it were, the difference across estimators in risk for that DGP is tiny, whereas the nested-prior estimator does much better than the fixed-effects estimator for some (plausible) DGPs; (iii) the maximal regret risk values are not affected by the additional DGP, so the minimax criterion still strongly favors the nested-prior estimator.

#### 4.2. Predictive distributions

Consider individual  $i = 1$ . Let  $w = \{y_{1t}\}_{t=T+1}^H$  denote his future earnings. The data available to him when he makes his decision are  $z = \{y_i^T\}_{i=1}^N$ , where  $y_i^T = \{y_{it}\}_{t=0}^T$ . He regards  $(z, w)$  as the realization of the random variable  $(Z, W)$ , which has subjective distribution  $Q$ , and he adopts the parametric model and nested prior in (6)–(8). We shall work with a linear loss function, as in (5), which gives the following prediction problem:  $\min_d r(Q, d)$  with

$$r(Q, d) = E_Q[[c_1 1(d(Z) \leq Y_{1,T+k}) + c_2 1(d(Z) > Y_{1,T+k})] \cdot |Y_{1,T+k} - d(Z)|]. \quad (9)$$

The solution, for given  $Q$ , has  $d_Q(z)$  equal to the  $c_1/(c_1 + c_2)$  quantile of the conditional distribution of  $Y_{1,T+k}$  given  $Z = z$ . Working with different values for  $c_1/(c_1 + c_2)$  will provide an indication of how well the approximate priors perform in producing predictive distributions. I would also like to evaluate their performance in explicit models of optimal consumption, but that is left for future work.

##### 4.2.1. Forecasting one period ahead

We shall use the same DGPs and approximate priors as in the parameter estimation problem. First consider absolute error loss, with  $c_1 = c_2 = 1$ .  $\mathcal{S}$  contains 15 DGPs, corresponding to the three values for  $E(\gamma)$  and the five uniform

Table 5  
 $\max_{Q \in \mathcal{Q}} [r(Q, d^a) - r(Q, d_Q)]; c_1 = 1, c_2 = 1; k = 1$

Uniform prior $N$	$T$			Point-mass prior	$T$		
	2	4	10		2	4	10
100	0.004	0.001	0.000	0.005	0.001	0.000	
1000	0.001	0.000	0.000	0.001	0.000	0.000	

distributions for  $\lambda$ . The maximal regret risks for forecasting  $Y_{1,T+1}$  are given in Table 5.

Consider  $N = 100$  and  $T = 2$ . The maximal regret risk for the uniform prior is attained for the DGP with  $E(\gamma) = 0.2$  and  $\lambda$  uniform on  $(0, 0.1)$ . There the mean absolute errors for the forecasts corresponding to the DGP prior, the uniform prior, and the point-mass prior are all close to 0.26. These are substantial, in forecasts of log earnings. The regret risk for the uniform prior, however, is only 0.004. The uniform prior appears to be very robust, with little to be gained from a more informative prior. This robustness also shows up in the coverage probability, which is 0.50.

These results for the forecasting problem are in sharp contrast to the results for estimating  $\gamma$ , where the estimate based on the uniform prior had poor performance for small values of  $\lambda$ . That poor performance motivated the point-mass prior, which appears to not be needed in the forecast problem. Nevertheless, the point-mass prior does fine. With  $N = 100$  and  $T = 2$ , its maximal regret risk is attained for the DGP with  $E(\gamma) = 0.8$  and  $\lambda$  uniform on  $(0.1, 0.3)$ . There the mean absolute errors for the forecasts corresponding to the DGP prior, the uniform prior, and the point-mass prior are all close to 0.30. The regret risk for the point-mass prior is only 0.005, and its coverage probability is 0.49.

We obtain similar results for the risk function in (9) with  $c_1 = 1$  and  $c_2 = 19$ , with the Bayes rule equal to the 0.05 quantile of the conditional predictive distribution of  $Y_{1,T+1}$ . When  $N = 100$  and  $T = 2$ , the maximal regret risk for the uniform prior is attained for the DGP with  $E(\gamma) = 0.5$  and  $\lambda$  uniform on  $(0, 0.1)$ . There the risks for the forecasts corresponding to the DGP prior, the uniform prior, and the point-mass prior are 0.71, 0.72, and 0.71. The regret risk for the uniform prior is only 0.01, and its coverage probability is 0.06. Now consider  $c_1 = 19$  and  $c_2 = 1$ , still with  $N = 100$  and  $T = 2$ . The Bayes rule equals the 0.95 quantile of the conditional predictive distribution of  $Y_{1,T+1}$ . The maximal regret risk for the uniform prior is attained for the DGP with  $E(\gamma) = 0.2$  and  $\lambda$  uniform on  $(0, 0.1)$ . There the risks for the forecasts corresponding to the DGP prior, the uniform prior, and the point-mass prior are 0.67, 0.68, and 0.67. The regret risk for the uniform prior is only 0.01, and its coverage probability is 0.94.

