Asymptotic Likelihood Based Prediction Functions

Cooley, Thomas F.

Working Paper No. 151
August 1988

ASYMPTOTIC LIKELIHOOD BASED PREDICTION FUNCTIONS

Thomas F. Cooley
University of Rochester

and

William R. Parke
University of North Carolina, Chapel Hill

Working Paper No. 151

First Draft December 1986

Revised: August 1988

We have received helpful comments from Sean Collins, Mark Feldman, Lars Hansen and two anonymous referees. The responsibility for errors is our own.

ABSTRACT

This paper develops asymptotic prediction functions that approximate the shape of the density of future observations and correct for parameter uncertainty. The functions are based on extensions to a definition of predictive likelihood originally suggested by Lauritzen and Hinkley. The prediction function is shown to possess efficiency properties based on the Kullback-Liebler measure of information loss. Examples of the application of the prediction function and the derivation of relative efficiency are shown for linear-normal models, non-normal models and ARCH models.

1. Introduction

Although prediction is often a primary goal of econometric research, problems of predictive inference have received relatively little attention in the literature. A glance at any econometrics text reveals only a few pages devoted to problems of prediction, the major concern being with problems of parametric estimation and inference. This neglect may stem from the fact that no one frequentist technique is accepted as universally appropriate for predictive inference. In practice the prediction problem is approached by a diverse collection of techniques whose properties are not always well understood. Recent papers by Fair (1980) and Brown and Mariano (1983,1984,1985) have furthered understanding of some common procedures for generating predictions, but a unified basis for evaluating them is still missing. The Bayesian viewpoint provides a consistent theory of prediction but implementation in complex problems is often difficult. Our objective in this paper is to suggest a class of likelihood based prediction functions that is widely applicable. The likelihood concept proposed has the advantage that it puts predictive inference on a consistent footing, a role similar to that played by the likelihood principle of estimation. The use of a formal definition of predictive likelihood also provides a reference point for the interpretation of existing approaches to prediction.

The properties of commonly used prediction methods have been studied in a number of papers that try to rationalize their performance in the context of models with well defined characteristics. Bianchi & Calzolari (1980) and Fair (1980) among others have studied the behavior of Monte Carlo predictors of various sorts to ascertain the contribution of different sources of uncertainty to prediction error. In a series of papers Mariano and Brown have compared the asymptotic properties of deterministic predictors, which replace structural disturbances by their expected values, with stochastic predictors based on drawings of the disturbances. The latter include straightforward Monte Carlo predictors as well as stochastic predictors based on the use of sample period residuals.

The approach taken in the current paper is in the same spirit as the research just cited. We emphasize accounting for the uncertainty due to stochastic disturbances and, particularly, the uncertainty due to the use of estimated parameters. In contrast to earlier research, however, we emphasize obtaining analytic prediction functions that approximate the entire distribution of future observations rather than focusing on the bias properties of alternative point predictors. We do this because many interesting econometric prediction problems are characterized by predictive distributions that are non-normal, and, hence, not well characterized by the mean and variance alone. Prediction functions that approximate well the distribution of future observations will be important for obtaining accurate confidence intervals or probability statements about predictions.

The basis for our approach to prediction functions is a definition of likelihood due originally to Lauritzen (1974) and Hinkley (1979). Their definition has been applied by Butler (1986) and Cooley, Parke and Chib (1987). In this paper we extend the Lauritzen-Hinkley definition in a way that permits direct application to more complex econometric problems. We also introduce the concepts of predictive consistency and first and second order predictive efficiency. These are shown to be necessary to discriminate among alternative prediction functions.

In the next section we review the prediction problem and the most commonly used predictors. The Lauritzen-Hinkley definition of predictive likelihood is presented. Our asymptotic likelihood prediction function is presented in Section 3 and a convenient form for application is derived. We introduce definitions of predictive consistency and efficiency based on the Kullback-Leibler information measure in Section 4. Section 5 shows the relationship between predictive likelihood functions and mean squared error prediction functions, while Section 6 extends the definition to cover the use of consistent, but possibly inefficient parameter estimates. Finally, the usefulness of our prediction function is illustrated in the context of regression models with non normal disturbances and autoregressive conditional heteroskedasticity (ARCH) models.

2. Prediction Functions

Suppose interest centers on predictions of a random variable y_i defined over the space Y. The m data period observations $(y_t : t=1,...,m)$ are denoted by y_d. The n future period observations that we wish to predict, $(y_t : t=m+1,...,m+n)$, are denoted by y_f. The most informative possible statement about the future is the density $f(y_f|y_d,\theta)$, where θ is a vector of true parameters contained in a paramter space Θ. Knowledge of $f(y_f|y_d,\theta)$ permits one to make a variety of point forecasts (mean, median or mode) and to construct confidence regions for predictions. Because θ is unknown, practical prediction procedures most often generate point estimates of y_f based on point estimates of θ and in some cases attempt to estimate the second moment of y_f.

We suppose that the model generating realizations of y_f can be represented as

$$y_f = g(x_f, u_f, \theta),$$

where x_f is a vector of exogenous variables, and u_f is a vector of stochastic disturbances. A predictor is defined by making specific assumptions about u_f and θ. Mariano and Brown define the <u>deterministic predictor</u> based on a consistent estimate $\hat{\theta}_d$ of θ to be

$$y_f = g(x_f, 0, \hat{\theta}_d),$$

where the error term is set equal to its expected value. An alternative to the deterministic predictor is the Monte Carlo predictor defined as

$$\underset{\sim}{y}_f = g(x_f, \underset{\sim}{u}_f, \hat{\theta}_d), \tag{2.1}$$

where the $\underset{\sim}{u}_f$ represent draws from some specified distribution of u_f and $\underset{\sim}{y}_f$ represents the corresponding set of realizations of y_f. A second form of Monte Carlo predictor that is often used (Muench et al.(1974), Fair (1980)) is defined by draws of both error terms and coefficients

$$\underset{\sim}{y}_f = g(x_f, \underset{\sim}{u}_f, \underset{\sim}{\hat{\theta}}_d), \tag{2.2}$$

where here $\underset{\sim}{\hat{\theta}}_d$ denotes drawings from the asymptotic distribution of the estimated coefficients.

Although interest typically focuses on the first and second moments of the distributions generated by (2.1) and (2.2), the entire distribution is of interest as an approximation to $f(y_f;\theta)$. Indeed, the Monte Carlo procedure described by (2.1), can be thought of as an attempt to capture the density

$$f(y_f;\hat{\theta}_d) \tag{2.3}$$

by drawing the error terms. The second Monte Carlo procedure attempts to weight the density (2.3) by drawings from the asymptotic distribution of the $\hat{\theta}_d$'s:

$$\int f(y_f;\underset{\sim}{\hat{\theta}}_d)\, e^{-1/2(\underset{\sim}{\hat{\theta}}_d-\hat{\theta}_d)^2/V(\hat{\theta}_d)} \tag{2.4}$$

The obvious drawback to (2.1) is that it ignores the uncertainty introduced by using $\hat{\theta}_d$, while (2.2) appears to take account of it, but does so in a way that is difficult to judge without reference to some theoretical standard. The procedures developed in this paper have a lot in common with Monte Carlo methods. They involve corrections to forecasting densities to account for parameter uncertainty and will typically be implemented by simulation but they are motivated theoretically in the following sections.

An alternative to the approaches just discussed is to eliminate the unknown parameters θ by the use of sufficient statistics. This is the basis of the notion of predictive likelihood that was originally suggested by Lauritzen (1974) and Hinkley (1979). The Lauritzen-Hinkley concept recognizes the central importance of $f(y_f;\theta)$ for problems of prediction, but uses sufficient statistics to eliminate the unknown parameter θ. Let S_d, S_f and S_{d+f} be sufficient reductions of Y_d, Y_f and their union respectively. Sufficiency ensures that the density $f(y_d;\theta)$ can be factored as

$$f(y_d;\theta) = f(y_d|S_d)\, f(S_d;\theta),$$

where $f(y_d|S_d)$ does not depend on θ. The Lauritzen-Hinkley definition of predictive likelihood exploits the fact that S_{d+f} is a function of S_f and S_d that does not depend on θ.

<u>Definition 1</u> (Lauritzen-Hinkley):

The predictive likelihood function is

$$\text{plik}(y_f|y_d) = f(y_f,S_d|S_{d+f}) = \frac{f(y_f;\theta)\ f(S_d;\theta)}{f(S_{d+f};\theta)} \quad .$$

This definition envisions treating $\text{plik}(y_f|y_d)$ as a likelihood function for the future observations y_f. In practical applications the plik could be used to order future values by their plausibility and to obtain confidence intervals for y_f. This definition has been applied to several econometric problems by Cooley, Parke and Chib (1987), but its applicability is limited. There are some problems for which there is no sufficient reduction of the data - probit models are one example. There are many other examples where minimal sufficient statistics exist but have unworkably complex distributions - logit models are an example. In the next section we develop an alternative definition that is applicable and easily implemented in these situations.

3. <u>Asymptotic Prediction Functions</u>

The limitations of the preceding definition of predictive likelihood are not insurmountable. First, we know that maximum likelihood estimates are asymptotically sufficient. These provide a solution to problems that do not admit sufficient statistics. Second, we can replace the (often intractable) exact distributions in the Lauritzen-Hinkley definition with asymptotic distributions. In the appendix we show that, using a series of asymptotically valid approximations, we arrive at definition in the following practical form:[1]

<u>Definition 2</u>: The asymptotic predictive likelihood function is

$$\text{plik}^a(y_f|\hat{\theta}_d) = f(y_f;\hat{\theta}_d) \cdot \exp\{\, w_1(y_f;\hat{\theta}_d) + w_2(y_f;\hat{\theta}_d) \,\}, \tag{3.1}$$

where

$$w_1(y_f;\hat{\theta}_d) = -\,\frac{1}{2}\,\nabla(y_f;\hat{\theta}_d)\; H(y_{d+f};\hat{\theta}_d)^{-1}\nabla(y_f;\hat{\theta}_d)'$$

$$w_2(y_f;\hat{\theta}_d) = \nabla(y_f;\hat{\theta}_d)\psi(\hat{\theta}_d) - \frac{1}{2}\,\text{tr}[\; H(y_f;\hat{\theta}_d)\; H(y_d;\hat{\theta}_d)^{-1}] \;,$$

$\nabla(y_f;\hat{\theta}_d)$ is the log gradient function of $f(y_f;\theta)$ evaluated at y_f and $\hat{\theta}_d$, $H(y_{d+f};\hat{\theta}_d)$ is the log Hessian of $f(y_{d+f};\theta)$, and $\psi(\hat{\theta}_d)$ is the $O(m^{-1})$ bias in the MLE $\hat{\theta}_d$.

