

NONPARAMETRIC PREDICTION OF ASSET RETURNS: FURTHER NEGATIVE RESULTS

F. X. Diebold, University of Pennsylvania, and R. M. Sella, Federal Reserve Board
Francis X. Diebold, 3718 Locust Walk, Philadelphia, PA 19104-6297

KEYWORDS: Locally weighted regression, nearest neighbor, forecasting, nonlinearity

1. Introduction

It is widely agreed that a variety of high-frequency asset returns are well described as linearly unpredictable, conditionally heteroskedastic, and unconditionally leptokurtic. (See for example Diebold (1988), Diebold and Nerlove (1988a), and Pagan and Schwert (1989).) Documentation of linear unpredictability may be traced at least to early work on efficient markets, such as Cootner (1963) and Fama (1964); similarly, leptokurtosis has been appreciated at least since Mandelbrot (1963). The early writers were also aware of the apparent occurrence of volatility clustering in asset returns, and the work of Engle (1982) on ARCH processes provided a tool for its formal study. It is now agreed that many time series of asset returns, while approximately uncorrelated, are not temporally independent; dependence arises through persistence in the conditional variance and perhaps in other conditional moments.

If the above characterization of asset price dynamics (linear conditional mean with nonlinearities working through the conditional variance) is correct, then the nonlinearities cannot be exploited to generate improved point predictions relative to linear models. It is not clear, however, that the ARCH effects are structural, i.e., that they are a characteristic of the true data-generating process (DGP). Instead, ARCH may indicate misspecification, serving as a proxy for neglected nonlinearities in the conditional mean. For an illustration of the difficulties involved in separating conditional-mean from conditional-variance dynamics, see Weiss (1986), who discusses ARCH and bilinearity. A finding of significant conditional-mean nonlinearity would be important for both theoretical and empirical work: theoretically, substantial challenge to our understanding of asset-price dynamics would be posed, and empirically, a source of improved point prediction relative to linear models would be provided.

Interestingly, recent empirical and theoretical results are consistent with the conjecture that nonlinearities may be present in asset-return conditional means. The empirical results may be categorized into two groups: (1) those using ideas from the theory of stochastic nonlinear time-series, and (2) those using ideas from the theory of deterministic chaotic systems. In the nonlinear time series area, a number of studies, including Domowitz and Hakkio (1985), Hinich and Patterson (1985, 1987), Weiss (1986), Engle, Lillien, and Robins (1987), Diebold and Pauly (1988), and others, appear to detect statistically significant nonlinearity in conditional means of various asset prices and other economic aggregates. Recent work on regime switching, including Flood and Garber (1983), Engel and Hamilton (1988) and Froot and Obstfeld (1989) is also squarely in the nonlinear tradition. Similar results have been obtained in the chaos literature, using tests based on estimated Lyapunov exponents and correlation dimensions, as developed in Brock, Dechert and Scheinkman (1987), *inter alia*. Scheinkman and LeBaron (1989), for example, find strong evidence of nonlinearity in common stock returns and

suggest that it could be exploited for improved point prediction. Similarly, Gallant, Hsieh and Tauchen (1988) and Hsieh (1989) report evidence of residual nonlinearity in exchange rates, after controlling for conditional heteroskedasticity. These empirical results are provocative, because they challenge us to take seriously the possible existence of nonlinear conditional-mean dynamics in asset prices.

In summary, there appears to be strong evidence, consistent with rigorous economic theory, that important nonlinearities may be operative in asset price determination. Upon further consideration, however, it becomes clear that the literature is not in satisfactory condition, due to a puzzle that immediately arises: Why is it that statistically significant rejections of linearity in asset returns routinely occur, while no nonlinear model has been found that can significantly outperform even the simplest linear model in out-of-sample forecasting? Because a number of factors may be operative, a number of explanations may be offered. One, of course, is that the nonlinearities present may be in even-ordered conditional moments, and therefore are not useful for point prediction. Second, in-sample nonlinearities such as outliers and structural shifts may be present, and may cause various linearity tests to reject, while nevertheless being of no use for out-of-sample forecasting. Third, very slight conditional-mean nonlinearities might be truly present and be detectable with large datasets, while nevertheless yielding negligible *ex ante* forecast improvement. In other words, *significance* of nonlinearity does not necessarily imply its *importance*. Finally, even if conditional-mean nonlinearities are present and *are* important, the overwhelming variety of plausible candidate nonlinear models makes determination of a good approximation to the DGP a difficult task. The seemingly large variety of parametric nonlinear models that have received attention lately (e.g., bilinear, threshold, exponential autoregressive, etc.) is in fact a very small subset of the class of plausible nonlinear DGP's.

