The Overlapping Data Problem

by

Ardian Harri

B. Wade Brorsen

Ardian Harri is an econometrician with Toyota Motors Finance in Torrance CA. B. Wade Brorsen is a regents professor and Jean & Patsy Neustadt chair in the Department of Agricultural Economics at Oklahoma State University.
The authors would like to thank Dr. Massimiliano Marcellino for his helpful comments regarding the lagged dependent variable case.

H:\brorsen\papers\overlap.wpd 09Aug01

The Overlapping Data Problem

Abstract

We consider the overlapping data problem. The conventional estimation approach with overlapping data is to use the Newey-West estimation procedure. When the standard assumptions hold, generalized least squares (GLS) is asymptotically efficient. Monte Carlo results show that the Newey-West procedure has considerably larger variances of parameter estimates and lower power than GLS. Hypothesis tests using the Newey-West procedure also have incorrect size even with sample sizes as large as one thousand. We also discuss possible estimation approaches when overlapping data occurs in conjunction with some other econometric problem.

Key words: autocorrelation, Monte Carlo, Newey-West, overlapping data

Introduction

Time series studies estimating multiple-period changes can use overlapping data in order to achieve greater efficiency (Gilbert). A common example is using annual returns when monthly data are available. A one-year change could be calculated from January to December, another from February to January, and so on. In this example the January to December and February to January changes would overlap for eleven months. The overlapping of observations creates a moving average (MA) error term and thus ordinary least squares (OLS) parameter estimates would be inefficient and hypothesis tests biased (Hansen and Hodrick). Past literature has recognized the presence of the moving average error term. Our paper seeks to improve econometric practice when dealing with overlapping data.

One way of dealing with the overlapping observations problem is to use a reduced sample in which none of the observations overlap. For the example given above, the reduced sample will have only one observation per year. Thus, for a 30-year period of monthly data only 30 annual changes or observations will be used instead of 249 (the maximum number of overlapping observations that can be created for this period) annual observations. This procedure will eliminate the autocorrelation problem but it is obviously highly inefficient. A second way involves using average data. For our example this means using the average of the 12 overlapping observations that can be created for each year. This procedure results in the same degree of data reduction and apparently 'uses' all the information. In fact, not only is it inefficient, it also, as Gilbert shows, does not eliminate the moving average error term and can complicate estimation. A third way is to use the overlapping data and to account for the moving average error term in hypothesis testing. Several heteroskedasticity and autocovariance consistent (HAC) estimators have been constructed that can provide asymptotically valid hypothesis tests when using data with overlapping observations. These HAC estimators include Hansen and Hodrick (HH) (1982), Newey-West (NW) (1987), and Andrews and Monahan (AM) (1992). A fourth way is to use OLS estimation with overlapping data which yields biased hypothesis tests. We argue that all of these methods are sometimes inferior to other methods.

To illustrate the enormity of the problem the number of empirical articles involving the use of overlapping data in regression analysis in three journals during 1996 were counted. The journals were, *The Journal of Finance*, *The American Economic Review*, and *The Journal of Futures Markets*. The methods of estimation are classified as OLS with non-overlapping data (OLSNO), OLS with the Newey-West (1987) variance covariance estimator, OLS with any of the other GMM estimators, and just OLS. The numbers are presented in Table 1.

Table 1 shows the number of empirical articles involving the use of overlapping data as a total and as a percentage of the total number of the empirical articles in the journal for that year. Most of the empirical articles that use overlapping data involve asset returns or economic growth. A common feature of these articles is that returns or growth are measured over a period that is longer than the observation period. For example, data are observed monthly and the estimation is done annually. As a result, the estimation involves temporal aggregation. There are several possible reasons to use aggregated data. The most common reason given is measurement error in independent variables. For example, Jones and Kaul (p. 469), state that they select "use of quarterly data on all variables as a compromise between the measurement errors in monthly data ...". Another reason could be the lack of normality in the nonaggregated data. Also, when some data are missing, using overlapping data allows using all of the data. Most authors provide no justification for using overlapping data, but there must be some advantage to using it or it would not be so widely used.

Table 1 also shows each of the estimation methods frequency of use. The OLSNO and Newey-West estimation methods are used most often. We defined OLSNO as estimation using non-overlapping observations. This means that the data exist to create overlapping observations but the researchers chose to work with non-overlapping observations. It might be more correct to say that OLSNO is used simply because it is not a practice to create overlapping data. The OLSNO method will yield unbiased and consistent parameter estimates and valid hypothesis tests. But it will be inefficient since it "throws away information."

The GLS estimation procedure derived in this paper could not be applied in every situation described by Table 1 where Newey-West or OLSNO estimation is used. An example would be when lagged values of the dependent variable or some other endogenous variable are used as an explanatory variable. In this case, as Hansen and Hodrick argue, the GLS estimates will be inconsistent since an endogeneity problem is created when the dependent and explanatory variables are transformed. For the specific case of overlapping data considered by Hansen and Hodrick, we have nothing to add to the previous literature (eg. Mark) that favors using the bootstrap to correct the small sample bias in the Hansen and Hodrick approach. In other cases of overlapping data and lagged dependent variables, it can be impossible to obtain consistent estimates. The number of cases where lagged values of the dependent variable are used as an explanatory variable is reported for two of the journals mentioned earlier. In *The Journal of Finance*, from a total of 26 articles reported in Table 1, only six include a lagged dependent variable as an explanatory variable (three with the Newey-West estimator and three with

OLSNO). For the *American Economic Review* only one (with the Newey-West estimator) of 14 articles included a lagged dependent variable.

In this paper we will discuss the general overlapping data problem and argue that there are situations when Newey-West and OLSNO are grossly inefficient ways of handling the overlapping data problem since the order of the MA process is known. This will be done by determining and comparing the small-sample properties of Newey-West, OLSNO, MLE, and GLS estimates. Unrestricted maximum likelihood estimation is included as an alternative to GLS to show what happens when the MA coefficients are estimated[1]. Also, the power and size of the hypothesis tests for the four methods of estimation will be compared. Monte Carlo simulation methods are used. Finally, we discuss ways of adapting the GLS estimation procedure to handle additional econometric problems such as additional autocorrelation, missing data, and heteroskedasticity. We find that GLS is not the preferred estimator when there are errors in the variables or when lagged dependent variables are included as regressors.

I. Theory

Estimation with multiple-period changes can use data with overlapping observations in order to ensure greater efficiency of estimates. Here, we consider only strictly exogenous explanatory variables. Other variations of the overlapping data problem are considered in Section VI.

---

[1] With normality, the GLS estimator is the maximum likelihood estimator. The true MLE would have the parameters of the moving average process be known rather than estimated. Such a restricted MLE should be considered with large sample sizes since it uses less storage than GLS.

To consider the overlapping data problem we start with the following regression equation:

$$y_t = \beta' x_t + u_t \tag{1}$$

where, y_t is the dependent variable, x_t is the vector of strictly exogenous independent variables, and u_t is the error term. Equation (1) represents the basic data that are then used to form the overlapping observations. The error terms, u_t, in (1) have the following properties: $\mathrm{E}[u_t] = 0$, $\mathrm{E}[u_t^2] = \sigma_u^2$, and $\mathrm{Cov}[u_t, u_s] = 0$ if $t \neq s$.

However, one might want to use aggregated data and instead of (1) estimate the following equation:

$$Y_t = \beta' X_t + e_t \tag{2}$$

where Y_t and X_t represent an aggregation of y_t and x_t respectively. To estimate (2) the overlapping observations are created by summing the original observations as follows:

$$Y_t = \sum_{j=t}^{t+k-1} y_j,\; X_t = \sum_{j=t}^{t+k-1} x_j,\; \text{and } e_t = \sum_{j=t}^{t+k-1} u_j \tag{3}$$

where k is the number of periods for which the changes are estimated. If n is the original sample size, then $n - k + 1$ is the new sample size. These transformations of the dependent and independent variables induce an MA process in the error terms of (2).

