

Consistent Covariance Matrix Estimation  
In Probit Models with Autocorrelated Errors\*

by

Arturo Estrella  
Federal Reserve Bank of New York  
New York, NY 10045  
Tel.: 212-720-5874, Fax: 212-720-1582  
E-mail: arturo.estrella@ny.frb.org

and

Anthony P. Rodrigues  
Federal Reserve Bank of New York  
New York, NY 10045  
Tel.: 212-720-5027, Fax: 212-720-1582  
E-mail: tony.rodrigues@ny.frb.org

April 1998 (GMM2.WPD)

JEL Codes: C22, C25

Keywords: Probit, autocorrelation, generalized method of moments

\*The views expressed in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System. We would like to thank Christopher Sims, Kenneth West, and colleagues at the Federal Reserve Bank of New York, including participants at a seminar, for useful comments and Marianna Giokas for excellent research assistance.

## Consistent Covariance Matrix Estimation In Probit Models with Autocorrelated Errors

### Abstract

Some recent time-series applications use probit models to measure the forecasting power of a set of variables. Correct inferences about the significance of the variables requires a consistent estimator of the covariance matrix of the estimated model coefficients. A potential source of inconsistency in maximum likelihood standard errors is serial correlation in the underlying disturbances, which may arise, for example, from overlapping forecasts. We discuss several practical methods for constructing probit autocorrelation-consistent standard errors, drawing on the generalized method of moments techniques of Hansen (1982), Newey-West (1987) and others, and we provide simulation evidence that these methods can work well.

# Consistent Covariance Matrix Estimation In Probit Models with Autocorrelated Errors

## 1. Introduction

Empirical macroeconomic applications of probit models (or more generally, models of variables with  $\{0,1\}$  outcomes) with time series data often yield prediction errors that are autocorrelated. Autocorrelated errors can arise even when the marginal distribution is specified correctly. Common examples include cases in which the dependent variable is a multiple period outcome such as a probability forecast. This autocorrelation implies that conventional standard errors derived from maximum likelihood estimates (MLEs) assuming independent observations are inconsistent. Thus, tests based on those standard errors will be have incorrect size and confidence intervals will be misleading.

We propose using autocorrelation-consistent standard errors for models with discrete outcomes in the presence of autocorrelation. One advantage of our approach is that these standard errors can be computed for the probit model in a straightforward way. Some earlier papers have addressed the main issue of this paper, but the results have not led to suggestions that are easy to use in practical applications. Examples of these papers are Gouriéroux, Monfort, and Trognon (1984), Bates and White (1988), Poirier and Ruud (1988), and Wooldridge (1994). Here, we compute an autocorrelation-consistent covariance matrix using the generalized method of moments (GMM) procedures suggested by Hansen (1982), Newey-West (1987) and Andrews (1991).

We also provide simulation evidence that these methods can work well, particularly in large samples when both the forecasting variables and the model disturbances are serially correlated. In most cases, the standard Hansen (1982) covariance estimator and the Newey-West (1987) estimator with optimal lag length produce test statistics of approximately the correct size. When there is information about the MA process of the discrete choice model errors, either of these estimators provides a good choice for computing standard errors.

The paper begins by constructing a model that incorporates the key features of the applications in the literature that would benefit from our approach. Using this model, we prove

the claims made above regarding consistency and relate our approach to others in the literature. Finally, we use simulations to obtain numerical estimates of the benefits of our approach.

## **2. Probit Models with Autocorrelated Errors**

Historically, most applications of probit have been cross-sectional, where autocorrelated errors would not naturally arise. Autocorrelation can arise in panel data problems, and Heckman (1981a, 1981b) has discussed approaches to this problem that exploit both the time series and cross section aspects of panel data. In addition, Avery, Hansen, and Hotz (1983) have a probit application with panel data, which requires a large cross-section relative to the time-series and is therefore not directly applicable to pure time-series data sets.

In contrast, our approach is motivated by problems with purely time series data where cross-sectional information cannot be used for identification. Recent macroeconomic literature contains examples of models of multi-period expectations with discrete outcomes, which are likely to have serially correlated expectation errors with a moving average form. These papers include Stock and Watson (1991), Estrella and Hardouvelis (1991), Bernard and Gerlach (1996), and Estrella and Mishkin (1997), all of whom are interested in forecasting recessions over a variety of horizons.

A natural way to approach the serial correlation problem in these papers is suggested by the procedures of Hansen (1982), Newey-West (1987) and Andrews (1991). However, while these authors have addressed the problem of covariance estimation in great generality, they did not discuss the particular issues involved in probit applications. Methods for estimating the covariance matrix of coefficients in purely time-series applications of probit are presented in Poirier and Ruud (1988) and Gouriéroux, Monfort, and Trognon (1984), but these papers are largely theoretical and use computationally complex methods to obtain the covariance matrix. Their methods require, for example, the numerical computation of single and double integrals of the normal density function. Our contribution is to present a family of computationally simple covariance matrix estimators, to demonstrate their asymptotic equivalence to the Poirier and Ruud (1988) and Gouriéroux, Monfort, and Trognon (1984) matrices (and, thus, their validity in large samples), and to investigate their small sample performance through simulation.

Our theoretical framework starts with the assumption that we observe a time series on a binary variable,  $y_t$ , whose values are determined by the unobserved variable,  $y_t^*$ , defined by the equation:

$$y_t^* = \beta'x_t + \epsilon_t$$

In the above,  $\beta$  is a vector of  $k$  coefficients and  $x_t$  is a vector of values of  $k$  independent variables. The observable variable  $y_t$  has only two possible values and is related to  $y_t^*$  through the relationship

$$y_t = 1 \text{ if } y_t^* > 0 \text{ and } y_t = 0 \text{ otherwise.}$$

The following distributional assumption motivates our use of probit models.

**Assumption A:** Conditional on all the  $x_t$ , the vector of disturbances  $\epsilon_t$  is multivariate normal with zero mean and covariance matrix  $\Sigma$ , with diagonal elements equal to one.

