NBER TECHNICAL WORKING PAPER SERIES

DOES CORRECTING FOR HETEROSCEDASTICITY HELP?

Frederic S. Mishkin

Technical Working Paper No. 88

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1990

Research support has been provided by the Faculty Research Fund of the Graduate School of Business, Columbia University. I thank Bob Cumby for helpful comments. This paper is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #88
May 1990

DOES CORRECTING FOR HETEROSCEDASTICITY HELP?

ABSTRACT

This paper conducts Monte Carlo simulation experiments to evaluate how well the Hansen-Hodrick-Newey-West-White methodology performs for a particular example in the literature. The conclusion from this exercise is that although correcting for the overlapping data does help produce better statistical inference in finite samples, correcting for heteroscedasticity can substantially worsen statistical inference even when heteroscedasticity is present in the data. The answer to the question posed in the title of the paper is that correcting for heteroscedasticity may not help produce better statistical inference, but rather can do the opposite.

Frederic S. Mishkin
Graduate School of Business
Uris Hall 619
Columbia University
New York, NY 10027
(212)854-3488

I. Introduction

Empirical analysis of financial market data now often involves estimating linear rational expectations models in which the standard errors of the coefficients are corrected for both serial correlation and heteroscedasticity in the error term. Recent examples in the literature include Campbell and Clarida (1987), Froot (1989), Hardouvelis (1988), and Mishkin (1988, 1990a, 1990b). The correction for serial correlation of the error term is necessary because the number of periods spanned by the forecast horizon or the holding period of a security's return is greater than the observation interval -- i.e., the data is overlapping. Heteroscedasticity is corrected for generally because, as stated by Cumby and Huizinga (1989), "the econometrician is unlikely to know a priori whether a given data set is conditionally heteroscedastic or not. Therefore, having a test which works well on both conditionally homoscedastic and conditionally heteroscedastic data is desireable."

The standard estimation procedure to handle the serial correlation and heteroscedasticity in the regression is to use the method outlined by Hansen and Hodrick (1980), with a modification due to White (1980) and Hansen (1982) that allows for heteroscedasticity[1] and a modification by Newey and West (1987) that insures the variance-covariance matrix is positive definite by imposing linearly declining weights on autocovariance matrices. Although the Hansen-Hodrick estimation procedure is valid asymptotically, it can lead to incorrect statistical inference in finite samples when the amount of data overlap is substantial. Furthermore, the correction for heteroscedasticity,

---

[1]The Hansen (1982) modification is the same numerically as that proposed by White (1980). Hansen's analysis assumes there is conditional heteroscedasticity, while White's uses other assumptions but only assumes there is unconditional heteroscedasticity rather than conditional heteroscedasticity. For further discussion of econometric issues for regressions with overlapping data and serial correlation, see Cumby, Huizinga and Obstfeld (1983).

although it helps produce asymptotically valid standard errors, may also not improve statistical inference in finite samples.

This brief paper conducts Monte Carlo simulation experiments to evaluate how well the Hansen-Hodrick-Newey-West-White methodology performs for a particular example in the literature. The conclusion from this exercise is that although correcting for the overlapping data does help produce better statistical inference in finite samples, correcting for heteroscedasticity can substantially worsen statistical inference even when heteroscedasticity is present in the data. The answer to the question posed in the title of the paper is that correcting for heteroscedasticity may not help produce better statistical inference, but rather can do the opposite.

II. The Monte Carlo Experiments

The particular empirical example analyzed here is from Mishkin (1990a and 1990b), which examines the ability of the term structure of interest rates to forecast changes in future inflation rates. Specifically, the null hypothesis that $\beta_{m,n} = 0$ is tested in the following regression.

(1) $$\pi_t^m - \pi_t^n = \alpha_{m,n} + \beta_{m,n}[i_t^m - i_t^n] + \eta_t^{m,n}$$

where,

π_t^m = the realized inflation rate over the next m months, starting at time t.

i_t^m = the m-month nominal interest rate at time t.

Monte Carlo experiments to evaluate the small sample properties of the t-test on $\beta_{m,n}$ are conducted as follows. The data generating process for the inflation rates and the $i_t^m - i_t^l$ spread variables are obtained from ARMA models whose parameters are estimated from the full sample periods in Mishkin (1990a and b). Lagrange-multiplier tests described by Engle [1982] reveal the presence of ARCH (autoregressive conditional heteroscedasticity) in the error terms and so the error terms are drawn from a normal distribution in which the variance follows an ARCH process whose parameters are also estimated from the relevant sample periods. The Monte Carlo results reported here assume that the error terms for the inflation and $i_t^m - i_t^l$ equations are independent.[2] Start-up values for AR terms in the times series models are obtained from the actual realized data from five and six years before the sample period (or at the start of the sample period if earlier data were unavailable), and then five years of draws from the random number generator produce start-up values for the error terms. Thirty year (360 observations) and three year (36 observations) sample sizes were produced using errors drawn from the distribution described above.