Despite a bit of notational complexity, (3.1) is a definition that can be easily implemented for common econometric prediction problems. The first and second derivatives of the log density are not usually difficult to compute, and (3.1) can often be incorporated into a Monte Carlo simulation strategy.

The elements of (3.1) have the following intuitive justification. The first term on the right hand side of (3.1) is simply the prediction function that would obtain if we knew the correct functional form of $f(y_f;\theta)$, but substituted consistent estimates for the unknown parameters. We will refer to this as the <u>certainty equivalence</u> (CEQ) prediction function (although it

---

[1] The definition and subsequent development apply to models with independent observations. Example 3 in Section 7 covers a case where a simple extension to dependent observations is possible. A more general extension would follow along similar lines.

should be noted that the term is wishful rather than descriptive as no equivalence exists). It is, as noted in the previous section, the form one is approximating via the Monte Carlo prediction procedures extensively analyzed by Mariano and Brown.

The factor $w_1(y_f;\hat{\theta}_d)$ corrects the certainty equivalence prediction function for parameter uncertainty. It typically puts more probability in the tails of a prediction function, where the log gradient $\nabla(y_f;\hat{\theta}_d)$ is largest. Loosely, this increase in the dispersion of the prediction function relative to the CEQ density recognizes that $y_f - \hat{y}_f$ will have a greater variance than $y_f - E(y_f)$. We formalize this idea in this paper.

The two terms of $w_2(y_f;\hat{\theta}_d)$ correct for two related problems. The first adjusts for asymptotic bias of order $O(m^{-1})$ in the m.l.e., and the second adjusts for the possibility that the second derivative matrix is not constant over y_f. Both elements could be derived by simply estimating the expectation of a Taylor series approximation to $g(y_f;\theta) = \log(f(y_f;\theta))$:

$$g(y_f;\hat{\theta}_d) - g(y_f;\theta) = \nabla_f(y_f;\hat{\theta}_d)\ (\hat{\theta}_d-\theta) + \frac{1}{2}\ (\hat{\theta}_d-\theta)' H_f(y_f;\hat{\theta}_d)\ (\hat{\theta}_d-\theta).$$

The expectation of this is zero for a linear-normal model, but in general it will not be.

4. Predictive Efficiency

Having proposed a candidate prediction function we now discuss how to evaluate it. Most common methods of evaluating forecasting errors (e.g. looking at mean-squared errors) are based on the first two moments. This can only make sense to the extent that predictive densities are well approximated by normal distributions. The non-normal distribution of the forecast errors for many econometric models motivates us to adopt a measure of predictive efficiency that is sensitive to the shape of the future density as well as its moments. That measure, the Kullback-Leibler information measure (Kullback(1959)), provides a natural metric for evaluating candidate prediction functions.

In this section, we formalize the information measure of predictive efficiency and then establish four results. First, we derive the information efficiency for the CEQ technique. Second, we establish the order of the relative efficiency gain that can be secured by adjusting the functional form to account for parameter uncertainty. Third, we construct an expansion useful for calculating the efficiency measure for particular prediction functions. Fourth, we show that the predictive likelihood approach yields unambiguous efficiency gains for an important class of location parameter models.

The Kullback-Leibler measure for a particular realized prediction function $f^*(y_f;\hat{\theta}_d)$ can be written as:

$$(4.1) \qquad I(f,f^*) = \int [g(y_f;\theta) - g^*(y_f;\hat{\theta}_d)]\, f(y_f;\theta)\, dy_f,$$

where $g(y_f;\theta) = \log(f(y_f;\theta))$ and $g^*(y_f;\hat{\theta}_d) = \log(f^*(y_f;\hat{\theta}_d))$. To abstract from the dependence of (4.1) on the particular realizations of y_d and $\hat{\theta}_d$, we will compute the expected information loss due to parameter uncertainty

$$(4.2) \qquad \bar{I}(f,f^*) = \int I(f,f^*)\, f(\hat{\theta}_d;\theta)\, d\hat{\theta}_d,$$

where $f(\hat{\theta}_d;\theta)$ is the density of $\hat{\theta}_d$.

The asymptotic properties of $\bar{I}(f,f^*)$ will prove both workable and interesting even though evaluating $\bar{I}(f,f^*)$ may prove difficult for many typical econometric applications.[2] Predictive consistency will be defined as

(4.3) $$\bar{I}(f,f^*) \to 0 \text{ as } m \to \infty.$$

This requires basically that $\hat{\theta}_d$ is consistent and that $f^*(y_f;\hat{\theta}_d)$ converges to $f(y_f;\theta)$ as $\hat{\theta}_d \xrightarrow{p} \theta$. While inappropriate simplifications such as applying a normal approximation to a nonnormal true density fail to achieve predictive consistency, a variety of functions, including CEQ and $plik^a(y_f|y_d)$, are predictive consistent.

Among predictive consistent functions, the first order predictive efficiency is the scalar $\lambda_1(f,f^*)$ in the expansion

(4.4) $$\bar{I}(f,f^*) = m^{-1}\lambda_1(f,f^*) + o(m^{-1}).$$

We can derive the first order predictive efficiency for the CEQ function under fairly general assumptions:

Proposition 1. We assume that: (i) The derivatives to order three of $g(y_f;\theta)$ with respect to θ exist, (ii) the derivatives to order two are bounded by integrable functions, and (iii) the third derivatives are uniformly bounded by a function with finite expectation. Then the CEQ first order prediction efficiency is given by:

(4.5) $$\lambda_1(f,\hat{f}) = -\frac{1}{2}\,\mathrm{tr}[\ V(\hat{\theta}_d)^{-1}\ E_Y(H(y_f;\theta))\ ]\ ,$$

---

[2] The expected value of $I(f,f^*)$ over the distribution of $\hat{\theta}_d$ is typically about as difficult to derive as is the expected value of $\hat{\theta}_d$ itself. For example, $\bar{I}(f,plik^a)$ can be derived precisely for linear-normal models and models with nonlinearities in variables as in Cooley, Parke, and Chib (1987).

where $V(\hat{\theta}_d)$ is the asymptotic variance-covariance matrix of $m^{1/2}(\hat{\theta}_d-\theta)$ and $E_Y(H(y_f;\theta))$ is the expectation over Y of $H(y_f;\theta)$.

Proof: Under these standard assumptions, we can expand

$$\bar{I}(f,\hat{f}) = \int_{\Theta}\int_{Y} (g(y_f;\theta)-g(y_f;\hat{\theta}_d))\ f(y_f;\theta)\ f(\hat{\theta}_d;\theta)\ dy_f\ d\hat{\theta}_d$$

as the sum of two expressions. The first,

$$-\int_{\Theta}\int_{Y} \nabla(y_f;\theta)(\hat{\theta}_d-\theta)\ f(y_f;\theta)\ f(\hat{\theta}_d;\theta)\ dy_f\ d\hat{\theta}_d,$$

equals zero because $\nabla(y_f;\theta)$ is independent of $\hat{\theta}_d-\theta$ and $\int_Y \nabla(y_f;\theta)\ f(y_f;\theta)\ dy_f = 0$. The second,

$$-\frac{1}{2}\int_{\Theta}\int_{Y} (\hat{\theta}_d-\theta)'\ H(y_f;\theta^*)\ (\hat{\theta}_d-\theta)\ f(y_f;\theta)\ f(\hat{\theta}_d;\theta)\ dy_f\ d\hat{\theta}_d,$$

where θ^* is between θ and $\hat{\theta}_d$, converges (multiplied by m) to (4.5).[3]

End of proof.

We now turn our attention toward efficiency improvements that can be secured by accounting for parameter uncertainty. Proposition 1 states that ignoring parameter uncertainty leads to $\bar{I}(f,\hat{f}) = O(m^{-1})$, and it will turn out that the best improvement generally available is $O(m^{-2})$. This bound on relative prediction efficiency is meaningful (as are the well known bounds on estimation efficiency) only if the class of alternatives is restricted by suitable regularity conditions. We gain some insight into establishing

---

[3] This convergence requires that the variance of $m^{1/2}(\hat{\theta}_d-\theta)$ converge to the asymptotic variance $V(\hat{\theta}_d)$. Exceptions to this technical condition, which is also incorporated into the notion of an asymptotic mean squared error, could be dealt with formally by truncating slightly the range of $\hat{\theta}_d$. Equivalently, we could have defined the average infomation loss (4.2) to be with respect to the asymptotic distribution of $\hat{\theta}_d$.

appropriate regularity conditions by considering a simple, but compelling example of superefficiency. The superefficiency example motivates us to require that any efficiency gain occur over a neighborhood N in the parameter space rather than for just certain true parameters. Given this last assumption, we show in Proposition 2 that the largest possible efficiency improvement is $O(m^{-2})$. We then contemplate the efficiency improvement achieved by the particular case $\text{plik}^a(y_f|\hat{\theta}_d)$.