This assumption implies that, conditional on the matrix of values of  $x_t$ ,  $t=1, \dots, T$ , the distribution of the vector of values of  $y_t^*$  is multivariate normal with mean elements  $\beta'x_t$  and covariance matrix  $\Sigma$ . We also assume that the variances of  $\epsilon_t$  are constant. As in the standard probit model, we cannot identify the value of this constant variance, but only the ratio of the parameters  $\beta$  to the square root of the variance. Hence we set the variance to be equal to 1, but allow the off-diagonal elements of  $\Sigma$  to be non-zero. It follows from the foregoing that

$$\Pr(y_t=1|x_t) = \Phi(\beta'x_t)$$

where  $\Phi$  represents the standard normal cumulative distribution function.

To complete the description of data, we also make

**Assumption B1:** The  $x_t$  are non-random and bounded, the univariate empirical distribution of sequences of  $x_t$  converges to a single limiting distribution, regardless of the starting time, and the joint empirical distribution of  $(x_t, x_{t+\alpha})$  converges to a limiting distribution that depends on  $\alpha$ .

Alternatively, we can hold

**Assumption B2:** The distribution of the  $x_t$  is stationary and ergodic.

The normality assumptions suffice to produce consistency of the coefficient covariance matrix estimators and are in line with most of the existing literature. However, normality is not necessary for some of the results, as argued by Poirier and Ruud (1988).

### 3. Covariance Matrix Estimation

If there were sufficient information about the variance-covariance structure of the errors in the model of section 2, the ideal estimation method would be full information maximum likelihood (see, e.g., Eichengreen, Watson and Grossman (1985)). In most applications, however, the precise form of the time series model for the disturbances is not known. In those cases, a method that does not impose a particular time series model is less prone to specification error. We consider two basic approaches to estimation: a quasi-maximum likelihood estimator (QMLE) that does not incorporate assumptions about the time series model followed by the disturbances and a GMM estimator that makes use of empirical moments of a function of the parameters. We argue that both the QMLE and its corresponding GMM estimator produce consistent coefficient estimates, and that the GMM methods for computing standard errors are consistent.

The QMLE method uses the likelihood function corresponding to independent observations, in other words a standard probit formulation. Define

$$F_t \equiv \Phi(\beta'x_t) , \quad f_t \equiv \Phi'(\beta'x_t)$$

and construct the log-likelihood function

$$\log L = \sum_{t=1}^T \{ y_t F_t + (1-y_t)(1-F_t) \}$$

This is the appropriate likelihood under the assumption that  $\text{Var}(\epsilon) = \mathbf{I}$ . Let  $u_t = y_t - F_t$  and  $w_t^2 = 1 / F_t(1-F_t)$ . Then the first order conditions which define the QMLE of  $\beta$  are:

$$\sum_{t=1}^T u_t w_t^2 f_t x_t = 0 \tag{1}$$

Amemiya (1976, 1981) has demonstrated this estimator is equivalent to a non-linear weighted least squares estimator that minimizes

$$\sum_{t=1}^T w_t^2 u_t^2 .$$

Poirier and Ruud (1988) and Gourieroux, Monfort, and Trognon (1984) prove that this QMLE is consistent under our Assumptions A and B1. Wooldridge (1994), as an implication of a more general result, shows that the QMLE is consistent under our Assumptions A and B2. In either case, we can use the QMLE to construct a GMM estimator of the covariance matrix  $\Sigma$ . Hansen's (1982) Theorem 3.1 can be then be applied to demonstrate the consistency of the covariance estimates.

To construct the GMM estimator, we select a set of orthogonality conditions that have zero expected values at the true parameters. Estimation proceeds by choosing parameter estimates to minimize the weighted distance of the orthogonality conditions from zero. A natural candidate for the orthogonality conditions is a weighted quadratic form of the vector of first order conditions from the QMLE above. To be more specific, define for each observation

$$h_t \equiv u_t w_t^2 f_t' x_t \text{ and for the whole sample } h \equiv \sum_t h_t .$$

Actual estimation involves selecting  $\beta$  to minimize the quadratic form  $h' W h$  for some choice of the weight matrix  $W$ . One possibility for the weighting matrix is  $W = \mathbf{I}$ , although any positive definite matrix  $W$  produces the first order conditions  $h = 0$ , which satisfy equation (1) and are thus equivalent to the QMLE.

More generally, estimates of the covariance matrix of either estimator are based on estimators of the covariance matrix of  $h$ . Building blocks are the sample autocovariances of  $h$ :

$$\hat{\Omega}_j = \frac{1}{T} \sum_{t=j+1}^T h_t h_{t-j}'$$

These autocovariances can be used, following Hansen (1982), Newey-West (1987) or Gallant (1987), to construct a family of estimators of the covariance of  $h$ :

$$\hat{S} = \hat{\Omega}_0 + \sum_{j=1}^m \lambda_j (\hat{\Omega}_j + \hat{\Omega}_j')$$

where  $\lambda_j = 1$  in Hansen (1982),  $\lambda_j = 1 - \frac{j}{m+1}$  in Newey-West (1987), and  $\lambda_j$  are Parzen weights in Gallant (1987). With a proper selection of  $m$ , this matrix is a consistent estimator of the covariance matrix of the orthogonality conditions. Note that for consistency, Newey and West argue that  $m$  should grow with the sample size, but at a fractional power of the sample size. The latter is a standard result in spectral analysis; see, for instance, Brillinger (1981). Andrews (1991) discusses optimal growth rates for  $m$  as the sample size increases.

The consistent covariance matrix for the GMM estimator obtained with weighting matrix  $W=I$  is, following Hansen (1982) and Newey-West (1987),  $V = \frac{1}{T}(H' H)^{-1} H' \hat{S} H (H' H)^{-1}$ , where  $H \equiv \frac{1}{T} \frac{\partial h}{\partial \beta} = \frac{1}{T} \sum_t \frac{\partial h_t}{\partial \beta}$  and  $\hat{S}$  is defined above. Newey and West (1987) and Gallant (1987) show also that the covariance estimator constructed using their weights is guaranteed to be positive definite.

