The Overlapping Data Problem

by

Ardian Harri[a] and B. Wade Brorsen[b]

[a]Department of Agricultural Economics, Mississippi State University, PO Box 5187, Mississippi State, MS 39762, U.S.A. Phone: (662) 325-2044, email harri@agecon.msstate.edu

[b]Department of Agricultural Economics, Oklahoma State University

Corresponding Author:
B. Wade Brorsen
Department of Agricultural Economics
Oklahoma State University
308 Ag Hall
Stillwater, OK 74075

Phone: (405) 744-6836
Fax: (405) 744-8210
E-mail: wade.brorsen@okstate.edu

Final Version. Published as

Harri, Ardian and B. Wade Brorsen, "The Overlapping Data Problem," *Quantitative and Qualitative Analysis in Social Sciences*, 2009, *3* (3), 78–115.

Abstract

Overlapping data are often used in finance and economics, but applied work often uses inefficient estimators. The article evaluates possible reasons for using overlapping data and provides a guide about which estimator to use in a given situation.

JEL Code: C13

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. 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, 1980). Past literature has recognized the presence of the moving average error term. Our article seeks to improve econometric practice when dealing with overlapping data by synthesizing and adding to the literature on overlapping data. We find limited statistical reasons for not using the disaggregate data and that the preferred estimation method can vary depending on the specific problem.

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 does not eliminate the moving average error term (Gilbert, 1986) and can

introduce autocorrelation not present in the original series (Working 1960). 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) (1980), Newey-West (NW) (1987), Andrews and Monahan (AM) (1990), and West (1997). A fourth way is to "transform the long-horizon overlapping regression into a non-overlapping regression of one-period returns onto a set of transformed regressors" Britten-Jones and Neuberger (B-JN) (2004). A final way is to use OLS estimation with overlapping data, which yields biased hypothesis tests.

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 and 2004 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 portion of articles using overlapping data increased from 1996 to 2004 (Table 1) so that the majority of articles in finance now use overlapping data. Most of the empirical articles that used overlapping data studied asset returns or economic growth. A common feature of these articles is that returns or growth are measured over a period longer than the observation period. For example, data are observed monthly and the estimation is done annually. Authors provide several possible reasons for using aggregated data. The most common reason given is

measurement error in independent variables. For example, Jones and Kaul (1996, p. 469), state that they select "use of quarterly data on all variables as a compromise between the measurement errors in monthly 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. Britton Jones and Neuberger (2004) contend, the use of overlapping data is based more on economic reasons rather than statistical ones. Here, we evaluate possible statistical reasons for using overlapping data.

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."

We first demonstrate that the commonly used Newey-West and OLSNO methods can be grossly inefficient ways of handling the overlapping data problem. This is 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]. Then, we consider possible statistical reasons for using overlapping data such as nonnormality, missing data, and errors in variables. Finally, we evaluate ways of handling overlapping data when there are economic reasons for doing so.

While Newey-West and OLSNO estimation provide inefficient estimates the GLS estimation cannot be applied in every situation involving overlapping data. 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 (1980) 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 little to add to the previous literature (eg. Mark, 1995) that favors using the bootstrap to correct the small sample bias in the Hansen and Hodrick approach. With a general multivariate time series model, often overlapping data cannot be used to recover estimates of the disaggregate process that generated the data. The percentage of cases where lagged values of the dependent variable are used as an explanatory variable is reported in Table 1. In *The Journal of Finance* less than 25 percent of articles include a lagged dependent variable as an explanatory variable (half with the Newey-West estimator and half with OLSNO). For the *American Economic Review* about 7 percent (all with the Newey-West estimator) of the articles included a lagged dependent variable. Thus, in most cases where nonoverlapping data are used, there are no lagged dependent variables and so more precise estimation methods are available.

1. The Strictly Exogenous Regressors Case

There are many variations on the overlapping data problem. We first consider the simplest case where the data represent aggregates and the explanatory variables are strictly exogenous. This is the most common case in the literature such as when annual data are used for dependent and independent variables and monthly data are available for both.

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

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

where y_t is the dependent variable, $\boldsymbol{x_t}$ 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:

$$\mathrm{E}[u_t] = 0,\ \mathrm{E}[u_t^2] = \sigma_u^2,\ \text{and}\ \mathrm{cov}[u_t, u_s] = 0 \quad \text{if } t \neq s.$$

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:

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

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).

Because 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:

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

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:

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

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

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

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

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

where, Ω is the correlation matrix. The correlation matrix, Ω, appears in Gilbert's article and the presence of a moving average error term is commonly recognized.

With Ω derived analytically the generalized least squares (GLS) parameter estimates and their variance-covariance matrix can be obtained as follows:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \tag{9}$$

and

$$\text{var}[\hat{\beta}] = \sigma_e^2(X'\Omega^{-1}X)^{-1} \tag{10}$$

where $X = (X_1',...,X_{n-k+1}')$ and $Y = (Y_1,...,Y_{n-k+1})$. Under these assumptions, the GLS estimator of the aggregate model 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. This case cannot explain why authors would choose to use overlapping data. The disaggregate model does not lose observations to aggregation so it would still be preferred in small samples.