Both the uniform prior and the point-mass prior exhibit a substantial degree of risk robustness when we are forecasting one period ahead. Next we shall see how this robustness holds up when we forecast ten periods ahead.

#### 4.2.2. Forecasting ten periods ahead

We shall consider three sample configurations:  $(N, T) = (100, 2)$ ,  $(100, 4)$ ,  $(1000, 2)$ , and three versions of the loss function in (5):  $(c_1, c_2) = (1, 1)$ ,  $(1, 19)$ ,  $(19, 1)$ .  $\mathcal{S}$  contains 15 DGPs, corresponding to  $E(\gamma) = 0.2, 0.5, 0.8$  and to the five uniform distributions for  $\lambda$ . The maximal regret risks for forecasting  $Y_{1, T+10}$  are given in Table 6.

Consider  $(N, T) = (100, 2)$  and  $(c_1, c_2) = (1, 1)$ . The maximal regret risk for the uniform prior is attained for the DGP with  $E(\gamma) = 0.8$  and  $\lambda$  uniform on  $(0, 0.1)$ . There the mean absolute errors for the forecasts corresponding to the DGP prior, the uniform prior, and the point-mass prior are 0.71, 0.81, and 0.76. These are very substantial, in forecasts of log earnings. The regret risk for the uniform prior is much smaller at 0.09, but still substantial. The point-mass prior does better than the uniform prior for this DGP, with a regret risk of 0.05. However, the maximal regret risk is smaller for the uniform prior than for the point-mass prior: 0.09 versus 0.13. This is also true at the other sample configurations: 0.02 versus 0.07, and 0.02 versus 0.09.

There are similar results for the risk function with  $(c_1, c_2) = (1, 19)$ , with the Bayes rule equal to the 0.05 quantile of the conditional predictive distribution of  $Y_{1, T+10}$ . When  $(N, T) = (100, 2)$ , the maximal regret risk for the uniform prior is attained for the DGP with  $E(\gamma) = 0.8$  and  $\lambda$  uniform on  $(0, 0.1)$ . There the risks for the forecasts corresponding to the DGP prior, the uniform prior, and the point mass prior are 2.0, 2.5, and 2.3. The regret risk for the uniform prior is 0.49, and it is 0.28 for the point-mass prior. As before, however, the maximal regret risk for the uniform prior is less than for the point-mass prior: 0.49 versus 1.4. In fact, for all the DGPs, loss functions, and sample configurations, the risk for the uniform prior is less than for the point-mass prior except when the DGP has  $\lambda$  uniform on  $(0, 0.1)$ . So the point-mass prior cannot be generally recommended for this forecast problem.

Table 6  
 $\max_{Q \in \mathcal{S}} [r(Q, d^a) - r(Q, d_Q)]; k = 10$

Uniform prior ( $c_1, c_2$ )	$(N, T)$			Point-mass prior	$(N, T)$		
	(100, 2)	(100, 4)	(1000, 2)		(100, 2)	(100, 4)	(1000, 2)
(1, 1)	0.09	0.02	0.02		0.13	0.07	0.09
(1, 19)	0.49	0.09	0.07		1.4	0.24	0.29
(19, 1)	0.47	0.08	0.12		1.1	0.27	0.34

## 5. Conclusion

We have considered the role of econometrics in decision making under uncertainty. This has led to a stress on predictive distributions. This has important consequences because econometric procedures are often evaluated without regard to predictive distributions. For example, in dynamic models for longitudinal data, the focus is often on autoregressive and variance-component parameters, corresponding to aspects of a serial covariance matrix. Various methods (quasi-maximum likelihood, minimum distance, generalized method of moments) may have desirable properties relative to these parameters but fail to provide predictive distributions.

