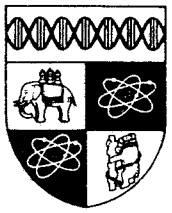
FINITE-SAMPLE PROPERTIES OF STOCHASTIC PREDICTORS IN NONLINEAR SYSTEMS: SOME INITIAL RESULTS*

by

Roberto S. Mariano**

Number 266

WARWICK ECONOMIC RESEARCH PAPERS



DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

FINITE-SAMPLE PROPERTIES OF STOCHASTIC PREDICTORS IN NONLINEAR SYSTEMS: SOME INITIAL RESULTS*

by

Roberto S. Mariano**

Number 266

July 1985

*Prepared during the Warwick Summer Workshop on Macroeconometric Modelling. Support from the Workshop, the ESRC Macroeconomic Modelling Bureau, and the Department of Economics at the University of Warwick is deeply appreciated and gratefully acknowledged.

** Department of Economics
University of Pennsylvania
McNeill Building
Philadelphia
Pennsylvania 19104
USA

1. INTRODUCTION AND SUMMARY

Many econometric models for forecasting and policy analysis consist of a statistically estimated system of nonlinear simultaneous stochastic equations. The distinguishing feature of these models is the nonlinearity of the solution for the endogenous variables in terms of model disturbances. Despite the widespread use of these models, there has been little formal analysis of predictions based on such models. Furthermore, practitioners' validation of such models has proceeded, for the most part, on an informal basis.

This paper covers a preliminary study of the finite-sample properties of predictors in nonlinear systems and as such provides an additional analytical treatment in parallel with the large-sample asymptotic analysis carried out recently by Mariano and Brown (1983a, 1983b, 1984, 1985).

The large-sample analysis, done by Mariano and Brown through asymptotic expansions for prediction error, points to potential shortcomings in the current practice of forecasting through deterministic simulations of the estimated model and suggests alternative forecasting procedure based on stochastic simulations. In particular, two alternative procedures are discussed in detail: the Monte Carlo and the residual-based stochastic predictors. In a deterministic treatment, the disturbances in the model are replaced by their expected values in the simulation of the model. In stochastic simulations, on the other hand, the disturbances are replaced by random proxies. For the Monte Carlo predictor, the proxies are random draws from an estimated parametric

distribution - typically the multivariate normal. The residual-based predictor, on the other hand, uses the calculated sample period residuals as stochastic proxies for the disturbances. The deterministic predictor is used commonly in practice. The Monte Carlo predictor has been used in the past, but to a much lesser extent than the deterministic predictor. The residual-based is a new procedure developed in Brown and Mariano (1984) as a device to reduce the computational burden and misspecification sensitivity in the Monte Carlo predictor.

The model, the alternative predictors, and their large-sample asymptotic properties are summarized in Section 2. For large samples, the asymptotic results lend support to the residual-based procedure as a viable alternative to the Monte Carlo and deterministic forecasting methods. In general, the deterministic predictor is asymptotically biased while the residual-based procedure, like the Monte Carlo, is asymptotically unbiased when the model is correctly specified and consistently estimated. If the distribution of the error process is misspecified, however, the Monte Carlo predictor becomes asymptotically biased while the residual-based remains unbiased provided the functional form is correctly specified and consistently estimated. When parameters are estimated efficiently and the number of Monte Carlo simulations does not exceed sample size, the residual-based predictor is efficient relative to the Monte Carlo in terms of asymptotic mean squared prediction error. These asymptotic results lead naturally to the development of model validation tests which are sensitized to deterioration in the predictive performance of the model. These specification tests, not summarized in this paper but discussed in detail in Mariano and Brown (1983b), are based on comparisons of stochastic simulations with actual observations over the sample period. They are asymptotically valid

significance tests and, as such, they provide a more formal model validation procedure than the common practice of calculating summary error statistics such as root mean squared percentage error and the like.

Beyond point predictions of the endogenous variables, or point estimates of their first moments, it would be of major interest to develop prediction regions which reflect the inherent variability in the endogenous variables. Prediction regions are difficult to obtain for nonlinear systems due to the general non-normality of the endogenous variables. Thus, unlike linear systems, the distribution of the endogenous variables is not characterized by their first two moments. First and second moments, however, can be utilized to form conservative prediction regions using Chebyshev-type inequalities. One of the primary advantages of both Monte Carlo and residual-based predictors over the deterministic is that the stochastic simulations used to estimate the first moment may also be used to assess the higher moments as well as the probability distribution itself of the endogenous variables. The asymptotic analysis of point predictors summarized here has been extended in Brown and Mariano (1984) to the estimation of the higher order moments and the cumulative distribution functions of the endogenous variables in the nonlinear system.

The applicability of this whole large-sample analysis of predictors in a practical modelling situation with a given sample size remains an open issue. One might suspect, for example, that the residual-based procedure may not perform as well in small samples. There are indications of this, for example, in Fisher and Salmon's (1984) stochastic simulations of the National Institute of Economic and Social Research and the London Business School models of the U.K. economy.

In an initial attempt to address this issue, we carry out in this paper an analytical investigation of the finite-sample properties of our alternative predictors. The exact bias and mean-squared prediction errors for the closed form, the Monte Carlo, and the residual-based predictors in a static nonlinear model are compared in Section 3. The use of antithetic variates to modify the Monte Carlo and residual-based predictors are also covered in this analysis, as well as the bootstrap procedure as an alternative to the residual based. The derived moment expressions lead to some relative efficiency comparisons. They show for example that the closed form, the Monte Carlo and the Monte Carlo antithetic predictors all have the same finite-sample bias and that, in terms of exact mean-squared prediction error, the Monte Carlo predictor, with or without antithetic variates, is less efficient than the closed form predictor for any sample size but with the inefficiency disappearing as the Monte Carlo replication size increases indefinitely. The use of antithetic variates may or may not improve the Monte Carlo procedure depending on how the nonlinearities in the system affect the covariance structure: specific conditions are given under which the use of antithetic variates leads to a smaller mean squared prediction error for the Monte Carlo predictor.