We begin by formally defining the predictive efficiency of $f^*(y_f;\hat{\theta}_d)$ relative to $f(y_f;\hat{\theta}_d)$ as

$$\bar{I}(\hat{f},f^*) = -\int h(y_f;\hat{\theta}_d)\, f(y_f;\theta)\, f(\hat{\theta}_d;\theta)\, dy_f\, d\hat{\theta}_d, \tag{4.6}$$

where $h(y_f;\hat{\theta}_d)$ is defined via

$$g^*(y_f;\hat{\theta}_d) = g(y_f;\hat{\theta}_d) + h(y_f;\hat{\theta}_d). \tag{4.7}$$

Note that the efficiency measures are additive in the sense that $\bar{I}(f,f^*) = \bar{I}(f,\hat{f}) + \bar{I}(\hat{f},f^*)$. The <u>second order relative predictive efficiency</u> is the scalar $\lambda_2(\hat{f},f^*)$ in the expansion

$$\bar{I}(\hat{f},f^*) = m^{-2}\lambda_2(\hat{f},f^*) + o(m^{-2}). \tag{4.8}$$

The regularity conditions we will introduce rule out certain instances of superefficiency. Consider the example of a prediction function $f^*(y_f;\hat{\theta}_d) = f(y_f;\theta^a)$, where θ^a is a fixed element of Θ. This choice entails zero information loss if θ happens to equal θ^a, but not for any other true parameters. It fails to attain even predictive consistency for $\theta \neq \theta^a$ because $f^*(y_f;\hat{\theta}_d) = f(y_f;\theta^a)$ for all $\hat{\theta}_d$ regardless of the actual true parameters. To rule out such cases, we require that the advantage of

$f^*(y_f;\hat{\theta}_d)$ over $f(y_f;\hat{\theta}_d)$ be reasonably uniform for true parameters in a neighborhood N (that does not shrink with increasing m). We incorporate this requirement via the average of $\bar{I}(\hat{f},f^*)$ over all true parameters $\bar{\theta}$ in N, which we write as

$$(4.9) \qquad \bar{I}_N = -\int_N \int_\Theta \left[\int_Y h(y_f;\hat{\theta}_d)\ f(y_f;\bar{\theta})\ dy_f \right] f(\hat{\theta}_d;\bar{\theta})\ d\hat{\theta}_d\ d\bar{\theta}\ .$$

Our strategy will be to show that $\bar{I}_N$ can be negative only if $\bar{I}(\hat{f},f^*) = O(m^{-2})$ almost everywhere in N.

Assume that there exists an α such that

$$(4.10) \qquad \int [h(y_f;\bar{\theta})]^2\ f(y_f;\bar{\theta})\ dy_f = O(m^{-\alpha}),$$

and

$$(4.11) \qquad \int [h(y_f;\bar{\theta})]^k\ f(y_f;\bar{\theta})\ dy_f = o(m^{-\alpha})\ ,\ k > 2.$$

These conditions are not restrictive because they are essentially if $f^*(y_f;\theta)$ is to integrate to unity. (See (4.17) below.) We will add to the usual assumption

$$(4.12) \qquad \int_Y \nabla(y_f;\theta)'\nabla(y_f;\theta)\ f(y_f;\theta)\ dy_f = O(1)$$

a similar, but higher order requirement that

$$(4.13) \qquad \int_Y (\nabla(y_f;\theta)'\nabla(y_f;\theta)+H(y_f;\theta))^2\ f(y_f;\theta)\ dy_f = O(1).$$

Proposition 2. If (4.10) is not met with $\alpha \geq 2$ almost everywhere in N, then the largest term in an expansion of (4.9) will be unambiguously positive.

Proof.[4] To evaluate the integral (4.7) over the Cartesian product $N \times \Theta$, we will integrate first over N and then over Θ. For a given $\hat{\theta}_d \in \Theta$, the integral over $\tilde{\theta} \in N$ can be written as the sum of two terms:

$$(4.14) \qquad - \int_N [\int_Y h(y_f;\hat{\theta}_d)\ f(y_f;\tilde{\theta})\ dy_f\ ]\ f(\hat{\theta}_d;\tilde{\theta})\ d\tilde{\theta}$$

$$- \int_N \{ \int_Y h(y_f;\hat{\theta}_d)\ [e^{g(y_f;\tilde{\theta})-g(y_f;\hat{\theta}_d)} - 1]\ f(y_f;\hat{\theta}_d)\ dy_f\ \}\ f(\hat{\theta}_d;\tilde{\theta})\ d\tilde{\theta}.$$

Expanding the element in square brackets in the second term shows that term to equal

$$(4.15) \qquad - \int_Y h(y_f;\hat{\theta}_d)\ \nabla(y_f;\hat{\theta}_d)\ f(y_f;\hat{\theta}_d)dy_f\ \int_N (\hat{\theta}_d-\tilde{\theta})f(\hat{\theta}_d;\tilde{\theta})d\tilde{\theta}$$

$$- \frac{1}{2} \int_Y h(y_f;\hat{\theta}_d)\ [\nabla(y_f;\hat{\theta}_d)^2+H(y_f;\hat{\theta}_d)]\ f(y_f;\hat{\theta}_d)dy_f\ \int_N (\hat{\theta}_d-\tilde{\theta})^2 f(\hat{\theta}_d;\tilde{\theta})d\tilde{\theta}$$

plus terms involving third and higher powers of $\hat{\theta}_d-\tilde{\theta}$.

We can rewrite the first term in (4.14) as

$$(4.16) \qquad \int_Y h(y_f;\hat{\theta}_d)\ f(y_f;\hat{\theta}_d)\ dy_f\ \int_N f(\hat{\theta}_d;\tilde{\theta})\ d\tilde{\theta}\ ,$$

where the second factor can be denoted $P(N|\hat{\theta}_d)$.[5] The first factor in (4.16) is unambiguously nonnegative because $\int \log(\pi/\pi^*)\ \pi \geq 0$ for any densities π and π^* with equality only for $\pi^* = \pi$ almost everywhere. The first factor is also the leading term in a Taylor series for the unit integral requirement

---

[4] For notational simplicity, we will use a scalar θ.

[5] The notation $P(N|\hat{\theta}_d)$ conditioning on $\hat{\theta}_d$ is not completely well defined because $\tilde{\theta}$ is not a random variable. It does, however, furnish a convenient shorthand description of the process of integrating over N. The same consideration motivates our use of the notation $E_N()$.

$\int_Y \exp(h(y_f;\hat{\theta}_d))\ f(y_f;\hat{\theta}_d)\ dy_f = 1$, which can be expanded as

(4.17) $$\int_Y [\ h(y_f;\hat{\theta}_d) + \frac{1}{2}\ (h(y_f;\hat{\theta}_d))^2 + \cdots\ ]\ f(y_f;\hat{\theta}_d)\ dy_f = 0.$$

Under (4.11), the remaining terms in (4.17) are $o(m^{-2})$.

We can now combine (4.15), (4.16), and (4.17) to write (4.14) as the product of $P(N|\hat{\theta}_d)$ and

(4.18) $$V_Y(\hat{h}) - COV_Y(\hat{h},\hat{\nabla})\cdot E_N(\hat{\theta}_d-\bar{\theta}|\hat{\theta}_d) - COV_Y(\hat{h},\hat{\nabla}^2+\hat{H})\cdot E_N((\hat{\theta}_d-\bar{\theta})^2|\hat{\theta}_d),$$

where $E_N(\hat{\theta}_d-\bar{\theta}|\hat{\theta}_d)$ and $E_N((\hat{\theta}_d-\bar{\theta})^2|\hat{\theta}_d)$ denote the order m^{-1} terms in the asymptotic conditional mean and variance of $\hat{\theta}_d-\bar{\theta}$ over N and $COV_Y(\cdot)$ and $V_Y(\cdot)$ denote integrals over y_f. The covariance inequality together with (4.10), (4.12), and (4.13) implies that $COV_Y(\hat{h},\hat{\nabla}) = O(m^{-\alpha/2})$ and $COV_Y(\hat{h},\hat{\nabla}^2+\hat{H}) =$
$O(m^{-\alpha/2})$. The two elements of (4.18) involving these factors are thus both $O(m^{-\alpha/2-1})$ while the unambiguously positive element $V_Y(\hat{h})$ is $O(m^{-\alpha})$. From this, we deduce that the largest term in an expansion of (4.14) will be unambiguously positive unless $\alpha \geq 2$.

Integrating (4.14) over $\hat{\theta}_d \epsilon \Theta$ will then produce (4.9). This last integral will be dominated by the unambiguously positive instances of $\alpha < 2$ if these have measure greater than zero over $\hat{\theta}_d \epsilon \Theta$.

End of proof.

The proof of Proposition 2 gives some guidance in constructing $h(y_f,\hat{\theta}_d)$. The unambiguously nonnegative term $V_Y(\hat{h})$ should be as small as possible and the covariances $COV_Y(\hat{h},\hat{\nabla})$ and $COV_Y(\hat{h},\hat{\nabla}^2+\hat{H})$ should be as large as possible. The function, $\nabla(y_f;\hat{\theta}_d)$ and $\nabla(y_f;\hat{\theta}_d)^2+H(y_f;\hat{\theta}_d)$ are clear candidates to form $h(y_f;\hat{\theta}_d)$ under these criteria. The asymptotic predictive likelihood function, which is originally motivated on distinctly different grounds, combines these two functions, weighting $\nabla(y_f;\hat{\theta}_d)$ by the asymptotic bias and weighting $\nabla(y_f;\hat{\theta}_d)^2+H(y_f;\hat{\theta}_d)$ by the asymptotic variance.

A more direct approach to calculating the second order relative efficiency of $\text{plik}^a(y_f|\hat{\theta}_d)$ and other prediction functions is possible if the conditions in Proposition 2 are strengthened slightly. The main conclusion of Proposition 2 is that reasonable candidate functions $h(y_f;\hat{\theta}_d)$ will be well behaved after multiplication by m. In practice, $h(y_f;\hat{\theta}_d)$ will generally be constructed by weighting a function of y_f by either the variance or bias of $\hat{\theta}_d$, both of which are proportional to m.