We have considered criteria for evaluating procedures, leading to risk robustness and minimax regret risk relative to a set  $\mathcal{S}$  of data-generating processes (DGPs). These criteria were stated without using large-sample approximations. This does not mean that approximations have no role to play. The numerical evaluation of our criteria is already nontrivial in the model of Section 4. Approximations may well be required to reduce the cost of computation in more general models. But such approximations should not be bound up in the definition of the evaluation criteria; better to state the rules of the game first, and then bring in approximations as necessary.

In order to construct predictive distributions, we have used Bayes procedures based on parametric models with approximate prior distributions. Nested priors can be used on parameter spaces of high dimension. The first stage of the prior incorporates qualitative restrictions such as exchangeability, and the second stage is quite diffuse. Special points in the parameter space, such as boundary points, can be accommodated with point-mass priors. A motivation for Bayes procedures is their finite sample optimality when based on the DGP.

In our application in Section 4, the DGPs were constructed by combining a parametric model with various distributions on the parameter space. The procedures were constructed as Bayes procedures for the same parametric model and two approximate prior distributions (uniform and point mass). Our evaluation criteria could, however, be applied in other cases. We need to specify some set  $\mathcal{S}$  of DGPs, but it need not be tied to a particular parametric model. Likewise, the procedures considered need not be limited to Bayes procedures for a parametric model. Bayes procedures are available for various nonparametric models, although the distinction is not a sharp one since we can work with parameter spaces of high dimension. Non-Bayes procedures can be considered, perhaps motivated as computationally cheaper approximations to Bayes procedures.

We have considered problems where the role of the econometrician is to provide advice to an individual on a portfolio choice or consumption decision. Related issues would arise in using data from job training experiments to advise an individual on whether he should enroll in a job training program. Or in using

data from clinical trials to advise an individual on choice of medical treatment. It may be fruitful to approach some social policy questions from the perspective of the econometrician providing the policy maker with predictive distributions, in which case our criteria for evaluating procedures may be relevant.

## Acknowledgements

An earlier version of this paper was presented at the Principles of Econometrics conference at the University of Wisconsin-Madison in May 1998. I am grateful to Charles Manski and Kenneth West for comments. Financial support was provided by the National Science Foundation.

## Appendix A. Computation

### A.1. Parameter estimation and one-period forecasts

The form of the likelihood function for  $\xi$  is

$$f(y | \mathcal{J}, \xi) = ch^{n/2} \exp[-(h/2)(y - X\beta)'(y - X\beta)],$$

where  $y$  is  $n \times 1$ ,  $X$  is  $n \times K$ ,  $\beta$  is  $K \times 1$ ,  $\xi' = (\beta', h, \lambda)$ , and  $c$  is a constant in the sense that it does not depend upon  $\xi$ .  $\mathcal{J}$  is a set of variables that we condition on throughout the analysis; it could include initial conditions in dynamic models. We use the natural conjugate prior for  $(\beta, h)$  conditional on  $\lambda$ :

$$\beta | \mathcal{J}, h, \lambda \sim \mathcal{N}(m, [hG(\lambda)]^{-1})$$

$$h | \mathcal{J}, \lambda \sim \mathcal{G}(a_1/2, 2a_2^{-1}).$$

( $\mathcal{G}(a, b)$  denotes a gamma distribution with shape parameter  $a$  and scale parameter  $b$ .) Let  $p(\beta, h | \lambda)$  denote the prior density for  $(\beta, h)$  conditional on  $\lambda$ . The posterior distribution of  $(\beta, h)$  conditional on  $\lambda$  is

$$\beta | \mathcal{J}, y, h, \lambda \sim \mathcal{N}(\bar{\beta}(\lambda), h^{-1}[X'X + G(\lambda)]^{-1}) \quad (\text{A.1})$$

$$h | \mathcal{J}, y, \lambda \sim \mathcal{G}((a_1 + n)/2, 2[q(\lambda)]^{-1}), \quad (\text{A.2})$$

where

$$\bar{\beta}(\lambda) = [X'X + G(\lambda)]^{-1}[X'y + G(\lambda)m],$$

$$q(\lambda) = a_2 + y'y + m'G(\lambda)m - \bar{\beta}(\lambda)'[X'X + G(\lambda)]\bar{\beta}(\lambda).$$