Asymptotically, there is no efficiency gain from using weighting matrices other than  $W=I$ . This follows because

$$\begin{aligned} H &= \frac{1}{T} \sum_t \frac{\partial h_t}{\partial \beta} = \frac{1}{T} \sum_t \frac{\partial (y_t - F_t) f_t x_t}{\partial \beta F_t (1 - F_t)} \\ &= \frac{1}{T} \sum_t \frac{-f_t^2 x_t x_t'}{F_t (1 - F_t)} + \frac{1}{T} \sum_t (y_t - F_t) \frac{\partial}{\partial \beta} \frac{f_t x_t}{F_t (1 - F_t)} \end{aligned}$$

Each of the elements in  $\frac{1}{T} \sum_t (y_t - F_t) \frac{\partial}{\partial \beta} \frac{f_t x_t}{F_t (1 - F_t)}$  has expectation zero and the variance is a

weighted sum of the covariance matrix of forecast errors so this term converges to zero. Thus, in

large samples, the sum  $H$  converges to  $H_0 = \lim \frac{1}{T} \sum_t \frac{-f_t^2 x_t x_t'}{F_t (1 - F_t)}$ . Thus

$T V_w = (H'WH)^{-1} W'H' \hat{S} HW(H'WH)^{-1}$  converges to  $H_0^{-1} S H_0^{-1}$  where  $S$  represents the limiting value of the matrix  $\frac{1}{T} \hat{S}$ .

As we noted in the introduction, other consistent estimates of the covariance matrix have been obtained by Poirier and Ruud (1988) and Gouriéroux, Monfort, and Trognon (1984). Our covariance matrix converges to the covariance matrices derived by these authors -- they are asymptotically equivalent. For example, consider the Poirier and Ruud (1988) matrix:

$$\Omega^{OP} \equiv \Delta_0^{-1} [\Delta_0 + 2 \sum_{j=1}^{\infty} \Delta_j] \Delta_0^{-1} = \Delta_0^{-1} [\Delta_0 + \sum_{j=1}^{\infty} (\Delta_j + \Delta_j')] \Delta_0^{-1}$$

with

$$\Delta_j \equiv E_x \left[ x x_{-j}' \frac{\phi(x' \beta_0) \phi(x_{-j}' \beta_0) \{ \Phi_2(x' \beta_0, x_{-j}' \beta_0, \rho_j) - \Phi(x' \beta_0) \Phi(x_{-j}' \beta_0) \}}{\Phi(x' \beta_0)(1 - \Phi(x' \beta_0)) \Phi(x_{-j}' \beta_0)(1 - \Phi(x_{-j}' \beta_0))} \right]$$

where  $\rho_j \equiv Cov(u_t, u_{t-j})$  and  $\Phi_2$  is the bivariate standard normal cumulative distribution.

From the above expression, it may be verified that  $\Delta_0 = H_0$ , where the latter is the limiting case of our  $H$  matrix. Further, consider our  $\hat{S} = \hat{\Omega}_0 + \sum_{j=1}^m \lambda_j (\hat{\Omega}_j + \hat{\Omega}_j')$  where

$$\hat{\Omega}_j = \frac{1}{T} \sum h_t h_{t-j}' = \frac{1}{T} \sum \frac{(y_t - F_t) f_t x_t x_{t-j}' f_{t-j} (y_{t-j} - F_{t-j})}{F_t (1 - F_t) F_{t-j} (1 - F_{t-j})}$$

Since  $E[(y_t - F_t)(y_{t-j} - F_{t-j})] = \Phi_2(x' \beta_0, x_{-j}' \beta_0, \rho_j) - \Phi(x' \beta_0) \Phi(x_{-j}' \beta_0)$  from Gouriéroux, Monfort, and Trognon (1984), the sum in  $\hat{\Omega}_j$  is a weighted sum of terms which converge to the Poirier-Ruud  $\Delta_j$ . Thus  $\frac{1}{T} \hat{S} \rightarrow \Delta_0 + \sum_{j=1}^{\infty} [\Delta_j + \Delta_j']$ .

However, the matrices proposed by Gourieroux, Monfort, and Trognon (1984) and Poirier and Ruud (1988) are computationally much more demanding than our estimator. Note that their estimators require that we know the correlations between disturbances across observations or that we are able to estimate them consistently from the observed covariances of  $(y-F)$ . In practice, our estimator is easier because an investigator only needs to compute the covariance between the sample moments,  $\frac{1}{T} \sum h_t h_{t-j}'$

To summarize, the QMLE/GMM interpretation of probit allows us to estimate the coefficients in the marginal probability distribution consistently, to construct a computationally straightforward non-singular covariance matrix based on time-series variation only, and, when the Newey-West method is employed, to ensure that the covariance matrix is positive semi-definite.

#### 4. Simulations

This section reports the results of experiments that test the accuracy of the asymptotic approaches outlined above in applications with moderate size samples. We focus on the performance of the QMLE and of three alternative GMM estimators that are representative of the available range of GMM estimators. The performance of these estimators is compared in two simulated estimation scenarios: when the lag length of the moving average process is specified correctly and when the lag length used in the covariance estimator is shorter than the correct lag length.

Our first set of simulations investigates the behavior of the four covariance estimators for the probit QMLE in a setting where the disturbance process is known. These simulations correspond most closely to an application where a probit model is used to forecast a discrete outcome over a period exceeding the interval between observations. Specifically, the disturbance process is assumed to be MA(1),

$$\epsilon_t = \frac{1}{\sqrt{1 + \gamma^2}} \eta_t + \frac{\gamma}{\sqrt{1 + \gamma^2}} \cdot \eta_{t-1}$$

and the observed explanatory variable is AR(1),  $x_t = \alpha \cdot x_{t-1} + v_t$ .