From the assumption that the original error terms were uncorrelated with zero mean, it follows that:

$$E[e_t] = E[\sum_{j=0}^{k-1} u_{t+j}] = \sum_{j=0}^{k-1} E[u_{t+j}] = 0. \tag{4}$$

Also, since the successive values of u_j are homoskedastic and uncorrelated, the unconditional variance of e_t is:

$$\mathrm{Var}[e_t] = \sigma_e^2 = E[e_t^2] = k\sigma_u^2. \tag{5}$$

Based on the fact that two different error terms, e_t and e_{t+s}, have k - s common original error terms, u, for any k - s > 0, the covariances between the error terms are:

$$\mathrm{cov}[e_t, e_{t+s}] = E[e_t e_{t+s}] = (k-s)\sigma_u^2 \quad \forall (k-s)>0. \tag{6}$$

Dividing by $k\sigma_u^2$ gives the correlations:

$$\mathrm{corr}[e_t, e_{t+s}] = \frac{k-s}{k} \quad \forall (k-s)>0. \tag{7}$$

Collecting terms we have as an example in the case of $n = k + 2$:

$$\mathbf{\Omega} = \begin{bmatrix} 1 & \frac{k-1}{k} & \cdots & \frac{k-s}{k} & \cdots & \frac{1}{k} & 0 & 0 \\ \frac{k-1}{k} & 1 & \frac{k-1}{k} & \cdots & \frac{k-s}{k} & \cdots & \frac{1}{k} & 0 \\ \cdots & \frac{k-1}{k} & 1 & \frac{k-1}{k} & \cdots & \frac{k-s}{k} & \cdots & \frac{1}{k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{k} & \cdots & \frac{k-s}{k} & \cdots & \frac{k-1}{k} & 1 & \frac{k-1}{k} & \cdots \\ 0 & \frac{1}{k} & \cdots & \frac{k-s}{k} & \cdots & \frac{k-1}{k} & 1 & \frac{k-1}{k} \\ 0 & 0 & \frac{1}{k} & \cdots & \frac{k-s}{k} & \cdots & \frac{k-1}{k} & 1 \end{bmatrix} \tag{8}$$

where, $\mathbf{\Omega}$ is the correlation matrix. The correlation matrix, $\mathbf{\Omega}$, appears in Gilbert's paper without a derivation, but we have not found it elsewhere, although the presence of a moving average error term is commonly recognized.

With $\mathbf{\Omega}$ derived analytically the generalized least squares (GLS) parameter estimates and their variance-covariance matrix can be obtained as follows:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\mathbf{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{\Omega}^{-1}\boldsymbol{Y} \qquad \textbf{(9)}$$

and

$$\mathrm{Var}[\hat{\boldsymbol{\beta}}] = \sigma_e^2(\boldsymbol{X}'\mathbf{\Omega}^{-1}\boldsymbol{X})^{-1}. \qquad \textbf{(10)}$$

where $\boldsymbol{X} = (X_1', \ldots, X_{n-k+1}')$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_{n-k+1})$. Under these assumptions, the GLS estimator will be best linear unbiased and asymptotically efficient. If errors are normally distributed, then GLS is efficient in small samples, standard hypothesis test procedures would be valid in small samples, and the GLS estimator would be the maximum likelihood estimator.

II. Alternative Estimation Methods

The next issue to be discussed is the OLSNO and Newey-West estimation methods and their inefficiency. We consider only Newey-West rather than the alternative GMM estimators. As Davidson and MacKinnon (p. 611) say "the Newey-West estimator is never greatly inferior to that of the alternatives." First a review of Newey-West's estimation method is presented. Parameter estimates are obtained by using OLS with overlapping data as follows:

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad \textbf{(11)}$$

and the variance of b is:

$$\mathrm{Var}[\boldsymbol{b}] = \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{12}$$

The OLS estimate b is unbiased and consistent but inefficient. The OLS estimate of $\boldsymbol{\sigma_e^2}$ is biased and inconsistent. To calculate Newey-West's autocorrelation consistent covariance matrix first the OLS residuals are obtained. Then the Newey-West's autocorrelation consistent estimator is calculated using the formula:

$$\boldsymbol{S} = \boldsymbol{S_0} + \frac{1}{n-k+1}\sum_{i=1}^{k-1}\sum_{t=i+1}^{n-k+1} w_i e_t e_{t-i}(\boldsymbol{x_t x'_{t-i}} + \boldsymbol{x_{t-i} x'_t}) \tag{13}$$

where,

$$\boldsymbol{S_0} = \frac{1}{n-k+1}\sum_{i=1}^{n-k+1} e_i^2 \boldsymbol{x_i x'_i} \tag{14}$$

is the White heteroskedasticity consistent estimator, $w_i = 1 - i / k$, is a scalar weight, and $n - k + 1$ is the sample size.

Then the autocorrelation consistent covariance matrix is estimated as:

$$\boldsymbol{V} = (n-k+1)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{S}(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{15}$$

The OLSNO method of estimation obtains parameter estimates using OLS with a reduced sample where the observations do not overlap. The OLS estimates of the variance are unbiased since with no overlap there is no autocorrelation. The OLSNO parameter estimates are less efficient than the GLS estimates because of the reduced number of observations used in estimation. For the example of one-year changes the number of observations in OLSNO estimation is 12 times less than the number of observations in GLS estimation.

The Newey-West estimator uses OLS with the overlapping data to obtain the parameter estimates which results in those parameter estimates being inefficient. In addition, the Newey-West estimator of the variance-covariance matrix is only consistent and thus the GLS estimator will provide more accurate hypothesis tests in small samples. While it is known that GLS is the preferred estimator, the loss from using one of the inferior estimators in small samples is not known. We use a Monte Carlo study to provide information about the small-sample differences among the estimators.

III. Monte Carlo Study

A Monte Carlo study was conducted to determine the size and power of the hypothesis tests when using overlapping data and GLS, OLSNO, Newey-West, and unrestricted MLE, estimation methods. The Monte Carlo study also provides a measure of the efficiency lost from using OLSNO, Newey-West, and when the MA coefficients are estimated. The mean and the variance of the parameter estimates are calculated to measure bias and efficiency. Mean-squared error (MSE) is computed for each method of estimation. To determine the size of the hypothesis tests, the percentage of the rejections of the true null hypotheses are calculated. To determine the power of the hypothesis tests the percentages of the rejections of false null hypotheses are calculated.

IV. Data and Procedure

Data are generated using Monte Carlo methods. A single independent variable x with an i.i.d. uniform distribution[2] (0,1) and error terms u with a standard normal distribution are generated. We also considered a N(0,1) for x but these results are not included in the paper since the conclusions did not change. The options RANUNI and RANNOR in SAS Version 6.11 are used. The dependent variable y is calculated based on the relation represented in equation (1). For simplicity β is assumed equal to one. The data set with overlapping observations of X and Y is created by summing the x's and y's as in (3).

The regression defined in (2) was estimated using the set of data containing X and Y. The number of replications is 2000. For each of the 2000 original samples, different vectors x and u are used. This is based on Edgerton's findings that using stochastic exogenous variables in Monte Carlo studies improves considerably the precision of the estimates of power and size of the hypothesis tests. Six sample sizes T are used, respectively, 30, 100, 200, 500, 1000, and 2000. Three levels of overlapping k-1 are used, respectively, 1, 11, and 29. The levels 1 and 29 are chosen to represent two extreme levels of overlapping of practical interest. The level 11 is chosen because it corresponds to using annual changes when monthly data are available.

The OLSNO, the Newey-West, and GLS estimates of β were obtained for each of the 2000 samples using PROC IML in SAS software version 6.12. The unrestricted MLE estimates of β were obtained using PROC ARIMA in SAS. The $\boldsymbol{\Omega}$ matrix to be used in GLS estimation

---

[2] When autocorrelation in x is large and the error term follows a first-order autoregressive process, Greene (1997, p.589) finds that the inefficiency of OLS relative to GLS increases when the x's are positively autocorrelated. Since many real-world datasets have explanatory variables that are positively autocorrelated, the inefficiency of OLS found here may be conservative.

was derived in equation (8). The Newey-West estimation was validated by comparing it with the available programmed estimator in SHAZAM software Version 7.0 using the OLS ... /AUTCOV option. The power of the tests are calculated for the null hypothesis $\beta = 0$.