For the residual-based predictor, the relative size of its bias and mean-squared error prediction error as compared to either the Monte Carlo or the closed form predictor is ambiguous. In general, the use of antithetic variates would alter the finite-sample moments of the residual-based predictor, with the direction of alteration again being ambiguous. Finally, the bootstrap predictor, which uses a sample from the empirical distribution of the residuals instead of a complete enumeration as in the residual-based, is shown to be less efficient than

the residual-based predictor: the additional sampling component in the bootstrap introduces another source of variation in the predictor. This inefficiency, however, disappears as the replication size in the bootstrap approaches complete enumeration.

Unfortunately, the finite-sample results reported in Section 3 do not cover relative efficiencies with respect to the deterministic predictor; neither do they include comparisons between the Monte Carlo or the closed form predictor, on one hand, and the residual-based predictor on the other. To get some partial answers for these, we consider the single-equation log linear model in Section 4 and analyse first and second moment expressions for the various predictors.

2. LARGE-SAMPLE ASYMPTOTIC RESULTS

This section introduces the model and the predictors to be discussed and summarizes the results contained in Mariano and Brown (1983a) and Brown and Mariano (1984, 1985) concerning the large-sample asymptotic properties of predictors in static as well as dynamic nonlinear systems.

2.1 The Model in the Static Case

In this section, we consider a model consisting of the following n simultaneous nonlinear stochastic equations:

$$f(y_t, x_t; \theta) = u_t \; ; \; t = 1, 2, \ldots, T \tag{2.1}$$

where f, y_t, and u_t are n x 1; x_t is m x 1; and θ is p x 1. The vector of functions, f, is completely known; θ denotes the unknown parameter vector; x_t represents nonstochastic exogeneous variables; and the u_t are unobservable disturbance terms for which we assume

$$u_t \sim \text{i.i.d. } N(0, I), \; t = 1, 2, \ldots, T. \tag{2.2}$$

We also assume that, as a mapping from y_t to u_t, (2.1) defines a locally unique inverse relationship

$$y_t = g(u_t, x_t; \theta). \tag{2.3}$$

This is the analog of the so-called reduced-form equations in the linear simultaneous equations model. In the nonlinear model, however, it is usually the case that $g(\cdot)$ cannot be written in closed form and, for a given set of

values of (u_t, x_t, θ), the corresponding value of y_t is calculated not directly from(2.3) but rather as the numerical solution to (2.1).

We will also assume that at least the first two moments of y_t are finite. We denote these moments by

$$\gamma(x_t; \theta) = Ey_t = \gamma_t \tag{2.4}$$

$$\Omega(x_t; \theta) = E\{(y_t - Ey_t)(y_t - Ey_t)'\} = \Omega_t .$$

2.2 Main Problem

Given a sample $\{(x_t, y_t): t = 1, 2, \ldots, T\}$, from which θ can be estimated, and given x_* for some time point, *, outside the sample period, the main problem is to predict y_*. Because of the static nature of the model, the solution to this problem also applies to multi-period prediction.

2.3 Predictors of Interest

We will analyze the large-sample asymptotic behavior of the following forecasting procedures.

1. Closed-form predictor:

$$\hat{y}_*^{(c)} = \gamma(x_*; \hat{\theta}) = \int_{u_*} g(u_*, x_*; \hat{\theta})\, pdf(u_*)du_* . \tag{2.5}$$

In general this integral cannot be evaluated analytically, since neither $g(\cdot)$ nor $\gamma(\cdot)$ is available in closed form. This integral has to be approximated through numerical integration techniques.

2. Deterministic predictor:

$$\hat{y}_*^{(d)} = g(0, x_*; \hat{\theta}). \tag{2.6}$$

Given x_* and $\hat{\theta}$, $\hat{y}_*^{(d)}$ is calculated as the solution to

$$f(\hat{y}_*^{(d)}, x_*; \hat{\theta}) = 0. \tag{2.7}$$

3. Monte Carlo stochastic predictor:

$$\hat{y}_*^{(m)} = \sum_{s=1}^{S} \hat{y}_{*s}^{(m)}/S, \tag{2.8}$$

where

$$\hat{y}_{*s}^{(m)} = g(\tilde{u}_s, x_*; \hat{\theta}) \tag{2.9}$$

$\tilde{u}_s$ = independent draws from N(0, I), s = 1, 2, ..., S. Again, $\hat{y}_{*s}^{(m)}$ is calculated numerically as the solution to

$$f(\hat{y}_{*s}^{(m)}, x_*; \hat{\theta}) = \tilde{u}_s. \tag{2.10}$$

Also, note that (2.8) is one way of approximating numerically the closed-form predictor (2.5) under the assumption that $u_* \sim N(0, I)$.

4. Residual-based stochastic predictor:

$$\hat{y}_*^{(r)} = \sum_{t=1}^{T} \hat{y}_{*t}^{(r)}/T \tag{2.11}$$

where

$$\hat{y}_{*t}^{(r)} = g(\hat{u}_t, x_*; \hat{\theta})$$

$$\hat{u}_t = f(y_t, x_t; \hat{\theta}). \tag{2.12}$$

2.4 Basic Approach

Under appropriate conditions, we derive asymptotic expansions of the following form for the various predictors under study:

$$\hat{y}_*-y_* = \sum_{j=0}^{q} e_j + O_p\{T^{-(q+1)/2}\}$$

$$(\hat{y}_* -y_*)\ (\hat{y}_*-y_*)' = \sum_{j=0}^{q} M_j + O_p\{T^{-(q+1)/2}\} \tag{2.13}$$

where

$$e_j = O_p(T^{-j/2})$$

$$M_j = O_p(T^{-j/2}). \tag{2.14}$$

From these expansions, define asymptotic moments (to order $T^{-q/2}$) as exact expectations of retained leading terms to order $T^{-q/2}$ in the appropriate asymptotic expansions. In particular, to order $T^{-q/2}$, asymptotic bias is

$$AB(\hat{y}_*) = E(\sum_{j=0}^{q} e_j) \tag{2.15}$$

and asymptotic mean squared prediction error is

$$\text{AMSPE}(\hat{y}_*) = E\Big(\sum_{j=0}^{q} M_j\Big). \tag{2.16}$$

We say that the predictor $\hat{y}_*$ is asymptotically unbiased if $AB(\hat{y}_*) = o(1)$, as T goes to infinity. Furthermore, relative asymptotic prediction efficiencies of competing predictors are measured in terms of their AMSPE's.

