Lecture Notes 8: Model Uncertainty

This discussion is designed to illustrate some ideas in decision theory and to relate them to regression analysis. Good references on this material include Berger (1985), Cox and Hinkley (1974) chapter 11, and French and Rios Insua (2000). This discussion is based on Brock, Durlauf, and West (2003).

1. Basic ideas

Suppose that a policymaker must choose a policy, indexed by \( i \) from some set of possible policies, \( I \). The policymaker is assumed to have available some information set \( F \), representing data that will be used in the evaluation. Payoffs to the policy are described by \( v(i, \theta) \) where \( \theta \) represents whatever unknown quantities affect the payoff. For example, \( \theta \) may represent parameters that determine the effects of the policy. Typically, \( \theta \) will also include innovations to the economy that have not been realized at the time the policy is chosen. From the perspective of a policymaker, uncertainty about \( \theta \) is the only source of uncertainty about the payoffs of a given policy.

As is standard, in such a context, one may attach an expected payoff to each of the possible policies. Letting \( \mu(\theta|F) \) denote the probability density that describes uncertainty about \( \theta \) given available data \( F \), the expected payoff to policy \( i \) is

\[
\int v(i, \theta) \mu(\theta|F) \tag{8.1}
\]

Optimal policy evaluation optimal policy choice therefore corresponds to

\[
\max_{i \in I} \int v(i, \theta) \mu(\theta|F_t) \tag{8.2}
\]
Uncertainty over $\theta$ can in turn be reexpressed as

$$
\mu(\theta|F) = \frac{\mu(\theta,F)}{\mu(F)} = \frac{\mu(F|\theta) \mu(\theta)}{\mu(F)}
$$

(8.3)

or

$$
\mu(\theta|F) \propto \mu(F|\theta) \mu(\theta)
$$

(8.4)

where “$\propto$” means “proportional to.” This latter formulation is often known as Bayes’ rule. The key idea is that the description of uncertainty about $\theta$ given what is known, $F$, also known as the posterior density, depends on two terms: the likelihood of the data given $\theta$ and $\mu(\theta)$, the prior density describing uncertainty about $\theta$. Notice that in our interpretation, this prior density represents the uncertainty about the unknown that exists before the data $F$ are employed. We do not assume that the unknowns are necessarily random (an assumption that may not be appealing when the unknowns are parameters that characterize the economy, but is of course natural when the unknowns are shocks), uncertainty about $\theta$ is defined relative to the policymaker.

We can restrict this general example as follows. Suppose that the available data are associated with a probability measure $\mu(F|\theta)$ where the form of $\mu$ is known, but the parameter vector $\theta$ is not. The posterior density of the parameter equals the likelihood function multiplied by the prior. If, when the number of observations increases, the likelihood becomes arbitrarily high at the true parameter value, the posterior will be dominated by the maximum likelihood estimate. This idea can be made rigorous.

2. Model Uncertainty
This basic framework may be employed to allow for model uncertainty. To see this, we start as follows. Suppose that there exists a set $M$ of possible models of the economy. For each model $m \in M$, one can describe uncertainty over $\theta$ as $\mu(\theta | F, m)$. “Standard” econometric practice consists of calculating quantities of this type. Incorporating model uncertainty means that one is explicit in treating the true model as another unknown variable. From the perspective of policy evaluation, a policymaker will not want to condition decisions on a particular model unless one knows that the model is true with probability 1. Under the assumption that $\theta$ contains all unknown variables that are part of the expected payoff calculation (1.1), accounting for model uncertainty means that the expected payoff to a policy should be calculated using $\mu(\theta | F)$ rather than $\mu(\theta | F, m)$. How does one move from the one conditional probability to the other? One treats $m$ as a random variable and integrates over it, i.e.

$$
\mu(\theta | F) = \sum_{m \in M} \mu(\theta | F, m) \mu(m | F)
$$

where

$$
\mu(m | F) = \frac{\mu(F | m) \mu(m)}{\mu(F)} \propto \mu(F | m) \mu(m)
$$

This approach is known as Bayesian model averaging (BMA). A nice overview of BMA is Wasserman (2000).

3. Relation to other work

The approach we advocate to incorporating model uncertainty may be usefully contrasted with a very important recent research program initiated by Hansen and Sargent (2005) that employs robust decision theory to account for the fact that a policymaker typically does not know the true model of the economy. Their framework differs from
ours in two respects. First, they only consider “local” model uncertainty. What this means is that they consider environments in which the true model is only known up to some set of models that are within a small $\epsilon$ of each other (how to define localness of models is not something we are concerned with here.) Second, they do not work with priors on the model space, i.e. $\mu(m)$. Rather, they engage in minimax analysis, in which the least favorable model in the space of potential models is assumed to be the “true” one for purposes of policy evaluation.