To evaluate how well the Hansen-Hodrick-White methodology performs, it is worth first looking at Monte Carlo results for the t-test of $\beta_{m,n} = 0$ using ordinary least squares estimates (OLS). Table 1 reports the results from Monte Carlo simulations with one thousand replications. Panel A contains results for a thirty year sample, which is a sample size frequently encountered in empirical work. Panel B contains results for a three year

---

[2]Mankiw and Shapiro [1986] and Stambaugh [1986] point out that if the data generating mechanism displays correlation between the regressor and the regressand because of contemporaneous correlation between their error terms, the small sample distribution of test statistics can be substantially affected. This problem is not an important one in the data here because the correlations between error terms for the inflation and $i_t^m - i_t^n$ equations are not statistically significant. Monte Carlo simulations which allowed for this correlation as in Mankiw and Shapiro [1986] and Stambaugh [1986], as well as simulations allowing for additional effects of past inflation on yield spreads, were not appreciably different from those reported in the text. Note that the dating convention for interest rates in this paper is off by one period from the conventional dating used in Mankiw and Shapiro [1986] and Stambaugh [1986]. Hence allowing for contemporaneous correlation of the error terms from the π_t^m and i_t^m ARIMA models in the Monte Carlo simulations involves allowing for the correlation between the i_t^m error term and one lag of the π_t^m error term in the notation of this paper.

Table 1

Monte Carlo Simulation Results
for t-test of $\beta_{m,n} = 0$
Using Ordinary Least Squares Estimates

m,n (months)	Critical values of t from Monte Carlos — Significance levels 50%	25%	10%	5%	1%	% Rejections using standard 5% critical value	% Rejections using standard 1% critical value
Panel A: 30 Year Sample (360 observations)							
3,1	0.65	1.05	1.56	1.92	2.56	4.4%	0.9%
6,3	0.83	1.38	2.07	2.49	3.36	11.8%	4.4%
9,6	0.92	1.62	2.37	2.70	3.52	17.7%	6.9%
12,6	1.04	1.75	2.70	3.25	4.12	20.1%	11.1%
24,12	2.39	4.00	5.57	6.52	8.70	57.4%	46.6%
36,12	2.72	4.64	6.45	7.78	10.51	62.7%	52.1%
48,12	2.85	4.89	7.11	8.46	11.24	63.3%	53.7%
60,12	2.59	4.71	6.77	7.94	10.53	59.9%	50.1%
Panel B: 3 Year Sample (36 observations)							
3,1	0.67	1.18	1.69	2.15	2.85	6.6%	1.8%
6,3	0.94	1.56	2.23	2.78	3.84	15.3%	6.3%
9,6	1.01	1.78	2.58	3.12	4.13	21.2%	10.0%
12,6	1.22	2.07	3.03	3.60	5.04	27.7%	15.8%
24,12	2.04	3.68	5.41	6.71	9.72	52.3%	39.3%
36,12	2.11	3.71	5.76	6.99	10.90	53.8%	41.5%
48,12	2.53	4.33	6.18	7.62	10.70	58.6%	49.3%
60,12	2.32	4.03	5.96	7.60	10.70	57.2%	44.2%

sample, corresponding to the sample length of the November 1979 to October 1982 period of a new Federal Reserve operating procedure, which has also frequently been studied in the literature.

As we can see in Panel A, not surprisingly, use of ordinary least squares produces quite misleading inference, even though the number of observations in the sample is over three hundred. The last two columns indicate how often the null hypothesis is rejected using a standard 5% critical t-value of 1.96 or a standard 1% critical t-value of 2.58, while the second through sixth columns provide more information on the finite sample distributions by reporting critical values for different marginal significance levels. As m and hence the degree of overlap increases, the bias of the OLS t-tests becomes progressively worse. By the time we reach m = 36 months, over 50% of the t-statistics indicate rejection of the null hypothesis when it is true using either standard 5% or 1% critical values. Furthermore, the appropriate 5% critical value of the finite sample distribution is around 8.0 rather than 2.0, while the 1% critical value is over 10.0. Panel B which contains results for the shorter three-year sample yields similar results.