V. Results

The means of the parameter estimates and their standard deviations as well as the MSE values for the three overlapping levels 1, 11, and 29, for the OLSNO, Newey-West, and GLS are presented in Tables 2, 3, and 4. The true standard deviations for the GLS estimation are lower than those for the OLSNO and Newey-West estimation. This shows that the Newey-West's and OLSNO parameter estimates are less efficient than the GLS estimates. The inefficiency is greater as the degree of overlapping increases and as the sample size decreases. For a sample size of 100 and overlapping level 29, the sample variance of the GLS estimates is 0.119 while the sample variance of the Newey-West and OLSNO estimates is 2.544 and 7.969 respectively. Besides the more efficient parameter estimates, the difference between the estimated and actual standard deviations of the parameter estimates are almost negligible for the GLS estimation regardless of sample size or overlapping level. The estimated standard deviations for the OLSNO estimation show no biases as expected, but the estimated standard deviations do vary from actual standard deviations in small samples. The Newey-West estimation tends to underestimate the actual standard deviations even for overlapping level 1. The degree of underestimation increases with the increase of overlapping level and as sample size decreases. Sometimes the estimated standard deviation is only one-fourth of the true value. The Newey-West covariance estimates have previously been found to be biased downward in small samples (eg. Nelson and Kim; Goetzmann and Jorion). The parametric bootstrap suggested by Mark can lead to tests with correct size, but still uses the inefficient OLS estimator.

The inferiority of the Newey-West and OLSNO parameter estimates compared to the GLS estimates is also supported by the MSE values computed for the three methods of estimation. Thus, for the sample size 100 and the overlapping level 29, the MSE for the GLS, Newey-West, and OLSNO estimation is respectively 0.12, 2.55, and 8.02.

The means of the parameter estimates and their standard deviations as well as the MSE values for the three overlapping levels 1, 11, and 29, for the unrestricted MLE are presented in Table 5. The results are similar to the results presented for the GLS estimation. However, in small samples the actual standard deviations of the MLE estimates are larger than those of the GLS estimates. As the degree of overlapping increases the sample size, for which the standard deviations for both methods are similar, also increases (e.g. from 100 for overlapping 1 to 1000 for overlapping 29).

The Newey-West and OLSNO estimation methods also perform considerably poorer than the GLS estimation in hypothesis testing. The results of the hypothesis tests are presented in Table 6. The Newey-West estimator rejects true null hypotheses far too often. In one extreme case, it rejected a true null hypothesis 50.0% of the time instead of the expected 5%. In spite of greatly underestimating standard deviations the Newey-West estimator has considerably less power than GLS except with the smallest sample sizes considered. While the OLSNO estimation has the correct size, the power of the hypothesis tests is much less than the power of the tests with GLS.

The results of the hypothesis tests for the unrestricted MLE are presented in Table 7. While the power of the hypothesis tests is similar to the power for the GLS estimation, the size is generally larger than the size for the GLS estimation. Unrestricted MLE tends to reject the true null hypotheses more often than it should. However, this problem is reduced or eliminated as larger samples are used, i.e. 500, 1000, 2000 observations. Table 7 also presents the number of iterations for each run, as well as the number/percentage of iterations that converge. The number/percentage of iterations that converge decreases as the degree of overlap increases and sample size decreases. Given the convergence problems, as shown in Table 7, it can be concluded that, when MLE is chosen as the method of estimating (2), the MA coefficients should be restricted rather than estimated unless the sample size is quite large. On the other hand, the GLS estimator tends to run into storage problems when the sample size is around 2500 observations with the 64 MB RAM computer used here. MLE provides an alternative estimation method that does not create a storage problem.

VI. Variations on the Overlapping Data Problem

In practice, overlapping data often occur at the same time as some other econometric problems. Since the solutions are not obvious, we now discuss how the properties and estimation methods would need to change with changes in the assumptions. Also, if the explanatory variables were strictly exogenous, no observations were missing, and the errors were distributed normally as assumed so far, there would be no need to use overlapping data since the disaggregate model could be estimated.

Nonnormality. The GLS estimator does not assume normality, so estimates with GLS would remain best linear unbiased and asymptotically efficient. The hypothesis tests derived depend on normality. Hypothesis tests based on normality would still be valid asymptotically provided the assumptions of the central limit theorem hold. As the degree of overlapping increases, the residuals would approach normality, so nonnormality would be less of a concern. The Newey-West estimator is also only asymptotically valid. The GLS transformation of the residuals might also speed the rate of convergence toward normality since it is "averaging" across more observations than the OLS estimator used with Newey-West.

We estimated (2) with two correlated x's and with the error term u following a t-distribution with four degrees of freedom. Results are reported in Table 8. The main difference with the previous results is the increased standard deviations for all methods of estimation. Proportionally, the increase in standard deviations is slightly larger for Newey-West and OLSNO. Thus, the Monte Carlo results support our hypothesis that the advantages of GLS would be even greater in the presence of nonnormality. This can also be seen from the hypothesis test results presented in Table 8. The power of the three methods of estimation is reduced with the biggest reduction occurring for the Newey-West and OLSNO. Finally, the increase of the standard deviations and the resulting reduction in power of hypothesis tests, is larger when the correlation between the two x's increases. This is true for the three methods of estimation.

Missing observations. Missing observations can be a reason to use overlapping data. It is not unusual in studies of economic growth to have key variables observed only every five or ten years at the start of the observation period, but every year in more recent years. Using overlapping data allows using all of the data.

When some observations are missing, one can derive the correlation matrix in (8) as if all observations were available and then delete the respective rows and columns for the missing overlapping observations. The Newey-West estimator assumes autocovariance stationarity and so available software packages that include the Newey-West estimator would not correctly handle missing observations. It should, however, be possible to modify the Newey-West estimator to handle missing observations.

Varying levels of overlap. It is not uncommon in studies of hedging to consider different hedging horizons which leads to varying levels of overlap (i.e. k is not constant). This introduces heteroskedasticity of known form in addition to the autocorrelation. In this case it is easier to work with the covariance matrix than the correlation matrix. The covariance matrix is σ_u^2 times a matrix that has the number of time periods (the value of k_t) used in computing that observation down the diagonal. The off diagonal terms would then be the number of time periods for which the two observations overlap. Allowing for the most general case of different overlap between every two consecutive observations, the unconditional variance of e_t (given in (5)) now is:

$$\mathrm{Var}[e_t] = \sigma_e^2 = E[e_t^2] = k_t\sigma_u^2. \qquad \textbf{(16)}$$

Previously, two different error terms, e_t and e_{t+s}, had k - s common original error terms, u, for any k - $s > 0$. Now, they may have less than k - s common u's and there no longer is a monotonic decreasing pattern of the number of the common u's as e_t and e_{t+s} get further apart. We let k_{ts}

represent the number of common u's (overlapping periods) between e_t and e_{t+s}. Therefore, the covariances between the error terms e_t and e_{t+s}, are:

$$\operatorname{cov}[e_t, e_{t+s}] = E[e_t e_{t+s}] = (k_{ts})\sigma_u^2 \qquad \textbf{(17)}$$

The example covariance matrix with $n = s + 2$ is then:

$$\boldsymbol{\Sigma} = \sigma_u^2 \begin{bmatrix} k_1 & k_{12} & k_{13} & \dots & k_{1s} & 0 & 0 \\ k_{21} & k_2 & k_{23} & \dots & \dots & k_{2s} & 0 \\ \dots & k_{32} & k_3 & k_{34} & \dots & \dots & k_{3s} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & k_{ts} & \dots & k_{t(t-2)} & k_{t(t-1)} & k_t \end{bmatrix} \qquad \textbf{(18)}$$

where, $k_{ts} = k_{st}$. The standard Newey-West procedure does not handle varying levels of overlap since it assumes autocovariance stationarity.

Lagged dependent variables. The case of overlapping data and a lagged dependent variable (or some other variable that is not strictly exogenous) was a primary motivation for Hansen and Hodrick's estimator. In the usual case of autocorrelation and a lagged dependent variable, ordinary least squares estimators are inconsistent. Hansen and Hodrick consider the case where aggregate data are used for the dependent variable, the lagged variables are disaggregate, and the lag length is longer than the length of the overlap. OLS is consistent in this case and

Newey-West can be used to calculate standard errors in large samples. In small samples, the bootstrap[3] should be used to conduct hypothesis tests.

Engle shows, for the case where the first lag of the dependent variable is used as an explanatory variable, that the use of OLS with aggregated data could lead to biases of either sign and almost any magnitude. Generalized least squares estimates are also inconsistent, but consistent estimates can be obtained using the maximum likelihood methods developed for time-series models. As Marcellino (1996, 1999) has shown, when nonoverlapping data are used, estimates of the parameters of the disaggregated process can no longer be recovered. With nonoverlapping data, the time-series process can be quite different than the original process.