Let  $s$  be a  $K \times 1$  vector; it is a constant in that it does not depend upon  $\xi$ , but it may depend upon  $\mathcal{F}$  or  $y$ .  $s'\beta$  has a posterior  $t$ -distribution conditional on  $\lambda$ :

$$s'\beta | \mathcal{F}, y, \lambda \sim s'\bar{\beta}(\lambda) + [s'[X'X + G(\lambda)]^{-1}s]^{1/2}[q(\lambda)/(a_1 + n)]^{1/2}t(a_1 + n). \tag{A.3}$$

Suppose that, conditional on  $\xi$ , we have the following predictive distribution for a future value  $D$ :

$$D | \mathcal{F}, y, \xi \sim \mathcal{N}(s'\beta, h^{-1}).$$

Then the posterior predictive distribution for  $D$  conditional on  $\lambda$  is

$$D | \mathcal{F}, y, \lambda \sim s'\bar{\beta}(\lambda) + [s'[X'X + G(\lambda)]^{-1}s + 1]^{1/2}[q(\lambda)/(a_1 + n)]^{1/2}t(a_1 + n). \tag{A.4}$$

This result will help us to obtain posterior predictive distributions one period ahead.

The marginal likelihood for  $\lambda$  is

$$\begin{aligned} r(\lambda) &= \iint f(y | \mathcal{F}, \beta, h, \lambda) p(\beta, h | \lambda) d\beta dh \\ &= c[q(\lambda)]^{-(a_1+n)/2} [\det(X'X + G(\lambda))]^{-1/2} [\det G(\lambda)]^{1/2}. \end{aligned} \tag{A.5}$$

The prior distribution for  $\lambda$  may contain a mass point, say at  $\lambda = 0$ :

$$dp(\lambda) = \zeta d\delta_0(\lambda) + (1 - \zeta)\eta(\lambda) d\lambda,$$

with  $0 \leq \zeta \leq 1$ ,  $\int_A d\delta_0(\lambda) = 1(0 \in A)$ , and  $\int \eta(\lambda) d\lambda = 1$ . The posterior distribution for  $\lambda$  is given by

$$\begin{aligned} dp(\lambda | \mathcal{F}, y) &= \frac{r(\lambda) dp(\lambda)}{\int r(\lambda) dp(\lambda)} \\ &= \frac{\zeta r(0) d\delta_0(\lambda) + (1 - \zeta)r(\lambda)\eta(\lambda) d\lambda}{\zeta r(0) + (1 - \zeta)\int r(\lambda)\eta(\lambda) d\lambda}. \end{aligned}$$

Let  $g(t | \lambda) = \Pr(s'\beta \leq t | \mathcal{F}, y, \lambda)$ , which can be obtained from the cumulative distribution function (cdf) of a  $t$ -distribution using (A.3). Then the posterior distribution of  $s'\beta$  is obtained as follows:

$$\begin{aligned} \Pr(s'\beta \leq t | \mathcal{F}, y) &\equiv g(t) = \int g(t | \lambda) dp(\lambda | \mathcal{F}, y) \\ &= \frac{\zeta g(t | 0)r(0) + (1 - \zeta)\int g(t | \lambda)r(\lambda)\eta(\lambda) d\lambda}{\zeta r(0) + (1 - \zeta)\int r(\lambda)\eta(\lambda) d\lambda}. \end{aligned} \tag{A.6}$$

Likewise, with  $g(t|\lambda) = \Pr(D \leq t | \mathcal{J}, y, \lambda)$  obtained from the cdf of a  $t$ -distribution using (A.4), we can obtain  $g(t) = \Pr(D \leq t | \mathcal{J}, y)$  as in (A.6).

The integrals in (A.6) are calculated numerically using adaptive quadrature. Then the  $\kappa$ -quantile is obtained by numerically finding a solution to  $g(t_\kappa) - \kappa = 0$ . The program is written in Fortran 90 and uses the NAG Fortran Library.

#### A.1.1. Improper prior

Partition  $\beta' = (\beta'_1, \beta'_2)$ , where  $\beta_1$  is  $K_1 \times 1$  and  $\beta_2$  is  $K_2 \times 1$ . Suppose that the prior density for  $\beta$  conditional on  $(h, \lambda)$  is a  $\mathcal{N}(m_1, [hG_1(\lambda)]^{-1})$  density for  $\beta_1$  times a constant (on  $\mathcal{R}^{K_2}$ ) density for  $\beta_2$ . Set

$$m = \begin{pmatrix} m_1 \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} G_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then the previous results continue to hold with  $a_1 + n$  replaced by  $a_1 + n - K_2$  and with  $\det(G(\lambda))$  replaced by  $\det(G_1(\lambda))$ . If the prior distribution for  $h$  conditional on  $\lambda$  is improper, with a constant density for  $\log h$  (on  $\mathcal{R}$ ), then set  $a_1 = a_2 = 0$ .