Table 2 contains the Monte Carlo results for one-thousand replications of the t-test of $\beta_{m,n} = 0$ using the Hansen-Hodrick-Newey-West procedure in which there is no correction for heteroscedasticity.[3] Here there is a substantial improvement in the performance of the t-tests. For the Panel A, thirty-year sample, the t-tests when m = 12 months or less lead to fairly reliable inference. The percentage of rejections under the null are close to what would be expected using standard critical values, while the 5% critical value from the finite sample distribution is close to 2.0 and the 1% critical value is close to 2.5. When m = 24 months or greater, there is some deterioration in the

[3]The Hansen-Hodrick-Newey-West procedure for the t-tests in Tables 2 and 3 assumes that the error terms of the regressions have a MA process of order m-1 because of the overlap in the data: i.e., the procedure has linearly declining weights on the first m-1 autocovariance matrices with the autocovariance matrices at lag m and above set to zero.

Table 2

Monte Carlo Simulation Results
for t-test of $\beta_{m,n} = 0$
Using Hansen-Hodrick-Newey-West Without Correcting for Heteroscedasticity

m,n (months)	Critical values of $\underline{t}$ from Monte Carlos: Significance levels 50%	25%	10%	5%	1%	% Rejections using standard 5% critical value	% Rejections using standard 1% critical value
	Panel A: 30 Year Sample (360 observations)						
3,1	0.60	1.04	1.51	1.77	2.36	3.3%	0.5%
6,3	0.66	1.13	1.59	1.85	2.51	3.8%	0.9%
9,6	0.72	1.21	1.71	2.03	2.58	6.1%	1.0%
12,6	0.66	1.18	1.74	2.04	2.74	6.5%	1.6%
24,12	0.82	1.42	1.97	2.36	3.31	10.4%	3.3%
36,12	0.82	1.43	2.02	2.42	3.19	11.2%	3.7%
48,12	0.83	1.48	2.13	2.50	3.27	12.4%	4.1%
60,12	0.87	1.41	2.02	2.38	3.38	10.9%	3.3%
	Panel B: 3 Year Sample (36 observations)						
3,1	0.67	1.12	1.65	1.97	2.79	5.2%	1.4%
6,3	0.72	1.29	1.86	2.23	2.96	7.7%	1.9%
9,6	0.72	1.26	1.96	2.38	3.28	9.9%	3.4%
12,6	0.94	1.51	2.11	2.52	3.58	12.7%	4.8%
24,12	1.16	1.98	3.16	4.20	6.39	25.1%	16.0%
36,12	1.32	2.31	3.60	4.43	7.57	32.0%	19.3%
48,12	1.36	2.34	3.70	4.76	7.49	34.1%	20.6%
60,12	1.39	2.29	3.50	4.69	7.16	33.5%	20.5%

performance of the t-tests, but the performance is still reasonable. The percentage of rejections using the standard 5% critical value is around 10%, while it is around 3 to 4% using the standard 1% critical value. The appropriate 5% critical values from the finite sample distribution are now closer to 2.5 than to 2.0, while the appropriate 1% critical value is around 3.3.

The Panel B results for the shorter sample period indicate a much poorer performance of the t-statistics. The percentage of rejections under the null using the standard 5% critical value now rises above 30%, and the appropriate 5% critical values for the t-tests rises above 4.0. These results indicate that a high degree of data overlap relative to the length of the sample period leads to substantial differences between the finite and asymptotic sample distributions.

Since the data generating process for the variables in the regression equation do exhibit conditional heteroscedasticity, we might expect that the t-tests which correct for heteroscedasticity (and thus have the appropriate asymptotic distribution) would perform the best of all. Table 3, which reports the results for the one-thousand replications of the t-tests using the Hansen-Hodrick-Newey-West procedure correcting for heteroscedasticity with the method of White (1980) and Hansen (1982), indicates, however, that this is not the case. The t-tests correcting for heteroscedasticity have much worse properties than do those which do not correct for heteroscedasticity. A comparison of Tables 2 and 3 reveals that the percentage of rejections of the null hypothesis using standard critical values are often more than twice as high when the t-statistics are corrected for heteroscedasticity versus when they are not.[4] For the thirty-year sample, the percentage of rejections using the standard 5% critical value is as high as 20% when $\underline{m}$ = 60 months and is over

[4]This finding that correcting for heteroscedasticity increases the percentage of rejections is also found in Froot and Klemperer (1989) who report that the standard errors of coefficient corrected for heteroscedasticity are smaller than those which are not.