We consider three sample sizes of 100, 500, and 1000. Our parameter values for the disturbance process correspond to autocorrelations of  $-.5$  ( $\gamma=-1$ ),  $0$  ( $\gamma=0$ ), and  $.5$  ( $\gamma=1$ ), while autocorrelation in the explanatory variable ( $\alpha$ ) is either  $0$ ,  $.5$ , or  $.9$ . Our covariance estimators include the QMLE estimator, Newey-West (1987) with one lag, Hansen (1982) with one lag, and Newey-West (1987) with an estimated optimal lag length derived following Andrews (1991) with an estimated MA(1) process. Each simulation result was obtained with 10,000 iterations.

We compare the bias of the covariance estimators in the three panels of Table 1. Biases are computed by comparing the average standard errors from each method to the standard deviation of the simulation coefficients. The small sample (100 observation) bias calculations in panel A indicate that the performance of the covariance matrix estimators depends on the extent of the serial correlation in the data. When the disturbances are positively correlated ( $\gamma=1$ ), the Hansen (1982) covariance matrix estimator has smallest bias. Perhaps unsurprisingly, the QMLE has the least bias when the disturbances are not serially correlated. This estimator also tends to have the lowest bias when the explanatory variable is serially uncorrelated.

Medium (500 observation) and large sample (1000 observation) results show that the Hansen (1982) estimator is the least biased when the disturbances are positively correlated and, in most cases, when they are negatively correlated. The QMLE is typically best when there is no serial correlation, although all of the estimators have low bias in this case. The Hansen (1982) and Newey-West with optimal lag length estimators seem to have faster bias reductions as the sample size grows.

Figure 1 presents graphically most of the key information in the three panels of table 1 (results for  $\alpha=.5$  are omitted). This figure shows clearly that the small sample (100) bias tends to be large for all estimators, but that the QMLE tends to be preferable in a small sample when there is no autocorrelation in either the disturbance or the explanatory variable. In larger samples (500 and 1000), the Hansen estimator performs quite well. It tends to perform better than the other estimators regardless of the autocorrelation in the residual or the explanatory variable.

Table 2 provides another perspective on the performance of each estimator when used with the asymptotic normal approximation to construct confidence regions or hypothesis tests. The table reports the simulated coverage percentages -- the percentage of simulations where the

true coefficient fell in a confidence interval constructed using either 90%, 95%, or 99% level -- for each covariance estimator in the experiments reported above.

The small sample results indicate that the QMLE is closest to the theoretical coverage probabilities either if there is no residual serial correlation or if there is no serial correlation in the explanatory variable. Hansen (1982) and Newey-West estimators have slightly low coverage probabilities in these cases, implying that their confidence intervals are somewhat too small. Newey-West methods typically are closest to the theoretical coverage when the disturbances are negatively autocorrelated. No method works very well in a small sample when both the disturbance and the explanatory variable are positively autocorrelated, but the Hansen (1982) method does best when there is high positive serial correlation.

In the medium to large sample results, Hansen (1982) performs best when the explanatory variable is positively correlated, while the Newey-West methods are occasionally slightly better than the Hansen (1982) approach when the disturbance is negatively autocorrelated. Not surprisingly, the greatest gains to using the Hansen (1982) or Newey-West approaches arise when both the disturbance and the explanatory variable are highly autocorrelated. There is support for using the QMLE when the disturbance is not serially correlated, although the other consistent methods work about as well in large samples.

Finally, we consider the performance of the standard error estimators when the econometrician applies methods that could be consistent, but uses too short a lag length in Hansen (1982) or Newey-West (1987) or in the specification of the MA process in computing the optimal Newey-West lag length. Specifically, in the basic model above we allow disturbances to follow an MA(2) process:

$$\epsilon_t = \frac{1}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \eta_t + \frac{\gamma_1}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \cdot \eta_{t-1} + \frac{\gamma_2}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \cdot \eta_{t-2}$$

However, we assume that the econometrician implements Hansen (1982) or Newey-West (1987) using one lag, assuming that the errors follow a MA(1) process.

Bias results are presented in Table 3. Except for small samples when there is no correlation in the explanatory variable, the Hansen (1982) method generally produces the smallest bias. Although the Newey-West optimal length estimator is consistent in these simulations, it seems that the misspecification of the error process leads to an inappropriate choice of lag length, so the inconsistent Hansen (1982) estimator has lower bias in most of the simulations.

Table 4 presents evidence on the coverage performance of these four estimators. While the QMLE is closest to the theoretical coverage in small samples with no autocorrelation in the explanatory variable, the Hansen (1982) and Newey-West with optimal lag length estimators produce as good or considerably better coverage in all other cases.

We conclude this section with two comments regarding the robustness of the results reported in the figure and tables. First, the sign of the autocorrelation ( $\alpha$ ) in the explanatory variable does not materially affect the results of the paper; that is, with a negative autocorrelation, we obtain results that are very analogous to those reported. To take a specific example, if the signs of both autocorrelation parameters ( $\alpha$  and  $\gamma$ ) are reversed in the right-hand panels of figure 1, the results are remarkably similar to those in the figure.

The second comment is of potentially greater significance. In models not based on forward-looking forecasts, the error process may be autoregressive rather than moving average. With mild autocorrelation in the errors (e.g., AR(1) with a correlation of .5), simulation results available from the authors are qualitatively similar to those in the figure and table. If AR(1) errors are highly autocorrelated (e.g., correlation of .8), the bias introduced by using a finite lag length may be large, even in large samples with the Newey-West and Hansen estimators with the Andrews (1991) optimal lag length.

## **5. Conclusions**

In this paper we have argued that, under relatively unrestrictive conditions, standard probit coefficient estimates are consistent even in the presence of autocorrelation in the disturbances. This implies that the GMM methodology of Hansen (1982), Newey-West (1987) and others may be used to construct consistent estimates of the covariance matrix of the coefficient estimates. Although these results suggest that one may approach large sample results with confidence, we

also perform simulations to examine the performance of the autocorrelation-consistent covariance estimators in small samples.

Our simulation results suggest that the Hansen (1982) estimator consistently yields superior standard error estimates, particularly when there is substantial positive autocorrelation in both the disturbance and the explanatory variables. However, Hansen (1982) is not always preferred to the other standard error estimators, particularly if the sample size is small and there is little autocorrelation in the disturbance. In those circumstances, the QMLE works well. All standard error estimators exhibit substantial bias in small samples, particularly when the disturbances and explanatory variables are highly positively autocorrelated. This bias can drop off quickly with sample size for the consistent estimators, particularly in the case of the estimators based on Hansen (1982) and Newey-West (1987) with optimal lag length.