#### A.1.2. Computational simplifications

It is helpful to simplify  $q(\lambda)$ ,  $r(\lambda)$ ,  $s'[X'X + G(\lambda)]^{-1}s$ , and  $s'\bar{\beta}(\lambda)$ , since they will be evaluated repeatedly for different values of  $\lambda$ . In our application in Section 4, (6)–(8), we have

$$X = (I_N \otimes l_T \ R), \quad G(\lambda) = \begin{pmatrix} \lambda^{-1} I_N & 0 \\ 0 & M \end{pmatrix}, \quad m = \begin{pmatrix} 0 \\ d \end{pmatrix},$$

where  $l_T$  is a  $T \times 1$  vector of ones,  $R$  is  $n \times K_2$ ,  $M$  is  $K_2 \times K_2$ , and  $d$  is  $K_2 \times 1$ . So  $n = N \cdot T$  and  $K = N + K_2$ . In particular, our application has

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \text{with} \quad y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix},$$

$$R = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix} \quad \text{with} \quad R_i = \begin{pmatrix} y_{i0} & 1 & y_{i0} \\ \vdots & \vdots & \vdots \\ y_{i,T-1} & 1 & y_{i0} \end{pmatrix},$$

with  $K_2 = 3$ . Define  $v_i = \alpha_i - \tau_1 - \tau_2 y_{i0}$ , so that  $v_i | \{Y_{i0} = y_{i0}\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$ . Then  $\beta'_1 = (v_1, \dots, v_N)$ ,  $\beta'_2 = (\gamma, \tau_1, \tau_2)$ ,  $h = 1/\sigma^2$ , and  $\lambda = \sigma_v^2/\sigma^2$ .



Let  $\bar{y} = (I_N \otimes l'_T)y/T$  and  $\bar{R} = (I_N \otimes l'_T)R/T$ . Define

$$T_y = y'y, \quad T_r = R'R, \quad T_{ry} = R'y$$

$$B_y = T\bar{y}'\bar{y}, \quad B_r = T\bar{R}'\bar{R}, \quad B_{ry} = T\bar{R}'\bar{y}.$$

Define

$$\tilde{T}_y = a_2 + T_y + d'Md, \quad \tilde{T}_r = T_r + M, \quad \tilde{T}_{ry} = T_{ry} + Md.$$

Let  $L$  denote the eigenvectors of  $\tilde{T}_r$  relative to  $B_r$ , with eigenvalues  $\kappa_1, \dots, \kappa_{K_2}$ :

$$L'\tilde{T}_rL = \text{diag}\{\kappa_1, \dots, \kappa_{K_2}\}, \quad L'B_rL = I_{K_2}.$$

Let  $\omega = T\lambda/(T\lambda + 1)$ . Let  $\tilde{T}_{ry}^* = L'\tilde{T}_{ry}$ ,  $B_{ry}^* = L'B_{ry}$ , and  $\bar{R}^* = \bar{R}L$ . Let  $s_2^* = L's_2$ , where  $s' = (s'_1, s'_2)$  and  $s_1$  is  $N \times 1$ ,  $s_2$  is  $K_2 \times 1$ . Let

$$A = \text{diag}\{(\kappa_1 - \omega)^{-1}, \dots, (\kappa_{K_2} - \omega)^{-1}\}.$$

Straightforward algebra shows that

$$q(\lambda) = (\tilde{T}_y - \omega B_y) - (\tilde{T}_{ry}^* - \omega B_{ry}^*)'A(\tilde{T}_{ry}^* - \omega B_{ry}^*), \tag{A.7}$$

$$r(\lambda) = c(T\lambda + 1)^{-N/2} \left[ \prod_{j=1}^{K_2} (\kappa_j - \omega) \right]^{-1/2} q(\lambda)^{-(a_1+n)/2}, \tag{A.8}$$

$$s'[X'X + G(\lambda)]^{-1}s = \omega s'_1 s_1/T + (s_2^* - \omega \bar{R}^* s_1)'A(s_2^* - \omega \bar{R}^* s_1), \tag{A.9}$$

$$s'\bar{\beta}(\lambda) = \omega s'_1 \bar{y} + (s_2^* - \omega \bar{R}^* s_1)'A(\tilde{T}_{ry}^* - \omega B_{ry}^*). \tag{A.10}$$