In this paper, we contribute to a resolution of this puzzling behavior of asset returns by estimating nonparametrically the conditional-mean function of a sample of daily IBM stock returns. By so doing, we avoid the parametric model-selection problem, thereby expanding greatly the class of potential models. In section 2, we discuss various aspects of nonparametric functional estimation, and we highlight the locally weighted regression (LWR) procedure, which we use extensively. Section 3 contains empirical results; in particular, both in-sample LWR fits and out-of-sample LWR forecasts are compared to those arising from linear models. Section 4 concludes with a comparison to existing results.

2. Nonparametric Prediction

Nonparametric techniques may be used for estimation of a variety of densities and econometric functionals, including regression functions, first and higher-order derivatives of regression functions, conditional-variance functions, hazard and survival functions, etc. For surveys of various aspects of nonparametric and semiparametric estimation, see Ullah (1988) and Robinson

(1988). We shall generally be concerned with nonparametric estimation of conditional expectation, or regression, functions,

$$E(y|x) = \int yf(y|x)dy = \int y[f(y,x)/f(x)]dy,$$

which we use for nonparametric prediction. Because the meaning is obvious from context, we use lower-case letters for both random variables and their realizations, and we use "f" to denote all probability density functions. This is achieved by (explicit or implicit) substitution of nonparametric estimates of the underlying joint and marginal densities into the above expression. In our dynamic models, the stochastic conditioning vector x is composed of lagged dependent variables,

$$x_t = \{y_{t-1}, \dots, y_{t-p}\}.$$

We shall work with the very general nonlinear autoregressive structure,

$$y_t = g(y_{t-1}, \dots, y_{t-p}) + \varepsilon_t$$

$$E(\varepsilon_t | y_{t-1}, \dots, y_{t-p}) = 0,$$

$t = 1, \dots, T$, so that

$$\begin{aligned} E(y_{t+1} | y_t, \dots, y_{t-p+1}) \\ &= \int y_{t+1} f(y_{t+1} | y_t, \dots, y_{t-p+1}) dy_{t+1} \\ &= \int y_{t+1} [f(y_{t+1}, y_t, \dots, y_{t-p+1}) \\ &\quad / f(y_t, \dots, y_{t-p+1})] dy_{t+1}. \end{aligned}$$

The regression function estimates may be obtained by a variety of interrelated nonparametric methods, including kernel, series, and nearest-neighbor (NN) techniques, consistency results for which have been obtained in time-series environments by Robinson (1983), Gallant and Nychka (1987), and Yakowitz (1987), respectively, *inter alia*.

In this study we make use of a generalized NN technique, known as locally-weighted regression (LWR). NN methods proceed by estimating $g(x)$, at an arbitrary point $x = x^*$ in p -dimensional Euclidean space, via a weight function,

$$\hat{g}(x^*) = \sum_{t=1}^T w_{k_T}(x_t) y_t$$

where $w_{k_T}(x_t) = 1/k_T$ if x_t is one of the k_T nearest neighbors of x^* ,

and $w_{k_T}(x_t) = 0$ otherwise. The subscript "T" of k_T serves as a

reminder that the number of nearest neighbors used should depend on sample size, as discussed subsequently. The LWR estimator, as proposed by Cleveland (1979) and refined by Cleveland and Devlin (1988) and Cleveland *et al.* (1988), is an important generalization of the NN estimator. Like a NN estimator, LWR fits the surface at a point x^* as a function of the y values corresponding to the k_T nearest neighbors of x^* . Unlike

NN, however, LWR does not take $\hat{g}(x^*)$ as a simple average of those y values; rather, $\hat{g}(x^*)$ is the fitted value from a regression surface. This corresponds to a simple average only in the very unlikely case that the constant term is the sole regressor with explanatory power.

