EFFICIENT SEMIPARAMETRIC PREDICTION INTERVALS AND REGIONS

Bryan W. Brown[1]

The construction of prediction intervals and regions and their probability content for nonlinear systems with nonparametric disturbances is considered. The semiparametric efficiency bound for estimating the probability content of a known interval (region) and estimators that attain the bound are developed. Semiparametric efficient estimation of optimal prediction intervals (regions) which either (i) maximize probability content given interval length (region area) or (ii) maximize interval length (region area) given probability content is studied. The estimated probability content of (i) is found to have the same limiting behavior as if the interval (region) were known with certainty and hence attains the semiparametric efficiency bound. Further, the estimated probability of the estimated interval (region) approximates the true coverage probability to order $\sqrt{n}$ for (i) but order smaller than $\sqrt{n}$ for (ii). A Monte Carlo experiment is conducted to compare the new predictors to competitors.

Keywords: Semiparametric efficiency bound, optimal prediction intervals, prediction regions, nonlinear systems

1. Introduction

The conditional prediction problem involves developing knowledge of the distribution of the predictand variable(s) outside the sample period given certain conditioning information. The attribute of this distribution that has received the most attention is its center, as manifested by, say, the conditional mean. This point prediction problem is well-studied for linear models, where the problem reduces to finding the center of the distribution of the disturbances, typically zero, and substituting parameter estimates into the conditional mean function. In nonlinear (in the variables) models, this problem involves more complete knowledge of the distribution of the disturbances, and has been addressed through the use of simulation techniques to estimate the conditional mean using draws from an estimate of the distribution of the disturbances. This distribution could either be parametric as in Howrey and Kelejian (1971) or nonparametric as in Brown and Mariano (1984).

Beyond point prediction, we are interested in determining a range of values of the predictand variables and some measure of the probability of falling in the range. Fixing the probability at a given value, the problem becomes one of developing an appropriate interval in the univariate case and an appropriate region in the multivariate case. Again, for linear models, a great deal is known if, in addition, the disturbances are assumed to be normal. In this case, the predictand is conditionally normal and completely characterized by its mean and covariance matrix and the construction of prediction intervals and regions is straightforward and well-studied. In particular, the target intervals and regions can be represented as known functions of estimated parameters that are appropriate, at least asymptotically and, for some cases, in finite samples. Similarly, if the distribution of the disturbances is nonnormal but still parametrically specified, the intervals and regions can generally be represented as known functions of estimated parameters that are, at least asymptotically, appropriate.

Unfortunately, if the model is nonlinear in the variables or the distribution of the disturbances is nonparametric then the construction of the intervals and regions is somewhat

[1] Partial support from NSF grant SES-9905816 is gratefully acknowledged. The author wishes to thank Mahmoud El-Gamal and especially Roberto Mariano, my co-author on the related 1991 paper, for helpful discussions.

more complicated and less well studied. Although the construction of prediction intervals and regions is an obviously important topic and most predictive models are nonlinear in the variables, there has been very little work concerning the behavior of prediction intervals and regions in nonlinear simultaneous systems. Likewise there has been very little work on prediction intervals and regions in linear models when the distribution is not specified. An exception is the unpublished paper by Brown and Mariano (1991), which considers Monte Carlo simulation-based prediction intervals and regions when the distribution of the disturbances is known and residual-based prediction intervals and regions when the distribution in not specified.

The purpose of this paper is to develop techniques appropriate for construction of prediction intervals and regions when the disturbances are nonparametric and the structural model is nonlinear. The research presented in this paper builds on the residual-based techniques in Brown and Mariano, which are appropriate for models where the disturbances are independent of the regressors. The approach introduced in Brown and Newey (1998) for semiparametric efficient estimation of expectations is applied to develop semiparametric efficient estimates of the probability content of known intervals and regions for parametric models with nonparametric disturbance distribution. Beyond known intervals and regions, intervals and regions that have asymptotic optimal properties in the sense of minimal area for a given probability content or maximal probability content for a given area are developed. Although explicitly developed for nonlinear systems, the techniques should apply equally well to linear systems with nonparametric disturbance distribution.

The outline of the paper is as follows. In the second section, the basic model is introduced and relevant previous results reviewed. The construction and properties of efficient prediction intervals for semiparametric models, including optimal intervals, are presented in the third section. In the fourth section, efficient semiparametric prediction regions, including optimal regions, are developed. The relative behavior of the proposed optimal predictors and their competitors are presented and contrasted via a Monte Carlo study in the fifth section. The results are summarized and a number of promising possible extensions are discussed in the final section. For the purposes of this paper construction of a prediction interval or region is construed to include both the estimation of the interval or region given a probability content and estimation of the probability content given the interval or region.

2. Basic Concepts

In this paper, we will consider prediction in a static nonlinear system with independent errors. The data generating process for such a system can be represented

$$y = \pi(\varepsilon; x; \beta) \tag{1}$$

where y is a $g \times 1$ vector of endogenous variables, x is a $k \times 1$ vector of exogenous variables, ε is a $g \times 1$ disturbance vector, $\pi(\cdot)$ is a $g \times 1$ vector of known functions, and β is a $p \times 1$ vector of unknown parameters. The vector of driving variables (ε', x') are assumed to be jointly $i.i.d.$ with a joint distribution that satisfies independence of ε and x but is otherwise unspecified and unrestricted, except for some smoothness restrictions. Note the distribution of the disturbances ε are allowed to have a nonzero location parameter, say $\alpha = E[\varepsilon]$, so the parameter vector β does not include an intercept.

We assume that the relationship between y and ε, given x, is one-to-one. Thus, we will restrict our attention to models which have the unique inverse representation

$$\varepsilon = \rho(y; x; \beta): \tag{2}$$

where $\rho(\cdot)$ is a known function. This inverse representation is usually interpreted as the structural form of the model while the data generating process is the reduced form. A number of models fall within this framework including the linear and nonlinear regression models and the linear and nonlinear simultaneous equation models. In practice, the reduced form corresponding to a particular structure may not be available in closed form but can be obtained through numerical techniques.