Now we can obtain  $\Pr(s'\beta \leq t | \mathcal{J}, y)$  as in (A.6), using (A.7)–(A.10) to simplify the quadrature. Setting  $s_1 = 0$  and  $s'_2 = (1, 0, 0)$  gives  $s'\beta = \gamma$ . With  $D = Y_{1, T+1}$ , we can obtain  $g(t | \lambda) = \Pr(Y_{1, T+1} \leq t | \mathcal{J}, y, \lambda)$  from (A.4) by setting  $s'_1 = (1, 0, \dots, 0)$  and  $s'_2 = (y_{1T}, 1, y_{10})$ . Then obtain  $g(t) = \Pr(Y_{1, T+1} \leq t | \mathcal{J}, y)$  as in (A.6), using (A.7)–(A.10) to simplify the quadrature.

### A.1.3. Fixed effects

The fixed-effects estimator corresponds to a uniform prior for  $(\alpha_1, \dots, \alpha_N)$ . Let

$$X = (I_N \otimes l'_T R), \quad R = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix} \quad \text{with} \quad R_i = \begin{pmatrix} y_{i0} \\ \vdots \\ y_{i, T-1} \end{pmatrix}$$

$K_2 = 1$ ,  $\beta'_1 = (\alpha_1, \dots, \alpha_N)$ ,  $\beta_2 = \gamma$ ,  $h = 1/\sigma^2$ . The prior for  $(\beta, \log h)$  has constant density on  $\mathcal{R}^{N+K_2+1}$ . Define the within-group moments:

$$W_y = T_y - B_y, \quad W_r = T_r - B_r, \quad W_{ry} = T_{ry} - B_{ry}.$$

$s'\beta$  has a posterior  $t$ -distribution:

$$s'\beta | \mathcal{J}, y \sim s'b + [s'(X'X)^{-1}s]^{1/2}[q/(NT - N - K_2)]^{1/2}t(NT - N - K_2),$$

where

$$b = (X'X)^{-1}X'y = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \bar{y} - \bar{R}b_2 \\ W_r^{-1}W_{ry} \end{pmatrix},$$

$$s'(X'X)^{-1}s = s'_1s_1/T + (s_2 - \bar{R}'s_1)'W_r^{-1}(s_2 - \bar{R}'s_1),$$

$$q = W_y - W'_{ry}W_r^{-1}W_{ry}.$$

The posterior predictive distribution of  $Y_{1,T+1}$  is also a  $t$ -distribution:

$$Y_{1,T+1} | \mathcal{J}, y \sim s'b + [s'(X'X)^{-1}s + 1]^{1/2}[q/(NT - N - K_2)]^{1/2} \times t(NT - N - K_2),$$

with  $s'_1 = (1, 0, \dots, 0)$  and  $s_2 = y_{1T}$ .

### A.2. $m$ -period forecast

Return to the model with the nested prior, as in Section 4, (6)–(8). We shall show that the posterior predictive distribution of  $Y_{1,T+m}$  is a  $t$ -distribution, conditional on  $(\gamma, \lambda)$ . Then we can obtain the marginal distribution using two-dimensional quadrature. Let

$$X = (I_N \otimes l_T \quad R), \quad R = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix} \quad \text{with } R_i = \begin{pmatrix} y_{i0} & 1 & y_{i0} \\ \vdots & \vdots & \vdots \\ y_{i,T-1} & 1 & y_{i0} \end{pmatrix}.$$

Define  $v_i = \alpha_i - \tau_1 - \tau_2 y_{i0}$ , so that  $v_i | \{Y_{i0} = y_{i0}\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$ . Let  $\beta'_1 = (v_1, \dots, v_N)$ ,  $\beta'_2 = (\gamma, \tau_1, \tau_2)$ ,  $h = 1/\sigma^2$ , and  $\lambda = \sigma_v^2/\sigma^2$ . The model implies that

$$Y_{1,T+m} | \mathcal{J}, y, \beta, h, \lambda \sim \mathcal{N}\left(\frac{1 - \gamma^m}{1 - \gamma} s'\beta + \gamma^m y_{1T}, h^{-1} \frac{1 - \gamma^{2m}}{1 - \gamma^2}\right), \quad (\text{A.11})$$

with  $s'_1 = (1, 0, \dots, 0)$  and  $s'_2 = (0, 1, y_{10})$ .