## References

- Amemiya, Takeshi. 1976. The Maximum Likelihood, the Minimum Chi-Square and the Nonlinear Weighted Least-Squares Estimator in the General Qualitative Response Model. *Journal of the American Statistical Association* 71, 347-351.
- Amemiya, Takeshi. 1981. Qualitative Response Models: A Survey. *Journal of Economic Literature* 19, 1483-1536.
- Amemiya, Takeshi. 1985. *Advanced Econometrics*. Harvard University Press, Cambridge, MA.
- Andrews, Donald W.K. 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817-858.
- Avery, Robert B., Hansen, Lars Peter and Hotz, V. Joseph. 1983. Multiperiod Probit Models and Orthogonality Condition Estimation. *International Economic Review* 24, 21-35.
- Bates, Charles and White, Halbert. 1988. Efficient Instrumental Variables Estimation of Systems of Implicit Heterogeneous Nonlinear Dynamic Equations with Nonspherical Errors. In Barnett, W.A., Berndt, E.R. and White, H. (Eds.) *Dynamic Econometric Modeling: Proceedings of the Third International Symposium in Economic Theory and Econometrics*. Cambridge University Press, Cambridge.
- Bernard, Henri and Gerlach, Stefan. 1996. Does the Term Structure Predict Recessions? The International Evidence. Bank for International Settlements Working Paper No. 37.
- Brillinger, D.R. 1981. *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco.
- Eichengreen, Barry, Watson, Mark W. and Grossman, Richard S. 1985. Bank Rate Policy Under the Interwar Gold Standard: A Dynamic Probit Model. *The Economic Journal* 95, 725-745.
- Estrella, Arturo and Hardouvelis, Gikas. 1991. The Term Structure as a Predictor of Real Economic Activity. *Journal of Finance* 46, 555-576.
- Estrella, Arturo and Mishkin, Frederic S. 1998. Predicting U.S. Recessions: Financial Variables as Leading Indicators. *Review of Economics and Statistics* 80, 45-61.
- Gallant, A. Ronald. 1987. *Nonlinear Statistical Models*. Wiley, New York.
- Gourieroux, C., Monfort, A. and Trognon, A. 1984. Estimation and Test in Probit Models with Serial Correlation. In Florens, J.P., Mouchart, M., Raoult, J.P. and Simar, L. (Eds.).

- Alternative Approaches to Time Series Analysis. Publications des Facultés Universitaires Saint-Louis, Brussels.
- Hansen, Lars Peter. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029-1054.
- Heckman, James L. 1981a. Statistical Models for Discrete Panel Data. In Manski C. and McFadden, D. (Eds.) *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA.
- Heckman, James L. 1981b. The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time--Discrete Data Stochastic Process. In Manski C. and McFadden, D. (Eds.) *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA.
- Newey, Whitney K. and West, Kenneth D. 1987. A Simple-Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55, 703-708.
- Poirier, Dale J. and Ruud, Paul A. 1988. Probit with Dependent Observations. *Review of Economic Studies* 55, 593-614.
- Stock, James H. and Watson, Mark W. 1991. A Probability Model of the Coincident Economic Indicators. In Lahiri, Kajal and Moore, Geoffrey H. (Eds.) *Leading Economic Indicators: New Approaches and Forecasting Records*. Cambridge University Press, Cambridge.
- Wooldridge, Jeffrey. 1994. Estimation and Inference for Dependent Processes. *Handbook of Econometrics*. Elsevier, Amsterdam.

**Table 1: Bias of Several Probit Standard Error Estimators  
(percent of estimated standard error of coefficient)**

**Panel A: Sample size = 100**

		$\gamma$											
		-1				0				1			
		Probit	N-W	N-W opt	Hansen	Probit	N-W	N-W opt	Hansen	Probit	N-W	N-W opt	Hansen
	.0	-5.4	-8.2	-8.5	-8.3	-6.4	-9.4	-9.6	-10.0	-10.8	-11.5	-11.6	-9.9
$\alpha$	.5	-1.6	-6.4	-6.8	-8.7	-2.7	-6.1	-6.4	-7.1	-14.1	-15.0	-14.9	-12.6
	.9	12.5	1.8	-2.8	-6.7	-5.7	-10.1	-10.7	-11.4	-20.2	-17.0	-16.2	-10.8

Notes: The table reports estimated biases for standard errors from probit QMLE (Probit), Newey-West with one lag (N-W), Newey-West with Andrews (1991) optimal MA(1) lag (N-W opt), and the Hansen (1982) with one lag (Hansen). Estimated biases are reported as a percent of the estimated standard error of the simulated coefficients. Data were generated from an unobserved indicator

model  $y_t^* = \beta'x_t + \eta_t$  where  $y_t = \begin{cases} 1, & y_t^* \geq 0 \\ 0, & y_t^* < 0 \end{cases}$  is observed,  $\eta_t = \frac{1}{\sqrt{1+\gamma^2}}\epsilon_t + \frac{\gamma}{\sqrt{1+\gamma^2}} \cdot \epsilon_{t-1}$ , and  $x_t = \alpha \cdot x_{t-1} + v_t$ . The number of repetitions was 10000 with  $\beta=1$ .