We are interested in the conditional prediction of the endogenous variables for some outside sample observation, denoted by subscript τ, given the values of the exogenous variables. The prediction approaches introduced below will depend crucially on estimation of conditional expectations of known functions of the endogenous variables y_τ given the exogenous variables x_τ. Due to the independence assumption, such expectations have the canonical representation

$$\begin{aligned} \mu(\beta; h) &= E_{\beta;h}[g(y)|x = x_\tau] = \int g(\pi(\varepsilon; x_\tau; \beta)) f_\varepsilon(\varepsilon|h) d\varepsilon \\ &= \int g(\pi(\rho(z; \beta); x_\tau; \beta)) f_z(z|\beta; h) dz \\ &= \int m(z; \beta) f_z(z|\beta; h) dz \end{aligned} \tag{3}$$

where $g(\cdot)$ and hence $m(\cdot) = g(\pi(\rho(\cdot, \beta); x_\tau; \beta))$ are known $q \times 1$ functions, $f_\varepsilon(\cdot|h)$ is the density of ε, $f_z(z|\beta; h)$ is the density of z, and h is an unknown function to reflect the distribution free nature of the specification. The last line is the canonical expectation studied in Brown and Newey (1998) with restrictions implied by the form in the second line.

The essential complications in the estimation of μ are the unavailability of β and the inability to perform the indicated integration since the distribution is unspecified. The most natural response to these complications is to estimate β with, say, $\hat{\beta}$ and approximate the integral with an average, which does not necessitate specifying the distribution. Specifically, we propose the method of moments estimator

$$\hat\mu = n^{-1}\sum_{t=1}^{n} g(\pi(\hat\epsilon_t; x_\tau; \hat\alpha)) = n^{-1}\sum_{t=1}^{n} g(\pi(\rho(z_t;\hat\beta); x_\tau; \hat\alpha)), \tag{4}$$

where $\hat\epsilon_t = \rho(z_t;\hat\beta)$. The functions $g(\cdot)$ and $\pi(\cdot;\beta)$ are unrestricted except for some smoothness in the expectation of the latter, which will be imposed below. This is the residual-based estimator of the target expectation proposed by Brown and Mariano (1984).

It is instructive to examine several examples of residual-based estimators considered by Brown and Mariano. For point prediction we are interested in $\gamma(x_\tau) = E[y_\tau|x_\tau]$, whereupon $g(y) = y$, and the method of moments estimator is given by

$$\hat\gamma(x_\tau) = n^{-1}\sum_{i=1}^{n} \pi(\rho(z_i;\hat\beta); x_\tau; \hat\alpha). \tag{5}$$

In measuring predictive accuracy, the second conditional moment $\sigma(x_\tau) = E[(y_\tau - \gamma(x_\tau))(y_\tau - \gamma(x_\tau))|x_\tau]$ is important and may be estimated by

$$\hat\sigma(x_\tau) = n^{-1}\sum_{i=1}^{n} (\pi(\rho(z_i;\hat\beta); x_\tau; \hat\alpha) - \hat\gamma(x_\tau))(\cdot)'. \tag{6}$$

And the conditional distribution function $F_1(c;x_\tau) = E[1(y_1 \le c)|x_\tau]$, which is of direct interest below, has $g(y) = 1(y_1 \le c)$ with

$$\hat F_1(c;x_\tau) = n^{-1}\sum_{i=1}^{n} 1(\pi_1(\rho(z_i;\hat\beta); x_\tau; \hat\alpha) \le c) \tag{7}$$

as its method of moments estimator.

Suppose that the data are generated by a parametric model which satisfies the semiparametric assumptions and contains the truth. Such a model is called a parametric submodel since it is a subset of the model consisting of distributions satisfying the assumptions. Formally, we suppose

$$z \sim f(z|\beta_0; h(\eta_0)) \tag{8}$$

where η_0 is a finite-length vector of shape parameters for the true distribution $f(\cdot)$, and a zero subscript indicates the true parameter value. The set of parametric submodels, then, is defined as the set of distributions which satisfy the semiparametric assumptions and

$$f^i(z|\beta; h^i(\eta^i)) = f(z|\beta_0; h(\eta_0)) \tag{9}$$

for some $\theta^i = (\beta', \eta^{i\prime})' = (\beta_0', \eta_0^{i\prime})' = \theta_0^{i\prime}$ and all z, where η^i is a shape parameter for parametric submodel i. Note that the length of the shape parameter vector η^i and hence θ^i may differ for different parametric submodels.

Projections onto the spaces spanned by the scores of the parametric submodels are important in determining the semiparametric estimators and efficiency bounds. Let $S_\theta^i(z) = (S_\beta^i(z)', S_\eta^i(z)')'$ denote the scores of a parametric submodel. Define the nonparametric tangent set as the mean square closure of the union of all possible q-dimensional linear combinations of $S_\eta(z)$, i.e.

$$\mathcal{T} = \{t \in R^q : E[t't] < \infty;\ \exists B_j; S_{\eta j}(z)\ \text{s.t.}\ E[\|t - B_j S_{\eta j}(z)\|^2] = o(1)\}, \tag{10}$$

where B_j are constant matrices with q rows and $E[\cdot]$ denotes expectation at the truth. Newey (1989) has previously studied the estimation of the parameter vector β for the present model under the independence assumption and shown that the nonparametric tangent set is given by

$$T = \{t_1(\varepsilon) + t_2(x) : E[t_1(\varepsilon)] = E[t_2(x)] = 0\}; \tag{11}$$

where $t_1(\cdot)$ and $t_2(\cdot)$ are unrestricted functions except for the mean zero property. Note that the residual of the projection of the score $S_\beta(z)$, for any parametric submodel which includes the truth, on the nonparametric tangent set,

$$S(z) = S_\beta(z) - \mathrm{Proj}(S_\beta(z)|T); \tag{12}$$

is known as the efficient score for β, where $\mathrm{Proj}(g(z)|T)$ denotes the projection of $g(z)$ on T.