<u>Proposition 3</u>. Assume that (i) the derivatives to order three of $m\ h(y_f;\theta)$ with respect to θ exist (we will denote derivatives by subscripts,e.g. $h_\theta(y_f;\theta)$), (ii) the derivatives to order two of $m\ h(y_f;\theta)$ are bounded by integrable functions, (iii) the third derivatives of $m\ h(y_f;\theta)$ are uniformly bounded by a function with finite expectation, and (iv) (4.9) and (4.10) are satisfied for $\alpha = 2$. Then

$$\lambda_2(\hat{f},f^*) = \frac{1}{2}\,\underline{h}^2 - \psi\,\underline{h}_\theta - \frac{1}{2}\,\mathrm{tr}[\ V(\hat{\theta}_d)\ \underline{h}_{\theta\theta}\ ]\ , \tag{4.19}$$

where

$$\underline{h}^2 = \lim_{m\to\infty} \int_Y m^2\ h(y_f;\theta)^2\ f(y_f;\theta)\ dy_f$$

$$\underline{h}_\theta = \lim_{m\to\infty} \int_Y m\ h_\theta(y_f;\theta)\ f(y_f;\theta)\ dy_f$$

$$\underline{h}_{\theta\theta} = \lim_{m\to\infty} \int_Y m\ h_{\theta\theta}(y_f;\theta)\ f(y_f;\theta)\ dy_f \quad \text{and}$$

$m^{-1}\psi$ is the $O(m^{-1})$ bias in $\hat{\theta}_d$.

<u>Proof.</u> By expanding $h(y_f;\theta)$, we can write

$$- m^2 \int_\Theta \int_Y h(y_f;\hat{\theta}_d)\ f(y_f;\theta)\ f(\hat{\theta}_d;\theta)\ dy_f\ d\hat{\theta}_d$$

as the sum of three terms

$$- m^2 \int_Y h(y_f;\theta)\ f(y_f;\theta)\ dy_f$$

$$- m^2 \int_\Theta \int_Y h_\theta(y_f;\theta)\ (\hat{\theta}_d-\theta)\ f(y_f;\theta)\ f(\hat{\theta}_d;\theta)\ dy_f\ d\hat{\theta}_d$$

$$- \frac{1}{2} m^2 \int_\Theta \int_Y (\hat{\theta}_d-\theta)'\ h_{\theta\theta}(y_f;\theta^*)\ (\hat{\theta}_d-\theta)\ f(y_f;\theta)\ f(\hat{\theta}_d;\theta)\ dy_f\ d\hat{\theta}_d,$$

where θ^* is between θ and $\hat{\theta}_d$. The first term can be approximated by $\frac{1}{2}\ \underline{h}^2$ via the unit integral requirement (4.17) and assumption (iv). The other two terms converge in probability to the terms in (4.19).

End of proof.

For the particular prediction functions $\text{plik}^a(y_f|\hat{\theta}_d)$, the conclusion of Proposition 3 can be expressed in another way that may be useful by evaluating $\underline{h}^2$, $\underline{h}_\theta$, and $\underline{h}_{\theta\theta}$. We record these results as Proposition 4 below. One of these includes, as a special case, the classic location parameter prediction problem where we know the form of the density for y_f, but not its location.

<u>Proposition 4.</u> (a) If the bias in $\hat{\theta}_d$ is $o(m^{-1})$ (perhaps, because $\hat{\theta}_d$ is corrected for bias), then $\lambda_2(\hat{f},\text{plik}^a)$ can be written as:

$$(4.20)\qquad - \frac{1}{8}\ \text{vec}(H_d^{-1})'\ \{\ E_Y[(\text{vec}H_f)(\text{vec}H_f)'] - E_Y[\nabla_f'\otimes H_f\otimes\nabla_f]$$
$$+\ 2\ E_Y[H_f]_{\theta\theta} - E_Y[(\nabla_f'\otimes\nabla_f')\nabla_f]_\theta\ \}\ \text{vec}(H_d^{-1}),$$

where we are letting the arguments of H_d, H_f and ∇_f be implicit for notational simplicity.

(b) If, in addition, $E_Y[H_f]$ and $E_Y[(\nabla_f'\otimes\nabla_f')\otimes\nabla_f]$ are constant over θ, then

(4.21) $\lambda_2(\hat{f},plik^a) = -\frac{1}{8} E_Y[\ (tr(\ H_d^{-1}\ H_f\ ))^2\ ] + \frac{1}{8} E_Y[\ \nabla_f\ H_d^{-1}\ H_f\ H_d^{-1}\ \nabla_f']$

(c) If $H(y_f;\theta)$ is globally negative semi-definite, then $\lambda_2(\hat{f},plik^a)$ is unambiguously negative.

Proof: Appendix B.

In Section 7 of this paper, we give some results for specific models that illustrate the process of finding these expectations and applying Propositions 3 and 4.

5. Mean Squared Error Prediction Functions.

The information efficiency measures proposed in Section 4 emphasize the importance of correct functional form. In contrast, the most commonly used technique for evaluating predictions - mean squared error (MSE) analysis - is often used simply as a criterion for evaluating the forecasting error of point predictions without regard for functional form (e.g. Baillie (1981)).[6] Indeed, abstracting from nonlinear functional forms is regarded as a virtue of the technique.

We can represent an asymptotic MSE analysis in the setting of a particular functional form, thereby deriving a MSE prediction function that approximates $f(y_f;\theta)$. We base our asymptotic MSE analysis on a point

---

[6] We might add that mean squared error analysis rarely involves any attempt to actually minimize mean squared error. The focus is instead on simply assessing the error in some point forecast.

forecast $\hat{y}_f$, which is typically computed by simply setting unknown errors to zero. Because mean squared error analysis is concerned with only the second moment of $y_f-\hat{y}_f$, the natural functional form is a normal density

(5.1) $$g_{(}y_f;\hat{\theta}_d) \propto -\frac{1}{2}(y_f-\hat{y}_f)'V_f^{-1}(y_f-\hat{y}_f),$$

where V_f is the variance-covariance matrix of y_f. Dealing with parameter uncertainty requires the derivatives $D_f = \partial\hat{y}_f/\partial\hat{\theta}_d$. Treating D_f as constant over y_f leads to the approximate first and second derivatives

(5.2) $$\nabla_{(}y_f;\hat{\theta}_d) \approx (y_f-\hat{y}_f)'V_f^{-1}D_f$$

(5.3) $$H_f \approx D_f'V_f^{-1}D_f.$$

Using (5.2) and (5.3), $w_1(y_f;\hat{\theta}_d)$ becomes

(5.4) $$w_1(y_f;\hat{\theta}_d) = -\frac{1}{2}(y_f-\hat{y}_f)'V_f^{-1}D_f[H_d + D_f'V_f^{-1}D_f]^{-1}D_f'V_f^{-1}(y_f-\hat{y}_f).$$

If we ignore any asymptotic bias in $\hat{\theta}_d$ and treat H_f as constant over y_f, then the term $w_2(y_f;\hat{\theta}_d)$ in (3.1) is constant over y_f. We can combine (5.1) and (5.4) using the identity [Rao (1973, p. 33)]

(5.5) $$-V_f^{-1} - V_f^{-1}D_f[H_d + D_f'V_f^{-1}D_f]^{-1}D_f'V_f^{-1} = -[V_f - D_f'H_d^{-1}D_f]^{-1}$$

to form the mean squared error prediction function

(5.6) $$MSE(y_f|\hat{\theta}_d) \propto \exp\{-\frac{1}{2}(y_f-\hat{y}_f)'[V_f - D_fH_d^{-1}D_f']^{-1}(y_f-\hat{y}_f)\},$$

which incorporates the variance V_f of y_f and an approximate variance $-D_fH_d^{-1}D_f'$ due to parameter uncertainty, but fails to acknowledge both any non-normality of $f(y_f;\theta)$ and any nonlinearity in the parameter uncertainty.

This derivation emphasizes that (5.6) can be regarded as in the same family as $\text{plik}^a(y_f|\hat{\theta}_d)$, but subject to additional linearization. It is straightforward to demonstrate that, for nonnormal or nonlinear models, (5.6) does not satisfy even the predictive consistency criterion. Advocates of the MSE approach might respond to this failure to converge to zero information loss in large samples by substituting a quadratic loss function for $\bar{I}(f,f^*)$,

6. Inefficient Statistics.

In practical problems, interest will often center on data period parameter estimates that are consistent and asymptotically normal, but not asymptotically efficient. (Consider, for example, two stage least-squares.) Let $\tilde{\theta}_d$ be such an estimate with

$$m^{1/2}(\tilde{\theta}_d-\theta) \to N(0,V(\tilde{\theta}_d)),$$

where $V(\tilde{\theta}_d)-V(\hat{\theta}_d)$ is a positive semi-definite matrix. Let H_{d+f} denote V_m^{-1} + H_f, where mV_m is a consistent estimate of $V(\tilde{\theta}_d)$. We can then extend our asymptotic predictive likelihood definition to cover prediction functions based on $\tilde{\theta}_d$:

$$\text{(6.1)} \qquad \text{plik}^c(y_f|y_d) = f(y_f;\tilde{\theta}_d) \cdot \exp\{\ w_1(y_f;\tilde{\theta}_d) + w_2(y_f;\tilde{\theta}_d)\ \},$$

where

$$w_1(y_f;\tilde{\theta}_d) = -\frac{1}{2}\,\nabla(y_f;\tilde{\theta}_d)\ H(y_{d+f};\tilde{\theta}_d)^{-1}\nabla(y_f;\tilde{\theta}_d)'$$

$$w_2(y_f;\tilde{\theta}_d) = \nabla(y_f;\tilde{\theta}_d)\psi(\tilde{\theta}_d) - \frac{1}{2}\,\mathrm{tr}[\ H(y_f;\tilde{\theta}_d)\ H(y_d;\tilde{\theta}_d)^{-1}]\ .$$

This definition is further motivated in appendix A. This prediction function possesses many of the important features of (3.1), taking due account of the fact that the estimate $\tilde{\theta}_d$ is not asymptotically efficient.