2. 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 (1993, p. 611) say "the Newey-West estimator is never greatly inferior to that of the alternatives." With the Newey-West estimation method, parameter estimates are obtained by using OLS. The OLS estimate *b* is unbiased and consistent but inefficient. The OLS estimate of σ_e^2 is biased and inconsistent. The Newey and West (1987) autocorrelation consistent covariance matrix is computed using the OLS residuals.

The OLSNO estimation method estimates parameters using OLS with a reduced sample where the observations do not overlap. The OLSNO 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.

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.

3. 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 also computed. 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.

4. Monte Carlo Procedure

Data are generated using Monte Carlo methods. A single independent variable x with an i.i.d. uniform distribution (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 here since the conclusions did not change. The options RANUNI and RANNOR in SAS software 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 (1996) findings that using stochastic exogenous variables in Monte Carlo studies improves considerably the precision of the estimates of power and size. 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 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. The unrestricted MLE estimates of β were obtained using PROC ARIMA in SAS. The Ω matrix to be used in GLS estimation was derived in equation (8). The Newey-West estimation was validated by comparing it with the available programmed estimator in SHAZAM software using the OLS ... /AUTCOV option. The power of the tests are calculated for the null hypothesis $\beta = 0$.

5. Results for the Exogenous Regressor Case

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 demonstrates that the Newey-West 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. 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, 1993; Goetzmann and Jorion, 1993; Smith and Yadar 1996; Britten-Jones and Neuberger, 2004). The parametric bootstrap suggested by Mark (1995) and used by Irwin et al. (1996) 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 hypothesis testing results 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 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 replications as well as the number/percentage of replications that converge. Fewer replications converge as the degree of overlap increases and as 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.

6. Possible Statistical Reasons for Using Overlapping Data

If the explanatory variables were strictly exogenous, no observations were missing, and the errors were distributed normally as assumed so far, there are no statistical reasons to use overlapping data since the disaggregate model could be estimated. We now consider possible statistical reasons for using overlapping data.

6.1 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. In this case, the disaggregate model cannot be estimated so OLSNO is what has been used in the past.

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 and thus use GLS estimation. 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. From this discussion it can be argued that the case of missing observations is a statistical reason for using overlapping data that stands up to scrutiny but more efficient estimators are available than the often used OLSNO.

6.2 Nonnormality. The GLS estimator does not assume normality, so estimates with GLS would remain best linear unbiased and asymptotically efficient even under nonnormality. The hypothesis tests derived, however, 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. However, the GLS results are almost identical to the results from the disaggregate model. This means that lack of normality cannot be a valid statistical reason for using overlapping data when the disaggregate data are available.

6.3 *Errors in variables*. The most common reason authors give for using overlapping data is 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 9. 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 bias due to measurement error. This is true for all methods of estimation. GLS would be the preferred estimator since it is always superior to Newey-West and OLSNO in terms of MSE, especially as the overlapping level increases relative to the sample size.

In the case when x is autocorrelated, for relatively low level of overlap in relation to the sample size, Newey-West and OLSNO have the smaller MSE, as a result of smaller bias. As the degree of overlap increases relative to the same sample size, the GLS estimator would be preferred compared to Newey-West and OLSNO estimators based on smaller MSE as a result of the smaller variance. Thus the trade-off for the researcher is between less biased parameter estimates with Newey-West or OLSNO versus smaller standard deviations for the parameter estimates with GLS. However, the GLS transformation of the variables does not reduce further the measurement error producing estimates that are just barely less biased than the disaggregate

estimates. On the other hand, Newey-West and OLSNO standard errors are still biased with the bias increasing as the overlapping level increases. 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.

7. Possible Economic Reasons for Using Overlapping Data

The economic reasons for using overlapping data typically involve lagged variables. Even though in most cases where overlapping data are used, there are no lagged variables, it is important to consider the case because it is the case that has generated the most econometric research. The lagged variable may be strictly exogenous or be a lagged value(s) of the dependent variable. With lagged variables and overlapping data, GLS is generally inconsistent and so the situation has generated considerable research interest. A second economic reason for using overlapping data is the case of long-horizon regressions. Examples of long-horizon regressions include the case of expected stock return (as dependent variable) and dividend yield (as explanatory variable) and also the case of the GDP growth and nominal money supply. Since the economic reason for using overlapping data is increased prediction accuracy, we compare prediction errors when using overlapping data to those using disaggregate data.

7.1 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 (1980) estimator. In the textbook case of autocorrelation and a lagged dependent variable, ordinary least squares estimators are inconsistent.

Engle (1969) shows that when the first lag of the aggregated dependent variable is used as an explanatory variable, that using OLS and aggregated data could lead to bias of either sign and almost any magnitude. Generalized least squares is also inconsistent. Consistent estimates can be obtained using the maximum likelihood methods developed for time-series models.

With a lagged dependent variable in the right hand side (for simplicity we are using a lag order of one), equation (1) now becomes:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_1 + u_{t,} \qquad u_t \sim N(0,1) \tag{11}$$

where for simplicity $\alpha_0 = 0$ and $\alpha_2 = 1$. The value selected for α_1 is 0.5. To get the overlapping observations, for $k = 3$ apply (3) to (11) to obtain the equivalent model of (2) as:

$$Y_t = 0.5Y_{t-1} + X_t + e_t \tag{12}$$

where $Y_t = y_1 + y_{t-1} + y_{t-2}$, $X_t = x_t + x_{t-1} + x_{t-2}$, and $e_t = u_t + u_{t-1} + u_{t-2}$. The model in (12) also has the same variance-covariance matrix, described by (5) and (6), as our previous model in (2).