Similarly, define the tangent set $\mathcal{S}$ as the mean square closure of the union of all q-dimensional linear combinations of $S_\mu(z)$ for all regular parametric submodels satisfying the semiparametric assumptions, i.e.

$$\mathcal{S} = \{s \in R^q : E[s's] < \infty;\ \exists A_j, S_{\mu_j}(z) \text{ s.t. } E[\|s - A_j S_{\mu_j}(z)\|^2] = o(1)\}, \tag{13}$$

where A_j are constant matrices with q rows. As might be expected there is a close relationship between the two tangent sets. More compactly, we can write $\mathcal{S} = \{B s_\beta + \tau : \tau \in T$ and B is a constant $q \times p$ matrix$\}$. By definition, $B s_\beta = B s + B\,\mathrm{Proj}(s_\beta|T) = B s + \tau$ for $\tau \in T$, whereupon $\mathcal{S} = \{B s + \tau : \tau \in T\}$. Note that the two components are orthogonal, which implies that a projection onto $\mathcal{S}$ can be obtained as the sum of the projection onto the two components.

Since a distribution is not explicitly specified in obtaining $\hat{\theta}$, the estimator will be semiparametric if $\hat{\beta}$ is semiparametric. Specifically, $\hat{\beta}$ should remain consistent for any distribution satisfying the semiparametric assumptions. Accordingly, we make the following assumptions and obtain the accompanying Theorem. Proofs are given in the Appendix.

Assumption 1: $\hat{\beta}$ is asymptotically linear with influence function $\psi_\beta(z)$, $E[\psi_\beta(z)] = 0$, and $V_\beta = E[\psi_\beta(z)\psi_\beta(z)']$ finite.

Assumption 2: $M(\beta) = \partial E[m(z;\beta)]/\partial\beta'$ exists and continuous on a neighborhood of β_0.

Assumption 3: $n^{-1/2}\sum_{t=1}^{n}[\{m(z_t;\beta) - E[m(z;\beta)]\} - \{m(z_t;\beta_0) - \theta_0\}]$ stochastically equicontinuous at $\beta = \beta_0$.

Assumption 4: $V_m = E[(m(z;\beta_0) - \theta_0)(m(z;\beta_0) - \theta_0)']$ exists and finite.

Theorem 1: Suppose Assumptions 1-4 are satisfied, then

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0; V_\theta); \tag{14}$$

where $V_\theta = V_m + M V_\beta M'$ for $M = M(\beta_0)$.

This result demonstrates that the method of moments estimator is consistent and asymptotically normal under fairly standard conditions. Since we are comparing the estimators on the basis of asymptotic variance, Assumption 4, which assumes the existence of a variance is fairly innocuous. It will be satisfied for indicator functions such as used below. Assumption 3, the stochastic equicontinuity assumption, will be met if, for example, $(m(z_t;\beta) - E[m(z;\beta)])$ satisfies a central limit theorem throughout a neighborhood of β_0: In particular, if $m(\cdot)$ is an indicator function, as below, this condition will be met. And the continuous differentiability condition, Assumption 2, is the standard approach for obtaining derivative terms in the asymptotic expansion when the underlying functions are discontinuous.

Based on the limiting covariance matrix, alternative method of moments estimators based on different estimators $\hat{\beta}$ can be ranked in terms of the V_β. This suggests that a lower bound of some sort is attained if the estimator $\hat{\beta}$ is itself semiparametric efficient. The theorem below verifies this conjecture using additional notation and assumptions. For each parametric submodel, define the target parametric function in terms of the underlying parameters

$$\nu(\theta) = \nu(\beta; h(\eta)); \tag{15}$$

where $\theta = (\beta', \eta')'$ and we have dropped the superscript i indexing the various parametric families.

Assumption 5: For all parametric submodels, $E_\theta[\|\psi_\beta(z)\|^2]$ exists and continuous on a neighborhood of θ_0.

Assumption 6 For all regular parametric submodels, $\nu(\theta)$ differentiable and $E_\theta[\|m(z_t;\beta_0) - \nu_0\|^2]$ exists and continuous on a neighborhood of θ_0.

Theorem 2: Suppose Assumptions 1-6 are satisfied, regular $\hat{\beta}$ exists, $E[S\,S']$ exists and nonsingular, and $[V_m + M V_\beta^* M']$ is nonsingular, then $\hat{\nu}$ is regular and attains the semiparametric efficiency bound $V_\nu^* = V_m + M V_\beta^* M'$ for $\hat{\beta}$ semiparametric efficient.

Thus we see that the method of moments estimator attains the semiparametric efficiency bound when based on $\hat{\beta}$ semiparametric efficient, as was conjectured above. The above theorems provides a more direct alternative to related results in Brown and Newey (1998). The results there targeted $E_{\beta,h}[m(z;\beta)]$ with more general forms of $m(\cdot)$ and reduced to the present results under the assumption of independence and $m(z;\beta) = g(\rho(z;\beta);\beta)$. The basic difference is that condition (d) in Theorem 2 there, which guarantees asymptotic independence with respect to the nuisance parameters is not needed in the present context. In addition several of the conditions there can be combined into a single simpler condition. Finally, the direct approach taken here avoids the need to consider the theory of V-statistics, although it may be needed to develop the semiparametric efficient estimator of β.

3. Prediction Intervals

In this section, the estimation of prediction intervals and their probability content is considered. In the presentation of this section, we will not discuss the most usual approaches to constructing intervals such as intervals symmetric around the conditional mean or intervals with equal tail probabilities. Instead, the focus is on the construction and estimation of optimal intervals and regions, which will generally differ from the usual approaches. The approach considered in the following subsection can be easily adapted to handle the construction of intervals symmetric around the mean. In any event, for the cases where the usual approaches make the most sense and turn out to be optimal, such as the linear model with normal disturbances, the optimal approaches introduced below will turn out to be asymptotically equivalent.

3.1 Known Interval, Estimated Probability

We start by investigating the estimation of the probability of a known interval for, without loss of generality, the first endogenous variable. Consider the half-open interval $(\Delta_1; \Delta_2]$, and define

$$\begin{aligned} P^k(\Delta_1; \Delta_2; x_\tau) &= Pr[(\Delta_1 < y_1 \leq \Delta_2)|x_\tau] \\ &= E_z[1(\Delta_1 < \pi_1(\rho(z;\beta_0); x_\tau; \beta_0) \leq \Delta_2)|x_\tau] \\ &= \int 1(\Delta_1 < \pi_1(\rho(z;\beta_0); x_\tau; \beta_0) \leq \Delta_2) f_z(z;\beta_0;h_0) dz \end{aligned} \quad (16)$$

as the probability of y falling in the interval given x_τ. We use the half-open interval because the probability can then be written as the difference in two c.d.f.'s. Of course if the density is continuous, then the difference in the probability content between an open, closed, and half-open interval is zero.