2.5. Main Results for the Static Case

All orders of magnitude are under T going to infinity and all asymptotic moments are to order 1/T. Throughout, we assume that

$$T^{1/2}\,(\hat{\theta}-\theta) \longrightarrow N(0,\ \Psi). \tag{2.17}$$

Under additional smoothness conditions on $f(\cdot)$ and regularity assumptions about the distribution of y_t, as given, for example in Mariano and Brown (1983a) or Brown and Mariano (1984), the following results hold for the static case:

1. The asymptotic biases for the various predictors have the following orders of magnitude:

$$\begin{aligned} AB(\hat{y}_*^{(d)}) &= O(1) \\ AB(\hat{y}_*^{(c)}) &= O(1/T) \\ AB(\hat{y}_*^{(m)}) &= O(1/T) \\ AB(\hat{y}_*^{(r)}) &= O(1/T) \end{aligned} \tag{2.18}$$

2. For any consistent $\hat{\theta}$, the AMSPE for the closed-form and the Monte Carlo predictors are:

$$\text{AMSPE}\ (\hat{y}_*^{(c)}) = \Omega_* + \Gamma_* \Psi \Gamma_*'/T$$

$$\text{AMSPE}(\hat{y}_*^{(m)}) = \Omega_* + \Gamma_* \Psi \Gamma_*'/T + \Omega_*/S, \tag{2.19}$$

where Ψ is implicitly defined in (2.17) and Γ_* is defined in (2,21) below.

3. If $\hat{\theta}$ is the maximum likelihood estimator (MLE) of θ, or any other estimator asymptotically equivalent to MLE, then

$$\text{AMSPE}(\hat{y}_*^{(r)}) = \Omega_* + \Gamma_* \Psi \Gamma_*'/T + \{\Omega_* - (\Gamma_* - \overline{H})\ \Psi\ (\Gamma_* - \overline{H})'\}/T, \tag{2.20}$$

where

$$\Omega_* = \Omega(x_*;\theta),$$

$$\Gamma_* = \partial\gamma(x_*;\theta)/\partial\theta,$$

$$\overline{H} = \text{plim} \sum_{t=1}^{T} \{\partial h_t(\theta)/\partial\theta\}/T, \tag{2.21}$$

$$h_t(\theta) \equiv h(y_t, x_t, x_*; \theta)$$

$$\equiv g(f(y_t, x_t; \theta), x_*; \theta).$$

2.5 The Dynamic Case

Now, consider the model with structural equations

$$f(y_t, y_{t-1}, x_t; \theta) = u_t \tag{2.22}$$

$$u_t \sim \text{i.i.d. } N(0, I) \text{ for } t = 1, 2, \ldots, T$$

assumed to define a locally unique solution

$$y_t = g(u_t, y_{t-1}, x_t; \theta). \tag{2.23}$$

Given the same sample as before, namely observations on x_t and y_t for $t = 1,\ldots, T$, we now consider the problem of predicting

$$y_{T+1} = g(u_{T+1}, y_T, x_{T+1}; \theta) \tag{2.24}$$

and

$$y_{T+2} = g(u_{T+2}, y_{T+1}, x_{T+2}; \theta)$$

$$= g\{u_{T+2}, g(u_{T+1}, y_T, x_{T+1}), x_{T+2}; \theta\} \tag{2.25}$$

As before, we abstract from the problem of forecasting exogenous variables and assume that we know the values of x_{T+1} and x_{T+2}. In this dynamic model, multi-period-ahead prediction entails additional statistical complications beyond those encountered in one-period predictions - hence the inclusion of y_{T+2} in the discussion.

The predictors of interest are as defined before in the case of Y_{T+1}:

$$\hat{y}^{(d)}_{T+1} = g(0,\ y_T,\ x_{T+1};\ \hat{\theta})$$

$$\hat{y}^{(m)}_{T+1} = \sum_{s=1}^{S} g(\tilde{u}_s,\ y_T,\ x_{T+1};\ \hat{\theta})/S \qquad (2.26)$$

$$\hat{y}^{(r)}_{T+1} = \sum_{t=1}^{T} g(\hat{u}_t,\ y_T,\ x_{T+1};\ \hat{\theta})/T$$

where

$$\tilde{u}_s = \text{independent draws from } N(0,\ I) \text{ for } s = 1,\ 2,\ \ldots,\ S,$$

$$\hat{u}_t = f(y_t,\ y_{t-1};\ x_t;\ \hat{\theta}). \qquad (2.27)$$

For a two-period-ahead-prediction (for T even),

$$\hat{y}^{(d)}_{T+2} = g(0,\ \hat{y}^{(d)}_{T+1},\ x_{T+2};\ \hat{\theta})$$

$$\hat{y}^{(m)}_{T+2} = \sum_{s=1}^{S} g\{\tilde{u}_{s2},\ g(\tilde{u}_{s1},\ y_T,\ x_{T+1};\ \hat{\theta}),\ x_{T+2};\ \hat{\theta}\}/S \qquad (2.28)$$

$$\hat{y}^{(r)}_{T+2} = \sum_{t=1}^{T/2} g\{\hat{u}_t,\ g(\hat{u}_{T/2+t},\ y_T,\ x_{T+1};\ \hat{\theta}),\ x_{T+2};\ \theta\}/(T/2),$$

where $\hat{u}_t$ is as defined in (2.27), and

$$\tilde{u}_{si} = \text{independent draws from } N(0,I) \text{ for } i=1,\ 2 \text{ and } s=1,\ 2,\ \ldots,\ S.$$

For one-period dynamic prediction, the problem parallels the static case. The main difference lies in the presence of y_T in the reduced-form equation for y_{T+1}. Since prediction of y_{T+1} is made conditional on the

observed value of y_T, one way of pursuing our analysis is to consider the asymptotic behavior of dynamic predictors conditional on this observed value of y_T. Because of this conditioning process, the initial assumptions we make concerning the behavior of y_t and the smoothness of the structural equations now pertain to conditional distributions given y_T. These assumptions parallel closely those introduced by Brown and Mariano (1984) for the static case and are enumerated in detail in Brown and Mariano (1985). Asymptotic moment expressions for the deterministic and the Monte-Carlo and residual based stochastic predictors are the same as their counterparts for the static case, as summarized in the preceding section, except for the change in the interpretation of quantities like Ω_*, Ψ, and Γ_* in terms of conditional distributions given y_T.