Our work is motivated by the belief that model uncertainty is, in many macroeconomic contexts, global rather than local. Disagreements as to whether business cycles are better understood as generated by monetary versus real factors typically are associated with very different ways of modeling the macroeconomy. Hence, we believe that the local uncertainty assumption is probably not the most appropriate one for evaluating policies such as money growth rules; see Sims (2001) for a related critique. Notice that if one does not believe that the space of potential models is “narrow” in the sense defined by Hansen and Sargent, the minimax approach is likely to give highly unsatisfactory results. The reason is that the minimax assumption implies that policy evaluation will ignore posterior model probabilities. Hence a model with arbitrarily low posterior probability can determine the optimal policy so long it represents the “worst case” in terms of payoff calculations.

4. Relationship to the analysis of regressions

In this section, I want to discuss how a decision-theoretic perspective affects the interpretation of regressions used to inform policy decisions. I do this in the context of cross-country growth regressions. Here is a typical cross-country growth regression:

$$g_i = X_i \beta + Z_i \gamma + P_i \delta + \epsilon_i$$  \hspace{1cm} (8.7)

where $g_i$ is real per capita growth across some fixed time interval, $X_i$ is a set of regressors suggested by the Solow growth model (population growth, technological change, physical and human capital savings rates transformed in ways implied by the
model), $Z_i$ is a set of additional control variables suggested by new growth theories, $p_i$ is the policy variable of interest, and $\varepsilon_i$ is an error. The distinction between $X_i$ and $Z_i$ is important in econometric practice because while $X_i$ variables are essentially constant across empirical studies, there is no consensus on which $Z_i$ variables should be included.

What does it mean to use a regression of this type to evaluate a policy? Presumably, a policymaker wishes to compare the effects of setting a policy variable at some fixed level $\bar{p}$ to an alternative setting at $\bar{p}$. Supposing that the policymaker has a payoff function

$$V(y_i, R_i, p_i)$$

(8.8)

Where $R_i$ represents some set of characteristics of country $i$ that affect the policymaker’s assessments. One can think about the policy problem as comparing payoffs associated with the alternative policies. Supposing that the policymaker has a payoff function $V$, this means that the policy evaluation will amount to computing

$$EV(y_i, R_i, \bar{p}|D) - EV(y_i, R_i, \bar{p}|D).$$

(8.9)

In this representation, $E$ is an expected value operator and $D$ denotes all data available to the policymaker. I have written the payoff function in terms of levels of per capita output, but in fact if lagged per capita output is part of $R_i$, then this function can accommodate the case where the growth rate is the relevant argument in the payoff function for the policymaker.

From the perspective of the policy analysis, the key question is simple. How one can use regressions of the form (9.7) to inform calculations of (9.9)? This question is hardly an unusual one; indeed it is precisely this type of question that underlies the development of statistical decision theory commencing with the seminal work of Abraham Wald (1950). The relevance of eq. (9.7) for policy analysis is that it allows for the computation of the distribution of growth rates under alternative choices of the policy
variable $p_i$. These distributions matter only in how they affect the expected payoff of the policymaker.

Surprisingly, this is not how policy implications are usually drawn from growth regressions. Instead, one typically sees policy evaluations drawn as an implication of hypothesis tests made on the coefficient associated with the policy variable of interest. In the context of (9.7), this amounts to using the statistical significance of $\delta$ in (9.7) to determine whether one can recommend a change in the magnitude of $p_i$ in order to enhance growth in country $i$. A good example of this is the assessment of alternative policy variables in Barro and Sala-i-Martin (1995) chapter 12. In this survey of the empirical growth literature, the empirical evaluation of various policy variables in the growth process is virtually always related to statistical significance, usually as related to the 5% level. (Significance at the 10% level but not 5% is apparently considered to be sufficiently weak evidence that a variable can be ignored.)

There is a vast statistical literature debating the use of statistical significance levels in evaluating statistical models; much of this debate revolves around frequentist versus Bayesian approaches to statistical analysis. Our concerns are somewhat different. The question is whether the statistical significance of a variable provides much insight into calculations of (9.9). As the form of (9.9) makes clear, the general answer is no. In order for there to be such a relationship, it would be necessary for the payoff function to possess a functional form such that the implied policy recommendation would be of the form “implement the policy if the coefficient on the policy variable is statistically significant; otherwise do not implement the policy.” Suppose that the question is whether to move from $\overline{p}$ to $\overline{\overline{p}}$. Assume that the OLS estimate of $\delta$, $\hat{\delta}$, can be interpreted as it’s expected value and that the OLS variance of the parameter estimate of $\delta$ is the variance of the parameter. (There are conditions under which holds that need not concern us here; see Brock and Durlauf (2001) for discussion.) How would one find a decision-theoretic analog to the “statistical significance” rule that one should only implement a policy change if the t-statistic for the policy coefficient is greater than or equal to 2 and the sign of the policy change is the same as the coefficient estimate? Such a rule corresponds to the payoff function implicitly defined by
it will be the case that one only increases the policy variable from $\bar{p}$ to $\bar{p}$ if the t-statistic in the OLS regression is at least equal to 2 and the sign of the coefficient is positive. (The use of 2 versus some other value is immaterial.)