Marcellino (1996, 1999) discusses in detail the issues related to temporal aggregation of time-series models. Following his notation, (except that x and y are switched) let

$$g(L)y_t = f(L)x_t + s(L)u_t \qquad t = 1,2,...,T \tag{19}$$

represent a general autoregressive disaggregated model where L is the lag operator, $g(L)$, $f(L)$, and $s(L)$ are polynomials of orders g, f, and s in the lag operator, x_t is strictly exogenous, and u_t is a white noise (WN) process, $u_t \sim WN(0, \sigma_u)$. The overlapping observations are obtained using the following relation:

$$(1 + L + ... + L^{k-1})g(L)y_t = (1 + L + ... + L^{k-1})f(L)x_t + (1 + L + ... + L^{k-1})s(L)u_t \tag{20}$$

[3] Mark suggests bootstrapping the Newey-West t-statistic. However, recent research (eg. Kazimi and Brownstone; Coulibaly and Brorsen) suggests the asymptotically pivotal property is not as important in conducting bootstrap hypothesis tests as was once thought. So, it might work just as well to bootstrap the OLS parameter estimates directly.

or

$$G(L)\,Y_t \;=\; F(L)X_t \;+\; S(L)\,e_t \qquad t \;=\; k, k+1, \ldots, T \qquad \textbf{(21)}$$

where k is the order of the summation, and Y_t and X_t are the overlapping observations. Our previous results in (9) and (10) can be derived as a special case of (21). In most instances, $G(L)$=$g(L)$. If $s(L)$=1, then the covariance matrix is the same as in (5) and (6). While GLS estimates would not be consistent, consistent estimates can be obtained with the maximum likelihood methods developed for time-series models. When $s(L)$=1 the MA coefficients would be known and asymptotically efficient estimates would require restricting the MA coefficients.

Marcellino refers to the process of creating overlapping data as the first step of average sampling. The second step, that is often applied by past literature, is what Marcellino calls point-in-time sampling of order k to the overlapping data. In a point-in-time sampling process only the k^{th} Y_t and X_t observations of the process in (20) and (21), for our example, are retained:

$$G^*(B)\,Y_\tau \;=\; F^*(B)X_\tau \;+\; S^*(B)\,e_\tau \qquad \textbf{(22)}$$

where $Y_\tau = Y_{tk}$, and $B \;=\; L^{\tau}$. Our nonoverlapping observations are average sampling of the disaggregated process in (1). Marcellino derives the upper bounds of the autoregressive (AR), g, and moving average (MA), s, order for the aggregated process obtained by point-in-time or average sampling. Marcellino (1996) shows that “there is an aggregated MA component even with an original pure AR process” (p.13)[4]. Thus, if the autocorrelation in the error term in (21) is ignored in the estimation, as is usually done with OLSNO or Newey-West, parameters are

[4] See also Brewer (1973), Wei (1981), and Weiss (1984).

estimated inconsistently. To illustrate and confirm the theoretical results, an example is now provided.

Consider the disaggregated model given below:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + u_t, \qquad u_t \sim N(0,1) \tag{23}$$

where for simplicity $\alpha_0 = 0$ and $\alpha_2 = 1$. The value selected for α_1 is 0.5. For $k = 3$, the model usually estimated[5] is:

$$Y_t = \beta_0 + \beta_1 Y_{t-3} + \beta_2 X_t + \varepsilon_t \tag{24}$$

where $Y_t = y_t + y_{t-1} + y_{t-2}$, and $X_t = x_t + x_{t-1} + x_{t-2}$. As we will show, the error term in this model is an MA(1) and additional lags of X should be included. To get the overlapping observations apply (20) to (23) to get:

$$(1 + L + L^2)(1 - 0.5L)y_t = (1 + L + L^2)x_t + (1 + L + L^2)u_t \tag{25}$$

where $g(L) = (1 - 0.5L)$, $f(L) = 1$, and $s(L) = 1$, and therefore the model analogous to our previous model in (2) is

$$Y_t = 0.5\, Y_{t-1} + X_t + e_t. \tag{26}$$

The model in (26) also has the same variance-covariance matrix, described by (5) and (6), as our previous model in (2).

[5] The model considered by Hansen and Hodrick is $Y_t = \beta + \beta_1 y_{t-3} + \beta_2 x_{t-3}$ where the independent variables are disaggreate.

To obtain (24) we can start from (26), substitute for Y_{t-1} and then for Y_{t-2} to get:

$$Y_t = 0.5^3 Y_{t-3} + X_t + 0.5 X_{t-1} + 0.5^2 X_{t-2} + \varepsilon_t. \qquad \textbf{(27)}$$

The error term ε_t in (24) is a MA process of order four of the error term u_t in (23) with coefficients 1.5, 1.75, 0.75 and 0.25, $\varepsilon_t = u_t + 1.5u_{t-1} + 1.75u_{t-2} + 0.75u_{t-3} + 0.25u_{t-4}$. The MA process for ε_t can be derived simply by substituting for the original error term u_t or by following the procedure discussed by Marcellino (1996, 1999). Following Marcellino's procedure, the MA process, $N(L)$, for ε_t, can be derived using the following relation $N(L) = C(L) * S(L)$, where, for our example, $C(L) = (1 + 0.5\,L + 0.25\,L^2)$ and $S(L) = (1 + L + L^2)$.

If only the k^{th} Y_t and X_t observations are observed in practice (average sampling) then, X_{t-1}, and X_{t-2} are not observable. In this case, an analytical solution of (27) cannot be derived. To be consistent with our previous result, X is strictly exogenous and not autocorrelated. Based on the temporal aggregation literature (Brewer (1973), p.141, Weiss (1984), p. 272, and Marcellino (1996) p. 32), no analytical solution is possible unless x_t is generated by some autocorrelated process and the unobserved terms can be derived from the observed terms. However, based on the fact that the AR coefficient is the same whether point-in-time or average sampling is used, we know then that the AR coefficient is 0.125. The number of lags for X and the order of the MA process cannot be derived analytically. Therefore, we used Box-Jenkins methods to identify which lags to include in the model. We estimated the following models:

$$Y_\tau = 0.5^3 Y_{\tau-3} + \beta_2 X_\tau + \beta_3 X_{\tau-3} + v_\tau \qquad \textbf{(28)}$$

and

$$Y_\tau = 0.5^3 Y_{\tau-3} + X_t + 0.5X_{t-1} + 0.5^2 X_{t-2} + \eta_\tau. \qquad \textbf{(29)}$$

The model in (29) is sound theoretically in the sense that the unobserved lags for X are X_{t-1} and X_{t-2} and thus it makes sense to include them in the model. However, the model in (29) may not be feasible in practice. It uses nonoverlapping data for the Y, but it requires overlapping data on the X which may not always be available.

To confirm our analytic findings we estimated the models in (26), (27), (28), and (29) with MLE using PROC ARIMA in SAS software version 6.12 using a large Monte Carlo sample of 500,000 observations. The results are reported in Table 8. The empirical estimates of the AR and MA coefficients and the coefficients of the Xs for the models in (26) and (27) fully support our analytic findings. One potential problem with the model in (27) is the noise introduced by aggregation. The variable X_{t-1} includes x_{t-1}, x_{t-2}, and x_{t-3} , and X_{t-2} includes x_{t-2}, x_{t-3}, and x_{t-4}, while only x_{t-1} and x_{t-2} are relevant. This errors-in-variables problem biases parameter estimates toward zero. The noise introduced and the associated bias would be greater as the degree of overlap increases.

We estimated (28) with MLE and nonoverlapping data, while (29) is estimated using both overlapping and nonoverlapping data. Both models result in an ARMA(1,1) process with the AR coefficient 0.118 for (28) and 0.123 for (29) which are close to the analytical value of 0.125. The MA coefficient is the same for both models, 0.163 which provides support to the choice of these models. Higher lags of X for the model in (28) were not significant.

We also estimated (28) with Newey-West and OLSNO. The lagged value of the X is not included in the estimation in order to be consistent with the models usually estimated in the empirical literature. These models are the same as the model presented in (24). The parameter estimates were identical for both methods. The parameter estimates are 0.278 for the coefficient on $Y_{\tau-3}$, and 1.415 for the coefficient on X_{τ}. The parameter estimate for $Y_{\tau-3}$ is biased upwards for two reasons. First, $Y_{\tau-3}$ is correlated with the missing explanatory variable $X_{\tau-3}$. Also, the coefficient of $Y_{\tau-3}$ is capturing part of the effect of the missing MA term. Thus, our empirical estimates confirm the inconsistency of Newey-West and OLSNO.