For two-period-ahead prediction, a reformulation of the model and of the problem of predicting y_{T+2}, given observations up to time T, allows us to imbed this problem in the context of one-period dynamic prediction. This approach allows us to infer asymptotic moments for the prediction errors in the deterministic, the Monte-Carlo, and, in particular, the residual-based dynamic predictor defined in (2.28). These asymptotic moments differ from the one-period, and hence static, cases only in terms of the sample halving for the residual-based predictor. Thus, the same qualitative conclusions follow concerning the relative asymptotic efficiencies of predictors, namely:

1. The closed-form predictor, $\gamma_2(y_T, x_{T+1}, x_{T+2}; \hat{\theta})$, where

$$\gamma_2(\cdot) = E(y_{T+2}|y_T),$$

provides a lower bound for AMSPE within a class of stochastic predictors of y_{T+2} based on a consistent $\hat{\theta}$. However, this procedure is not feasible.

2. The Monte-Carlo and residual-based predictors of y_{T+2}, as defined previously, are both inefficient relative to the closed-form

predictor, in terms of AMSPEs.

3. With sufficiently numerous replications, the inefficiency of the Monte Carlo predictor can be made arbitrarily small.

4. The residual-based predictor of y_{T+2} is asymptotically efficient relative to the Monte-Carlo predictor when the Monte Carlo replication size (S) is less than or equal half the sample size. For m-period ahead prediction, the effective number of sample residuals used in the residual-based procedure (as defined here) would be T/m and the condition for the efficiency of the residual-based relative to the Monte-Carlo is $S \leq T/m$.

5. While a ranking can be established for the closed-form, the Monte Carlo, and the residual-based predictors as indicated above, their AMSPEs differ from each other by an amount which is small relative to total AMSPE. Thus, for large samples, the possible gain in AMSPE from using one of the three predictors rather than another is in fact small.

The above results indicate that the version of the residual-based procedure we have analyzed here suffers from a deterioration in precision (in terms of its AMSPE) as the forecast horizon is increased. Thus, alternatives to sample-splitting should be explored in order to improve the asymptotic precision of the residual-based procedure in multi-period ahead forecasting. One possibility is to sample from the empirical distribution of the residuals, as in the bootstrap. When predicting m-periods ahead, each replication requires m independent realizations for the disturbances. These may be proxied by m independent draws from the empirical distribution of the residuals. An extreme case of this procedure is complete enumeration, where all T^m alternative realizations are used. (The residual-based procedure as implemented for the static and one-period ahead dynamic cases is an example of complete enumeration.)

Alternatively, we should also consider the behavior of stochastic predictors, both Monte-Carlo and residual-based, when combined with variance-reduction techniques. Some references to earlier work along these lines are Calzolari (1979), where dramatic improvements are obtained in the precision of the Monte Carlo predictor through antithetic variates; Calzolari and Sterbenz (1981), where control variates are used in conjunction with Monte Carlo simulations to improve the estimates of the second moments of endogenous variables; and Fisher and Salmon (1984) where antithetic variates are used for the residual-based procedure in stochastic simulations of the National Institute of Economic and Social Research and the London Business School models of the U.K. economy. Our preliminary formal results indicate the possiblity of gains in the asymptotic precision of the Monte Carlo and residual-based predictors when combined with antithetic or (in some cases) control variates. Due to the nonlinearity of model solution with respect to the structural disturbances, however, the gains in precision may be of a smaller order of magnitude than have been obtained for estimators in linear models. Also, under certain conditions, the asymptotic efficiency of the residual-based predictor relative to the Monte Carlo is maintained when antithetic and control variates are used.

3. FINITE-SAMPLE ANALYSIS FOR THE GENERAL STATIC NONLINEAR

In this section, we consider a general static nonlinear model as described in the preceding sections and derive propositions concerning the first two finite-sample moments of the Monte Carlo and residual-based predictors. We also deal with modifications of these predictors through the use of antithetic variates and the application of the bookstrap in implementing the residual-based procedure. In the bootstrap, stochastic proxies are generated by sampling from the empirical uniform discrete distribution of the residuals instead of a complete enumeration as in the residual-based.

Note that any predictor, say $\hat{y}_*$, in the group mentioned above is independently distributed of u_*. Hence, its mean-squared prediction error is

$$\begin{aligned}
\text{MSPE}\ (\hat{y}_*) &= E\ (\hat{y}_*-y_*)\ (\hat{y}_*-y_*)' \\
&= E\ (\hat{y}_*-\gamma_*)\ (\hat{y}_*-\gamma_*)' + E(\hat{y}_*-\gamma_*)\ (\hat{y}_*-\gamma_*)' \\
&= \Omega_* + \text{MSE}\ (y_*) \\
&= \Omega_* + V(\hat{y}_*) + B(\hat{y}_*) \qquad (3.1)
\end{aligned}$$

where

$$\text{MSE}\ (\hat{y}_*) = E\ (\hat{y}_*-\gamma_*)(\hat{y}_*-\gamma_*)'$$

$$V(\hat{y}_*) = E(\hat{y}_* - E\hat{y}_*)(\hat{y}_* - E\hat{y}_*)' \tag{3.2}$$

$$B(\hat{y}_*) = (E\hat{y}_* - \gamma_*)(E\hat{y}_* - \gamma_*)'.$$

The second equality in (3.1) holds if and only if $\hat{y}_*$ and u_* are statistically independent and is appropriate for the static model but not for the dynamic model. The rest of the equations in (3.1) is a matter of notational definition. Measuring finite-sample efficiency in terms of MSPE, we can see from (3.1), then, that the relative prediction efficiency of any two predictors would be equivalent to their relative efficiency as estimators of γ_*. Furthermore, their relative prediction efficiency would depend only on their finite-sample covariance matrices, $V(\cdot)$, if these two predictors have the same finite-sample bias.

The first two propositions we present serve to formalize the intuitive notion that, in relation to the closed form predictor, the Monte Carlo predictor introduces additional stochastic noise in the calculation of the estimated mean function γ_* and hence must have a larger mean-squared prediction error than the closed form.

Proposition 1. The closed form predictor and the Monte Carlo stochastic predictor have the same finite-sample bias.