**Table 1: Bias of Several Probit Standard Error Estimators  
(percent of estimated standard error of coefficient)**

**Panel B: Sample size = 500**

		$\gamma$											
		-1				0				1			
		Probit	N-W	N-W opt	Hansen	Probit	N-W	N-W opt	Hansen	Probit	N-W	N-W opt	Hansen
	.0	-5.1	-4.4	-4.3	-3.2	-0.3	-1.1	-1.1	-1.2	-4.6	-3.8	-3.7	-2.6
$\alpha$	.5	2.6	0.0	-0.2	-1.8	-1.4	-2.3	-2.3	-2.6	-9.4	-5.9	-4.7	-1.8
	.9	14.7	7.0	2.5	-0.5	-0.8	-1.7	-1.8	-2.0	-17.3	-9.9	-6.4	-2.5

Notes: The table reports estimated biases for standard errors from probit QMLE (Probit), Newey-West with one lag (N-W), Newey-West with Andrews (1991) optimal MA(1) lag (N-W opt), and the Hansen (1982) with one lag (Hansen). Estimated biases are reported as a percent of the estimated standard error of the simulated coefficients. Data were generated from an unobserved indicator

model  $y_t^* = \beta'x_t + \epsilon_t$  where  $y_t = \begin{cases} 1, & y_t^* \geq 0 \\ 0, & y_t^* < 0 \end{cases}$  is observed,  $\epsilon_t = \frac{1}{\sqrt{1+\gamma^2}}\eta_t + \frac{\gamma}{\sqrt{1+\gamma^2}} \cdot \eta_{t-1}$ , and  $x_t = \alpha \cdot x_{t-1} + v_t$ . The number of repetitions was 10000 with  $\beta=1$ .

**Table 1: Bias of Several Probit Standard Error Estimators  
(percent of estimated standard error of coefficient)**

**Panel C: Sample size = 1000**

		$\gamma$											
		-1				0				1			
		Probit	N-W	N-W opt	Hansen	Probit	N-W	N-W opt	Hansen	Probit	N-W	N-W opt	Hansen
	.0	-2.3	-1.0	-0.8	0.5	0.4	-0.0	-0.1	-0.1	-4.5	-3.1	-2.8	-1.4
$\alpha$	.5	4.7	2.2	1.7	0.2	-0.5	-0.8	-0.8	-0.9	-10.5	-6.3	-4.4	-1.9
	.9	14.7	7.4	2.6	0.1	-1.0	-1.3	-1.4	-1.5	-16.2	-8.6	-4.3	-1.2

Notes: The table reports estimated biases for standard errors from probit QMLE (Probit), Newey-West with one lag (N-W), Newey-West with Andrews (1991) optimal MA(1) lag (N-W opt), and the Hansen (1982) with one lag (Hansen). Estimated biases are reported as a percent of the estimated standard error of the simulated coefficients. Data were generated from an unobserved indicator

model  $y_t^* = \beta'x_t + \epsilon_t$  where  $y_t = \begin{cases} 1, & y_t^* \geq 0 \\ 0, & y_t^* < 0 \end{cases}$  is observed,  $\epsilon_t = \frac{1}{\sqrt{1+\gamma^2}}\eta_t + \frac{\gamma}{\sqrt{1+\gamma^2}} \cdot \eta_{t-1}$ , and  $x_t = \alpha \cdot x_{t-1} + v_t$ . The number of repetitions was 10000 with  $\beta=1$ .

**Table 2: Coverage Probabilities (Percent) for Several Probit Standard Error Estimators**  
**Panel A: Simulated Sample Size equals 100.**

		<u>Probit MLE</u>			<u>Newey-West</u>			<u>Newey-West (Opt)</u>			<u>Hansen</u>		
Theoretical Level:		<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>
<u><math>\alpha</math></u>	<u><math>\gamma</math></u>												
<b>0</b>	<b>-1</b>	99	95	90	99	94	88	99	93	88	98	93	88
<b>0</b>	<b>0</b>	99	95	90	98	94	88	98	93	88	98	93	87
<b>0</b>	<b>1</b>	99	94	88	98	92	86	98	92	86	98	93	87
<b>.5</b>	<b>-1</b>	99	96	91	99	94	89	98	94	89	98	93	88
<b>.5</b>	<b>0</b>	99	96	91	99	94	89	99	94	88	99	94	88
<b>.5</b>	<b>1</b>	98	93	87	98	92	85	98	92	85	98	92	86
<b>.9</b>	<b>-1</b>	100	98	95	99	96	91	98	94	89	98	93	88
<b>.9</b>	<b>0</b>	99	95	90	98	93	88	98	93	87	98	93	87
<b>.9</b>	<b>1</b>	97	90	82	96	90	83	97	90	84	97	92	86

Notes: The table reports simulated coverage probabilities for 90%, 95%, and 99% confidence intervals obtained with standard errors from probit QMLE, Newey-West with one lag, Newey-West with Andrews (1991) optimal lag, and the Hansen (1982) with one lag.

Data were generated from an unobserved indicator model  $y_t^* = \beta'x_t + \epsilon_t$  where  $y_t = \begin{cases} 1, & y_t^* \geq 0 \\ 0, & y_t^* < 0 \end{cases}$  is observed,  $x_t = \alpha \cdot x_{t-1} + v_t$ ,

and  $\epsilon_t = \frac{1}{\sqrt{1 + \gamma^2}} \eta_t + \frac{\gamma}{\sqrt{1 + \gamma^2}} \cdot \eta_{t-1}$ . The number of repetitions was 10000 and  $\beta=1$ .

**Table 2: Coverage Probabilities (Percent) for Several Probit Standard Error Estimators**  
**Panel B: Simulated Sample Size equals 500.**

		<u>Probit MLE</u>			<u>Newey-West</u>			<u>Newey-West (Opt)</u>			<u>Hansen</u>		
Theoretical Level:		<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>
<u><math>\alpha</math></u>	<u><math>\gamma</math></u>												
<b>0</b>	<b>-1</b>	99	94	88	99	94	88	99	94	88	99	94	89
<b>0</b>	<b>0</b>	99	95	90	99	95	90	99	95	90	99	95	90
<b>0</b>	<b>1</b>	99	94	89	99	94	89	99	94	89	99	95	89
<b>.5</b>	<b>-1</b>	99	96	91	99	95	90	99	95	90	99	95	89
<b>.5</b>	<b>0</b>	99	95	90	99	95	89	99	94	89	99	94	89
<b>.5</b>	<b>1</b>	98	93	87	98	94	88	99	94	88	99	94	89
<b>.9</b>	<b>-1</b>	100	98	94	99	96	92	99	95	91	99	95	90
<b>.9</b>	<b>0</b>	99	95	90	99	94	89	99	94	89	99	94	89
<b>.9</b>	<b>1</b>	97	90	83	98	92	86	98	93	88	99	94	89

Notes: The table reports simulated coverage probabilities for 90%, 95%, and 99% confidence intervals obtained with standard errors from probit QMLE, Newey-West with one lag, Newey-West with Andrews (1991) optimal lag, and the Hansen (1982) with one lag.