If Y_t and X_t are observed in every time period t, the order of the lagged variables as well as the AR and MA orders can be derived analytically. For a detailed discussion of the issues related to temporal aggregation of time series see Marcellino (1996, 1999)[3].

If Y_t, or X_t, or none of them are observed in every time period t, then either lagged values for aggregate Y, (Y_{t-1} and Y_{t-2}), or X, (X_{t-1} and X_{t-2}), or both Y and X are not observable. The model usually estimated[4] in this second situation is:

$$Y_\tau = 0.5^k Y_{\tau-1} + \beta_2 X_\tau + \beta_3 X_{\tau-1} + \upsilon_\tau \tag{13}$$

where, τ represents every k^{th} observation. Assuming the data are observed every time period (13) is equivalent to[5]:

$$Y_t = 0.5^3 Y_{t-3} + X_t + 0.5X_{t-1} + 0.5^2 X_{t-2} + \varepsilon_t \qquad (14)$$

The resulting error term ε_t in (14) is a MA process of order four of the error term u_t in (11) with coefficients 1.5, 1.75, 0.75 and 0.25, $\varepsilon_t = u_t + 1.5u_{t-1} + 1.75u_{t-2} + 0.75u_{t-3} + 0.25u_{t-4}$.

One potential problem with the model in (14) is the noise introduced by aggregation. The variable X_{t-1} and X_{t-2} include x_{t-1}, x_{t-2}, x_{t-3}, and x_{t-4}, while only x_{t-1}, and x_{t-2}, are relevant as shown by the model in (12). 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. The errors-in-variables problem is even bigger for the model in (13) where x_{t-5}, is also included in the model through $X_{\tau-1}$.

An analytical solution for the β-s in (13) cannot be derived. This is because 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.

Finally, another possible model using nonoverlapping observations for Y and overlapping observations for X is:

$$Y_\tau = 0.5^3 Y_{\tau-1} + X_t + 0.5X_{t-1} + 0.5^2 X_{t-2} + \eta_\tau \qquad (15)$$

In general, in cases when nonoverlapping data are used, as also is the case of Hansen and Hodrick's estimator, Marcellino (1996, 1999) shows that 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.

We estimated the models in (12), (13), and (15) with MLE employing PROC ARIMA in SAS software using a large Monte Carlo sample of 500,000 observations. There is no need to estimate the model in (14) since the model in (12) can be estimated when overlapping data are available for both Y and X. The results are reported in Table 10. The empirical estimates of the AR and MA coefficients and the coefficients of the Xs for the models in (12) fully support the analytic findings. The parameter estimates for the exogenous variables in (15) are similar to the analytical values. On the other hand, the parameter estimates for the exogenous variables in (15) are very different from the analytical values derived for either (12) or (14) because of the different lagged values of the exogenous variable included in the model. Both models in (13) and (15) result in an ARMA(1,1) process with the AR coefficient 0.118 for (13) and 0.123 for (15). The MA coefficient is the same for both models, 0.163. As noted above, the AR and MA coefficients for (13) and (15) are different from the respective coefficients of the disaggregate model.

With overlapping data and a lagged dependent value as an explanatory variable where the lag is less than the level of overlap, the only consistent estimation method is maximum likelihood. Maximum likelihood provides consistent estimates when the explanatory variables are predetermined whether or not they are strictly exogenous. When overlapping data are used for both the dependent and independent variables the parameters of the aggregate model are the

same as those of the disaggregate model. When nonoverlapping data are used for the dependent or the independent or both dependent and independent variables the parameters of the aggregate model cannot be used to recover those of the disaggregate model.

7.2 The Case of Long-Horizon Regressions with Overlapping Observations

The underlying data-generated processes commonly assumed are:

$$y_t = \alpha + x_{t-1}\beta + u_t \tag{16}$$

$$x_t = \mu + \rho x_{t-1} + \nu_t \tag{17}$$

$$E[(u_t, \nu_t,)'(u_t, \nu_t,)] = \sum = [\sigma_{11}^2 \; \sigma_{12}; \sigma_{21}\sigma_{22}^2] \tag{18}$$

The effect of the assumption in (17) is that the covariance between the error terms in (16) that are one period apart contains additional terms as compared to (6):

$$\text{cov}[e_t, e_{t+1}] = E[e_t \, e_{t+1}] = (k-1)\sigma_u^2 + \beta^2(1+\rho)'\sigma_v^2 \tag{19}$$

The other covariance terms are as in (6).

The long-horizon variables, Y_t and X_{t-1} are created as in (3). Then the estimated models include:

$$Y_t = \alpha + X_{t-k-1}\beta + e_t \tag{20}$$

As Valkanov (2003, p. 205) argues "Intuitively, the aggregation of a series into a long-horizon variable is thought to strengthen the signal, while eliminating the noise." Several studies

have attempted to provide more efficient estimators of the standard errors compared to the ordinary least squares estimates. Valkanov (2003) suggests rescaling of the *t*-statistic by the square root of the sample size, Hjalmarsson (2004) suggests rescaling the *t*-statistic by the square root of the level of aggregation, *k* and Hansen and Tuypens (2004) suggest rescaling the *t*-statistic by the square root of ⅔ of the level of aggregation, *k*.