Proof. For $s = 1, 2, \ldots, S$,

$$E[g(\tilde{u}_s, x_*; \hat{\theta}) \mid \hat{\theta}] = \gamma(x_*; \hat{\theta}) \tag{3.3}$$

since $\tilde{u}_s$ is independent of $\hat{\theta}$. Therefore

$$E\,(\hat{y}_*^{(m)}\,|\hat{\theta}) = \sum_s E\,[g(\tilde{u}_s,\, x_*;\, \hat{\theta})\,|\hat{\theta}]/S$$

$$= \gamma(x_*;\, \hat{\theta}) \qquad (3.4)$$

and, hence,

$$E\,(\hat{y}_*^{(m)}) = E[E(y_*^{(m)}\,|\hat{\theta}] = E\,[\gamma(x_*;\hat{\theta})] = E\hat{y}_*^{(c)}. \qquad (3.5)$$

Q.E.D.

Proposition 2. In terms of mean-squared prediction error,

$$MSPE\,(\hat{y}_*^{(m)}) - MSPE\,(\hat{y}_*^{(c)}) = E\,[\Omega\,(x_*;\, \hat{\theta})]/S] > 0. \qquad (3.6)$$

Consequently, for any finite sample size T, the Monte Carlo predictor is less efficient than the closed form predictor. However, this inefficiency disappears as the Monte Carlo replication size, S, tends to infinity.

Proof. Since $\hat{y}_*^{(m)}$ and $\hat{y}_*^{(c)}$ have the same mean (by Proposition 1), it follows that the difference in their MSPE is equal to the difference in their variances. Now,

$$V\,(\hat{y}_*^{(m)}) = E[V(\hat{y}_*^{(m)}\,|\hat{\theta})] + V[E(\hat{y}_*^{(m)}\,|\hat{\theta})]. \qquad (3.7)$$

From (3.4), the second term in (3.7) is

$$V[E(\hat{y}_*^{(m)}|\hat{\theta})] = V[\gamma(x_*;\ \hat{\theta})] = V(\hat{y}_*^{(c)}) . \tag{3.8}$$

Since $\tilde{u}_s$ and $\hat{\theta}$ are statistically independent, the second term in (3.7) simplifies as follows

$$E[V(\hat{y}_*^{(m)}|\hat{\theta})] = E[V(\Sigma_s g(\tilde{u}_s,\ x_*;\ \hat{\theta})/S|\hat{\theta})]$$

$$= E[\Omega\ (x_*;\ \hat{\theta})/S] . \tag{3.9}$$

Q.E.D.

Another way of summarizing the argument in Propositions 1 and 2 is as follows. We can decompose the estimation error in $\hat{y}_*^{(m)}-\gamma_*$ as

$$\hat{y}_*^{(m)}-\gamma_* = (\hat{y}_*^{(m)}-\hat{\gamma}_*) + (\hat{\gamma}_*-\gamma_*)$$

$$= (\hat{y}_*^{(m)}-\hat{y}_*^{(c)}) + (\hat{y}_*^{(c)}-\gamma_*) \tag{3.10}$$

where

$$\gamma_* = \gamma(x_*;\ \theta)$$

$$\hat{\gamma}_* = \gamma(x_*;\ \hat{\theta}) = \hat{y}_*^{(c)} . \tag{3.11}$$

The first term represents the additional variation due to the Monte Carlo approximation to the multiple integral $\hat{\gamma}_*$ and is the source of inefficiency in the Monte Carlo predictor relative to the closed form.

Since $E(y_*^{(m)}|\hat{\theta}) = \hat{\gamma}_*$, the first term has zero mean and, thus, the Monte Carlo and the closed form predictors have the same finite-sample bias. Furthermore, the cross-product of the two terms in (3.10) has zero expectation:

$$\begin{aligned} E[(\hat{\gamma}_*-\gamma_*)(\hat{y}_*^{(m)}-\hat{\gamma}_*)'] &= EE[(\hat{\gamma}_*-\gamma_*)(\hat{y}_*^{(m)}-\hat{\gamma}_*)'|\hat{\theta}] \\ &= E\{(\hat{\gamma}_*-\gamma_*)\, E[(\hat{y}_*^{(m)}-\hat{\gamma}_*)'|\hat{\theta}]\} \\ &= 0 \end{aligned} \tag{3.12}$$

with the last equality holding because $E(y_*^{(m)}|\hat{\theta}) = \hat{\gamma}_*$.

Thus, we have

$$\begin{aligned} MSE(\hat{y}_*^{(m)}) &= MSE(\hat{y}_*^{(c)}) + E(y_*^{(m)}-\hat{\gamma}_*)(\hat{y}_*^{(m)}-\hat{\gamma}_*)' \\ &= MSE(\hat{y}_*^{(c)}) + E\{E(\hat{y}_*^{(m)}-\hat{\gamma}_*)(\hat{y}_*^{(m)}-\hat{\gamma}_*)'|\hat{\theta}\} \\ &= MSE(\hat{y}_*^{(c)}) + E[\Omega(x_*;\hat{\theta})/S] \end{aligned} \tag{3.13}$$

which is a re-statement of (3.6) in Proposition 2.

Now, consider the Monte Carlo predictor with antithetic variates:

$$\hat{y}_*^{(ma)} = \sum_s [\hat{g}(\tilde{u}_s) + \hat{g}(-\tilde{u}_s)]/2S$$

$$= (\hat{y}_*^{(m)} + \hat{y}_*^{(m')})/2 \qquad (3.13)$$

where

$$\hat{g}(u) = g(u, x_*; \hat{\theta}), \text{ for any } u,$$

$$\hat{y}_*^{(m')} = \sum_s \hat{g}(-\tilde{u}_s)/S. \qquad (3.14)$$

Proposition 3. The following relationships hold for the Monte Carlo predictor with antithetic variates:

$$E(y_*^{(ma)} | \hat{\theta}) = \gamma(x_*; \hat{\theta}) = \hat{\gamma}_* \qquad (3.15)$$

$$V(\hat{y}_*^{(ma)}) = [V(\hat{y}_*^{(m)}) + COV(\hat{y}_*^{(m)}, \hat{y}_*^{(m')})]/2 \qquad (3.16)$$

$$MSE(\hat{y}_*^{(ma)}) = MSE(\hat{y}_*^{(c)}) + (1/2S) E[\Omega(x_*; \hat{\theta}) + E\{\hat{g}(\tilde{u}_s)\hat{g}'(-\tilde{u}_s) | \hat{\theta}\}]. \qquad (3.17)$$