With overlapping data and a lagged dependent value as an explanatory variable the only consistent estimation method is maximum likelihood with (26). Unlike GLS, maximum likelihood provides consistent estimates when the explanatory variables are predetermined whether or not they are strictly exogenous. Also, the model in (26) has the familiar ARMA process, with the AR order the same as the AR order of the disaggregated model (in our case (23)) and MA order *k*-1.

Additional source of autocorrelation. In practice there may be sources of autocorrelation in addition to that caused by the overlapping data problem. Mathematically, this would imply that u_t in (1) is autocorrelated. If the disaggregated process is an MA process, then the procedure developed in the lagged dependent variable section can be applied straight forward. If the error term in (1) follows an ARMA process then the same procedure can be applied with slight modification. Assume that u_t in (1) follows the process:

$$m(L)u_t = h(L)\xi_t \qquad \textbf{(30)}$$

where ξ_t is a white noise (WN) process, $\xi_t \sim WN(0,\sigma_\xi)$. Aggregation of (1) to obtain the overlapping observations

$$(1 + L + ... + L^{k-1}) y_t \; = \; (1 + L + ... + L^{k-1}) x_t \; + \; (1 + L + ... + L^{k-1}) u_t \qquad \textbf{(31)}$$

introduces the same level k of aggregation to (30), which now becomes:

$$(1 + L + ... + L^{k-1}) m(L) u_t \; = \; (1 + L + ... + L^{k-1}) h(L) \xi_t \qquad \textbf{(32)}$$

or

$$M(L) e_t \; = \; H(L) E_t . \qquad \textbf{(33)}$$

Then, the procedures discussed in the lagged dependent variable case can be applied with respect to (30) to obtain the order and the values of the AR and MA coefficients in (33) to be used in estimating (2). In this case, maximum likelihood methods for estimating a regression with ARMA errors can be used.

Heteroskedasticity. If the residuals in the disaggregated data (u_t in (1)) are heteroskedastic, then estimation is more difficult. Define σ^2_{ut} as the time-varying variance of u_t and σ^2_{et} as the time-varying variance of e_t . Assume the u_t's are independent and thus $\sigma^2_{et} = \sum_{j=0}^{k-1} \sigma^2_{ut-j}$. For simplicity, assume that σ^2_{ut} depends only on x_t. If σ^2_{ut} is assumed to be a linear function of x_t $(\sigma^2_{ut} = \gamma' x_t)$ then the function aggregates nicely so that $\sigma^2_{et} = \sum_{j=0}^{k-1} \gamma' x_{t-j} = \gamma' X_t$. But, if multiplicative heteroskedasticity is assumed $(\sigma^2_{ut} = \exp(\gamma' x_t))$ then $\sigma^2_{et} = \sum_{k-1}^{j=0} \exp(\gamma' x_{t-j})$ and there is no way to consistently estimate γ using only aggregate data (nonoverlapping data also have the same problem).

The covariance between e_t and e_{t+s} for any $k - s \geq 0$ would be

$$\text{Cov}\,(e_t, e_{t+s}) = \sum_{j=s}^{k-1} \sigma^2_{u(t-j)}. \quad \textbf{(34)}$$

Since the correlation matrix, $\boldsymbol{\Omega}$ is known, as given by (8), the covariance matrix can be derived using the relation:

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Omega}\,\boldsymbol{\Gamma} \quad \textbf{(35)}$$

where $\boldsymbol{\Gamma} = \left[\gamma' X_1,\ \gamma' X_2,\ \ldots,\ \gamma' X_T\right] \times I_T$. A feasible generalized least squares estimator can then be developed using (16). It might be reasonable to use (9) as the first stage in a FGLS estimation that corrected for heteroskedasticity.

Errors in variables. The most common reason authors give for using overlapping data is a problem with errors in the explanatory variables. Errors in the explanatory variables causes parameter estimates to be biased toward zero, even asymptotically. Using overlapping data reduces this problem, but the problem is only totally removed as the level of overlap, k, approaches infinity.

We added to the x in (1) a measurement error, ω, that is distributed normally with the same variance as the variance of x, $\omega \sim N(0, 1/12)$. We then conducted the Monte Carlo study with x not being autocorrelated and also with x being autocorrelated with an autoregressive coefficient of 0.8. In addition to estimating (2) with GLS, Newey-West, and OLSNO, we also estimated (1) using the disaggregate data. The results are reported in Table 10. The estimation was performed only for two sample sizes, respectively 100 and 1000 observations. In the case when x is not autocorrelated, there is no gain in using overlapping observations, in terms of reducing the measurement error. This is true for all methods of estimation.

In the case when x is autocorrelated, the largest reduction in measurement error occurs when Newey-West and OLSNO are used. Moreover, the bias is always larger for GLS estimates compared to Newey-West and OLSNO estimates. The reduction in the measurement error because of using overlapping observations is confirmed by comparing the Newey-West and OLSNO estimates to the disaggregate estimates. The GLS transformation of the variables does not reduce further the measurement error. Instead it almost totally offsets the error reduction effect of the aggregation process that creates the overlapping observations. This can be seen from the results of Table 10 where the GLS estimates are just barely less biased than the disaggregate estimates. Therefore, the GLS estimation is not an appropriate estimation method if the reason for using overlapping data is errors in the variables. Newey-West standard errors are still biased, so the preferred estimation method in the presence of large errors in the variables would be OLS with overlapping data and with standard errors calculated using Monte Carlo methods.

Imperfect overlap. Sometimes observations overlap, but they do not overlap in the perfect way assumed here and so the correlation matrix is no longer known. An example would be where the dependent variable represents six months returns on futures contracts. Assume that there are four different contracts in a year, the March, June, September, and December contracts. Then, the six-month returns for every two consecutive contracts would overlap while, the six-months returns between say March and September contracts would not overlap. Two six-month returns for, say the March contract, that overlap for three months would be perfectly correlated for these three months. The six-month returns for the March and June contracts would overlap for three months, but they would not be perfectly correlated during these three months, since the March and June contract are two different contracts. Let

$$Cov(u_{jt}, u_{st+m}) = \begin{cases} \sigma_{js} \text{ if } m = 0 \\ 0 \text{ otherwise} \end{cases} \quad \textbf{(36)}$$

be the covariance between the monthly returns *m* months (or days if disaggregated data are daily data) apart for the March and June contracts where u_{jt} and u_{st} are the error term from regression models with disaggregate data for the March and June contract. Then,

$$Var(u_{jt}) = Var(u_{st}) = \sigma_u^2, \quad Var(e_{jt}) = Var(e_{st}) = k\sigma_u^2 \quad \textbf{(37)}$$

and

$$Cov(e_{jt}, e_{st-m}) = k_{js}\sigma_{js} \quad \textbf{(38)}$$

where k_{js} is the number of overlapping months between the March and June contracts and $\sigma_{js} = \rho_i \sigma_u^2$ where ρ_i (i = 1, 2) is the correlation between the u's for two consecutive contracts with maturities three (ρ_1) and six (ρ_2) months apart. The covariance matrix for (2) with n = 12, in this case is:

$$
\boldsymbol{\Sigma} = \sigma_u^2 \begin{bmatrix}
k & \frac{k-1}{k} & \frac{k-2}{k}\rho_1 & \frac{k-3}{k}\rho_1 & \frac{k-4}{k}\rho_2 & \frac{k-5}{k}\rho_2 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{k-1}{k} & k & \frac{k-1}{k} & \frac{k-2}{k}\rho_1 & \frac{k-3}{k}\rho_1 & \frac{k-4}{k}\rho_2 & \frac{k-5}{k}\rho_2 & 0 & 0 & 0 & 0 & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\frac{k-3}{k}\rho_1 & \frac{k-2}{k}\rho_1 & \frac{k-1}{k} & k & \frac{k-1}{k} & \frac{k-2}{k}\rho_1 & \frac{k-3}{k}\rho_1 & \frac{k-4}{k}\rho_2 & \frac{k-5}{k}\rho_2 & 0 & 0 & 0 \\
\frac{k-4}{k}\rho_2 & \frac{k-3}{k}\rho_1 & \frac{k-2}{k}\rho_1 & \frac{k-1}{k} & k & \frac{k-1}{k} & \frac{k-2}{k}\rho_1 & \frac{k-3}{k}\rho_1 & \frac{k-4}{k}\rho_2 & \frac{k-5}{k}\rho_2 & 0 & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{k-5}{k}\rho_2 & \frac{k-4}{k}\rho_2 & \frac{k-3}{k}\rho_1 & \frac{k-2}{k}\rho_1 & \frac{k-1}{k} & k
\end{bmatrix} \quad \textbf{(39)}
$$

Nonparametric methods. Sometimes authors want to use nonparametric methods that assume independence. In this case the only general solutions we can propose are to use nonoverlapping data, switch to a parametric method, or use Monte Carlo hypothesis testing procedures such as bootstrapping.