Following Brown and Newey (1998), the efficient estimate of this conditional expectation under the independence assumption is given by the average

$$\hat{P}^k = n^{-1} \sum_{i=1}^{n} 1(\Delta_1 < \pi_1(\hat{z}_i; x_\tau; \hat{\beta}) \leq \Delta_2) \quad (17)$$

where $\hat{z}_i = \rho(y_i; x_i; \hat{\beta})$ and $\hat{\beta}$ is a semiparametrically efficient estimator. And the asymptotic limiting behavior of the estimator is given by application of Theorem 1 is

$$n^{1/2}(\hat{P}^k - P^k) \xrightarrow{d} N(0; P^k(1 - P^k) + \bar{P}^k V_\beta \bar{P}^{k\prime}) \quad (18)$$

where $\bar{P}^k = \partial E_z[1(\Delta_1 < \pi_1(\rho(z;\beta); x_\tau; \beta) \leq \Delta_2)]/\partial\beta'|_{\beta=\beta_0}$. By Theorem 2, the covariance matrix of this limiting distribution represents the semiparametric efficiency bound for estimation of the probability content of the known interval $(\Delta_1; \Delta_2]$ for $V_\beta = V_\beta^*$ and is attained when $\hat{\beta}$ is semiparametric efficient.

3.2 Optimal Probability Interval (Given Length)

If the interval $(\Delta_1; \Delta_2]$ is arbitrarily chosen, then it can likely be improved upon. Specifically, we can often find an interval of similar length $A = \Delta_2 - \Delta_1$ that has higher probability content. Suppose that the distribution of y given x is unimodal, then we can formalize this notion by choosing the interval of given length A that has highest probability content:

$$\max_{c_1;c_2} P^k(c_1;c_2;x_\tau;\beta_0;h_0);\ s.t.:\ c_2 - c_1 = A. \qquad (19)$$

If the conditional density $f_{y|x_\tau}(\cdot|x_\tau;\beta_0;h_0)$ exits and is continuous, then the first order conditions for this optimization are $f_{y|x_\tau}(c_2|x_\tau;\beta_0;h_0) = f_{y|x_\tau}(c_1|x_\tau;\beta_0;h_0)$ together with the side condition $c_2 - c_1 = A$, which implicitly defines the unique solutions $c_1^* = c_1(A;x_\tau;\beta_0;h_0)$ and $c_2^* = c_1^* + A$. Substitution of the optimal interval into the probability function yields $P^*(A;x_\tau;\beta_0;h_0) = P^k(c_1^*;c_1^* + A;x_\tau;\beta_0;h_0)$ as the probability content of the optimal interval.

In order to apply the method of moments approach to estimate the probability content of the optimal interval, we must first estimate the nuisance parameter c_1^*. Let $\hat f_{y|x_\tau}(\cdot x_\tau)$ denote a consistent estmator, such as the kernel, of $f_{y|x_\tau}(\cdot\beta_0;h_0)$ and take $\hat c_1 = \hat c_1(A;x_\tau)$ as the solution to the implicit function $\hat f_{y|x_\tau}(\hat c_1 + A;x_\tau) = \hat f_{y|x_\tau}(\hat c_1;x_\tau)$. Then a feasible estimator of the probability content of the estimated optimal interval can be estimated by

$$\hat P^* = n^{-1}\sum_{i=1}^{n} 1(\hat c_1 < \pi(\rho(y_i;x_i;\hat\beta);x_\tau;\hat\beta) \le \hat c_1 + A). \qquad (20)$$

Interestingly, the limiting distribution of this estimator is the given by

$$n^{1/2}(\hat P^* - P^*) \xrightarrow{d} N(0;P^*(1-P^*) + P_\beta^{*\prime} V P_\beta^*)$$

where $P_\beta^* = \partial E[1(c_1^* < \pi_1(\rho(z;\beta);x_\tau;\beta) \le c_1^* + A)|x_\tau]/\partial\beta'|_{\beta=\beta_0}$, which is the same as if c_1^* were known with certainty. Thus, for $\hat\beta$ semiparametric efficient, $\hat P^*$ is the semiparametric efficient estimator of the probability content of the true optimal interval, which is unknown but consistently estimated by $(\hat c_1; \hat c_1 + A]$

Ultimately, of course, we are interested in the coverage probability of the estimated interval relative to the estimated probability. For y outside the estimation sample, which is appropriate for outside sample prediction, and hence independent of $(\hat c_1; \hat c_1 + A]$, we can show

$$\begin{aligned} \Pr[\hat c_1 < y \le \hat c_1 + A|x_\tau] &= E[E[1(\hat c_1 < \pi(\varepsilon;x_\tau;\beta_0) \le \hat c_1 + A)\,|\,\hat c_1;x_\tau]|x_\tau] \qquad (21)\\ &= P^* + E[o_p(n^{-1/2})] = P^* + o(n^{-1/2}) \end{aligned}$$

provided $\hat f_{y|x_\tau}(\cdot x_\tau) = f_{y|x_\tau}(\cdot x_\tau;\beta_0;h_0) + o_p(n^{-1/4})$ and hence $\hat c_1 = c_1^* + o_p(n^{-1/4})$. Combining the two results, we find that

$$\begin{aligned} \hat P^* &= \Pr[\hat c_1 < y \le \hat c_1 + A\,|x_\tau] + n^{-1/2} N(0;P^*(1-P^*) + P_\beta^{*\prime} V P_\beta^*) + o_p(n^{-1/2}) \qquad (22)\\ &= \Pr[\hat c_1 < y \le \hat c_1 + A\,|x_\tau] + O_p(n^{-1/2}) \end{aligned}$$

with the discrepancy between the true and estimated probability of the estimated interval resulting from estimating the true probability content of the true optimal interval.