Consequently, the Monte Carlo predictor with antithetic variates has the same finite sample bias as the standard Monte Carlo and the closed form predictors and is less efficient than the closed form predictor with the inefficiency converging to zero as S approaches infinity. Furthermore, the Monte-Carlo-antithetic predictor would be more efficient

relative to the standard Monte Carlo predictor if and only if the two terms composing it in (3.13) have a covariance which is smaller than the variance of the standard Monte Carlo predictor; that is, if and only if

$$\mathrm{COV}(\hat{y}_*^{(m)}, \hat{y}_*^{(m')}) < \mathrm{V}(\hat{y}_*^{(m)}) . \tag{3.18}$$

In particular, this inequality would be satisfied if $\hat{y}_*^{(m)}$ and $\hat{y}_*^{(m')}$ are negatively correlated.

Proof. (3.15) and (3.16) follow directly from (3.13) and the fact that $\hat{y}_*^{(m)}$ and $\hat{y}_*^{(m')}$ have the same probability distribution since $-\tilde{u}_s$ has the same distribution as $\tilde{u}_s$ and is also independent of $\hat{\theta}$.

A calculation similar to (3.10) and (3.12) leads to (3.17):

From

$$\hat{y}_*^{(ma)} - \gamma_* = (\hat{y}_*^{(ma)} - \hat{\gamma}) + (\hat{\gamma}_* - \gamma_*), \tag{3.19}$$

we get

$$\mathrm{MSE}(\hat{y}_*^{(ma)}) = \mathrm{MSE}(\hat{y}_*^{(c)}) + \mathrm{E}(\hat{y}_*^{(ma)} - \hat{\gamma}_*)\ (\hat{y}_*^{(ma)} - \hat{\gamma}_*)' \tag{3.20}$$

because of (3.15) and the result that

$$\mathrm{E}[(\hat{y}_*^{(ma)} - \hat{\gamma}_*)\ (\hat{\gamma}_* - \gamma_*)' \,|\, \hat{\theta}] = 0. \tag{3.21}$$

Now, given $\hat{\theta}$, we have

$$E[\hat{y}_*^{(ma)} - \hat{\gamma}_*)(\hat{y}_*^{(ma)} - \hat{\gamma}_*)' | \hat{\theta}]$$

$$= (1/4)V(\hat{y}_*^{(m)} | \hat{\theta}) + (1/4)V(\hat{y}_*^{(m')} | \hat{\theta})$$

$$+ (1/2)\ COV[(\hat{y}_*^{(m)}, \hat{y}_*^{(m')}) | \hat{\theta}]. \qquad (3.22)$$

(3.17) now follows from

$$V(\hat{y}_*^{(m)} | \hat{\theta}) = V(\hat{y}_*^{(m')} | \hat{\theta}) = \Omega(x_*; \hat{\theta})/S \qquad (3.23)$$

and

$$COV[(\hat{y}_*^{(m)}, \hat{y}_*^{(m')}) | \hat{\theta}] = E[g(\tilde{u}_s)g'(-\tilde{u}_s) | \hat{\theta}]/S. \qquad (3.24)$$

Finally, note that the third term in (3.17) is the expected value of (3.22) and, hence, is positive definite - thus showing that $\hat{y}_*^{(ma)}$ is inefficient relative to $\hat{y}_*^{(c)}$.

Q.E.D.

Now, consider the residual-based predictor with and without antithetic variates:

$$\hat{y}_*^{(r)} = \sum_t \hat{g}(\hat{u}_t)/T$$

and

$$\hat{y}_*^{(ra)} = \sum_t [\hat{g}(\hat{u}_t) + \hat{g}(-\hat{u}_t)]/2T$$

$$= [y_*^{(r)} + y_*^{(r')}]/2 . \tag{3.25}$$

As in the case of the Monte Carlo predictor, the use of antithetic variates would add nothing but duplicative calculation if $g(.)$ is an even function of u. In general, unlike the result for the Monte Carlo predictor, we now have

$$E(\hat{y}_*^{(r)} | \hat{\theta}) \neq \gamma(x_*; \hat{\theta}) \tag{3.26}$$

since the conditional distribution of $\hat{u}_t$ given $\hat{\theta}$ would not be the same as the unconditional distribution of u_t with $\theta = \hat{\theta}$. Consequently, the finite sample bias of the residual-based predictor would differ from that of the closed form and the Monte Carlo predictors. Furthermore, the finite-sample efficiency of the residual based predictor relative to the closed form is ambiguous. Very simply, we can write

$$MSPE(\hat{y}_*^{(r)}) = MSPE(\hat{y}_*^{(c)}) + E(\hat{y}_*^{(r)} - \hat{\gamma}_*)(\hat{y}_*^{(r)} - \hat{\gamma}_*)'$$

$$+ E(\hat{y}_*^{(r)} - \hat{\gamma}_*)(\hat{\gamma}_* - \gamma_*)'$$

$$+ E(\hat{\gamma}_* - \gamma_*)(\hat{y}_*^{(r)} - \hat{\gamma})' \tag{3.27}$$

and the last two terms, coming from the cross product in the error decomposition

$$\hat{y}_*^{(r)} - \gamma_* = (\hat{y}_*^{(r)} - \hat{\gamma}_*) + (\hat{\gamma}_* - \gamma_*), \tag{3.28}$$

are not necessarily equal to zero as in the Monte Carlo predictor.

For the residual-based predictor with antithetics, the two components $\hat{y}_*^{(r)}$ and $\hat{y}_*^{(r')}$ in (3.25) do not necessarily have the same distribution unless $\hat{u}_t$ and $-\hat{u}_t$ are identically distributed. In general then, the use of antithetics would change the finite-sample of the residual-based predictor and no uniform statement can be made about the direction of change. However, if $\hat{u}_t$ and $-\hat{u}_t$ have the same probability distribution, some additional comparisons can be made.