We now discuss the procedure in some detail. We compute the LWR estimate of the surface at a point x^* , $\hat{g}(x^*)$, as follows.

Let ξ be a smoothing constant such that $0 < \xi \leq 1$, and let $k_T = \text{int}(\xi \cdot T)$, where $\text{int}(\cdot)$ rounds down to the nearest integer. Then rank the x_t^* 's by Euclidean distance from x^* ; call these $x_1^*, x_2^*, \dots, x_{k_T}^*$.

Thus, x_1^* is closest to x^* , x_2^* is second closest to x^* , and so on. Let

$\lambda(a, b)$ measure Euclidean distance; then $\lambda(x^*, x_{k_T}^*)$ is the

Euclidean distance from x^* to its k_T th closest neighbor,

$$\lambda(x^*, x_{k_T}^*) = \left[\sum_{j=1}^p (x_{k_T,j}^* - x_j^*)^2 \right]^{1/2}.$$

Form the neighborhood weight function,

$$v_t(x_t^*, x^*, x_{k_T}^*) = C[\lambda(x_t^*, x^*) / \lambda(x^*, x_{k_T}^*)],$$

where $C(\cdot)$ is the tricube function,

$$C(u) = \begin{cases} (1-u)^3 & \text{for } u < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The value of the regression surface at x^* is then computed as

$$\hat{y}^* = \hat{g}(x^*) = x^* \hat{\beta},$$

where

$$\hat{\beta} = \underset{t=1}{\text{argmin}} \left[\sum_{t=1}^T v_t(y_t - x_t^* \beta)^2 \right].$$

The LWR procedure, exactly as described above, is used in our subsequent empirical work. Obviously, it reflects a number of judgmental decisions, such as use of Euclidean norm and tricube neighborhood weighting, as well as locally linear (as opposed to higher-order, such as quadratic) fitting. The Euclidean norm has obvious geometric appeal, as does the tricube weight function, which produces a smooth, gradual decline in weight with distance from x^* . Locally linear fitting is also highly reasonable (and computationally feasible) in the present context. See Cleveland *et al.* (1988) for further discussion.

Of greater interest is the choice of ξ , which determines the number of nearest neighbors used, and hence the degree of smoothing. Consistency of NN estimators (and hence LWR) requires that the number of nearest neighbors used go to infinity with sample size, but at a slower rate, i.e.,

$$\lim_{T \rightarrow \infty} k_T = \infty$$

$$\lim_{T \rightarrow \infty} (k_T/T) = 0.$$

This implicitly creates a "window" whose width becomes smaller as sample size goes to infinity, but at a slower rate. In this way the shrinking window nevertheless contains progressively more neighbors, so that bias is reduced along with variance. Similar issues arise in kernel and series estimation. In the kernel case, the window width corresponds to bandwidth, which must shrink with sample size but at a slower rate. Finally, in series estimation, the window width corresponds (inversely) to the number of included series terms; again, consistency requires that the truncation point increase with sample size, but at a slower rate.

It is interesting to note that earlier-discussed LWR rule of

$\xi \cdot T$ for selecting the number of nearest neighbors does not satisfy the second regularity condition for consistency. In any finite sample such as ours, however, this is of no consequence, since there exists a selection rule such as T^α , $\alpha < 1$, which does satisfy the regularity condition and results in use of an identical number of nearest neighbors. For example, in a sample of size 800, the $\xi \cdot T$ rule with $\xi = .5$ selects approximately 400 nearest neighbors, as does the T^α rule with $\alpha = .9$. Moreover, in the empirical work that follows, we explore a wide range of ξ values.

3. Empirical Analysis

The data are daily IBM stock returns, measured close-of-day, as reported in the University of Chicago's CRSP database, exactly as used in White (1988). We have 1517 observations, corresponding to the first business day of the second quarter of 1974 through the last business day of the first quarter of 1980. We examine both in-sample "fit" and out-of-sample predictive performance of the LWR nonparametric conditional mean estimator. Observations 8 through 1017 (74Q2-78Q1) are used for in-sample analysis, while observations 1018 through 1517 (78Q2-80Q1) are reserved for out-of-sample forecast comparison. Our out-of-sample forecasts are completely ex ante, using LWR estimates formed recursively in real time, using only information actually available.