Two scenarios are considered. The first one is the scenario where the returns are unpredictable which implies that $\beta = 0$. The second is when returns are predictable, so that $\beta \neq 0$ (studies usually assume $\beta = 1$). The first scenario is exactly the case considered in Section 1, so the conclusions from Section 1 apply here.

For the case when returns are predictable, $\beta = 1$, we perform Monte Carlo simulations to compare the standard errors of the GLS estimator and the OLS standard errors rescaled by the square root of the sample size (Valkanov, 2003), by the square root of the level of aggregation (Hjalmarsson, 2004) and by the square root of ⅔ of the level of aggregation (Hansen and Tuypens, 2004). We use two levels of aggregation k=12 with sample sizes, 50, 100, 250, 500 and 1000 and k=75 with sample sizes, 200, 250, 500, 750 and 1000. 5000 replications are done for each case. We conduct Monte Carlo simulations for commonly used assumptions of $\rho = 1$ (in equation (19)), and $\sigma_{12} = \sigma_{21} = 0.9$ and $\sigma_{12} = \sigma_{21} = -0.9$. We also conduct simulations by changing ρ to 0.9 and σ_{12} to 0.5 and 0.1. In addition, we assume $\alpha = \mu = 0$.

A summary of the results from the above simulation follows. In general, all estimators and rescaling approaches produce very good power against the alternative hypothesis $\beta = 0$. However, this is not true for the size of the tests. Size is the critical issue when using an approximation like this because a test should be conservative so that if the test rejects the null

hypothesis, the researcher can be confident that the conclusion is correct. For the high absolute values of σ_{12} considered above, respectively 0.9 and -0.9, the most promising approach is the one suggested by Hansen and Tuypens where standard errors are rescaled by the square root of ⅔ of the level of aggregation. However, this approach still produces correct test sizes only when the ratio level of aggregation/sample size is close to 1/10. Test sizes are greater than the nominal size when the ratio level of aggregation/sample size is less than 1/10 and less than the nominal size when the ratio level of aggregation/sample size is greater than 1/10. Our simulations suggest that better test sizes are produced when the following adjustment is applied to the Hansen and Tuypens approach. When the ratio is less than 1/10, the adjustment is as follows: $\sqrt{\frac{2}{3}(0.9k-1)}$ and when the ratio is greater than 1/10 the adjustment is: $\sqrt{\frac{2}{3}k\left(0.9+\frac{k}{n}\right)}$. Test sizes for the Hansen and Tuypens (HT) and our modified version of HT approach are reported in Table 11. Table 11 also reports test sizes for the cases when σ_{12} equals 0.5 and 0.1. In the case when $\sigma_{12} = 0.5$, the modified Hjalmarsson rescaling of the standard errors by the square root of the level of aggregation produces better test sizes for different sample/aggregation level combination. In this case, for ratios less than 1/10, the adjustment is: $\sqrt{(0.9k-1)}$ and for ratios greater than 1/10 the adjustment is: $\sqrt{k\left(0.9+\frac{k}{n}\right)}$. Finally, when $\sigma_{12} = 0.1$, the GLS standard errors produce good test sizes. This is not surprising since a small σ_{12} brings us closer to the case of exogenous independent variables discussed in Section 1. Therefore, no modified rescaling is needed in this last case. However, we also report in Table 11 the test sizes for the unmodified Hjalmarsson rescaling as a comparison to the GLS test sizes.

7.3 Prediction accuracy. In this section we compare the prediction accuracy of the aggregate and disaggregate models. We also compare the prediction accuracy for the Hansen and Hodrick (HH) model for HH given in endnote 4. The disaggregate model with the HH estimator is similar to the aggregate model with the only change being the dependent variable is disaggregate (one-period change). Numerous authors have argued for using overlapping data when predicting multiperiod changes. Bhansali (1999), however, reviews the theoretical literature and finds no theoretical advantage to using overlapping data when the number of lags is known. Marcellino, Stock, and Watson (2006) consider a number of time series and find no empirical advantage to using overlapping data even when using pretesting to determine the number of lags to include. Since the models here are known, theory predicts no advantage to using overlapping data.

The Monte Carlo simulation involves generating 50,000 sample pairs. The first sample is used to obtain the parameter estimates for the aggregate and disaggregate models for all cases. The second sample is used to obtain predicted values by utilizing the parameter estimates for the first sample. Dynamic forecasting is used to obtain predicted values of the lagged values in the disaggregate model.

The means and standard deviations of the 50,000 root-mean-squared forecast errors are reported in Table 12. Two levels of aggregation, 12 with five sample sizes, 50, 100, 250, 500, and 1000, and 75 with sample sizes 100, 250, 500, 750, and 1000, are used in the simulations.

In the case of long horizon regressions the aggregate and disaggregate model are roughly equal in forecast accuracy as expected. In the case of the HH model for low levels of aggregation (level 12) the differences between the aggregate and the disaggregate models are small. With the

high level of aggregation, the disaggregate model sometimes outperforms the aggregate one. If the economic goal is prediction, overlapping data might be preferred if they make calculations easier since there is little difference in forecast accuracy.