In particular, $\mathrm{plik}^c(y_f;\tilde{\theta}_d)$ may well secure an efficiency gain over the corresponding plug-in function $\tilde{f} = f(y_f;\tilde{\theta}_d)$ also based on the inefficient estimates $\tilde{\theta}_d$. In terms of first order efficiency, direct extensions of Propositions 1 and 2 show that $\lambda_1(f,\mathrm{plik}^c) = \lambda_1(f,\tilde{f})$ and that $\mathrm{plik}^c(y_f|y_d)$ is first order inefficient relative to $\mathrm{plik}^a(y_f|y_d)$ to the extent that

$$\lambda_1(\hat{f},\tilde{f}) = - \frac{1}{2}\,\mathrm{tr}[\ E_Y(H(y_f;\theta)\ (V(\tilde{\theta}_d)-V(\hat{\theta}_d))\ ].$$

is positive. Extensions of Propositions 3 and 4, however, would suggest that the CEQ $f(y_f;\tilde{\theta}_d)$ may be second order inefficient relative to $\mathrm{plik}^c(y_f|\hat{\theta}_d)$. Thus, from a practical standpoint, if sufficient motivation exists to favor calculation of only inefficient parameter estimates, $\mathrm{plik}^c(y_f;\tilde{\theta}_d)$ still incorporates a useful adjustment for parameter uncertainty.

7. Examples.

The definition of asymptotic predictive likelihood and the concept of predictive efficiency developed above are useful only to the extent that they sharpen our understanding of practical prediction problems. In this section we consider examples that extend well known results for linear-normal models to models with nonnormal disturbances and to ARCH models.

Example 1. Linear-Normal Model. Before considering more complex models, it is helpful to examine the asymptotic efficiency concepts for a linear regression model because Definitions 1 and 2 coincide and exact information losses can be computed. We can write that model as

$$y_i = x_i\beta + \varepsilon_i, \quad \varepsilon_i \sim N(0,\sigma^2), \tag{7.1}$$

for regressors x_i, parameters β, and known σ^2. Cooley, Parke and Chib (1987) show that, for a single future observation, $\text{plik}^a(y_f|\hat{\theta}_d) \propto \exp(-\frac{1}{2}\varepsilon_f^2/(\sigma^2+\tau^2))$, where the variance component $\tau^2 = \sigma^2 x_f(x_d'x_d)^{-1}x_f'$ corrects for parameter uncertainty. The comparison between the two efficiencies $\bar{I}(f,\hat{f}) = \frac{1}{2}\,\tau^2/\sigma^2$ and $\bar{I}(f,\text{plik}^a) = \frac{1}{2}\log(1+\tau^2/\sigma^2)$ can be put into the present framework by expanding the second of these as $\frac{1}{2}\,\tau^2/\sigma^2 - \frac{1}{4}\,(\tau^2/\sigma^2)^2 + \cdots$. The first order asymptotic efficiencies equal $\frac{1}{2}\lim_{m\to\infty} m\,\tau^2/\sigma^2$ in both cases. Correcting for parameter uncertainty secures the second order efficiency gain $\frac{1}{4}\lim_{m\to\infty} m\,(\tau^2/\sigma^2)^2$, which is reflected in the curvature of the log function. That efficiency gain is likely to be most important for difficult forecasting problems where the first order efficiency loss is also important.

<u>Example 2</u>. Nonnormal Model. Suppose that the errors are drawn from a t distribution with ν degrees of freedom, where σ and ν are known. (For the variance to exist we require that $\nu > 2$.) The predictive likelihood function takes account of the relatively fat tails in the t distribution. If we let ζ denote $(y_{m+1}-x_{m+1}\hat{\beta}_d)/\sigma$, then for a single future observation

$$(7.2)\quad \text{plik}^a(y_f|\hat{\beta}_d) \propto \frac{1}{\sigma}\left(1+\zeta^2/\nu\right)^{-(\nu+1)/2}\cdot\exp\{\,w_1(y_f;\hat{\theta}_d) + w_2(y_f;\hat{\theta}_d)\,\},$$

where

$$w_1(y_f;\hat{\theta}_d) = -\frac{1}{2}\,\frac{(\nu+1)^2}{\nu^2}\,\frac{x_f\,H_{d+f}^{-1}\,x_f'}{(1+\zeta^2/\nu)^2}\,\zeta^2$$

and

$$w_2(y_f;\hat{\theta}_d) = \frac{(\nu+1)}{\nu}\,\frac{x_f\,H_d^{-1}\,x_f'}{(1+\zeta^2/\nu)^2}\,(1-\zeta^2/\nu).$$

The correction for parameter uncertainty $w_1(y_f;\hat{\theta}_d)$ increases the dispersion of the plik by adding to the density in the tails, where ζ^2 is greatest. Relative to the linear-normal model in Example 1, however, the true density already has fat tails, and the denominator $(1+\zeta^2/\nu)^2$ in $w_1(y_f;\hat{\beta}_d)$ moderates the extent of the correction in the extreme tails. The term $w_2(y_f;\hat{\beta}_d)$ adds a lesser correction for a nonconstant second derivative matrix.

Proposition 4 provides the information efficiency calculations for this model. If we let Γ denote $\lim_{m\to\infty} m\, x_f H_d^{-1} x_f'$, then

$$\lambda_1(f,\hat{f}) = \frac{1}{2}\left\{\frac{\nu+4}{\nu}\right\}^{-\frac{1}{2}} \frac{\nu+1}{\nu}\left\{1 - \frac{1}{\nu+2}\right\}\Gamma .$$

This figure ranges from the value of $\frac{1}{2}\Gamma$ for Example 1 ($\nu=\infty$) to $1.95\cdot\frac{1}{2}\Gamma$ for $\nu=2$, revealing the extent to which fatter tails (smaller ν) lead to a more difficult forecasting problem. The second order relative efficiency

$$\lambda_2(\hat{f},\mathrm{plik}^a) = -\frac{1}{8}\left[\frac{\nu+8}{\nu}\right]^{-\frac{1}{2}}\left[\frac{\nu+1}{\nu}\right]^2\left[1 + \frac{(\nu-1)}{(\nu+6)} - \nu\,\frac{3 + 6/(\nu+4)}{(\nu+6)^2}\right]\Gamma^2$$

is clearly negative so that the predictive likelihood correction for parameter uncertainty lowers the information loss.

Unknown error distribution parameters such as σ^2 and ν also present interesting forecasting problems. If σ^2 in Example 1 is unknown, a direct analysis of the sampling distribution of $\zeta = (y_f - x_f\hat{\beta}_d)/s$, where s is the sample period estimate of σ, shows the appropriate prediction function to be of the functional form of a t distribution. Definition 1 yields precisely this result using a χ^2 distribution for the sample variance s^2 (Cooley, Parke, and Chib (1987a)). The advantage of $\text{plik}^a(y_f|s^2)$ is that only $g(y_f;\theta)$ and its derivatives are needed because Definition 2 is based (see Appendix A) on the asymptotically valid normal approximation $m^\sigma(s^2-\sigma^2) \overset{d}{\to} N(0,2)$. For the simple model $y_i \sim N(0,\sigma^2)$ that abstracts from uncertainty about β,

$$\text{plik}^a(y_f|s^2) \propto -\frac{1}{2}\, y_f^2/s^2 - \frac{1}{4}\,(m+1)^{-1}\,(y_f^2/s^2-1)^2.$$

This function is identical to the first two terms in a series expansion of the logarithm of a t density for y_f, expanding $-\frac{1}{2}\,(\nu+1)\,\log(1+(y_f^2/s^2)/\nu)$ about $-\frac{1}{2}\,(\nu+1)\,\log(1+1/\nu)$. The correction for parameter uncertainty in $\text{plik}^a(y_f|s^2)$ thus captures the essential features of a t distribution.

The information efficiency calculations for this model are a special case of those for Example 3 below. We simply note here that $\text{plik}^a(y_f|s^2)$ will be second order efficient relative to the CEQ $f(y_f;s^2)$ to the extent that a t distribution is more appropriate for $(y_f-x_f\hat{\beta}_d)/s$ than is a normal distribution.

<u>Example 3</u>. The most interesting features of models with unknown error distribution parameters can be demonstrated in the context of the autoregressive conditional heteroscedasticity (ARCH) model. Following Engle (1980), we emphasize the essential aspects of this model using a simple ARCH model without regressors

$$y_t \sim N(0,h_t),$$

where $h_t = z_t\alpha$ for $z_t = (1,y_{t-1}^2,\ldots,y_{t-p}^2)$ and $\alpha = (\alpha_0,\alpha_1,\ldots,\alpha_p)$. (The function h_t follows Engle's notation and is not related to $h(y_f;\theta)$.) This model emphasizes the dispersion of the future density rather than its mean.

The predictive likelihood function again approximates the functional form of a t distribution.[7] For one period ahead (so that f denotes m+1),

$$\log(\mathrm{plik}^{a}(y_f|\hat{\alpha}_d)) \propto -\frac{1}{2}\frac{y_f^2}{h_f} - \frac{1}{8}\frac{z_f H_{d+f}^{-1} z_f'}{h_f^2}\left\{\frac{y_f^2}{h_f} - 1\right\}^2 .$$

In this approximation to a t distribution, the "degrees of freedom"

$$\nu_m = \frac{1}{2}\frac{h_f^2}{z_f H_{d+f}^{-1} z_f'} .$$

will be proportional to the data period sample size because H_{d+f} grows at rate m, but will also depend on the particular z_f vector. That vector appears in both the numerator $h_f^2 = (z_f\hat{\alpha}_d)^2$ and in the denominator, which essentially equals $V(z_f\hat{\alpha}_d)$. If, for example, the elements of $\hat{\alpha}_d$ are negatively correlated so that H_{d+f}^{-1} has negative off-diagonal elements, ν_m will be smaller (and the correction for parameter uncertainty will be greater) for a vector z_f with a single large element than for z_f with more equally sized elements.

The formal information efficiency calculations also reflect the dependence of ν_m on z_f. The first order information efficiency is

$$\lambda_1(f,\hat{f}) = 1/2\ \nu_\infty^{-1},$$

where $\nu_\infty = \lim_{m\to\infty} m^{-1}\ \nu_m$ depends on z_f. Proposition 3 shows that

---

[7] We are omitting the term $w_2(y_f;\hat{\theta}_d)$ on two grounds. First, the asymptotic bias is not known for the ARCH model, making an analytic implementation impossible. Second, the two terms in $w_2(y_f;\hat{\theta}_d)$ cancel for Example 3 and will largely offset in this case as well.