VII. Conclusions

We have explored using the GLS estimator when working with overlapping data. When explanatory variables are strictly exogenous the GLS estimator is superior to the commonly used estimators. The alternative estimators that were compared with the GLS estimator were, the Newey-West estimator and ordinary least squares with nonoverlapping data (OLSNO) and unrestricted MLE. Unrestricted MLE tends to reject the true null hypotheses more often than it should. However, this problem is reduced or eliminated as larger samples are used, i.e. at least 1000 observations. GLS can exhaust computer memory limits when the sample size is large. MLE can be used instead since it requires less memory. In some computer packages, restricted MLE may be easier to implement than GLS.

There is a gain in the efficiency of the parameter estimates when the GLS estimator is used instead of the other two estimators. The gain in efficiency increased with the level of overlap. With overlapping of 29 time periods, the MSE of Newey-West was roughly 20 times larger than the MSE of GLS. The MSE of the OLSNO estimator was even larger.

The Newey-West estimator rejected true null hypotheses too often. This problem persisted even with sample sizes of 1,000. The power of the Newey-West hypotheses tests also was much slower to converge to one than the power of the GLS estimator. While hypothesis tests with the OLSNO estimator had the correct size, they had considerably lower power than either of the other two estimators.

We evaluated ways of obtaining parameter estimates when our original assumptions are relaxed. Several of these are especially important since they provide the motivation for using overlapping data in the first place. Others are important because they are commonly faced in empirical work. If the motivation for using overlapping data is missing observations or nonnormality then GLS is still the preferred estimator. When lagged dependent variables are used as explanatory variables, GLS is inconsistent. The usual Newey-West and OLSNO estimators are consistent when disaggregate data are used as regressors. When aggregate data are used as regressors, consistent parameter estimates can sometimes be obtained with maximum likelihood. In other cases, aggregation makes it impossible to recover the parameters of the disaggregate model. When the reason for using overlapping data is to reduce bias due to errors in the variables, GLS is nearly as biased as the disaggregate model. We suggest using OLS to estimate parameters and using Monte Carlo methods to calculate standard errors when there are errors in the variables.

Overlapping data are often used in finance and in studies of economic growth. Many of the commonly used estimators are either inefficient or yield biased hypothesis tests. The appropriate estimator to use with overlapping data depends on the situation.

Table 1. Number of Articles Using Overlapping Data, 1996.

Journal	Number of articles					Total number of empirical articles in the journal	Percentage of articles with overlapping data
	OLSNO	N-W	Other[a]	OLS	Total		
J. Finance	16	8	8	-	26	55	47.3
Amer. Econ. Rev.	10	3	2	-	14	77	18.2
J. Fut. Mkts.	12	3	5	2	19	43	44.2

Note: The sum of the columns 2 through 5 may be larger than the total in column 6 since some articles use more than one method of estimation.
[a] These include HH and AM estimators.

Table 2. Parameter Estimates, Standard Deviations, and MSE for OLSNO, Newey-West, and GLS Estimation (Overlapping 1).

	GLS Estimation			Newey-West Estimation			Non-overlapping Estimation		
Sample Size	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE
30	0.981	0.639 [a] 0.663 [b]	0.440	0.971	0.631 [a] 0.808 [b]	0.654	0.970	0.893 [a] 0.930 [b]	0.865
100	1.005	0.348 [a] 0.345 [b]	0.119	0.996	0.374 [a] 0.423 [b]	0.179	0.997	0.490 [a] 0.497 [b]	0.247
200	0.993	0.246 [a] 0.244 [b]	0.060	0.993	0.269 [a] 0.303 [b]	0.092	0.989	0.346 [a] 0.345 [b]	0.119
500	1.001	0.155 [a] 0.154 [b]	0.024	1.003	0.172 [a] 0.189 [b]	0.036	1.001	0.219 [a] 0.218 [b]	0.048
1000	1.001	0.110 [a] 0.109 [b]	0.012	0.997	0.122 [a] 0.134 [b]	0.018	1.005	0.155 [a] 0.156 [b]	0.024
2000	1.002	0.077 [a] 0.082 [b]	0.007	0.998	0.086 [a] 0.098 [b]	0.010	1.002	0.110 [a] 0.116 [b]	0.014

Note: The sample sizes are the sizes for samples with overlapping observations.
[a] These are the estimated standard deviations of the parameter estimates.
[b] These are the actual standard deviations of the parameter estimates.

Note: The model estimated is $Y_t = \beta' X_t + e_t$ where Y_t and X_t represent some aggregation of the original disaggregated variables. For simplicity β is chosen equal to 1. The model is estimated using Monte Carlo methods involving 2000 replications. The errors for the original process are generated from a standard normal distribution and are homoskedastic and not autocorrelated. As a result of the aggregation, e_t follows an MA process with the degree of the process depending on the aggregation level applied to y and x.

Table 3. Parameter Estimates, Standard Deviations, and MSE for OLSNO, Newey-West, and GLS Estimation (Overlapping 11).

	GLS Estimation			Newey-West Estimation			Non-overlapping Estimation		
Sample Size	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE
30	1.001	0.647 [a] 0.647 [b]	0.418	1.032	0.665 [a] 1.878 [b]	3.527	1.220	2.940 [a] 4.601 [b]	21.216
100	0.998	0.348 [a] 0.359 [b]	0.129	1.003	0.651 [a] 1.047 [b]	1.096	1.008	1.256 [a] 1.308 [b]	1.711
200	0.994	0.245 [a] 0.236 [b]	0.056	0.989	0.527 [a] 0.698 [b]	0.487	0.993	0.871 [a] 0.895 [b]	0.802
500	1.005	0.155 [a] 0.155 [b]	0.024	1.005	0.363 [a] 0.455 [b]	0.207	1.026	0.540 [a] 0.542 [b]	0.294
1000	0.997	0.110 [a] 0.112 [b]	0.013	1.004	0.262 [a] 0.315 [b]	0.099	1.002	0.382 [a] 0.390 [b]	0.152
2000	0.995	0.078 [a] 0.077 [b]	0.006	0.999	0.189 [a] 0.223 [b]	0.050	0.999	0.270 [a] 0.272 [b]	0.074

Note: The sample sizes are the sizes for samples with overlapping observations.
[a] These are the estimated standard deviations of the parameter estimates.
[b] These are the actual standard deviations of the parameter estimates.

Table 4. Parameter Estimates, Standard Deviations, and MSE for OLSNO, Newey-West, and GLS Estimation (Overlapping 29).

	GLS Estimation			Newey-West Estimation			Non-overlapping Estimation		
Sample Size	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE
30	0.996	0.648 [a] 0.668 [b]	0.446	0.996	0.539 [a] 2.204 [b]	4.858	-- [c]	-- [c] -- [c]	-- [c]
100	1.005	0.349 [a] 0.345 [b]	0.119	1.077	0.711 [a] 1.595 [b]	2.551	1.233	2.228 [a] 2.823 [b]	8.023
200	0.996	0.245 [a] 0.248 [b]	0.062	1.016	0.694 [a] 1.216 [b]	1.478	0.988	1.467 [a] 1.571 [b]	2.469
500	1.005	0.155 [a] 0.158 [b]	0.025	1.029	0.523 [a] 0.726 [b]	0.528	1.025	0.867 [a] 0.893 [b]	0.798
1000	1.004	0.110 [a] 0.110 [b]	0.012	1.011	0.394 [a] 0.496 [b]	0.246	1.010	0.605 [a] 0.611 [b]	0.374
2000	1.002	0.077 [a] 0.078 [b]	0.006	1.002	0.290 [a] 0.343 [b]	0.118	1.004	0.427 [a] 0.425 [b]	0.181

Note: The sample sizes are the sizes for samples with overlapping observations.
[a] These are the estimated standard deviations of the parameter estimates.
[b] These are the actual standard deviations of the parameter estimates.
[c] These values cannot be estimated because of the very small number of observations.

Table 5. Parameter Estimates, Standard Deviations, and MSE for the Maximum Likelihood Estimates Assuming the MA Coefficients are Unknown for Three Levels of Overlapping (1, 11, and 29).