Given interval length, the motivation for using the optimal interval for making a probability statement is clear. The complication is that we must estimate the endpoints of the interval as well as the probability content. Although the endpoint estimators converge to their targets at a rate slower than $\sqrt{n}$, the estimated probability $\hat P^*$ will converge to the probability content of the estimated interval at a $\sqrt{n}$ rate. Moreover, it is easy to see that $\sqrt{n}(\hat P^* - \Pr[\hat c_1 < y \le \hat c_1 + A\,|x_\tau])$ will attain a lower bound when $\sqrt{n}(\hat P^* - P^*)$ attains a lower bound. Thus, $\hat{P}^*$ provides a semiparametric efficient estimator for the probability content of the true and estimated optimal intervals.

Attaining a faster than $n^{1/4}$ rate of convergence for the kernel estimator $\hat{f}_{y|x_\tau}(\cdot|x_\tau)$ may be a problem if y_τ and/or x_τ are of high dimension. The dimensionality intoduced by the conditioning variables can be eliminated, however, by using a restricted kernel estimator. Due to the independence assumption, we have $f_{y|x}(y|x) = f_\varepsilon(\rho(y;x;\beta)) \, |\det(\partial\rho(y;x;\beta)/\partial y)|$ and the corresponding estimator

$$\hat{f}_{y|x}(y|x) = \hat{f}_\varepsilon(\rho(y;x;\hat{\beta})) \left| \det\left(\partial\rho\left(y;x;\hat{\beta}\right)/\partial y \right) \right|, \tag{23}$$

where $\hat{f}_\varepsilon(\cdot)$ is the kernel estimator of the density of ε and will not suffer from the dimensionality of x. If ε is a long vector, with more than three elements, then we will need to utilize higher-order kernels to attain the required rate of convergence. An added benefit of using the so-restricted kernel is that the corresponding c.d.f. estimator has the properties of a smoothed unconditional c.d.f. estimator and will have a $\sqrt{n}$ rate of convergence.

3.3 Optimal Length Interval (Given Probability)

The dual to the above optimization with respect to the interval is probably of more interest. Specifically, for a given probability content, we can choose the interval to be minimal length. Continuing to assume unimodal behavior this problem can be formalized as

$$\min_{c_1;c_2} A = c_2 - c_1; \ \ s.t.: P^k(c_1;c_2;x_\tau;\beta_0;h_0) = P \tag{24}$$

which has as first order conditions $f_{y|x}(c_2;x_\tau;\beta_0;h_0) = f_{y|x}(c_1;x_\tau;\beta_0;h_0)$ together with the side condition $P^k(c_1;c_2;x_\tau;\beta_0;h_0) = P$. Let A^* denote the value of A at the minimum, then $c_2^* = c_1^* + A^*$ at the minimum and $c_1^* = c_1(A^*;x_\tau;\beta_0;h_0)$, which is the same as before, by the first-order conditions. Substitution into the side condition yields A^* as the unique solution to the implicit equation

$$P = P^k(c_1(A^*;x_\tau;\beta_0;h_0); c_1(A^*;x_\tau;\beta_0;h_0) + A^*;x_\tau;\beta_0;h_0) \tag{25}$$

while $c_1^* = c_1(A^*;x_\tau;\beta_0;h_0)$ and $c_2^* = c_1^* + A^*$. This is easily seen as the inverse function to the solution of the maximum probability given length problem, presented in the previous subsection, and in a certain sense is a quantile.

The optimal endpoints can be estimated directly by $\hat{c}_1^* = \hat{c}_1(\hat{A}^*;x_\tau)$ and $\hat{c}_2^* = \hat{c}_1^* + \hat{A}^*$ where the corresponding estimated interval length $\hat{A}^*$ solves the implicit system

$$P = \hat{F}_{y|x_\tau}(\hat{c}_1(\hat{A}^*;x_\tau) + \hat{A}^*;x_\tau) - \hat{F}_{y|x_\tau}(\hat{c}_1(\hat{A}^*;x_\tau);x_\tau) \tag{26}$$

and $\hat{F}_{y|x_\tau}(\cdot|x_\tau)$ is the smoothed estimated c.d.f. corresponding to $\hat{f}_{y|x_\tau}(\cdot|x_\tau)$. More directly, we can use the difference in the empirical c.d.f.'s and take $\hat{A}^*$ as the solution to

$$\hat{A}^* = \sup_A \left(n^{-1} \sum_{i=1}^{n} 1(\hat{c}_1(A;x_\tau) < \pi(\rho(y_i;x_i;\hat{\beta});x_\tau;\hat{\beta}) \le \hat{c}_1(A;x_\tau) + A) \le P \right) \tag{27}$$

which is the inverse of $\hat{P}^*$, the estimated probability function, given in the previous subsection. The supremum is used since the empirical probability function is a step function and only asymptotically one-to-one.

In either case, analogous to above, we can show that

$$n^{1/2}(\hat{A}^* - A^*) \xrightarrow{d} N(0; f_{y|x}(c_1^*; x_\tau; \beta_0; h_0)^{-2}\{P(1-P) + P_\beta V_\beta P_\beta\}) \tag{28}$$

where now $P_\beta = \partial E_z[1(c_1^* < \pi_1(\rho(z;\beta); x_\tau; \beta) \le c_1^* + A^*)]/\partial\beta'|_{\beta=\beta_0}$ with the limiting covariance matrix representing a semiparametric efficiency bound for estimation of the optimal length. And for outside sample prediction, provided $\hat{f}_{y|x_\tau}(\cdot|x_\tau) = f_{y|x_\tau}(\cdot|x_\tau; \beta_0; h_0) + o_p(n^{-1/4})$, we similarly find that the probability content of the estimated optimal interval with given probability content is given by

$$\begin{aligned}
\Pr[\hat{c}_1^* < y \le \hat{c}_1^* + \hat{A}^* | x_\tau] &= E[E[1(\hat{c}_1^* < \pi(\varepsilon; x_\tau; \beta_0) \le \hat{c}_1^* + \hat{A}^*) | \hat{c}_1^*; \hat{A}^*; x_\tau] | x_\tau] \\
&= P + E[f_{y|x}(c_1^*; x_\tau; \beta_0; h_0)(\hat{A}^* - A^*) + o_p(n^{-1/2}) | x_\tau] \\
&= P + E[n^{-1/2} N(0; P(1-P) + P_\beta V_\beta P_\beta) + o_p(n^{-1/2}) | x_\tau] = P + o(n^{-1/2}).
\end{aligned} \tag{29}$$

Thus, although the endpoints of the estimated optimal interval $(\hat{c}_1; \hat{c}_2]$ have a slower than $\sqrt{n}$ rate of convergence, the probability content of the estimated interval converges to the ostensible probability at a faster than $\sqrt{n}$ rate.