Let  $e'_1 = (0, \dots, 0)$ ,  $e'_2 = (1, 0, 0)$ , so that  $e'\beta = \gamma$ . Let  $\bar{\gamma}(\lambda) = e'\bar{\beta}(\lambda)$ . As in (A.1), we have

$$\beta | \mathcal{J}, y, h, \lambda \sim \mathcal{N}(\bar{\beta}(\lambda), h^{-1}[X'X + G(\lambda)]^{-1}).$$

Hence

$$s'\beta | \mathcal{J}, y, h, \lambda, e'\beta \sim \mathcal{N}(\mu_1(\gamma, \lambda), h^{-1}d_1(\lambda)), \quad (\text{A.12})$$

with

$$\mu_1(\gamma, \lambda) = s' \bar{\beta}(\lambda) + [c_2(\lambda)/c_3(\lambda)][\gamma - \bar{\gamma}(\lambda)],$$

$$d_1(\lambda) = c_1(\lambda) - c_2(\lambda)^2/c_3(\lambda),$$

and

$$c_1(\lambda) = s'[X'X + G(\lambda)]^{-1}s, \quad c_2(\lambda) = s'[X'X + G(\lambda)]^{-1}e,$$

$$c_3(\lambda) = e'[X'X + G(\lambda)]^{-1}e.$$

There is a simplified formula for  $c_1(\lambda)$  in (A.9), and there are similar formulas for  $c_2(\lambda)$  and  $c_3(\lambda)$ :

$$c_2(\lambda) = \omega s'_1 e_1/T + (s_2^* - \omega \bar{R}^{*'} s_1)' A(e_2^* - \omega \bar{R}^{*'} e_1),$$

$$c_3(\lambda) = \omega e'_1 e_1/T + (e_2^* - \omega \bar{R}^{*'} e_1)' A(e_2^* - \omega \bar{R}^{*'} e_1),$$

where  $e_2^* = L'e_2$ .

Eqs. (A.11) and (A.12) imply that

$$Y_{1,T+m} | \mathcal{J}, y, \gamma, h, \lambda \sim \mathcal{N}(\mu_2(\gamma, \lambda), h^{-1}d_2(\gamma, \lambda)),$$

with

$$\mu_2(\gamma, \lambda) = \frac{1 - \gamma^m}{1 - \gamma} \mu_1(\gamma, \lambda) + \gamma^m y_{1T},$$

$$d_2(\gamma, \lambda) = \left( \frac{1 - \gamma^m}{1 - \gamma} \right)^2 d_1(\lambda) + \frac{1 - \gamma^{2m}}{1 - \gamma^2}.$$

Eqs. (A.1) and (A.2) imply that

$$h | \mathcal{J}, y, \lambda, \gamma \sim \mathcal{G}((a_1 + n + 1)/2, 2[\tilde{q}(\gamma, \lambda)]^{-1}),$$

with  $\tilde{q}(\gamma, \lambda) = q(\lambda) + (\gamma - \bar{\gamma}(\lambda))^2/c_3(\lambda)$ . Hence

$$Y_{1,T+m} | \mathcal{J}, y, \gamma, \lambda \sim \mu_2(\gamma, \lambda) + [d_2(\gamma, \lambda)]^{1/2} [\tilde{q}(\gamma, \lambda)/(a_1 + n + 1)]^{1/2} \\ \times t(a_1 + n + 1).$$

The posterior distribution of  $\gamma$  conditional on  $\lambda$  is a  $t$ -distribution:

$$\gamma | \mathcal{J}, y, \lambda \sim \bar{\gamma}(\lambda) + [c_3(\lambda)]^{1/2} [q(\lambda)/(a_1 + n)]^{1/2} t(a_1 + n).$$

Let  $p(\gamma | \mathcal{J}, y, \lambda)$  denote the density function. Then we have

$$g(t | \lambda) = \Pr(Y_{1,T+m} \leq t | \mathcal{J}, y, \lambda) \\ = \int \Pr(Y_{1,T+m} \leq t | \mathcal{J}, y, \gamma, \lambda) p(\gamma | \mathcal{J}, y, \lambda) d\gamma.$$

The integral over  $\gamma$ , for a given value of  $\lambda$ , is evaluated by quadrature. Then we obtain  $g(t) = \Pr(Y_{1,T+m} \leq t \mid \mathcal{J}, y)$  as in (A.6), again using quadrature.