Proposition 4. Suppose $\hat{u}_t$ and $-\hat{u}_t$ are identically distributed. Then

$$E(\hat{y}_*^{(r)}) = E(\hat{y}_*^{(ra)}) \tag{3.29}$$

and $MSPE(\hat{y}_*^{(ra)}) < MSPE(\hat{y}_*^{(r)})$ if and only if

$$COV(\hat{y}_*^{(r)}, \hat{y}_*^{(r')}) < V(\hat{y}_*^{(r)}). \tag{3.30}$$

Proof. Under the assumption of identical distributions for $\hat{u}_t$ and $-\hat{u}_t$, $y_*^{(r)}$ and $\hat{y}_*^{(r')}$ would also have identical distributions and (3.29) then follows directly from (3.25). Furthermore,

$$V(\hat{y}_*^{(ra)}) = (1/2)[V(\hat{y}_*^{(r)}) + COV(\hat{y}_*^{(r)}, \hat{y}_*^{(r')})]. \tag{3.31}$$

The second part of the proposition now follows from (3.29) and (3.31).

Q.E.D.

Another modification of the residual-based procedure is the bootstrap predictor which we define as

$$\hat{y}_*^{(b)} = \sum_{j=1}^{R} g(w_j, x_*, \hat{\theta})/R \tag{3.32}$$

where $w_1, w_2, \ldots, w_R$ are independent random draws (without replacement) from the empirical distribution of the residuals.

Proposition 5. For the bootstrap predictor,

$$E\hat{y}_*^{(b)} = E\hat{y}_*^{(r)} \tag{3.33}$$

and

$$MSPE(\hat{y}_*^{(b)}) = MSPE(\hat{y}_*^{(r)}) + E[V(\hat{y}_*^{(b)} | \hat{\theta})]. \tag{3.34}$$

Thus, in comparison with the residual-based, the bootstrap predictor has the same finite-sample bias and a larger mean squared prediction error. The difference in MSPEs is strictly positive definite for any replication size R but approaches zero as R approaches T.

Proof. The conditional distribution of w_j given $\hat{\theta}$, for any $j = 1, 2, \ldots, R$, is discrete with uniform mass of $1/T$ at the points $\hat{u}_t$ for $t=1, 2, \ldots, T$. Thus

$$E[g(w_j, x_*; \hat{\theta}) | \hat{\theta}] = \Sigma_t g(\hat{u}_t, x_*; \hat{\theta})/T = \hat{y}_*^{(r)} \tag{3.25}$$

and

$$E\hat{y}_*^{(b)} = EE(\hat{y}_*^{(b)} | \hat{\theta}) = E\hat{y}^{(r)}.$$

Furthermore,

$$V(\hat{y}_*^{(b)}) = EV(\hat{y}_*^{(b)} | \hat{\theta}) + V[E(\hat{y}_*^{(b)} | \theta)]$$

$$= EV(\hat{y}_*^{(b)} | \hat{\theta}) + V(\hat{y}_*^{(r)}), \text{ by (3.35)}, \tag{3.36}$$

thus proving (3.34). Finally, to prove that

$$EV(\hat{y}_*^{(b)} | \hat{\theta}) \to 0 \quad \text{as} \quad R \to T, \tag{3.37}$$

we note that $V(\hat{y}_*^{(b)} | \hat{\theta})$ is the variance of the mean based on a random sample of size R drawn without replacement from a finite population of size T. For $R = T$, this sample mean will always equal the population mean.

Q.E.D.

4. EXACT MOMENTS FOR PREDICTORS IN THE LOG-LINEAR UNIVARIATE MODEL

We now consider, as a special case, the log-linear univariate model

$$\text{Log } y_t = \alpha+\beta x_t+\varepsilon_t,\ \varepsilon_t \sim \text{iid } N(0,\sigma^2). \tag{4.1}$$

This is one of the simplest nonlinear models which we can use to illustrate the analysis in the paper. For example, Wallis (1979) and Calzolari and Corsi (1977) have used this model in their discussion of the bias in the deterministic predictor. Chesher, in his comments in a recent conference, also derived the finite-sample biases for the deterministic, Monte Carlo and residual-based predictors in this model.

This model has a closed form solution

$$y_t = \exp(\sigma u_t+\alpha+\beta x_t),\ u_t = \varepsilon_t/\sigma \sim \text{iid } N(0,1) \tag{4.2}$$

so that y_t has a log-normal distribution which we denote by

$$y_t \sim \text{LogN}(\alpha+\beta_t,\sigma^2) \tag{4.3}$$

and a mean function

$$\gamma_t = Ey_t = \exp(\sigma^2/2 + \alpha+\beta x_t). \tag{4.4}$$

The predictors of y_* are

$$\hat{y}_*^{(d)} = \exp(\hat{\alpha}+\hat{\beta}x_*)$$

$$\hat{y}_*^{(c)} = \exp(\hat{\sigma}^2/2+\hat{\alpha}+\hat{\beta}x_*)$$

$$= \hat{y}_*^{(d)} \cdot \exp(\hat{\sigma}^2/2)$$

$$\hat{y}_*^{(m)} = \hat{y}_*^{(d)} \cdot [\sum_{s=1}^{S} \exp(\hat{\sigma}\,\tilde{u}_s)/S]$$

$$\hat{y}_*^{(m)} = y_*^{(d)} \cdot [\sum_{t=1}^{T} \exp(\hat{\sigma}\,\hat{u}_t)/T] \qquad (4.5)$$

$$= \hat{y}_*^{(d)} \cdot [\sum_{t} \exp(\hat{\varepsilon}_t)/T]$$

$$\hat{y}_*^{(ma)} = y_*^{(d)} \cdot [\sum_{s=1}^{S} \{\exp(\hat{\sigma}\,\tilde{u}_s)+\exp(-\hat{\sigma}\tilde{u}_s)\}/(2S)]$$

$$y_*^{(ra)} = y_*^{(d)} \cdot [\sum_{t=1}^{T} \{\exp(\hat{\varepsilon}_t)+ \exp(-\hat{\varepsilon}_t)\}/(2T)]$$

$$\hat{y}_*^{(b)} = \hat{y}_*^{(d)} \cdot [\sum_{j=1}^{R} \exp(w_j)/R]$$

where

$$\hat{u}_s = \text{independent draws from } N(0,1) \text{ for } s = 1,2,\ldots;S,$$