3.4 Multimodal Distributions

In the above discussion, we assumed that the conditional distribution of y_1 was unimodal. If the distribution is multimodal then the approach will need some modification. Specifically, we must entertain the possibility that the interval will be discontiguous with one sub-interval for each mode. The general optimality approach will still work with either the total length of the intervals set and the probability content maximized or the probability content set and the total length of the intervals minimized. The first-order conditions will be the same with the density the same at all the endpoints of the subintervals. It turns out that there is a more direct approach that works for both the unimodal and multimodal optimal prediction interval and also for choosing the optimal prediction region for the multivariate case. Accordingly, I turn now to the prediction region problem.

4. Efficient Prediction Regions

In this section, the estimation of prediction regions and their probability content is considered. The objective, of course, is to make conditional probability statements regarding a vector of endogenous variables rather than a scalar endogenous variable as in the previous section. As in the previous section, we will only discuss the construction and estimation of optimal intervals. There are both advantages and disadvantages for prediction regions for a complete vector compared to a vector prediction intervals applied to each element of the vector. The advantage is that area of the region implied by the union of the univariate intervals is almost ivariabley larger than an optimally constructed region. The disadvantage is that prediction intervals are much easier to interpret and more easily understood by the uninitiated.

4.1 Known Region

We first examine the estimation of the probability content of a known region. Suppose R denotes some region in the space of feasible values of y. Then, analogous to the interval case, the probability content of the region is given by

$$\begin{aligned} P^k(R; x_\tau; \beta_0; h_0) &= \Pr[(y \in R)|x_\tau] \\ &= E[1(\pi(\varepsilon; x_\tau; \beta_0) \in R)|x_\tau] \\ &= E[1(\pi(\rho(z; \beta_0); x_\tau; \beta_0) \in R)|x_\tau]. \end{aligned} \tag{30}$$

The method of moments estimator of the probability content $P^k(R; x_\tau; \beta_0; h_0)$ is the residual-based estimator $\hat{P}^k = n^{-1}\sum_i 1(\pi(\hat{\varepsilon}_i; x_\tau; \hat{\beta}) \in R)$ which, by Theorem 1, will have the limiting behavior

$$n^{1/2}(\hat{P}^k - P^k) \xrightarrow{d} N(0; P^k(1 - P^k) + \dot{P}^k V_\beta \dot{P}^{k\prime}), \tag{31}$$

where $\dot{P}^k = \partial E[1(\pi(\rho(z; \beta); x_\tau; \beta) \in R)|x_\tau]/\partial\beta'|_{\beta=\beta_0}$. With some regularity, as indicated by Theorem 2, the covariance matrix of this estimator is the semiparametric efficiency bound for estimation of the probability content of the known region when $V_\beta = V_\beta^*$. This efficiency bound will be attained by the method of moments estimator if $\hat{\beta}$ is semiparametric efficient.

4.2 Optimal Probability Region (Given Area)

Consider the choice of an optimal region given that the region has a given area or volume. First we need to give some structure to the choice of the optimal set. Let $\mathcal{A}$ denote the set of Borel-measurable sets with volume A, then our problem is choosing from among $\mathcal{A}$ the set R^* with maximal probability measure. That is,

$$R^* = \arg\max_{R \in \mathcal{A}} P^k(R; x_\tau; \beta_0; h_0). \tag{32}$$

where $\mathcal{A} = \{B \in \mathcal{B}(\mathbb{R}^g) : V(B) = A\}$. Under sufficient smoothness, a necessary condition that the maximizing set must satisfy is that it be a member of the level sets of the density, which are defined by $R(q) = \{y : f_{y|x}(y; x_\tau; \beta_0; h_0) \geq q\}$ for any choice of the level q.

Note that $V(R(q))$, the volume of such regions, is monotonically decreasing in q. If the monotonicity is strict then we can find q^* as the solution to the implicit function $A = V(R(q)) = \int 1(f_{y|x}(y; x_\tau; \beta_0; h_0) \geq q)dy$. More generally, if the monotonicity is not strict, then we have

$$q^* = q(A;\beta_0;h_0) = \inf_q \{ \int 1(f_{y|x}(y;x_\tau;\beta_0;h_0) \ge q)dy \le A\} \tag{33}$$

and the optimal region is given by $R^* = \{y : f_{y|x}(y;x_\tau;\beta_0;h_0) \ge q(A;\beta_0;h_0)\}$. Substitution from the definitions of q^* and correspondingly R^* into (30) yields

$$\begin{aligned} P^*(A;x;\beta_0;h_0) &= \Pr[(y \in R^*)|x_\tau] \\ &= E[1(f_{y|x}(\pi(\varepsilon;x_\tau;\beta_0);x_\tau;;\beta_0;h_0) \ge q(A;\beta_0;h_0))|x_\tau] \\ &= E[1(f_{y|x}(\pi(\rho(z;\beta_0);x_\tau;\beta_0);x_\tau;;\beta_0;h_0) \ge q(A;\beta_0;h_0))|x_\tau] \end{aligned} \tag{34}$$

as the probability content of the optimal region.

Operationally, we need to estimate q^* and hence R^*, and the probability content of the latter. The complication with the first is the need to perform a multidimensional intergral. This may be avoided by transforming to expectations and using averages

$$\hat q = \inf_q \{ n^{-1} \sum_{t=1}^{n} [1(\hat f_{y|x}(y;x_\tau) \ge q)/\hat f_{y|x}(y;x_\tau)] \le A\} \tag{35}$$

where $\hat f_{y|x}(\cdot)$ is a consistent estimators of the multivariate conditional density. The optimal region with area A may be estimated by $\hat R^* = \{y : \hat f_{y|x}(y;x_\tau) \ge \hat q\}$ and the corresponding probability by

$$\hat P^* = n^{-1} \sum_{t=1}^{n} 1(\hat f_{y|x}(\pi(\rho(z_t;\hat\beta);x_\tau;\hat\beta);x_\tau) \ge \hat q). \tag{36}$$

In practice, $\hat q$ as given by (35) can be obtained by a binary search since the estimated volume is also monotonic by definition.