8. Special Cases of Overlapping Data

There are several special cases of overlapping data that do not fit any of the standard procedures. Since the solutions are not obvious, we now discuss how to handle overlapping data in the presence of varying levels of overlap, imperfect overlap, seasonal unit roots, additional sources of autocorrelation, heteroscedasticity, and generalized autoregressive conditional heteroskedasticity (GARCH).

8.1 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 variation of the missing data problem 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^2_u 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:

$$Var[e_t] = \sigma_e^2 = E[e_t^2] = k_t \sigma_u^2 \tag{21}$$

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:

$$\text{cov}[e_t, e_{t+s}] = E[e_t e_{t+s}] = (k_{ts})\sigma_u^2 \tag{22}$$

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

$$\Sigma = \sigma_u^2 \begin{bmatrix} k_1 & k_{12} & k_{13} & \cdots & k & 0 & 0 \\ k_{21} & k_2 & k_{23} & \cdots & \cdots & k_{2s} & 0 \\ \cdots & k_{32} & k_3 & k_{34} & \cdots & \cdots & k_{3s} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & k_{ts} & \cdots & k_{t(t-2)} & k_{t(t-1)} & k \end{bmatrix} \tag{23}$$

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

8.2 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-

month returns between say March and September contracts would not overlap. 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

$$\operatorname{cov}(u_{jt}, u_{st+m}) = \begin{matrix} \sigma_{js} \text{ if } m = 0 \\ 0 \text{ otherwise} \end{matrix} \tag{24}$$

be the covariance between the monthly returns *m* months (or days if disaggregate 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,

$$\operatorname{var}(u_{jt}) = \operatorname{var}(u_{st}) = \sigma_{jt}^2, \operatorname{var}(e_{jt}) = \operatorname{var}(e_{st}) = k\sigma_u^2 \tag{25}$$

and

$$\operatorname{cov}(e_{jt}, e_{st-m}) = k_{js}\sigma_{js} \tag{26}$$

where k_{js} is the number of overlapping months between the March and June contracts and $\sigma_{js} = \rho_i \sigma_u^2$ where $\rho_1 (i = 1,2)$ is the correlation between the *u*'s for two consecutive contracts with maturities three (ρ_i) and six (ρ_2) months apart. The covariance matrix for (2) with *n* = 12, in this case is:

$$
\Sigma = \sigma_u^2 \begin{bmatrix}
k & \frac{k-1}{k} & \frac{k-2}{k}\rho_1 & \frac{k-3}{k}\rho_1 & \frac{k-4}{k}\rho_2 & \frac{k-5}{k}\rho_2 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{k-1}{k} & k & \frac{k-1}{k} & \frac{k-2}{k}\rho_1 & \frac{k-3}{k}\rho_1 & \frac{k-4}{k}\rho_2 & \frac{k-5}{k}\rho_2 & 0 & 0 & 0 & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{k-3}{k}\rho_1 & \frac{k-2}{k}\rho_2 & \frac{k-1}{k} & k & \frac{k-1}{k} & \frac{k-2}{k}\rho_1 & \frac{k-3}{k}\rho_1 & \frac{k-4}{k}\rho_2 & \frac{k-5}{k}\rho_2 & 0 & 0 & 0 \\
\frac{k-4}{k}\rho_2 & \frac{k-3}{k}\rho_1 & \frac{k-2}{k}\rho_1 & \frac{k-1}{k} & k & \frac{k-1}{k} & \frac{k-2}{k}\rho_1 & \frac{k-3}{k}\rho_1 & \frac{k-4}{k}\rho_2 & \frac{k-5}{k}\rho_2 & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{k-5}{k}\rho_2 & \frac{k-4}{k}\rho_2 & \frac{k-3}{k}\rho_2 & \frac{k-2}{k}\rho_2 & \frac{k-1}{k}\rho_2 & k
\end{bmatrix} \quad (27)
$$

8.3. Seasonal Unit Roots. The seasonal difference model of Box and Jenkins (1970), which is called a seasonal unit root model in more recent literature, uses data which are in some sense overlapping, but do not create an overlapping data problem if correctly specified. For annual data, the seasonal root model is

$$
\begin{aligned}
y_t &= ax_t + u_t \\
u_t &= u_{t-12} + e_t
\end{aligned} \quad (28)
$$

where e_t is i.i.d. normal. In this case, the disaggregrate model

$$
y_t - y_{t-12} = \alpha(x_t - x_{t-12}) + e_t
$$

has no autocorrelation. In this example, twelfth differencing leads to a model that can be estimated using overlapping data and ordinary least squares. Seasonal unit roots have largely been used when the research objective was forecasting (e.g. Clements and Hendry 1997). One problem with the seasonal unit root model is that it is often rejected in empirical work (e.g. McDougall 1995). Another is that it implies that each month has its own independent unit root process and so each month's price can wander aimlessly away from the prices of the other months. Such a model seems implausible for most economic time series. Hyllebert et al. (1990) suggest that the seasonal unit roots may be cointegrated, which can overcome the criticism of one month's price moving aimlessly away from another month's price. Wang and Tomek (2007) present another challenge to the seasonal unit root model since they argue that commodity prices should not have any unit roots. While a seasonal unit root model may be an unlikely model, if it is the true model, it does not create an overlapping data problem.