This is a very special case and embodies several unintuitive assumptions. First, it is necessary that the policymaker only care about the component of growth affected by the control variable, rather than the effect of the control variable on growth per se. In other words, the policymaker considers the effect of the policy in isolation from all other determinants of growth. Second, it is necessary that the policymaker only care about the mean and variance of the effect on growth of the policy. This does not seem like an obvious assumption. For example, one can imagine that political stability issues would render a policymaker more sensitive to negative growth rates than positive growth rates. Third, it is necessary that there be a 2 to 1 tradeoff between the mean and standard deviation of the growth effect in the payoff function. This is where the significance level for the t-statistic is implicitly embedded in the payoff function. The point, of course, is that there is no reason to expect any of these assumptions to hold in practice.

5. Regressions and model averaging

From the perspective of model uncertainty, the use of one regression to evaluate policy amounts to calculating $\mu(\delta|F,m)$. One may think of this computation as nothing more than eliminating the dependence of $\mu(\delta|F,m)$ on $m$ by integrating out this additional conditioning variable. Since the number of models is discrete this amounts to computing

$$\mu(\delta|F) = \sum_{m \in M} \mu(\delta | F,m) \mu(m|F),$$  \hspace{1cm} (8.11)
Using Bayes rule, this expression may be rewritten as

\[ \mu(\delta|F) \propto \sum_{m \in M} \mu(\delta|F,m) \mu(F|m) \mu(m), \]  

which provides some insight into the difference between the BMA approach and conventional practice. Rather than condition on a single \( m \) in computing the posterior density, the BMA approach takes the posterior density \( \mu(\delta|F,m) \) for each model and computes a particular weighted average. The weights assigned to each model consist of two components, \( \mu(m) \) which is the prior probability assigned to a given model and \( \mu(F|m) \) which is the posterior probability of the data given a particular model. This latter term is nothing more than the likelihood function.

One can compute the posterior mean and variance of the parameter \( \delta \) using these formulas; as originally due to Leamer (1978), these are

\[ E(\delta|F) = \sum_{m \in M} \mu(m|F) E(\delta|F,m) \]  

and

\[ \text{var}(\delta|F) = E(\delta^2|F) - \left( E(\delta|F) \right)^2 = \sum_{m \in M} \mu(m|F) \left( \text{var}(\delta|m,F) + \left( E(\delta|F,m) \right)^2 \right) - \left( E(\delta|F) \right)^2 \]  

respectively.

These formulas illustrate how model uncertainty affects a given parameter estimate. First, the posterior mean of the parameter is a weighted average of the posterior means across each model. Second, the posterior variance is the sum of two terms. The first term is a weighted average of the variances for each model. The second term reflects the variance across models of the expected value for \( \delta \); these differences reflect the fact that the models are themselves different. This second variance term captures how model uncertainties.
uncertainty increases the variance associated with a parameter estimate relative to conventional calculations; see Draper (1995) for additional discussion.

Bayesian model averaging has been applied to growth regressions by Brock and Durlauf (2001), Doppelhofer, Miller and Sala-I-Martin (2004) and Fernandez, Ley, and Steel (2001) to deal with the question of what regressors to include in a given growth regression. Over 100 potential variables have been used in different growth regressions, (Durlauf and Quah (1999)) so the question of what variables to include is quite tricky. The problem, in one sense, is that growth theories are open ended, in that one theory typically says little about the validity of another; see Brock and Durlauf (2001) for an elaborated argument.

Model averaging is by no means the only way to evaluate the robustness of a regressor. Extreme bounds analysis, due to Edward Leamer, equates robustness with invariance of some property of a regressor coefficient across models. For example, one might require the sign to be invariant or the sign and statistical significance. The latter version implicitly, in the policy context, uses a loss function of the form

$$\inf_{m \in M} \mathbb{E} \left( \delta \left( \bar{p} - \bar{p} \right) \right) \left[ \text{var} \left( \delta \left( \bar{p} - \bar{p} \right) \right)^{1/2} \right] m$$

(8.15)

which means that model uncertainty is treated differently from other types of uncertainty. Reasons why this might make sense are given in Brock, Durlauf, and West (2003) and in Hansen and Sargent (2005), among other places.
Bibliography