Note that the nuisance parameters $\hat f_{y|x}(\cdot)$ and hence $\hat q$ will both be consistent but have slower than $n^{-1/2}$ rates of convergence. Nonetheless, the limiting behavior of the probability content estimator is given by

$$n^{1/2}(\hat P^* - P^*) \xrightarrow{d} N(0; P^*(1 - P^*) + \bar P^{*\prime} V \bar P^*) \tag{37}$$

where P^* is defined immediately above and $\bar P^* = \partial E[1(f_{y|x}(\pi(\rho(z;\beta);x_\tau;\beta);x_\tau;;\beta_0;h_0) > q(A;\beta_0;h_0))|x_\tau]/\partial\beta'|_{\beta=\beta_0}$, which is the same as if the optimal region were known with certainty. Furthermore, for $\hat\beta$ semiparametric efficient, the method of moments estimator will attain the semiparametric efficiency bound for the estimation of the probability content of the known optimal region. Analogous to the interval results, we find that $\hat P^*$ converges to $\Pr[y \in \hat R^*|x_\tau]$, the coverage probability of the estimated region, at the rate $n^{-1/2}$ and moreover that $n^{1/2}(\hat P^* - \Pr[y \in \hat R^*|x_\tau])$ attains a semiparametric efficiency bound, provided that $\hat f_{y|x}(\cdot)$ and $\hat q$ converge to their targets at a rate faster than $n^{1/4}$.

4.3 Optimal Area Region (Given Probability)

We are likely more interested in the dual problem of choosing a region with given probability so as to minimize the area or volume of the region. Let $\mathcal{P}$ denote the set of Borel-measurable sets with probability measure P, then the problem is choosing from among $\mathcal{P}$ the set R^*

with minimal volume. Formally, we have

$$\mathcal{R}^* = \arg\min_{\mathcal{R}\in\mathcal{P}} \mathcal{V}(\mathcal{R}), \tag{38}$$

where $\mathcal{P} = \{B \in \mathcal{B}(\mathbb{R}^g) : \mathcal{M}(B) = A\}$ and $\mathcal{M}(\cdot)$ is probability measure. As above, under sufficient smoothness conditions, the minimizing set must be a member of the level sets of the density. Since the probability measure $\mathcal{M}(\mathcal{R}(q))$ is also monotonic increasing in q, we have

$$q^* = q(P;\beta_0;h_0) = \inf_q \left\{ \int 1(f_{y|x}(y;x_\tau;\beta_0;h_0) \ge q) f_{y|x}(y;x_\tau;\beta_0;h_0)dy \le P \right\} \tag{39}$$

and the optimal region is given by $\mathcal{R}^* = \{y : f_{y|x}(y;x_\tau;\beta_0;h_0) \ge q(P;\beta_0;h_0)\}$. Substitution into the volume operator yields $A^*(P;\beta_0;h_0) = \mathcal{V}(\mathcal{R}(q^*)) = \int 1(f_{y|x}(y;x_\tau;\beta_0;h_0) \ge q(P;\beta_0;h_0))dy$.

Operationally, we do not need to estimate A^* since it will be given directly as a result of estimating q^*. Substitution of a sample average for the expectation in (39) and solving for q in terms of P yields the following estimator for q^*:

$$\hat{q}^* = \inf_q \left\{ n^{-1} \sum_{t=1}^n 1(\hat{f}_{y|x}(\pi(\rho(z_t;\hat\beta);x_\tau;\hat\beta);x_\tau) \ge q) \le P \right\}. \tag{40}$$

But this estimator is just the approximate inverse function for the estimated probability given by (27) in the previous subsection. Corresponding to the results for intervals, we find that

$$n^{1/2}(\hat{q}^* - q^*) \xrightarrow{d} N(0; f_{y|x}(q^*;x_\tau;\beta_0;h_0)^{-2}\{P(1-P) + \bar{P}'V\bar{P}\}) \tag{41}$$

where now $\bar{P} = \partial E[1(f_{y|x}(\pi(\rho(z_t;\beta);x_\tau;\beta);\beta_0;h_0) \ge q^*)]/\partial\beta'|_{\beta=\beta_0}$. Given our estimator of q^*, the estimated region is given directly as $\hat{\mathcal{R}}^* = \{y : \hat{f}_{y|x}(y;x_\tau) > \hat{q}^*\}$. Note that q plays much the same role as a quantile in the univariate case.

In the end, we are interested in $\Pr[\hat{f}_{y|x}(\pi(\rho(z_t;\hat\beta);x_\tau;\hat\beta);x_\tau) > \hat{q}^*|x_\tau]$, the coverage probability of the estimated region, relative to the given probability P. Following the development in the previous section for intervals, provided $\hat{f}_{y|x_\tau}(\cdot|x_\tau) = f_{y|x_\tau}(\cdot|x_\tau;\beta_0;h_0) + o_p(n^{-1/4})$, we have

$$\begin{aligned} \Pr[\hat{f}_{y|x}(y;x_\tau) > \hat{q}^*|x_\tau] &= E[E[1(\hat{f}_{y|x}(\pi(\varepsilon;x_\tau;\beta_0);x_\tau) > \hat{q}^*|\hat{q}^*;x_\tau]|x_\tau] \\ &= P + E[f_{y|x}(q^*;x_\tau;\beta_0;h_0)(\hat{q}^* - q^*) + o_p(n^{-1/2})|x_\tau] \\ &= P + E[n^{-1/2}N(0;P(1-P) + \bar{P}'V\bar{P}) + o_p(n^{-1/2})|x_\tau] = P + o(n^{-1/2}). \end{aligned} \tag{42}$$

Thus, although the boundries of the estimated prediction region have slower than $\sqrt{n}$ rates of convergence, the probability content of the estimated region may converge to its ostensible value at a rate faster than $\sqrt{n}$, as was the case with intervals.