$$\hat{\varepsilon}_t = \text{Log } y_t - \hat{\alpha} - \hat{\beta}x_t \qquad (4.6)$$

$$\hat{u}_t = \hat{\varepsilon}_t/\hat{\sigma}$$

w_j = independent draws from the discrete uniform distribution over $\{\hat{\varepsilon}_t : t=1,2,\ldots,T\}$ for $j=1,2,\ldots R$,

and $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}^2$ are the usual least squares estimators so that

$$\hat{\alpha}+\hat{\beta}x_* \sim N(\alpha+\beta x_*, \sigma^2\nu^2) \tag{4.7}$$

for $\nu^2 = (1,x_*)(X'X)^{-1}\begin{pmatrix}1\\x^*\end{pmatrix}$

$$X = \begin{pmatrix}1 & x_1\\ \vdots & \\ 1 & x_T\end{pmatrix}$$

and

$$(T-2)\hat{\sigma}^2/\sigma^2 \sim \chi^2(T-2), \text{ independent of } \hat{\alpha} \text{ and } \hat{\beta}. \tag{4.8}$$

We see from (4.5) that

$$\hat{y}_*^{(d)} \sim \text{LogN}(\alpha+\beta x_*, \sigma^2\nu^2) \tag{4.9}$$

and hence

$$E\hat{y}_*^{(d)} = \exp(\alpha+\beta x_*+\sigma^2\nu^2/2)$$

$$E(\hat{y}_*^{(d)})^2 = \exp\{2(\alpha+\beta x_*+\sigma^2\nu^2)\}. \tag{4.10}$$

For all the other predictions listed in (4.5), we can write each one of them as

$$\hat{y}_* = y_*^{(d)} \cdot Z \tag{4.11}$$

where Z changes across predictors but is always independently distributed of $\hat{y}_*^{(d)}$.

For the closed form predictor, we then have

$$E\hat{y}_*^{(c)} = [E\hat{y}_*^{(d)}] \cdot E[\exp(\hat{\sigma}^2/2)]$$

$$= [E\hat{y}_*^{(d)}] \cdot \phi[\sigma^2/(2(T-2))] \tag{4.12}$$

$$E(\hat{y}_*^{(c)}) = E(\hat{y}_*^{(d)})^2 \cdot \phi[\sigma^2/(T-2)]$$

where $\phi(\cdot)$ is the moment generating function of a central Chi-squared distribution with $(T-2)$ degrees of freedom.

For the Monte Carlo predictor, it can be directly verified that

$$E(\hat{y}_*^{(m)} | \hat{T}) = \hat{y}_*^{(c)} \tag{4.13}$$

so that

$$E\hat{y}_*^{(m)} = E\hat{y}_*^{(c)}. \tag{4.14}$$

Also,

$$E(\hat{y}_*^{(m)})^2 = E(\hat{y}_*^{(d)})^2 \cdot E(\Sigma_s \exp(\tilde{\varepsilon}_s)/S)^2 \tag{4.15}$$

$$= E(\hat{y}_*^{(d)})^2 \cdot \{\phi[2\sigma^2/(T-2)] + (S-1)\phi[\sigma^2/(T-2)]\}/S$$

since

$$E(\Sigma_s \exp(\tilde{\varepsilon}_s))^2 = E[E(\Sigma_s \exp(\tilde{\varepsilon}_s))^2 | \hat{\sigma}]$$

$$= E[S\exp(2\hat{\sigma}^2) + S(S-1)\exp(\hat{\sigma}^2)] \qquad (4.16)$$

$$= S\phi(2\sigma^2/(T-2)) + S(S-1)\phi(\sigma^2/(T-2))$$

with the second equality holding because

$$\exp(\tilde{\varepsilon}_s) | \hat{\sigma} \sim \text{i.i.d. } \text{LogN}(0,\hat{\sigma}^2). \qquad (4.17)$$

For the residual-based predictor, the vector of least squares residuals has a multivariate normal distribution:

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_T) \sim N(0, \sigma^2 Q)$$

$$Q = I - X(X'X)^{-1}X' = (q_{ij})$$

so that

$$\hat{\varepsilon}_t \sim \text{LogN}(0, \sigma^2 q_{tt})$$

and

$$E(y_*^{(r)}) = E(y_*^{(d)}) \cdot [\Sigma_t (e^{\sigma^2 q_{tt}/2}/T]$$

$$E(y^{(r)})^2 = E(y_*^{(d)})^2 \cdot [\Sigma_t \exp(2\sigma^2 q_{tt}) + 2 \Sigma_{i \neq j} \exp(\sigma^2 p_{ij}/2)]$$

$$p_{ij} = q_{ii} - 2q_{ij} + q_{jj}.$$

With regard to antithetic variates, this is a special case where $\hat{\varepsilon}_t$ and $-\hat{\varepsilon}_t$ have the same probability distribution. Thus the use of antithetic variates will not affect the first moment of not only the Monte Carlo predictor but also the residual-based predictor.

SELECTED REFERENCES

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

Brown, B.W. and R.S. Mariano, "Asymptotic Behavior of Predictors in Dynamic Nonlinear Simultaneous Systems." Unpublished; April, 1985.

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

Calzolari, G. and P. Corsi, "Stochastic Simulation as a Validation Tool for Econometric Models," Models for Regional Planning and Policy-Making (edited by A. Straszak and B.V. Wagle), (1977), 359-369.

Calzolari, G. "Antithetic Variates to Estimate the Simulation Bias in Nonlinear Models," Economic Letters 4 (1979), 323-328.

Calzolari, G. and F. Sterberz, "Efficient Computation of Reduced Form Variances in Nonlinear Econometric Models," Unpublished; Pisa: Centro Scientifico IBM, 1981.

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

Fisher, P. and M. Salmon, "On Evaluating the Importance of Nonlinearity in Large Macroeconometric Models", Unpublished; University of Warwick, 1984.

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

Mariano, R.S. and B.W. Brown, "Prediction-Based Tests for Misspecification in Nonlinear Simultaneous Systems," Studies in Econometrics, Time Series, and Multivariate Statistics (edited by S. Karlin, T. Amemiya, and L. Goodman), (1983b) 131-151. Academic Press.

Wallis, K.F. Topics in Applied Econometrics - Second Edition. Blackwell, 1979.