8.4 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 below 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:

$$m(L)u_t = h(L)\xi_t \tag{29}$$

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

$$(1+L+...+L^{k-1})y_t = (1+L+...+L^{k-1})x_t + (1+L+...+L^{k-1})u_t \quad (30)$$

introduces the same level *k* of aggregation to (31), which now becomes:

$$(1+L+...+L^{k-1})m(L)u_t = (1+L+...+L^{k-1})h(L)\xi_t \quad (31)$$

or

$$M(L)e_t = H(L)E_t \quad (32)$$

Then, the procedures discussed in the lagged dependent variable case can be applied with respect to (31) to obtain the order and the values of the AR and MA coefficients in (32) to be used in estimating (2). In this case, maximum likelihood methods for estimating a regression with ARMA errors can be used.

8.5 *Heteroskedasticity*. If the residuals in the disaggregated data (u_t in (1)) are heteroskedastic, then estimation is more difficult. Define σ_{ut}^2 as the time-varying variance of u_t and σ_{et}^2 as the time-varying variance of e_t. Assume the u_t's are independent and thus $\sigma_{et}^2 = \sum_{j=0}^{k-1} \sigma_{ut-j}^2$. For simplicity, assume that σ_{ut}^2 depends only on x_t. If σ_{ut}^2 is assumed to be a linear function of $x_t (\sigma_{ut}^2 = \gamma' x_t)$ then the function aggregates nicely so that $\sigma_{et}^2 = \sum_{j=0}^{k-1} \gamma' x_{t-j} = \gamma' X_t$. But, if multiplicative heteroskedasticity is assumed $\sigma_{ut}^2 = \exp(\gamma' x_t)$ then $\sigma_{et}^2 = \sum_{j=0}^{k-1} \exp(\gamma' x_{t-j})$ 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

$$\mathrm{Cov}(e_t, e_{t+s}) = \sum_{j=s}^{k-1} \sigma^2_{u(t-j)} \tag{33}$$

The correlation matrix, Ω is known, as given by (8), so the covariance matrix can be derived using the relation:

$$\Sigma = \Gamma' \Omega \Gamma \tag{34}$$

where $\Gamma = [\gamma' X_1,\ \gamma' X_2 \ldots, \gamma' X_T]$x I_T. A feasible generalized least squares estimator can then be developed using (12). It might be reasonable to use (9) as the first stage in a FGLS estimation that corrected for heteroskedasticity.

8.6 Overlapping data and GARCH processes. With financial data, it is common that the disaggregate model would follow a GARCH process (e.g. Yang and Brorsen 1993). The combination of overlapping data and GARCH processes besides the MA process in the mean also introduces additional autocorrelation in the second moment. As an example, let us assume a GARCH(1,1) process for the volatility of the disaggregate process and an overlapping level *k*=1. Then the diagonal elements of the covariance matrix for the aggregate model would follow a GARCH(2,2) process while the first off-diagonal elements would follow a GARCH(1,1) process. Thus the elements of the covariance matrix are correlated. The appropriate estimator in such a case could be the topic of future research.

9. Conclusions

We have evaluated different statistical and economic reasons for using overlapping data. These reasons are especially important since they provide the motivation for using overlapping data.

With strictly exogenous regressors as well as other standard assumptions, GLS is vastly superior to Newey-West and OLSNO. The Newey-West estimator gave hypothesis tests with incorrect size and low power even with sample sizes as large as 1,000. 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. If overlapping data were the only econometric problem, there would appear to be little reason to use overlapping data at all since the disaggregate model could be estimated. The practice of estimating a model with both monthly and annual observations, for example, would not have any apparent advantage.

We evaluated several statistical reasons for using overlapping data. If the motivation for using overlapping data is missing observations then GLS is the preferred estimator. Errors in variables with autocorrelated explanatory variables can be a reason to use overlapping data, but even with the extreme case considered, the advantage is small. When overlapping data are used due to nonnormality or errors in variables that are not autocorrelated, then GLS is still preferred compared to Newey-West or OLSNO. However, the GLS estimator provides no improvement compared to the disaggregate model. The GLS estimator would be easier to implement than the Newey-West estimator for varying levels of overlap or imperfect overlap.

We also evaluated economic reasons for using overlapping data. One such economic reason involves regressions of long-horizon asset returns with overlapping data as in the case of asset returns explained by dividend yields. In this case we propose a modified rescaling of the

errors that produces correct test sizes for different sample sizes and level of aggregation. Another economic reason is the case when lagged dependent variables are used as explanatory variables. In this case the GLS estimator is inconsistent. 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.

It can be reasonable to use overlapping data when the goal is to predict a multi-period change. Results showed no advantage in terms of prediction accuracy from directly predicting the multi-period change rather than using a disaggregate model and a multi-step forecast. But the aggregate model could be preferred if it were more convenient to use.

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, but authors could do much better than the methods they presently use.

Endnotes

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.

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.

3. See also Brewer (1973), Wei (1981), and Weiss (1984).

4. The model considered by Hansen and Hodrick is $Y_t = \beta + \beta_1 Y_{t-3} + \beta_2 Y_{t-4}$.

5. (14) is obtained by substituting for Y_{t-1} and then for Y_{t-2} in (12).

6. We also conducted Monte Carlo simulations for the transformations proposed by Britten-Jones and Neuberger (2004) to this model. Results are not reported since the β estimates from the BJN transformations were inconsistent.