Table 3

Monte Carlo Simulation Results
for t-test of $\beta_{m,n} = 0$
Using Hansen-Hodrick-Newey-West Correcting for Heteroscedasticity

m,n (months)	Critical values of t from Monte Carlos — Significance levels 50%	25%	10%	5%	1%	% Rejections using standard 5% critical value	% Rejections using standard 1% critical value
			Panel A: 30 Year Sample (360 observations)				
3,1	0.59	1.02	1.50	1.85	2.37	3.7%	0.6%
6,3	0.68	1.17	1.68	2.03	2.63	6.0%	1.2%
9,6	0.73	1.29	1.79	2.07	3.03	7.3%	2.1%
12,6	0.73	1.26	1.78	2.14	3.05	7.0%	2.6%
24,12	0.89	1.64	2.41	2.79	4.02	17.1%	6.8%
36,12	0.91	1.61	2.56	3.09	4.33	17.6%	9.8%
48,12	0.93	1.64	2.54	3.17	4.74	17.5%	9.3%
60,12	0.96	1.75	2.63	3.34	5.29	20.0%	10.4%
			Panel B: 3 Year Sample (36 observations)				
3,1	0.67	1.19	1.79	2.24	3.08	7.7%	2.9%
6,3	0.89	1.58	2.35	2.95	4.19	17.0%	7.8%
9,6	1.01	1.78	2.69	3.53	5.55	21.6%	11.1%
12,6	1.11	2.06	3.26	4.35	7.08	27.1%	17.0%
24,12	1.99	3.80	6.65	8.53	13.91	50.8%	39.7%
36,12	2.77	5.46	8.76	11.44	20.19	63.5%	52.6%
48,12	2.74	5.20	8.64	10.81	18.61	63.7%	52.7%
60,12	2.93	5.31	9.06	11.49	21.31	63.6%	55.4%

10% using a standard 1% critical value. Critical values for the finite sample distribution are now above 3.0 for the 5% level and above 4.0 for the 1% level in Panel A. In the shorter sample, Panel B, the results are even worse, with the percentage of rejections using either the standard 5% or 1% critical values often above fifty percent, and the appropriate 5% and 1% critical values are sometimes above 10.0 and 20.0 respectively.

III. Conclusions

The conclusion from the Monte Carlo experiments is that correcting for the serial correlation induced by overlapping data is important to achieving correct statistical inference. However, surprisingly, correcting for heteroscedasticity does not help produce better statistical inference. The results thus suggest that correcting for heteroscedasticity when analyzing financial data may not always be the best research strategy. Furthermore, they indicate the importance of conducting Monte Carlo simulations in order to obtain correct statistical inference when the degree of data overlap is high relative to the number of observations in the sample period, especially when there is a correction for heteroscedasticity.

References

Campbell, John Y. and Richard Clarida, "The Term Structure of Euromarket Interest Rates: An Empirical Investigation," Journal of Monetary Economics 19 (January 1987): 25-44.

Cumby, Robert J. and John Huizinga, "Testing the Autocorrelation Structure of Disturbances in Ordinary Least Squares and Instrumental Variables Regressions," mimeo. January 1989, forthcoming Econometrica.

---------------------------------- and Maurice Obstfeld, "Two Step, Two Stage Least Squares Estimation in Models with Rational Expectations, Journal of Econometrics 21 (1983): 333-355.

Engle, Robert F., "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation," Econometrica 50 (July 1982): 987-1007.

Froot, Kenneth, "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates," Journal of Finance 64 (June 1989): 283-305.

Froot, Kenneth, and Klemperer, Paul D., "Exchange Rate Pass-Through When Market Share Matters," American Economic Review 79 (September 1989): 637-54.

Hansen, Lars P., "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica L (1982), 1029-54

Hansen, Lars, and Hodrick, Robert, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates," Journal of Political Economy 88 (1980): 829-53.

Hardouvelis, Gikas A., "The Predictive Power of the Term Structure During Recent Monetary Regimes," Journal of Finance 43 (June 1988): 339-56.

Mankiw, N. Gregory and Matthew D. Shapiro, "Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models," Economics Letters XX (1986), 139-45.

Frederic S. Mishkin, "The Information in the Term Structure: Some Further Results," Journal of

Applied Econometrics 3 (October-December 1988): 307-14.

------------------, "What Does the Term Structure Tell Us About Future Inflation," Journal of Monetary Economics (January 1990a).

------------------, "The Information in the Longer-Maturity Term Structure About Future Inflation," Quarterly Journal of Economics (1990b).

Newey, W. and West Kenneth, "A Simple, Positive Definite, Heteroskedasdicity and Autocorrelation Consistent Covariance Matrix," Econometrica LIII (1987), 703-708.

Stambaugh, Robert F., "Bias in Regressions with Lagged Stochastic Regressors," Graduate School of Business, University of Chicago, (January 1986).

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and Direct tests for Heteroskedasticity," Econometrica 48 (1980): 817-38.