Our calculations are all conditional on the last few values of y_t. As Phillips (1979) notes, this introduces a minor dependency between the distribution of $\hat{\theta}_d$ and the last few values of y_t.

$$\lambda_2(\hat{f},plik^a) = -\ 23/2\ \nu_\infty^{-2}$$

The efficiency gain from correcting for parameter uncertainty thus depends on both the data period sample size and the particular vector z_f.

Predictive likelihood forecasts two or more periods ahead for an ARCH model recognize that the variance of y_{m+2} depends on the realization of y_{m+1}. This dynamic aspect of the problem is incorporated into

$$\log(plik^a(y_f|\hat{\alpha}_d)) \propto -\frac{1}{2}\sum_{i=m+1}^{m+n}\frac{y_{m+1}^2}{h_{m+i}^2} - \frac{1}{8}\,\zeta'\ z_f\ H_{d+f}^{-1}\ z_f'\ \zeta\ .$$

where ζ is the $n \times 1$ vector with elements $\zeta_i = (y_{m+i}^2/h_{m+i}-1)^2/h_{m+i}$, $i=1,\ldots,n$. If $n = 2$, then y_{m+1} appears (via z_f) in both the 2×2 matrix $z_f\ H_{d+f}^{-1}\ z_f'$ and in h_{m+2}. This joint predictive density for y_{m+1} and y_{m+2} thus directs that the dispersion for y_{m+2} be a function of the entire range of values for y_{m+1} weighted by their predictive likelihoods.

8. Conclusions.

The asymptotic predictive likelihood approach analyzed in this paper is closely related to Monte Carlo forecasting approaches discussed in Section 2. Monte Carlo procedures account for parameter uncertainty by drawing coefficients from an asymptotic distribution. The predictive likelihood approach, on the other hand, suggests a correction to the forecasting density. The correction can be implemented easily via stochastic simulation with a weighting determined from the correction terms in Definition 2. Consequently, although the calculations in the examples seem cumbersome, implementing these prediction functions via simulation is quite feasible.

The information measure of predictive efficiency derived in Section 4 help to identify the effects of various specification and estimation issues on predictive accuracy. Predictive consistency requires the correct functional form for the model. First order efficiency rests on the efficiency of the estimated parameters. Asymptotic estimation bias and corrections for parameter uncertainty affect second order efficiency. This is one explanation of why parameter uncertainty appears not to matter much in practice in most applications and is usually neglected.

Appendix A

Motivation for Predictive Likelihood Definitions

In situations where sufficient reductions of the data do not exist, we can exploit the fact that well-behaved maximum likelihood estimates are asymptotically sufficient (Cox and Hinkley (1974, p. 307)). Replacing the sufficient statistics S_d and S_{d+f} in Definition 1 by the MLE's $\hat{\theta}_d$ and $\hat{\theta}_{d+f}$ leads to the alternative definition:

$$\text{(A.1)} \qquad \text{plik}^1(y_f|\hat{\theta}_d) = f(y_f,\hat{\theta}_d|\hat{\theta}_{d+f}) = \frac{f(y_f;\theta)\ f(\hat{\theta}_d;\theta)}{f(\hat{\theta}_{d+f};\theta)}$$

where $f(\hat{\theta}_d;\theta)$ and $f(\hat{\theta}_{d+f};\theta)$ are exact finite sample distributions of the MLE's. For econometric problems of any complexity these exact finite sample distributions are intractable. This consideration leads us to:

$$\text{(A.2)} \qquad \text{plik}^2(y_f|\hat{\theta}_d) = \frac{f(y_f;\theta)\ f^a(\hat{\theta}_d;\theta)}{f^a(\hat{\theta}_{d+f};\theta)} ,$$

where $f^a(\cdot;\cdot)$ denotes an asymptotic density. $\hat{\theta}_{d+f}$ in the denominator of (A.2) is determined jointly by $\hat{\theta}_d$ and y_f just as S_{d+f} in Definition 1 is a function of S_d and y_f. The predictive likelihood value measures the joint compatibility of y_f and $\hat{\theta}_d$ with a common $\hat{\theta}_{d+f}$.

A further simplification eliminates the need to compute $\hat{\theta}_{d+f}$ for each possible y_f. We can relate $\hat{\theta}_d$ and $\hat{\theta}_{d+f}$ via

(A.3) $$\nabla(y_{d+f};\hat{\theta}_{d+f})' - \nabla(y_{d+f};\hat{\theta}_d)' = H(y_{d+f};\theta)(\hat{\theta}_{d+f}-\hat{\theta}_d) + O_p(m^{-1/2}).$$

Using the fact that $\nabla(y_{d+f};\hat{\theta}_{d+f}) = 0$ and $\nabla(y_d;\hat{\theta}_d) = 0$ (by the definitions of $\hat{\theta}_{d+f}$ and $\hat{\theta}_d$) and the independence of y_d and y_f,

(A.4) $$\hat{\theta}_{d+f} = \hat{\theta}_d - [H(y_{d+f};\theta)]^{-1}\,\nabla(y_f;\hat{\theta}_d)' + O_p(m^{-3/2}).$$

We will use the asymptotic distributions

(A.5a) $$g^a(\hat{\theta}_d;\theta) = -\frac{1}{2}\,(\hat{\theta}_d-\psi_d-\theta)'H(y_d;\theta)(\hat{\theta}_d-\psi_d-\theta),$$

(A.5b) $$g^a(\hat{\theta}_{d+f};\theta) = -\frac{1}{2}\,(\hat{\theta}_{d+f}-\psi_{d+f}-\theta)'H(y_{d+f};\theta)(\hat{\theta}_{d+f}-\psi_{d+f}-\theta),$$

where ψ is the $O(m^{-1})$ bias and $\psi_{d+f} = \psi_d + o(m^{-1})$. We match these quadratic forms with a Taylor series approximation to $g(y_f;\theta)$:

(A.6) $$g(y_f;\theta) = g(y_f;\hat{\theta}_d) - \nabla(y_f;\hat{\theta}_d)(\hat{\theta}_d-\theta)$$
$$+ \frac{1}{2}\,(\hat{\theta}_d-\theta)'H(y_f;\hat{\theta}_d)(\hat{\theta}_d-\theta) + O_p(m^{-3/2}).$$

Finally, substituting (A.4) into (A.5a), adding (A.6), and subtracting (A.5b) leaves three terms that are not constant with respect to y_f:

$$\frac{1}{2}\,\nabla(y_f;\hat{\theta}_d)\;H^{-1}(y_{d+f};\theta)\;\nabla(y_f;\hat{\theta}_d)'$$

$$+ \nabla(y_f;\hat{\theta}_d)\psi_d + \frac{1}{2}(\hat{\theta}_d-\theta)'H(y_f;\hat{\theta}_d)(\hat{\theta}_d-\theta)$$

The first is the basis for $w_1(y_f;\hat{\theta}_d)$ using the estimate $H(y_{d+f};\hat{\theta}_d)$ of $H(y_{d+f};\theta)$. The second and third terms yield $w_2(y_f;\hat{\theta}_d)$ via the estimate $-H(y_d;\hat{\theta}_d)^{-1}$ of $E[(\hat{\theta}_d-\theta)(\hat{\theta}_d-\theta)']$.

Equation (6.1) requires a suitable joint estimate $\tilde{\theta}_{d+f}$ based on $\tilde{\theta}_d$ and y_f. The analytic tractability of the approximate density

$$f^a(\tilde{\theta}_d,\theta) \propto \exp\{-\tfrac{1}{2}(\tilde{\theta}_d-\theta)'V_m^{-1}(\tilde{\theta}_d-\theta\}$$

suggests letting the joint estimate $\tilde{\theta}_{d+f}$ be computed by maximizing

$$g(\hat{\theta}_d,y_f;\theta) = g(y_f;\theta) - \tfrac{1}{2}(\tilde{\theta}_d-\theta)'V_m^{-1}(\tilde{\theta}_d-\theta)$$

over θ for fixed y_f and $\hat{\theta}_d$. An asymptotic expansion similar to (A.3) shows that

(A.8) $$\tilde{\theta}_{d+f} = \tilde{\theta}_d + [V_m^{-1}-H(y_f;\tilde{\theta}_d)]^{-1}\nabla(y_f;\tilde{\theta}_d)' + O_p(m^{-3/2}).$$

(6.1) follows from the previous analysis with (A.4) replaced by (A.8).

Appendix B

Proof of Proposition 4

We would like to apply Proposition 3 for

(B.1) $$h = \frac{1}{2}\nabla_f H_{d+f}^{-1}\nabla_f' + \frac{1}{2}\,\mathrm{tr}\, H_f H_d^{-1}.$$

For simplicity, we are letting the function arguments be implicit. Let $Z = \mathrm{vec}H_d^{-1}$, and note that $H_{d+f}^{-1} = H_d^{-1} + o_p(m^{-1})$.

The term $\frac{1}{2}\underline{h}^2$ equals $\frac{1}{8}Z'E[A]Z$, where

(B.2) $A = (\nabla'\otimes\nabla')(\nabla\otimes\nabla) + \nabla'\otimes\nabla'\otimes(\mathrm{vec}H)' + (\mathrm{vec}H)\otimes\nabla\otimes\nabla + (\mathrm{vec}H)(\mathrm{vec}H)'$.