The in-sample results appear in table 1. We perform a sensitivity analysis with respect to ξ , exploring a wide range of values from .10 through 10. Note that $\xi = 1$ does not correspond to a linear autoregression, because the observations are still weighted. Rather, our algorithm is such that as ξ approaches infinity, the linear autoregression emerges. (In practice, $\xi = 10$ produces an approximately linear autoregression.) It is apparent that, for all numbers of included lags, in-sample MSPE decreases with ξ and is minimized at the smallest ξ value of .1. The MSPE associated with the optimal ξ choice is always lower (by a very wide margin) than the random walk (RW) MSPE. Similarly mean absolute prediction error (MAPE) is also decreasing in ξ and is minimized at $\xi = .1$. Like MSPE, the MAPE associated with the optimal ξ choice is always smaller than the random walk MAPE. It is interesting to note the general tendency for both MSPE and MAPE to decrease and then level off as more lags ($p=3, p=5, p=7$) are included. Although this need not happen (different nearest neighbors are used, in general, for different p), it is intuitively reasonable by analogy to the fact that inclusion of additional regressors in an OLS regression must lower (or, at worst, leave unchanged) the sum of squared residuals.

We now turn to the out-of-sample analysis. Again, we estimate nonparametric autoregressions of order 3, 5 and 7 using the LWR procedure, with values of the smoothing parameter ξ ranging from 0.1 through 10.0 for each p , corresponding to use of roughly 125 nearest neighbors (with neighborhood weighting) through "all" nearest neighbors (with no neighborhood weighting). Forecast horizons explored are $K = 1, 4, 8$ and 12. For each ξ, p , and K combination, out-of-sample forecasts are computed by recursively re-estimating the relevant conditional mean in real time. This is continued until the sample is exhausted, resulting in a sequence of 500 ex ante forecasts for each p and K .

Out-of-sample results are contained in tables 2 (MSPE) and 3

(MAPE). A distinct pattern emerges: The random walk fares much better, indicating that the in-sample loss reduction may be the spurious result of overfitting. Out-of-sample loss reductions using LWR (with the best-performing ξ value) generally don't exist, and on the few occasions when they do, they are much smaller than those of the in-sample analysis. Moreover, loss is consistently minimized for large ξ values, which produce a linear autoregressive fit. These qualitative conclusions hold regardless of the choice of p and K . In truly ex ante forecasting, in which even the ξ value must be chosen by the investigator (based upon a combination of prior information and previous sample information), the scope for improved prediction appears extremely limited.

4. Summary and Concluding Remarks

Using a powerful nonparametric prediction technique, we find no evidence of nonlinearities exploitable for enhanced out-of-sample prediction of daily IBM stock returns. Our results corroborate those of White (1988), who obtains identical results using the same asset returns but a different prediction technique (neural networks), Diebold and Nason (1989), who use different asset returns (exchange rates) but the same prediction technique, Prescott and Stengos (1988), who use different asset returns (gold) and a different prediction technique (kernel estimators), and Meese and Rose (1989) who use the same technique in a structural exchange rate environment. Taken together, these results constitute fairly strong evidence against the existence of asset price nonlinearities that are exploitable for improved point prediction.

The research could of course be extended in a number of directions. The analysis could be made completely ex ante by choosing ξ in real time by cross validation, and multivariate generalizations might be undertaken. Computational considerations render some of these extensions infeasible at the present time. More importantly, however, such extensions would appear highly likely to reinforce, rather than alter, the basic tenor of our conclusions.

References

- Brock, W.A., Dechert, W.D. and Scheinkman, J.A. (1987), "A Test for Independence Based on the Correlation Dimension," SSRJ Working Paper #8702, Department of Economics, University of Wisconsin, Madison.
- Cleveland, W.S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829-836.
- Cleveland, W.S. and Devlin, S.J. (1988), "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, 83, 596-610.
- Cleveland, W.S., Devlin, S.J. and Grosse, E. (1988), "Regression by Local Fitting: Methods, Properties and Computational Algorithms," *Journal of Econometrics*, 37, 87-114.
- Cootner, P. (ed.) (1964), *The Random Character of Stock Market Prices*, Cambridge, Mass.: MIT Press.
- Diebold, F.X. (1988), *Empirical Modeling of Exchange Rate Dynamics*. New York: Springer-Verlag.