5. Sampling Experiment

In the previous two sections, we have examined the asymptotic behavior of the various prediction interval and regions. It is of obvious interest whether or not the asymptotic properties also obtain in small samples. In this section we shall attempt to address this issue by conducting a sampling experiment for a nonlinear simultaneous system. In order to keep the calculation problem manageable, the model is extremely simplified and has some rather special properties. As a result, the findings regarding the small sample performance of the various estimators are not necessarily applicable to more realistic models. Nevertheless, the study should give some indication of the relative performance of the alternative procedures in small samples. In particular, we are interested in how quickly the large sample relative efficiencies assert themselves in this model.

For the sampling experiment, we utilize the following two-equation nonlinear model

$$
\begin{aligned}
y_{t1} &= \beta_1 + \beta_2 x_t + u_{t1} \\
y_{t2} &= \beta_3 + \beta_4 y_{t1}^2 + \beta_5 x_t + u_{t2}
\end{aligned}
$$

where $(u_{t1}, u_{t2})' \sim i.i.d.\ N(0, \Sigma)$, $\Sigma = (\sigma_{ij})$. A special feature of this model is the availability of a closed-form solution:

$$
\begin{aligned}
y_{t1} &= \beta_1 + \beta_2 x_t + u_{t1} \\
y_{t2} &= \beta_3 + \beta_4 (\beta_1 + \beta_2 x_t + u_{t1})^2 + \beta_5 x_t + u_{t2}.
\end{aligned}
$$

As a result, the moments of y_τ are readily obtainable in closed form.

(To be completed)

6. Concluding Remarks

In this paper we have studied alternative procedures for obtaining prediction intervals and/or regions in nonlinear simultaneous systems with independent disturbances. The need for attaching probability values to our predictions is of obvious importance and has received considerable attention in both the linear regression model and linear simultaneous equation model. Although a substantial fraction of the models used for prediction are nonlinear simultaneous systems, very little attention has been given to what procedures might reasonably be used to generate probability values. The work presented in this paper applies the results on semiparametric efficient estimation of expectation functions by Brown and Newey (1998) to the prediction interval/region problem and thereby extends the previous unpublished work of Brown and Mariano (1991) on prediction intervals and regions. The latter was only partially semiparametric since the parameter estimator was assumed to be equivalent to maximum likelihood.

The construction of prediction intervals is examined in Section 3. The approach of Brown and Newey is applied to obtain the semiparametric efficiency bound for estimation of the probability content of a known interval. The limiting distribution of the method of moments estimator of the probability content is developed and shown to obtain the efficiency bound when based on semiparametric efficient parameter estimates. Optimal prediction intervals that maximize the probability content of the interval given the interval length are studied and feasible estimators of the interval and their probability content developed. The feasible estimators of the probability content is shown to be asymptotically equivalent to an estimator based on the true optimal interval. The estimated probability is shown to differ from the coverage probability of the estimated interval by terms of order and, moreover, the difference attains a lower bound when the method of moments probability estimator is based on a semiparametric efficient estimator of ¯. The dual problem of minimizing the interval length given the probability content of the interval is also considered. A feasible estimator of such an interval is developed its asymptotic behavior examined. The coverage probability of the estimated interval is shown to differ from the ostensible probability by term of order smaller than $\sqrt{n}$.

The construction of prediction regions is studied in Section 4. The limiting behavior of the method of moments estimator of the probability content of a known region is developed and shown to attain the semiparametric efficiency bound when based on semiparametric efficient estimates of the parameters. The construction and estimation of optimal regions which maximize probability content given region area or volume is examined. The optimal regions are shown to be level sets of the conditional density of the endogenous variables given the exogenous variables. Feasible estimators that are based on nonparametric estimators of the density and method of moments estimators of the probability content are devised and shown to attain the semiparametic efficiency bound for estimating the probability content of a known optimal region when based on semiparametric efficient estimates of ¯. Regions which minimize the area or volume of the region given the probability are also studied and feasible estimators devised. The asymptotic behavior of the feasible estimators of the region is developed. As with the interval case, the coverage probability of the estimated interval is shown to differ from the ostensible probability by term of order smaller than $\sqrt{n}$.

The results of a sampling experiment are presented in Section 5. (To be completed)

There are a number of directions in which the research outlined in this section can be extended. The formal model analyzed is i.i.d., while most predictive models are dynamic in nature. It appears that the results can be applied pretty much directly to stationary models with i.i.d. and independent innovations, but a number of details need to be worked out to formalize the extension. Another interesting extension is to fully nonparametric prediction intervals and regions. The approach outlined above can be utilized to calculate optimal unconditional prediction intervals for models with no systematic component. That is, we have no model for y but seek to construct optimal prediction intervals and regions using estimated distributions. It appears that such estimated intervals and regions will also be asymptotically independent of the nuisance parameters of the density estimator. This result should be of great interest and needs to be worked out in greater detail.

References

Andrews, D. (1994): "Empirical Process Methods in Econometrics," in Handbook of Econometrics, Vol. 4, ed. by R. Engle and D. McFadden, Amsterdam: North Holland, 2247-2294.

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

Brown, B. and R. Mariano (1989), "Predictors in Dynamic Nonlinear Models: Large-Sample Behavior," Econometric Theory, Vol. 5, No.3.

Brown, B. and R. Mariano (1991), "Prediction Intervals and Regions in Nonlinear Simultaneous Systems," Mimeo, Department of Economics, Rice University.

Brown, B., and W. Newey (1998): "E¢ cient Semiparametric Estimation of Expectations," Econometrica 66, 453-464.

Howrey, P., and H. Kelejian, "Simulation Versus Analytical Solutions: The Case of Econometric Models," in T.H. Naylor, ed., Computer Simulation Experiments with Models of Economic Systems, (New York: J. Wiley, 1971).

Mariano, R., and B. Brown, "Asymptotic Behavior of Predictors in a Nonlinear Simultaneous System," International Economic Review, 24 (1983), 523-536.

Newey, W. (1990): "Semiparametric E¢ ciency Bounds," Journal of Applied Econometrics 5, 99-135.

Newey, W. (1994): "The Asymptotic Variance of Semiparametric Estimators," Econometrica 62, 1349-1382.

Newey, W., and D. McFadden (1994): "Estimation and Inference in Large Samples," in Handbook of Econometrics, Vol. 4, ed. by R. Engle and D. McFadden, New York: North Holland, 2111-2445.