10. References

Andrews, D. W. K., J.C. Monahan (1990). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. Econometrica, 60, 953-966.

Bhansali, R. J. (1999). Parameter Estimation and Model Selection for Multistep Prediction of a Time Series: A Review, in Asympototics, Nonparametrics, and Time Series. S. Ghosh, editor, Marcel Dekker, New York, pp. 201-225.

Box, G., and G. Jenkins (1970). Time Series Analysis and Forecasting, and Control. Holden-Day, San Francisco.

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

Britten-Jones, M., and A. Neuberger (2004). Improved Inference and Estimation in Regression with Overlapping Observations. EFA 2004 Maastricht Meetings Paper No. 4156. Available at SSRN: http://ssrn.com/abstract=557090.

Clements, M.P., and D.F. Hendry (1997). An Empirical Study of Seasonal Unit Roots in Forecasting. International Journal of Forecasting, 13, 341-355

Coulibaly, N., and B.W. Brorsen (1999). Monte Carlo Sampling Approach to Testing Nonnested Hypotheses: Monte Carlo Results. Econometric Reviews, 18, 195-209.

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

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

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

Gilbert, C. L. (1986). Testing the Efficient Market Hypothesis on Averaged Data. Applied Economics, 18, 1149-1166.

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

Greene, W. H. (1997). Econometric Analysis, Third Edition. Macmillian Publishing Company, New York.

Hansen, C. S., and B. Tuypens (2004). Long-Run Regressions: Theory and Application to US Asset Markets. Working Paper, Zicklin School of Business, Baruch College, New York, NY.

Hansen, L. P., and R. J. Hodrick (1980). Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis. Journal of Political Economy, 88, 829-853.

Hjalmarsson, E. (2004). On the Predictability of Global Stock Returns. Working Paper, Department of Economics, Yale University.

Hyllebert, S., R.F. Engle, C.W.J. Granger, and S.B. Yoo. (1990). Seasonal Integration and Cointegration. Journal of Econometrics, 44, 215-238.

Irwin, S.H., C.R. Zulauf, and T.E. Jackson (1996). Monte Carlo Analysis of Mean Reversion in Commodity Futures Prices. American Journal of Agricultural Economics, 78, 387-399.

Jones, C. M., and G. Kaul (1996). Oil and the Stock Markets. The Journal of Finance, 51, 463-491.

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

Marcellino, M. (1999). Some Consequences of Temporal Aggregation in Empirical Analysis. Journal of Business and Economic Statistics, 17, 129-136.

Marcellino, M. (1996). Some Temporal Aggregation Issues in Empirical Analysis. Discussion Paper, University of California Department of Economics: San Diego, California.

Marcellino, M., J.H. Stock, and M.W. Watson (2006). A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. Journal of Econometrics, 135, 499-526.

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

McDougall, R.S. (1995). The Seasonal Unit Root Structure in New Zealand Macroeconomic Variables. Applied Economics, 27, 817:827

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

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

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

Valkanov, R. (2003). Long-Run Regressions: Theoretical Results and Applications. Journal of Financial Economics, 68, 201–232.

Wang, D., and W.G. Tomek. (2007). Commodity Prices and Unit Root Tests. American Journal of Agricultural Economics, 89, 873-889.

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

Weiss, A. A. (1984). Systematic Sampling and Temporal Aggregation in Time Series Models. Journal of Econometrics, 26, 271-281.

West, K. D. (1997). Another Heteroskedasticity and Autocorrelation-Consistent Covariance Matrix Estimator. Journal of Econometrics, 76, 171-191.

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

Working, H. (1960). Note on the Correlation of First Difference Averages in a Random Chain. Econometrica, 28, 916-918.

Yang, S.R. and B.W. Brorsen (1993). Nonlinear Dynamics of Daily Futures Prices: Conditional Heteroskedasticity or Chaos? Journal of Futures Markets, 13,175-191.

Table 1. Number of Articles Using Overlapping Data, 1996-2004.

Journal	Year	Number of articles					Total number of empirical articles in the journal	Percentage of articles with overlapping data
		OLSNO	NW	Other[a]	OLS	Total		
J. Finance	1996	16	8	8	-	26	55	47.3
	2004	23	16	16	3	45	71	63.4
Amer. Econ. Rev.	1996	10	3	2	-	14	77	18.2
	2004	19	4	2	-	20	109	18.3
J. Fut. Mkts.	1996	12	3	5	2	19	43	44.2
	2004	18	5	5	5	26	44	59.1

Note: The sum of the columns 3 through 6 may be larger than the total in column 7 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 x and y.

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 Replications	Replications 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					Disaggregate Estimation				
		Parameter Estimates	Standard Deviations	MSE	Power	Size	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	1.012	0.977 [a] 1.029 [b]	1.058	0.192	0.050
	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	0.970	0.517 [a] 0.525 [b]	0.276	0.467	0.052
	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	1.008	0.236 [a] 0.233 [b]	0.054	0.983	0.055
	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	1.004	0.166 [a] 0.163 [b]	0.027	1	0.039
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]	-- [c]	-- [c]	1.010	0.833 [a] 0.870 [b]	0.757	0.239	0.053
	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	1.001	0.502 [a] 0.523 [b]	0.274	0.516	0.053
	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	1.007	0.233 [a] 0.233 [b]	0.054	0.988	0.052
	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	1.002	0.164 [a] 0.166 [b]	0.027	1	0.040
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]	-- [c]	-- [c]	0.995	0.680 [a] 0.726 [b]	0.527	0.319	0.051
	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	1.020	0.466 [a] 0.486 [b]	0.237	0.599	0.041
	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	1.009	0.228 [a] 0.229 [b]	0.052	0.988	0.046
	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	1.001	0.164 [a] 0.168 [b]	0.028	1	0.061

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, 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).