- Diebold, F.X. and Nason, J.M. (1989), "Nonparametric Exchange Rate Prediction?," *Journal of International Economics*, forthcoming.
- Diebold, F.X. and Nerlove, M. (1989a), "The Dynamics of Exchange Rate Volatility: A Multivariate Latent-Factor ARCH Model," *Journal of Applied Econometrics*, 4, 1-22.
- Diebold, F.X. and Pauly, P. (1988), "Endogenous Risk in a Rational-Expectations Portfolio-Balance Model of the Deutschmark / Dollar Rate," *European Economic Review*, 32 (1), 27-54.
- Domowitz, I. and Hakkio, C.S. (1985), "Conditional Variance and the Risk Premium in the Foreign Exchange Market," *Journal of International Economics*, 19, 47-66.
- Engel, C. and Hamilton, J.D. (1988), "Long Swings in Exchange Rates: Are They in the Data and do Markets Know it?," Manuscript, Department of Economics, University of Virginia.
- Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.
- Engle, R.F., Lillien, D.M. and Robbins, R.P. (1987), "Estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391-408.
- Fama, E.F. (1965), "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34-105.
- Flood, R. and Garber, P. (1983), "A Model of Stochastic Process Switching," *Econometrica*, 51, 537-564.
- Froot, K. A. and Obstfeld, M. (1989), "Exchange Rate Dynamics Under Stochastic Regime Shifts: A Unified Approach," NBER Working Paper #2835.
- Gallant, A.R., Hsieh, D. and Tauchen, G. (1988), "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974-1983," Manuscript, Graduate School of Business, University of Chicago.
- Gallant, A.R. and Nychka, D.W. (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-390.
- Hinich, M. and Patterson, D. (1985), "Evidence of Nonlinearity in Daily Stock Returns," *Journal of Business and Economic Statistics*, 3, 69-77.
- Hinich, M. and Patterson, D. (1987), "Evidence of Nonlinearity in the Trade-by-Trade Stock Market Return Generating Process," Department of Government, University of Texas at Austin.
- Hsieh, D.A. (1989), "Testing for Nonlinear Dependence in Foreign Exchange Rates: 1974-1983," Manuscript, Graduate School of Business, University of Chicago, (revised). Forthcoming, *Journal of Business*.
- Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394-419.
- Meese, R.A. and Rose, A.K. (1989), "An Empirical Assessment of Nonlinearities in Models of Exchange Rate Determination," Manuscript, School of Business, University of California, Berkeley.
- Pagan, A.R. and Schwert, G.W. (1989), "Alternative Models for Conditional Stock Volatility," NBER Working Paper #2955.
- Prescott, D.M. and Stengos, T. (1988), "Do Asset Markets Overlook Exploitable Nonlinearities? The Case of Gold", Manuscript, Department of Economics, University of Guelph.
- Robinson, P.M. (1983), "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis*, 4, 185-207.
- Robinson, P.M. (1988), "Semiparametric Econometrics: A Survey," *Journal of Applied Econometrics*, 3, 35-51.
- Scheinkman, J.A. and LeBarron, B. (1989), "Nonlinear Dynamics and Stock Returns," Manuscript, Graduate School of Business, University of Chicago (revised). Forthcoming, *Journal of Business*.
- Ullah, A. (1988), "Nonparametric Estimation of Econometric Functionals," *Canadian Journal of Economics*, 21, 625-658.
- Weiss, A.A. (1986), "ARCH and Bilinear Time Series Models: Comparison and Combination," *Journal of Business and Economic Statistics*, 4, 59-70.
- White, H. (1988), "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns," Working Paper #88-20, UCSD.
- Yakowitz, S.J. (1987), "Nearest Neighbor Methods for Time Series Analysis," *Journal of Time Series Analysis*, 8, 235-247.