	Overlapping 1			Overlapping 11			Overlapping 29		
Sample Size	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE
30	0.975	0.622 [a] 0.624 [b]	0.391	1.019	0.541 [a] 0.833 [b]	0.694	- [c]	- [c] - [c]	- [c]
100	1.010	0.343 [a] 0.347 [b]	0.120	0.998	0.311 [a] 0.374 [b]	0.140	0.991	0.281 [a] 0.455 [b]	0.207
200	0.989	0.243 [a] 0.247 [b]	0.061	0.995	0.230 [a] 0.256 [b]	0.065	0.984	0.216 [a] 0.278 [b]	0.078
500	0.990	0.154 [a] 0.156 [b]	0.025	0.990	0.149 [a] 0.158 [b]	0.025	0.986	0.145 [a] 0.165 [b]	0.027
1000	0.991	0.112 [a] 0.109 [b]	0.013	0.991	0.107 [a] 0.112 [b]	0.013	0.990	0.105 [a] 0.112 [b]	0.013
2000	0.995	0.078 [a] 0.077 [b]	0.006	0.995	0.076 [a] 0.078 [b]	0.006	0.995	0.075 [a] 0.080 [b]	0.006

Note: The sample sizes are the sizes for samples with overlapping observations.
[a] These are the estimated standard deviations of the parameter estimates.
[b] These are the actual standard deviations of the parameter estimates.
[c] These values cannot be estimated because of the very small number of observations.

Table 6. Power and Size Values of the Hypothesis Tests for OLSNO, Newey-West, and GLS Estimation (Overlapping 1, 11, 29).

Degree of Overlapping	Sample Size	GLS Estimation		Newey-West Estimation		Non-overlapping Estimation	
		Power	Size	Power	Size	Power	Size
1	30	0.319	0.052	0.366	0.135	0.181	0.044
	100	1	0.043	0.500	0.090	0.500	0.052
	200	1	0.042	1	0.081	1	0.049
	500	1	0.053	1	0.078	1	0.052
	1000	1	0.049	1	0.075	1	0.056
	2000	1	0.058	1	0.089	1	0.072
11	30	0.315	0.044	0.500	0.492	0.045	0.044
	100	1	0.056	0.434	0.254	0.111	0.046
	200	1	0.039	0.486	0.169	0.194	0.045
	500	1	0.048	0.500	0.124	0.455	0.050
	1000	1	0.053	1	0.104	0.500	0.051
	2000	1	0.046	0.997	0.094	0.958	0.049
29	30	0.340	0.049	0.500	0.500	-- [a]	-- [a]
	100	1	0.044	0.500	0.417	0.070	0.056
	200	1	0.055	0.449	0.291	0.070	0.046
	500	1	0.061	0.500	0.176	0.203	0.044
	1000	1	0.050	0.500	0.132	0.364	0.055
	2000	1	0.059	0.885	0.113	0.646	0.051

Note: The sample sizes are the sizes for samples with overlapping observations.
[a] These values cannot be estimated because of the very small number of observations.

Table 7. Power and Size Values of the Hypothesis Tests for the Maximum Likelihood Estimates Assuming the MA Coefficients are Unknown for Three Levels of Overlap (1, 11, and 29).

Degree of Overlap	Sample Size	Total Number of Iterations	Iterations that Converge		Power [b]	Size [b]
			Number	Percentage		
1	30	1000	999	99.9	0.331	0.070
	100	1000	1000	100	0.827	0.047
	200	1000	1000	100	0.982	0.058
	500	1000	1000	100	1.000	0.060
	1000	1000	1000	100	1.000	0.062
	2000	1000	1000	100	1.000	0.051
11	30	1400	994	71.0	0.476	0.252
	100	1000	995	99.5	0.884	0.109
	200	1000	1000	100	0.980	0.085
	500	1000	998	99.8	0.998	0.075
	1000	1000	1000	100	1.000	0.069
	2000	1000	1000	100	1.000	0.056
29	30	-- [a]	-- [a]	-- [a]	-- [a]	-- [a]
	100	1600	970	60.6	0.814	0.254
	200	1200	1027	85.6	0.980	0.135
	500	1200	1082	90.2	1.000	0.081
	1000	1100	1066	96.9	1.000	0.078
	2000	1000	932	93.2	1.000	0.060

Note: The sample sizes are the sizes for samples with overlapping observations.
[a] These values cannot be estimated because of the very small number of observations.
[b] These are calculated based on the number of replications that converged.

Table 8. Parameter Estimates, Standard Deviations, MSE, and Power and Size of Hypothesis Tests for OLSNO, Newey-West, and GLS Estimation with Two Xs and Nonnormal Errors(Overlapping 1, 11, and 29).

Degree of Overlap	Sample Size	GLS Estimation					Newey-West Estimation					Non-overlapping Estimation				
		Parameter Estimates	Standard Deviations	MSE	Power	Size	Parameter Estimates	Standard Deviations	MSE	Power	Size	Parameter Estimates	Standard Deviations	MSE	Power	Size
1	30	1.014	0.953 [a] 1.003 [b]	1.007	0.208	0.046	0.997	0.898 [a] 1.267 [b]	1.606	0.288	0.152	1.049	1.334 [a] 1.794 [b]	3.220	0.201	0.128
	100	0.969	0.498 [a] 0.510 [b]	0.261	0.494	0.053	0.969	0.526 [a] 0.621 [b]	0.386	0.460	0.095	0.999	0.700 [a] 0.875 [b]	0.766	0.342	0.111
	500	1.008	0.226 [a] 0.223 [b]	0.050	0.988	0.051	1.005	0.249 [a] 0.273 [b]	0.074	0.956	0.082	0.996	0.317 [a] 0.390 [b]	0.152	0.832	0.117
	1000	1.004	0.159 [a] 0.155 [b]	0.024	1	0.042	1.001	0.177 [a] 0.192 [b]	0.037	0.999	0.070	1.002	0.225 [a] 0.286 [b]	0.082	0.971	0.121
11	30	1.019	0.943 [a] 0.943 [b]	0.890	0.202	0.049	0.977	0.830 [a] 2.585 [b]	6.684	0.579	0.541	-- [c]	-- [c] -- [c]	-- [c]	-- [a]	-- [a]
	100	0.994	0.507 [a] 0.523 [b]	0.274	0.498	0.052	0.998	0.915 [a] 1.482 [b]	2.196	0.338	0.244	0.944	2.059 [a] 2.230 [b]	4.975	0.072	0.051
	500	1.008	0.226 [a] 0.225 [b]	0.051	0.993	0.049	1.010	0.524 [a] 0.663 [b]	0.439	0.517	0.138	1.035	0.810 [a] 0.828 [b]	0.687	0.236	0.056
	1000	1.003	0.159 [a] 0.159 [b]	0.025	1	0.042	1.022	0.378 [a] 0.457 [b]	0.209	0.734	0.107	1.016	0.557 [a] 0.568 [b]	0.323	0.432	0.057
29	30	1.014	0.935 [a] 0.995 [b]	0.990	0.193	0.056	1.014	0.654 [a] 2.614 [b]	6.833	0.629	0.611	-- [c]	-- [c] -- [c]	-- [c]	-- [a]	-- [a]
	100	1.009	0.507 [a] 0.543 [b]	0.294	0.513	0.046	0.995	0.911 [a] 2.328 [b]	5.420	0.505	0.455	0.982	4.919 [a] 9.052 [b]	81.94	0.063	0.059
	500	1.010	0.226 [a] 0.225 [b]	0.051	0.989	0.050	0.958	0.759 [a] 1.041 [b]	1.085	0.335	0.177	0.950	1.350 [a] 1.385 [b]	1.920	0.103	0.052
	1000	1.000	0.160 [a] 0.162 [b]	0.026	1	0.058	1.008	0.570 [a] 0.739 [b]	0.547	0.464	0.143	1.023	0.898 [a] 0.904 [b]	0.818	0.200	0.056

Note: The sample sizes are the sizes for samples with overlapping observations.

[a] These are the estimated standard deviations of the parameter estimates.

[b] These are the actual standard deviations of the parameter estimates.

[c] These values cannot be estimated because of the very small number of observations.

Table 9. Parameter Estimates of Different Models for the Case of the Lagged Dependent Variable.