Table 10. Parameter Estimates of Different Models for the Case of the Lagged Dependent Variable.

Equation Number	Method of Estimation	Data	Estimated Model
(12)	MLE	Overlapping	$Y_t = 0.0016 + 0.496Y_{t-1} + 1.0065X_t + \varepsilon_t + \varepsilon_{t-1} + 0.99999\varepsilon_{t-2}$
(13)	MLE	Nonoverlapping	$Y_\tau = 0.019 + 0.118Y_{\tau-3} + 1.413X_\tau + 0.342X_{\tau-3} + \varepsilon_\tau + 0.163_{\tau-3}$
(15)	MLE	Y Nonoverlapping X Overlapping	$Y_\tau = 0.019 + 0.123Y_{\tau-1} + 1.002X_{t-1} - 1 + 0.489_{t-2} + 0.251X_{t-3} + \varepsilon_\tau + 0.163\varepsilon_{\tau-1}$

Note: The models in Table 10 are estimated using a large Monte Carlo sample of 500,000 observations. The unrestricted maximum likelihood estimates are obtained using PROC ARIMA in SAS.

Table 11. Size of Hypotheses Tests (β=1) for the Hansen and Tuypens (HT) and Hjalmarsson (H) rescaling and our modified HT and H approaches for Long-Horizon Return Regressions (Overlapping 11 and 74).

		$\sigma_{12} = 0.9$			
		Overlapping level 11			
Sample Size	**50**	**100**	**250**	**500**	**1000**
Size of HT	0.0622	0.0504	0.0314	0.0302	0.0280
Size of Modified HT[a]	0.0554	0.0504[b]	0.0464	0.0430	0.0420
		Overlapping level 74			
Sample Size	**200**	**250**	**500**	**750**	**1000**
Size of HT	0.0782	0.0724	0.0570	0.0458	0.0402
Size of Modified HT[a]	0.0522	0.0538	0.0524	0.0458[b]	0.0516
		$\sigma_{12} = 0.5$			
		Overlapping level 11			
Sample Size	**50**	**100**	**250**	**500**	**1000**
Size of H	0.0784	0.0608	0.0394	0.0308	0.0316
Size of Modified H[c]	0.0510	0.0608[b]	0.0574	0.0498	0.0488
		Overlapping level 74			
Sample Size	**200**	**250**	**500**	**750**	**1000**
Size of H	0.0962	0.0822	0.0668	0.0510	0.0504
Size of Modified H[c]	0.0576	0.0526	0.0550	0.0510[b]	0.0564
		$\sigma_{12} = 0.1$			
		Overlapping level 11			
Sample Size	**50**	**100**	**250**	**500**	**1000**
Size of H	0.1014	0.0782	0.0558	0.0480	0.0446
Size of GLS	0.0506	0.0530	0.0510	0.0468	0.0446
		Overlapping level 74			
Sample Size	**200**	**250**	**500**	**750**	**1000**
Size of H	0.1200	0.1096	0.0900	0.0734	0.0730
Size of GLS	0.0518	0.0520	0.0524	0.0504	0.0524

Note: The sample sizes are the sizes for samples with overlapping observations.
[a] This is the modified Hansen and Tuypens rescaling.
[b] No modification is performed when the ratio equals 1/10.
[c] This is the modified Hjalmarsson rescaling.

Table 12. Prediction accuracy for Long Horizon Regressions and Hansen and Hodrick Model

Aggregation Level	Sample Size	Long Horizon Regressions				Hansen and Hodrick Model			
		Aggregate Model		Disaggregate Model		Aggregate Model		Disaggregate Model	
		Mean[a]	SD[b]	Mean[a]	SD[b]	Mean[a]	SD[b]	Mean[a]	SD[b]
12	50	20.95	11.89	20.15	11.85	99.45	103.7	95.21	96.71
	100	21.87	8.17	21.74	8.35	59.46	61.05	58.70	58.68
	250	23.31	5.50	23.38	5.62	31.29	27.96	32.36	26.85
	500	23.93	4.05	24.06	4.18	21.34	14.62	23.19	13.50
	1000	24.33	2.90	24.49	3.00	16.48	7.34	18.98	6.53
75	100	95.14	66.35	74.46	42.64	1294.59	1475.41	1688.32	1425.21
	250	85.45	35.47	80.23	31.41	840.01	796.57	1034.06	1026.34
	500	86.22	25.31	84.72	24.69	563.64	495.89	706.27	618.32
	750	86.14	20.26	85.45	19.95	452.19	367.49	576.10	467.36
	1000	86.89	17.77	86.54	17.66	393.16	293.33	503.94	374.08

[a] These are means of the 50,000 estimated root-mean-squared forecast errors.
[b] These are the standard deviations of the 50,000 estimated root-mean-squared forecast error.