Note that $\nabla H\nabla'\nabla H\nabla = Z'(\nabla'\otimes\nabla')(\nabla\otimes\nabla)Z$ (Neudecker (1969)).[8] Generalizing Pfanzagl (1973, p. 997], $E[(\nabla'\otimes\nabla')\nabla]_{\theta}$ can be written as

$$E[(\nabla'\otimes\nabla')(\nabla\otimes\nabla)] + E[(\nabla'\otimes\nabla')(\mathrm{vec}H)'] + E[(\mathrm{vec}H)(\nabla\otimes\nabla)] + E[\nabla'\otimes H\otimes\nabla]$$

so that

(B.3) $$E[A] = E[(\mathrm{vec}H)(\mathrm{vec}H)'] - E[\nabla'\otimes H\otimes\nabla] + E[(\nabla'\otimes\nabla')\nabla]_{\theta} .$$

The term $\frac{1}{2}\,\mathrm{tr}[\ V(\hat{\theta}_{d})\ \underline{h}_{\theta\theta}\ ]$ equals $\frac{1}{4}\ Z'E[B]\ Z$, where

(B.4) $$B = 2(\mathrm{vec}H)(\mathrm{vec}H)' + G\otimes\nabla + \nabla'\otimes G' + F$$

and G and F denote the third and fourth derivative matrices. Differentiating E[H] twice yields

$$E[H]_{\theta\theta} = E[F] + E[G\otimes\nabla] + E[\nabla'\otimes G'] + E[(\mathrm{vec}H)(\mathrm{vec}H)'] + E[\nabla'\otimes H\otimes\nabla].$$

E[B] thus reduces to

(B.5) $$E[B] = E[\ (\mathrm{vec}H)(\mathrm{vec}H)' - \nabla'\otimes H\otimes\nabla\ ] + E[H]_{\theta\theta}.$$

Subtracting twice (B.5) from (B.3) yields (4.19). Parts B and C are immediate consequences of (4.20).

End of proof.

---

[8] There are several possible arrangements of higher order derivatives. We are working with arrangements that are compatible with a square fourth derivative matrix.

References

Akaike, H.(1973): "Information Theory and an Extension of the Maximum Likelihood Principle," in B.N. Petrov and F. Csaki, eds., Second International Symposium on Information Theory (Budapest: Akademiai Kiado).

Baillie, R.T.(1981): "Prediction from the Dynamic Simultaneous Equation Model with Vector Autoregressive Errors," Econometrica, 49 , 1331-1337.

Bianchi, C. and G. Calzolari (1980): "The One-Period Forecast Errors in Nonlinear Econometric Models," International Economic Review 21 , 201-208.

Billingsley, P.(1979): Probability and Measure, 1st Edition (New York: John Wiley).

Brown, B. and R. Mariano (1985): "Asymptotic Behavior of Predictors in Dynamic Nonlinear Simultaneous Systems," Mimeo .

Brown, B. and R. Mariano (1984): "Residual-Based Procedures for Prediction and Estimation of a Nonlinear Simultaneous System," Econometrica, 52 , 321-343.

Butler, R.W.(1986): "Predictive Likelihood Inference with Applications," Journal of the Royal Statistical Society, Series B, 48, 1-38.

Cooley, T. F. and W. R. Parke(1987): "Likelihood and Other Approaches to Prediction in Dynamic Models," Journal of Econometrics, 35 , 119-142.

Cooley, T.F., W.R. Parke and S. Chib (1987): "Predictive Efficiency for Simple Non-Linear Models" reproduced, University of Rochester, forthcoming Journal of Econometrics.

Cox, D.R. and D.V. Hinkley (1974): Theoretical Statistics (London: Chapman and Hall).

Efron, B.(1975): "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)," Annals of Statistics, 3, 1189-1242.

Engle R.F. (1983): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," Econometrica 50, 987-1007.

Fair, R.C. (1980):"Estimating the Expected Predictive Accuracy of Econometric Models," International Economic Review 21, 355-378.

Hinkley, D. (1979):"Predictive Likelihood," The Annals of Statistics, 7, 718-728.

Kullback, S.(1959): Information Theory and Statistics (New York: John Wiley and Sons).

Lauritzen, S.L.(1974): "Sufficiency, Prediction and Extreme Models," Scandinavian Journal of Statistics, 1 , 128-134.

Lehman, E.L. (1983): Theory of Point Estimation (New York: John Wiley & Sons).

Mariano, R. and B. Brown (1983): "Asymptotic Behavior of Predictors in a Nonlinear Simultaneous System," International Economic Review, 21, 523-536.

Muench, T., A. Rolnick, N. Wallace, and W. Weiler (1974): "Tests for Structural Change and Prediction Intervals for the Reduced Forms of Two Structural Models of the U.S.," Annals of Economic and Social Measurement, 3/3, 491-519.

Neudecker, H.(1969): "Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products," Journal of the American Statistical Association, 64, 953-963.

Pfanzagl, J.(1973): "Asymptotic Expansions Related to Minimum Contrast Estimators," Annals of Statistics, 1, 993-1026.

Phillips, P.C.B. (1979):" " Journal of Econometrics, .

Rao, C.R.(1963): "Criteria of Estimation in Large Samples," Sankhyā, 25, 189-206.

Serfling, R.J.(1980): Approximation Theorems of Mathematical Statistics (New York: John Wiley & Sons).

Rochester Center for Economic Research
University of Rochester
Department of Economics
Rochester, NY 14627

1987–88 DISCUSSION PAPERS

WP#68 RECURSIVE UTILITY AND OPTIMAL CAPITAL ACCUMULATION, I: EXISTENCE, by Robert A. Becker, John H. Boyd III, and Bom Yong Sung, January 1987

WP#69 MONEY AND MARKET INCOMPLETENESS IN OVERLAPPING–GENERATIONS MODELS, by Marianne Baxter, January 1987

WP#70 GROWTH BASED ON INCREASING RETURNS DUE TO SPECIALIZATION by Paul M. Romer, January 1987

WP#71 WHY A STUBBORN CONSERVATIVE WOULD RUN A DEFICIT: POLICY WITH TIME–INCONSISTENT PREFERENCES by Torsten Persson and Lars E.O. Svensson, January 1987

WP#72 ON THE CONTINUUM APPROACH OF SPATIAL AND SOME LOCAL PUBLIC GOODS OR PRODUCT DIFFERENTIATION MODELS by Marcus Berliant and Thijs ten Raa, January 1987

WP#73 THE QUIT–LAYOFF DISTINCTION: GROWTH EFFECTS by Kenneth J. McLaughlin, February 1987

WP#74 SOCIAL SECURITY, LIQUIDITY, AND EARLY RETIREMENT by James A. Kahn, March 1987

WP#75 THE PRODUCT CYCLE HYPOTHESIS AND THE HECKSCHER–OHLIN–SAMUELSON THEORY OF INTERNATIONAL TRADE by Sugata Marjit, April 1987

WP#76 NOTIONS OF EQUAL OPPORTUNITIES by William Thomson, April 1987

WP#77 BARGAINING PROBLEMS WITH UNCERTAIN DISAGREEMENT POINTS by Youngsub Chun and William Thomson, April 1987

WP#78 THE ECONOMICS OF RISING STARS by Glenn M. MacDonald, April 1987

WP#79 STOCHASTIC TRENDS AND ECONOMIC FLUCTUATIONS by Robert King, Charles Plosser, James Stock, and Mark Watson, April 1987

WP#80 INTEREST RATE SMOOTHING AND PRICE LEVEL TREND–STATIONARITY by Marvin Goodfriend, April 1987

WP#81 THE EQUILIBRIUM APPROACH TO EXCHANGE RATES by Alan C. Stockman, revised, April 1987

WP#82 INTEREST-RATE SMOOTHING
by Robert J. Barro, May 1987

WP#83 CYCLICAL PRICING OF DURABLE LUXURIES
by Mark Bils, May 1987

WP#84 EQUILIBRIUM IN COOPERATIVE GAMES OF POLICY FORMULATION
by Thomas F. Cooley and Bruce D. Smith, May 1987

WP#85 RENT SHARING AND TURNOVER IN A MODEL WITH EFFICIENCY UNITS OF HUMAN CAPITAL
by Kenneth J. McLaughlin, revised, May 1987

WP#86 THE CYCLICALITY OF LABOR TURNOVER: A JOINT WEALTH MAXIMIZING HYPOTHESIS
by Kenneth J. McLaughlin, revised, May 1987

WP#87 CAN EVERYONE BENEFIT FROM GROWTH? THREE DIFFICULTIES
by Herve' Moulin and William Thomson, May 1987

WP#88 TRADE IN RISKY ASSETS
by Lars E.O. Svensson, May 1987

WP#89 RATIONAL EXPECTATIONS MODELS WITH CENSORED VARIABLES
by Marianne Baxter, June 1987

WP#90 EMPIRICAL EXAMINATIONS OF THE INFORMATION SETS OF ECONOMIC AGENTS
by Nils Gottfries and Torsten Persson, June 1987

WP#91 DO WAGES VARY IN CITIES? AN EMPIRICAL STUDY OF URBAN LABOR MARKETS
by Eric A. Hanushek, June 1987

WP#92 ASPECTS OF TOURNAMENT MODELS: A SURVEY
by Kenneth J. McLaughlin, July 1987

WP#93 ON MODELLING THE NATURAL RATE OF UNEMPLOYMENT WITH INDIVISIBLE LABOR
by Jeremy Greenwood and Gregory W. Huffman

WP#94 TWENTY YEARS AFTER: ECONOMETRICS, 1966-1986
by Adrian Pagan, August 1987

WP#95 ON WELFARE THEORY AND URBAN ECONOMICS
by Marcus Berliant, Yorgos Y. Papageorgiou and Ping Wang, August 1987

WP#96 ENDOGENOUS FINANCIAL STRUCTURE IN AN ECONOMY WITH PRIVATE INFORMATION
by James Kahn, August 1987

WP#97 THE TRADE-OFF BETWEEN CHILD QUANTITY AND QUALITY: SOME EMPIRICAL EVIDENCE
by Eric Hanushek, September 1987

WP#98 SUPPLY AND EQUILIBRIUM IN AN ECONOMY WITH LAND AND PRODUCTION
by Marcus Berliant and Hou-Wen Jeng, September 1987

WP#99 AXIOMS CONCERNING UNCERTAIN DISAGREEMENT POINTS FOR 2-PERSON BARGAINING PROBLEMS
by Youngsub Chun, September 1987