## References

- Abowd, J., Card, D., 1989. On the covariance structure of earnings and hours changes. *Econometrica* 57, 411–445.
- Anscombe, F., Aumann, R., 1963. A definition of subjective probability. *Annals of Mathematical Statistics* 34, 199–205.
- Berger, J., 1984. The robust Bayesian viewpoint. In: Kadane, J. (Ed.), *Robustness of Bayesian Analyses*, North-Holland, Amsterdam, pp. 63–124.
- Berger, J., 1990. Robust Bayesian Analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference* 25, 303–328.
- Berger, J., 1994. An overview of robust Bayesian analysis. *Test* 3, 5–59.
- Blackwell, D., Girshick, M., 1954. *Theory of Games and Statistical Decisions*. Wiley, New York.
- Caballero, R., 1990. Consumption puzzles and precautionary savings. *Journal of Monetary Economics* 25, 113–136.
- Carroll, C., 1997. Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics* 112, 1–55.
- Chamberlain, G., Hirano, K., 1997. Predictive distributions based on longitudinal earnings data. Unpublished manuscript, Harvard University.
- Deaton, A., 1991. Saving and Liquidity Constraints. *Econometrica* 59, 1221–1248.
- Edwards, W., Lindman, H., Savage, L.J., 1963. Bayesian statistical inference for psychological research. *Psychological Review* 70, 193–242.
- Ferguson, T., 1967. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.
- Geweke, J., Keane, M., 1996. An empirical analysis of the male income dynamics in the PSID: 1968–1989. Unpublished manuscript, University of Minnesota.
- Good, I.J., 1952. Rational decisions. *Journal of the Royal Statistical Society Series B* 14, 107–114.
- Hause, J., 1977. The covariance structure of earnings and the on-the-job training hypothesis. *Annals of Economic and Social Measurement* 6, 335–365.
- Hirano, K., 1998. Semiparametric Bayesian models for dynamic earnings data, in *Essays on the econometric analysis of panel data*. Ph.D. Thesis, Harvard University.
- Hubbard, R.G., Skinner, J., Zeldes, S., 1994. The importance of precautionary motives in explaining individual and aggregate saving. *Carnegie–Rochester Conference Series on Public Policy*, Vol. 40, pp. 59–125.
- Hubbard, R.G., Skinner, J., Zeldes, S., 1995. Precautionary saving and social insurance. *Journal of Political Economy* 103, 360–399.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- Koenker, R., Bassett, G., 1982. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43–61.
- Lillard, L., Willis, R., 1978. Dynamic aspects of earnings mobility. *Econometrica* 46, 985–1012.
- MaCurdy, T., 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics* 18, 83–114.
- Manski, C., 1981. Learning and decision making when subjective probabilities have subjective domains. *The Annals of Statistics* 9, 59–65.
- Morris, C., 1983a. Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association* 78, 47–55.

- Morris, C., 1983b. Parametric empirical Bayes confidence intervals. In: Box, G., Leonard, T., Wu, C. (Eds.), *Scientific Inference, Data Analysis, and Robustness*. Academic Press, New York, pp. 25–50.
- Pratt, J., 1965. Bayesian interpretation of standard inference statements. *Journal of the Royal Statistical Society Series B* 27, 169–192.
- Robbins, H., 1964. The empirical Bayes approach to statistical decision problems. *Annals of Mathematical Statistics* 35, 1–20.
- Savage, L.J., 1954. *The Foundations of Statistics*. Wiley, New York.
- Simon, H., 1957. *Models of Man, Social and Rational*. Wiley, New York.
- Skinner, J., 1988. Risky income, life cycle consumption, and precautionary savings. *Journal of Monetary Economics* 22, 237–255.
- Wald, A., 1950. *Statistical Decision Functions*. Wiley, New York.
- Wasserman, L., 1992. Recent methodological advances in robust Bayesian inference. In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), *Bayesian Statistics Vol. 4*. Oxford University Press, Oxford, pp. 483–490.