Equation Number	Method of Estimation	Data	Estimated Model
(26)	MLE	Overlapping	$Y_t = 0.0016 + 0.496\,Y_{t-1} + 1.0065\,X_t + \varepsilon_t + \varepsilon_{t-1} + 0.99999\,\varepsilon_{t-2}$
(27)	MLE	Overlapping	$Y_t = 0.078 + 0.108\,Y_{t-3} + 1.007\,X_t + 0.493\,X_{t-1} + 0.234\,X_{t-2} + \varepsilon_t + 1.471\,\varepsilon_{t-1} + 1.69\,\varepsilon_{t-2} + 0.69\,\varepsilon_{t-3} + 0.219\,\varepsilon_{t-4}$
(29)	MLE	Overlapping	$Y_\tau = 0.019 + 0.123\,Y_{\tau-3} + 1.002\,X_{t-1} + 0.489\,X_{t-2} + 0.251\,X_{t-3} + \varepsilon_\tau + 0.163\,\varepsilon_{\tau-1}$
(28)	MLE	Nonoverlapping	$Y_\tau = 0.015 + 0.118\,Y_{\tau-3} + 1.413\,X_\tau + 0.342\,X_{\tau-3} + \varepsilon_\tau + 0.163\,\varepsilon_{\tau-3}$
(28)	Newey-West OLSNO	Nonoverlapping	$Y_\tau = 0.278\,Y_{\tau-3} + 1.415\,X_\tau + \varepsilon_\tau$

Note: The models in Table 9 are estimated using a large Monte Carlo sample of 500,000 observations. The unrestricted maximum likelihood estimates are obtained using PROC ARIMA while the Newey-West and OLSNO estimates are obtained using PROC IML in SAS.

Table 10. Parameter Estimates, Standard Deviations, and MSE, for GLS, Newey-West, OLSNO, and the Disaggregate Estimation with Measurement Errors in X (Overlapping 1, 11, and 29).

Correlation of X	Sample Size	Degree of Overlap	GLS Estimation			Newey-West Estimation			Non-overlapping Estimation			Disaggregate Estimation		
			Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE	Parameter Estimates	Standard Deviations	MSE
0	100	1	0.494	0.252 [a] 0.252 [b]	0.320	0.493	0.269 [a] 0.311 [b]	0.354	0.494	0.360 [a] 0.361 [b]	0.389	0.494	0.250 [a] 0.250 [b]	0.318
		11	0.509	0.252 [a] 0.263 [b]	0.310	0.512	0.479 [a] 0.739 [b]	0.784	0.503	0.952 [a] 1.028 [b]	1.303	0.510	0.239 [a] 0.251 [b]	0.303
		29	0.495	0.253 [a] 0.254 [b]	0.320	0.480	0.501 [a] 1.185 [b]	1.675	0.390	1.789 [a] 2.310 [b]	5.709	0.497	0.222 [a] 0.223 [b]	0.303
	1000	1	0.499	0.079 [a] 0.077 [b]	0.257	0.502	0.088 [a] 0.095 [b]	0.257	0.501	0.112 [a] 0.111 [b]	0.261	0.499	0.079 [a] 0.077 [b]	0.257
		11	0.502	0.079 [a] 0.080 [b]	0.255	0.499	0.189 [a] 0.227 [b]	0.303	0.497	0.277 [a] 0.281 [b]	0.332	0.501	0.079 [a] 0.080 [b]	0.255
		29	0.499	0.079 [a] 0.078 [b]	0.257	0.517	0.285 [a] 0.364 [b]	0.366	0.509	0.441 [a] 0.445 [b]	0.440	0.499	0.078 [a] 0.077 [b]	0.257
0.8 [c]	100	1	0.718	0.191 [a] 0.199 [b]	0.119	0.816	0.174 [a] 0.214 [b]	0.080	0.816	0.218 [a] 0.223 [b]	0.084	0.716	0.190 [a] 0.198 [b]	0.120
		11	0.731	0.187 [a] 0.196 [b]	0.111	0.931	0.187 [a] 0.302 [b]	0.096	0.934	0.337 [a] 0.351 [b]	0.127	0.721	0.181 [a] 0.187 [b]	0.113
		29	0.730	0.186 [a] 0.194 [b]	0.110	0.963	0.174 [a] 0.429 [b]	0.186	0.966	0.536 [a] 0.701 [b]	0.493	0.720	0.166 [a] 0.174 [b]	0.109
	1000	1	0.735	0.058 [a] 0.060 [b]	0.074	0.833	0.055 [a] 0.065 [b]	0.032	0.832	0.066 [a] 0.067 [b]	0.033	0.734	0.058 [a] 0.060 [b]	0.074
		11	0.733	0.058 [a] 0.062 [b]	0.075	0.940	0.071 [a] 0.086 [b]	0.011	0.941	0.096 [a] 0.097 [b]	0.013	0.732	0.058 [a] 0.062 [b]	0.075
		29	0.736	0.058 [a] 0.061 [b]	0.073	0.954	0.091 [a] 0.116 [b]	0.016	0.950	0.135 [a] 0.138 [b]	0.021	0.735	0.057 [a] 0.060 [b]	0.074

Note: The sample sizes are the sizes for samples with overlapping observations.

[a] These are the estimated standard deviations of the parameter estimates.

[b] These are the actual standard deviations of the parameter estimates.

[c] The x is generated as follows: $x_t = x_{0t} + \omega_t$, where $x_{0t} \sim uniform\ (0, 1)$ and $\omega_t \sim N\ (0, 1/12)$.

References

Andrews, D. W. K., and J. C. Monahan , “An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,” *Econometrica*, 60, No. 4, (1990), 953-966.

Brewer, K. R. W., “Some Consequences of Temporal Aggregation and Systematic Sampling from ARMA and ARMAX Models,” *Journal of Econometrics*, 1, (1973), 133-154.

Coulibaly, N., and B. W. Brorsen. “Monte Carlo Sampling Approach to Testing Nonnested Hypotheses: Monte Carlo Results.” *Econometric Reviews*, 18, No. 2 (1999), 195-209.

Davidson, R., and J. G. MacKinnon. *Estimation and Inference in Econometrics*. New York: Oxford University Press, 1993.

Edgerton, D. L., “Should Stochastic or Non-stochastic Exogenous Variables Be Used in Monte Carlo Experiments?” *Economics Letters*, 53, (1996), 153-159. (November).

Engle, R. F., “Biases from Time-Aggregation of Distributed Lag Models,” Ph.D. Thesis, Cornell University, University Microfilms, Ann Arbor, Michigan (1969).

Gilbert, C. L., “Testing the Efficient Market Hypothesis on averaged data,” *Applied Economics*, 18 (1986), 1149-1166.

Goetzmann, W. N., and P. Jorion. “Testing the Predictive Power of Dividend Yields,” *Journal of Finance*, 48, (1993), 663-680.

Greene, W. H., *Econometric Analysis*, Third Edition. New York: Macmillian Publishing Company, 1997.

Hansen, L. P., and R. J. Hodrick, “Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis,” *Journal of Political Economy*, 88, No.5, (1980), 829-853.

Jones, C. M., and G. Kaul, “Oil and the Stock Markets,” *The Journal of Finance*, 51, No. 2, (1996), 463-491.

Kazimi, C., and D. Brownstone. “Bootstrap Confidence Bands for Shrinkage Estimators.” *Journal of Econometrics,* 90(1999):99-127.

Marcellino, M., “Some Temporal Aggregation Issues in Empirical Analysis,” Discussion Paper 96-39, University of California, San Diego, Department of Economics, 1996.

Marcellino, M., “Some Consequences of Temporal Aggregation in Empirical Analysis,” *Journal of Business and Economic Statistics*, 17, No. 1, (1999), 129-136.

Mark, N. C. “Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability and Overshooting,” *American Economic Review*, 85, (1995), 201-218.

Nelson, C. R., and M. J. Kim, “Predictable Stock Returns: The Role of Small Sample Bias,” *Journal of Finance*, 48, (1993), 641-661.

Newey, W. K., and K. D. West, “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55 (1987), 703-708.

Searle, S. R., G. Casella, and C. E. McCulloch. *Variance Components*. New York: John Wiley & Sons, Inc., 1992.

Wei, W. W. S., “Effect of Systematic Sampling on ARIMA Models,” *Communications in Statistical-Theoretical Mathematics*, 10, (1981), 2389-2398.

Weiss, A. A., “Systematic Sampling and Temporal Aggregation in Time Series Models,” *Journal of Econometrics*, (1984), 271-281.

White, H., “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” *Econometrica*, 48, (1980), 817-838.