WP#100 MONEY AND INFLATION IN THE AMERICAN COLONIES: FURTHER EVIDENCE ON THE FAILURE OF THE QUANTITY THEORY
by Bruce Smith, October 1987

WP#101 BANK PANICS, SUSPENSIONS, AND GEOGRAPHY: SOME NOTES ON THE "CONTAGION OF FEAR" IN BANKING
by Bruce Smith, October 1987

WP#102 LEGAL RESTRICTIONS, "SUNSPOTS", AND CYCLES
by Bruce Smith, October 1987

WP#103 THE QUIT-LAYOFF DISTINCTION IN A JOINT WEALTH MAXIMIZING APPROACH TO LABOR TURNOVER
by Kenneth McLaughlin, October 1987

WP#104 ON THE INCONSISTENCY OF THE MLE IN CERTAIN HETEROSKEDASTIC REGRESSION MODELS
by Adrian Pagan and H. Sabau, October 1987

WP#105 RECURRENT ADVERTISING
by Ignatius J. Horstmann and Glenn M. MacDonald, October 1987

WP#106 PREDICTIVE EFFICIENCY FOR SIMPLE NONLINEAR MODELS
by Thomas F. Cooley, William R. Parke and Siddhartha Chib, October 1987

WP#107 CREDIBILITY OF MACROECONOMIC POLICY: AN INTRODUCTION AND A BROAD SURVEY
by Torsten Persson, November 1987

WP#108 SOCIAL CONTRACTS AS ASSETS: A POSSIBLE SOLUTION TO THE TIME-CONSISTENCY PROBLEM
by Laurence Kotlikoff, Torsten Persson and Lars E. O. Svensson, November 1987

WP#109 EXCHANGE RATE VARIABILITY AND ASSET TRADE
by Torsten Persson and Lars E. O. Svensson, Novmeber 1987

WP#110 MICROFOUNDATIONS OF INDIVISIBLE LABOR
by Vittorio Grilli and Richard Rogerson, November 1987

WP#111 FISCAL POLICIES AND THE DOLLAR/POUND EXCHANGE RATE: 1870-1984
by Vittorio Grilli, November 1987

WP#112 INFLATION AND STOCK RETURNS WITH COMPLETE MARKETS
by Thomas Cooley and Jon Sonstelie, November 1987

WP#113 THE ECONOMETRIC ANALYSIS OF MODELS WITH RISK TERMS
by Adrian Pagan and Aman Ullah, December 1987

WP#114 PROGRAM TARGETING OPTIONS AND THE ELDERLY
by Eric Hanushek and Roberton Williams, December 1987

WP#115 BARGAINING SOLUTIONS AND STABILITY OF GROUPS
by Youngsub Chun and William Thomson, December 1987

WP#116 MONOTONIC ALLOCATION MECHANISMS
by William Thomson, December 1987

WP#117 MONOTONIC ALLOCATION MECHANISMS IN ECONOMIES WITH PUBLIC GOODS
by William Thomson, December 1987

WP#118 ADVERSE SELECTION, AGGREGATE UNCERTAINTY, AND THE ROLE FOR MUTUAL INSURANCE COMPANIES
by Bruce D. Smith and Michael J. Stutzer, February 1988

WP#119 INTEREST ON RESERVES AND SUNSPOT EQUILIBRIA: FRIEDMAN'S PROPOSAL RECONSIDERED
by Bruce D. Smith, February 1988

WP#120 INTERNATIONAL FINANCIAL INTERMEDIATION AND AGGREGATE FLUCTUATIONS UNDER ALTERNATIVE EXCHANGE RATE REGIMES
by Jeremy Greenwood and Stephen D. Williamson, February 1988

WP#121 FINANCIAL DEREGULATION, MONETARY POLICY, AND CENTRAL BANKING
by Marvin Goodfriend and Robert G. King, February 1988

WP#122 BANK RUNS IN OPEN ECONOMIES AND THE INTERNATIONAL TRANSMISSION OF PANICS
by Peter M. Garber and Vittorio U. Grilli, March 1988

WP#123 CAPITAL ACCUMULATION IN THE THEORY OF LONG RUN GROWTH
by Paul M. Romer, March 1988

WP#124 FINANCIAL INTERMEDIATION AND ENDOGENOUS GROWTH
by Valerie R. Bencivenga and Bruce D. Smith, March 1988

WP#125 UNEMPLOYMENT, THE VARIABILITY OF HOURS, AND THE PERSISTENCE OF "DISTURBANCES": A PRIVATE INFORMATION APPROACH
by Bruce D. Smith, March 1988

WP#126 WHAT CAN BE DONE WITH BAD SCHOOL PERFORMANCE DATA?
by Eric Hanushek and Lori Taylor, March 1988

WP#127 EQUILIBRIUM MARKETING STRATEGIES: IS THERE ADVERTISING, IN TRUTH?
by Ignatius Horstmann and Glenn MacDonald, revised, March 1988

WP#128 REAL EXCHANGE RATE VARIABILITY UNDER PEGGED AND FLOATING NOMINAL EXCHANGE RATE SYSTEMS: AN EQUILIBRIUM THEORY
by Alan C. Stockman, April 1988

WP#129 POST-SAMPLE PREDICTION TESTS FOR GENERALIZED METHOD OF MOMENT ESTIMATORS
by Dennis Hoffman and Adrian Pagan, April 1988

WP#130 GOVERNMENT SPENDING IN A SIMPLE MODEL OF ENDOGENOUS GROWTH
by Robert J. Barro, May 1988

WP#131 FINANCIAL DEVELOPMENT, GROWTH, AND THE DISTRIBUTION OF INCOME
by Jeremy Greenwood and Boyan Jovanovic, May 1988

WP#132 EMPLOYMENT AND HOURS OVER THE BUSINESS CYCLE
by Jang-Ok Cho and Thomas F. Cooley, May 1988

WP#133 A REFINEMENT AND EXTENSION OF THE NO-ENVY CONCEPT
by Dimitrios Diamantaras and William Thomson, May 1988

WP#134 NASH SOLUTION AND UNCERTAIN DISAGREEMENT POINTS
by Youngsub Chun and William Thomson, May 1988

WP#135 NON-PARAMETRIC ESTIMATION AND THE RISK PREMIUM
by Adrian Pagan and Y. Hong, May 1988

WP#136 CHARACTERIZING THE NASH BARGAINING SOLUTION WITHOUT PARETO-OPTIMALITY
by Terje Lensberg and William Thomson, May 1988

WP#137 SOME SIMULATION STUDIES OF NON-PARAMETRIC ESTIMATORS
by Y. Hong and A. Pagan, June 1988

WP#138 SELF-FULFILLING EXPECTATIONS, SPECULATIVE ATTACKS AND CAPITAL CONTROLS
by Harris Dellas and Alan C. Stockman, June 1988

WP#139 APPROXIMATING SUBOPTIMAL DYNAMIC EQUILIBRIA: AN EULER EQUATION APPROACH
by Marianne Baxter, June 1988

WP#140 BUSINESS CYCLES AND THE EXCHANGE RATE SYSTEM: SOME INTERNATIONAL EVIDENCE
by Marianne Baxter and Alan C. Stockman, June 1988

WP#141 RENT SHARING IN AN EQUILIBRIUM MODEL OF MATCHING AND TURNOVER
by Kenneth J. McLaughlin, June 1988

WP#142 CO-MOVEMENTS IN RELATIVE COMMODITY PRICES AND INTERNATIONAL CAPITAL FLOWS: A SIMPLE MODEL
by Ronald W. Jones, July 1988

WP#143 WAGE SENSITIVITY RANKINGS AND TEMPORAL CONVERGENCE
by Ronald W. Jones and Peter Neary, July 1988

WP#144 FOREIGN MONOPOLY AND OPTIMAL TARIFFS FOR THE SMALL OPEN ECONOMY
by Ronald W. Jones and Shumpei Takemori, July 1988

WP#145 THE ROLE OF SERVICES IN PRODUCTION AND INTERNATIONAL TRADE: A THEORETICAL FRAMEWORK
by Ronald W. Jones and Henryk Kierzkowski, July 1988

WP#146 APPRAISING THE OPTIONS FOR INTERNATIONAL TRADE IN SERVICES
by Ronald W. Jones and Frances Ruane, July 1988

WP#147 SIMPLE METHODS OF ESTIMATION AND INFERENCE FOR SYSTEMS CHARACTERIZED BY DETERMINISTIC CHAOS
by Mahmoud El-Gamal, August 1988

WP#148 THE RICARDIAN APPROACH TO BUDGET DEFICITS
by Robert J. Barro, August 1988

WP#149 A MODEL OF NOMINAL CONTRACTS
by Bruce D. Smith, August 1988

WP#150 A BUSINESS CYCLE MODEL WITH PRIVATE INFORMATION
by Bruce D. Smith, August 1988

WP#151 ASYMPTOTIC LIKELIHOOD BASED PREDICTION FUNCTIONS
by Thomas F. Cooley, August 1988

To order copies of the above papers complete the attached invoice and return to Christine Massaro, W. Allen Wallis Institute of Political Economy, RCER, 109B Harkness Hall, University of Rochester, Rochester, NY 14627. Three (3) papers per year will be provided free of charge as requested below. Each additional paper will require a \$5.00 service fee which must be enclosed with your order. For your convenience an invoice is provided below in order that you may request payment from your institution as necessary. Please make your check payable to the **Rochester Center for Economic Research. Checks must be drawn from a U.S. bank and in U.S. dollars**.

W. Allen Wallis Institute for Political Economy

Rochester Center for Economic Research, Working Paper Series

OFFICIAL INVOICE

Requestor's Name ______________________________

Requestor's Address ______________________________

____

____

____

Please send me the following papers free of charge (**Limit: 3 free per year**).

WP# ______ WP# ______ WP# ______

I understand there is a \$5.00 fee for each additional paper. Enclosed is my check or money order in the amount of \$__________. Please send me the following papers.

WP# ______ WP# ______ WP# ______

WP# ______ WP# ______ WP# ______

WP# ______ WP# ______ WP# ______

WP# ______ WP# ______ WP# ______