Data were generated from an unobserved indicator model  $y_t^* = \beta'x_t + \epsilon_t$  where  $y_t = \begin{cases} 1, & y_t^* \geq 0 \\ 0, & y_t^* < 0 \end{cases}$  is observed,  $x_t = \alpha \cdot x_{t-1} + v_t$ ,

and  $\epsilon_t = \frac{1}{\sqrt{1 + \gamma^2}} \eta_t + \frac{\gamma}{\sqrt{1 + \gamma^2}} \cdot \eta_{t-1}$ . The number of repetitions was 10000 and  $\beta=1$ .

**Table 2: Coverage Probabilities (Percent) for Several Probit Standard Error Estimators  
Panel C: Simulated Sample Size equals 1000.**

		<u>Probit MLE</u>			<u>Newey-West</u>			<u>Newey-West (Opt)</u>			<u>Hansen</u>		
Theoretical Level:		<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>
<u><math>\alpha</math></u>	<u><math>\gamma</math></u>												
<b>0</b>	<b>-1</b>	99	94	89	99	95	89	99	95	90	99	95	90
<b>0</b>	<b>0</b>	99	95	90	99	95	90	99	95	90	99	95	90
<b>0</b>	<b>1</b>	99	94	89	99	94	89	99	94	89	99	95	90
<b>.5</b>	<b>-1</b>	99	96	91	99	96	91	99	95	91	99	95	90
<b>.5</b>	<b>0</b>	99	95	90	99	95	90	99	95	90	99	95	90
<b>.5</b>	<b>1</b>	98	93	86	99	94	88	99	94	88	99	95	89
<b>.9</b>	<b>-1</b>	100	97	94	99	96	92	99	95	91	99	95	90
<b>.9</b>	<b>0</b>	99	95	90	99	95	89	99	95	89	99	95	89
<b>.9</b>	<b>1</b>	97	90	83	98	93	87	99	94	88	99	95	90

Notes: The table reports simulated coverage probabilities for 90%, 95%, and 99% confidence intervals obtained with standard errors from probit QMLE, Newey-West with one lag, Newey-West with Andrews (1991) optimal lag, and the Hansen (1982) with one lag.

Data were generated from an unobserved indicator model  $y_t^* = \beta'x_t + \epsilon_t$  where  $y_t = \begin{cases} 1, & y_t^* \geq 0 \\ 0, & y_t^* < 0 \end{cases}$  is observed,  $x_t = \alpha \cdot x_{t-1} + v_t$ ,

and  $\epsilon_t = \frac{1}{\sqrt{1 + \gamma^2}} \eta_t + \frac{\gamma}{\sqrt{1 + \gamma^2}} \cdot \eta_{t-1}$ . The number of repetitions was 10000 and  $\beta=1$ .

**Table 3: Bias of Several Probit Standard Error Estimators - Misspecified Error Process  
(percent of estimated standard error of coefficient)**

		<u>Probit MLE</u>	<u>Newey-West</u>	<u>Newey-West (Opt)</u>	<u>Hansen</u>
<u>sample</u>	<u><math>\alpha</math></u>				
<b>100</b>	<b>.0</b>	-12.1	-14.4	-14.2	-13.2
<b>100</b>	<b>.5</b>	-22.7	-21.6	-20.8	-17.7
<b>100</b>	<b>.9</b>	-26.5	-21.3	-19.7	-14.5
<b>500</b>	<b>.0</b>	-8.8	-7.0	-6.6	-4.7
<b>500</b>	<b>.5</b>	-16.7	-11.4	-8.4	-5.8
<b>500</b>	<b>.9</b>	-26.7	-16.6	-8.4	-7.4
<b>1000</b>	<b>.0</b>	-7.3	-4.9	-4.1	-2.4
<b>1000</b>	<b>.5</b>	-17.0	-10.6	-5.7	-4.3
<b>1000</b>	<b>.9</b>	-27.1	-17.0	-7.5	-7.9

Notes: The table reports estimated biases for standard errors from probit QMLE (Probit), Newey-West with one lag (N-W), Newey-West with Andrews (1991) optimal MA(1) lag (N-W opt), and the Hansen (1982) with one lag (Hansen). Estimated biases are reported as a

percent of the estimated standard error of the simulated coefficients. Data were generated from  $y_t^* = \beta'x_t + \epsilon_t$  where  $y_t^* = \begin{cases} 1, & y_t^* \geq 0 \\ 0, & y_t^* < 0 \end{cases}$

is observed,  $\epsilon_t = \frac{1}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \eta_t + \frac{\gamma_1}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \cdot \eta_{t-1} + \frac{\gamma_2}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \cdot \eta_{t-2}$ , and the explanatory variable follows  $x_t = \alpha \cdot x_{t-1} + v_t$ . The

number of repetitions was 10000,  $\beta=1$ , and  $\gamma_1=\gamma_2=1$ .

**Table 4: Coverage Probabilities (Percent) for Several Probit Standard Error Estimators - Misspecified Error Process**

		<u>Probit MLE</u>			<u>Newey-West</u>			<u>Newey-West (Opt)</u>			<u>Hansen</u>		
Theoretical Level:		<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>	<u>99</u>	<u>95</u>	<u>90</u>
<u>sample</u>	<u><math>\alpha</math></u>												
<b>100</b>	<b>0</b>	99	94	88	98	92	86	98	92	86	98	92	86
<b>100</b>	<b>.5</b>	98	90	82	96	89	82	97	89	82	97	90	83
<b>100</b>	<b>.9</b>	96	87	79	96	89	82	97	89	83	97	91	85
<b>500</b>	<b>0</b>	98	93	87	98	93	88	98	93	88	98	94	89
<b>500</b>	<b>.5</b>	97	90	83	98	92	85	98	93	87	98	93	88
<b>500</b>	<b>.9</b>	95	85	77	97	90	83	98	93	87	98	93	87
<b>1000</b>	<b>0</b>	98	93	88	99	94	88	99	94	89	99	95	89
<b>1000</b>	<b>.5</b>	97	90	83	98	92	86	98	94	88	99	94	89
<b>1000</b>	<b>.9</b>	94	85	77	97	90	83	98	93	87	98	93	87

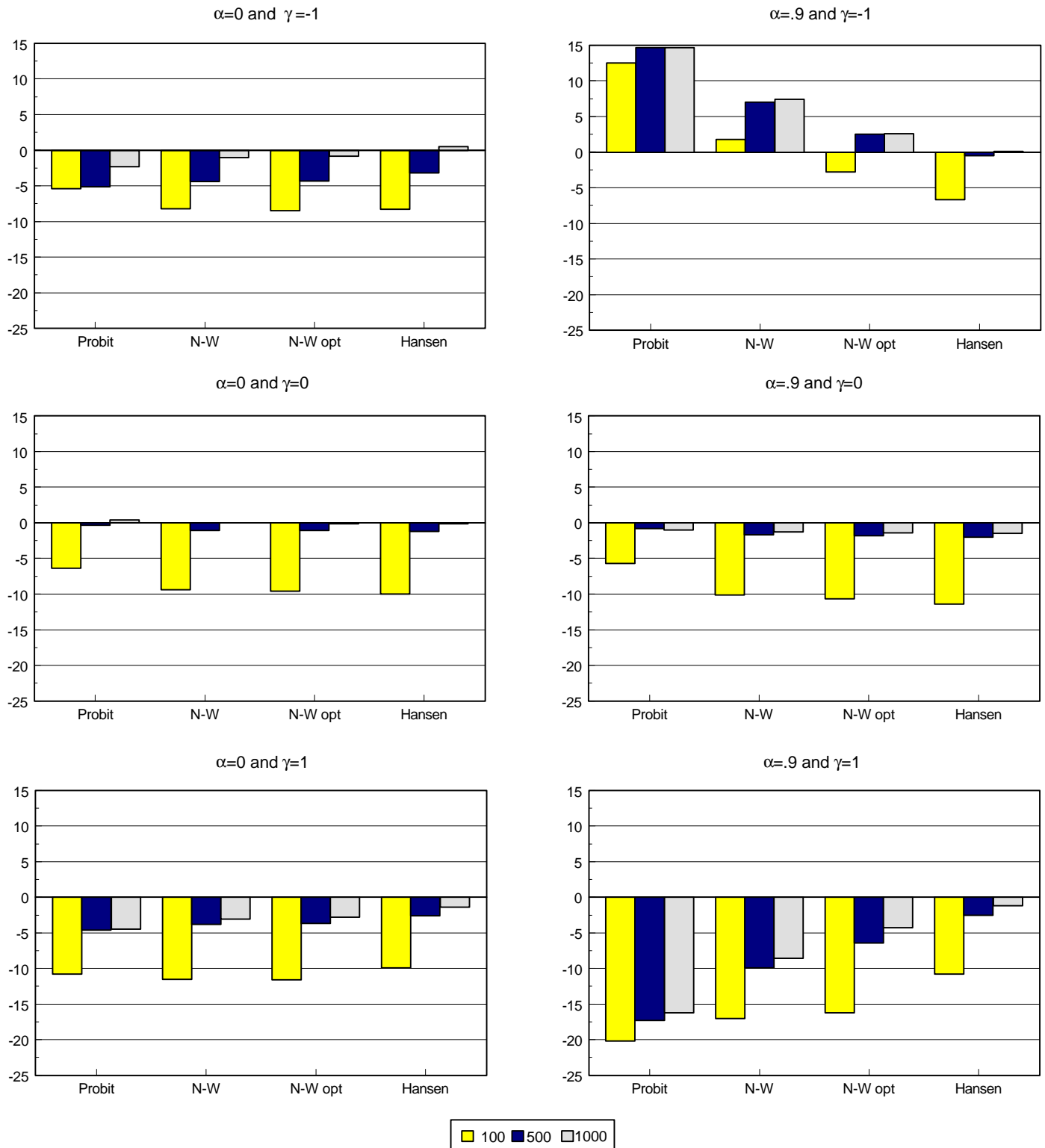
Notes: The table reports simulated coverage probabilities for 90%, 95%, and 99% confidence intervals obtained with standard errors from probit QMLE, Newey-West with one lag, Newey-West with Andrews (1991) optimal lag, and the Hansen (1982) with one lag.

Data were generated from  $y_t^* = \beta'x_t + \epsilon_t$  where  $y_t^* = \begin{cases} 1, & y_t^* \geq 0 \\ 0, & y_t^* < 0 \end{cases}$  is observed, the explanatory variable follows  $x_t = \alpha \cdot x_{t-1} + v_t$ ,

and  $\epsilon_t = \frac{1}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \eta_t + \frac{\gamma_1}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \cdot \eta_{t-1} + \frac{\gamma_2}{\sqrt{1 + \gamma_1^2 + \gamma_2^2}} \cdot \eta_{t-2}$ . The number of repetitions was 10000,  $\beta=1$ , and  $\gamma_1=\gamma_2=1$ .

Figure 1

## Bias of Probit Standard Error Estimators (as a percentage of the true standard error)



Notes: The chart displays estimated biases for standard errors from probit QMLE (Probit), Newey-West with one lag (N-W), Newey-West with Andrews (1991) optimal MA(1) lag(N-W opt), and the Hansen (1982) with one lag (Hansen). Estimated biases are reported as a percent of the estimated standard error of the simulated coefficients. Data were generated from an unobserved indicator model with one AR(1) explanatory variable with parameter  $\alpha$  and an MA(1) disturbance with parameter  $\gamma$ . The coefficient of the explanatory variable is one and the variance of the explanatory variable and disturbance are one.