Journal of International Economics 28 (1990) 315–332. North-Holland

NONPARAMETRIC EXCHANGE RATE PREDICTION?

Francis X. DIEBOLD*

University of Pennsylvania, Philadelphia, PA 19104-6297, USA

James A. NASON

University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada

Received June 1989, revised version received August 1989

Conditional heteroskedasticity is frequently found in the prediction errors of linear exchange rate models. It is not clear whether such conditional heteroskedasticity is a characteristic of the true data-generating process, or whether it indicates misspecification associated with linear conditional-mean representations. We address this issue by estimating nonparametrically the conditional-mean functions of ten major nominal dollar spot rates, 1973–1987, which are used to produce in-sample and out-of-sample nonparametric forecasts. Our findings bode poorly for recent conjectures that exchange rates contain nonlinearities exploitable for enhanced point prediction.

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. Documentation of linear unpredictability may be traced at least to early work on efficient markets, such as Cootner (1964) and Fama (1965); 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) 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.

*Diebold gratefully acknowledges financial support from the National Science Foundation and the University of Pennsylvania Research Foundation. An earlier version of this paper was presented at the 1988 North American Winter meetings of the Econometric Society, New York. We thank the participants there, as well as an anonymous referee, Buz Brock, Dee Dechert, Blake LeBaron, Andy Rose, and seminar participants at Boston College, Carnegie-Mellon, Wisconsin, Washington, LSU, and Stonybrook for useful comments. We especially thank William S. Cleveland, who generously provided his software, and Roberto Sella, who provided outstanding research assistance.

All of these results are manifest in exchange rates, which are the focus of the present study. The well-known work of Meese and Rogoff (1983a, b) provides graphic illustration of the failure of a variety of economic models to outperform a simple random walk in out-of-sample prediction.[1] Extensive reviews of linear unpredictability and leptokurtosis in exchange rates are contained in Westerfield (1977), Boothe and Glassman (1987), and Diebold and Nerlove (1989a). Conditional heteroskedasticity, in the form of ARCH and related effects, has also been repeatedly documented in exchange rates. Diebold (1988), for example, examines seven nominal dollar spot rates and finds little linear predictability but strong ARCH effects in all of them.

A number of explanations have been advanced for the leptokurtosis and volatility clustering, as well as reduction of excess kurtosis under temporal aggregation. The phenomena may be jointly explained, for example, by subordinating exchange rates to a process dictating information arrival. The seminal work of Clark (1973) shows that subordination to an i.i.d. information-arrival process produces leptokurtic returns. As mentioned by Diebold (1986) and formalized by Gallant, Hsieh and Tauchen (1988), Clark's analysis may be extended by allowing for persistence in the information-arrival process, which produces ARCH-like movements in higher-order conditional moments. Finally, Diebold (1988) uses central limit theorems for dependent observations to show that, under quite general conditions (not requiring, for example, existence of the fourth unconditional moment), the unconditionally fat-tailed behavior associated with stationary ARCH effects diminishes with temporal aggregation.

If the above characterization of exchange rate 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.[2] A finding of significant conditional-mean nonlinearity would be important for both theoretical and empirical work: theoretically, a 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

[1]Their candidate models include a flexible price monetary model, a sticky price monetary model, a sticky price monetary model with current account effects, six univariate time series models, a vector autoregressive model, and the forward rate.

[2]For an illustration of the difficulties involved in separating conditional-mean from conditional-variance dynamics, see Weiss (1986), who discusses ARCH and bilinearity.

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 Lyaponov 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.

A number of recent theoretical developments are beginning, implicitly or explicitly, to address this challenge. Sims (1984), for example, shows that general-equilibrium asset-pricing models imply martingale asset-price behavior *only* at arbitrarily short horizons. Thus, economic theory cannot rule out the possibility of nonlinear dependence in conditional means (as well as in higher-order conditional moments) of asset returns. Moreover, substantial progress has been made in general equilibrium models with time-varying risk, such as Abel (1988), Hodrick (1987), Baldwin and Lyons (1988), and Nason (1988).

In summary, there appears to be strong evidence, consistent with rigorous economic theory, that important nonlinearities may be operative in exchange rate determination. Upon further consideration, however, it becomes clear that the literature is not in satisfactory condition, owing to a puzzle that immediately arises: Why is it that while statistically significant rejections of linearity in exchange rates routinely occur, 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.[3] 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, and exponential autoregressive) is in fact a very small subset of the class of plausible nonlinear DGPs.

In this paper we contribute to a resolution of this puzzling behavior of exchange rates by estimating conditional-mean functions nonparametrically. By so doing, we avoid the parametric model-selection problem; the class of potential models entertained is expanded greatly. In section 2 we discuss various aspects of nonparametric functional estimation, and the locally weighted regression procedure, which we use extensively, is highlighted. 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.

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.[4] We shall generally be concerned with nonparametric estimation of conditional expectation, or regression, functions,

$$\mathrm{E}(y|x)=\int y f(y|x)\,\mathrm{d}y=\int y[f(y,x)/f(x)]\,\mathrm{d}y, \tag{1}$$

which we use for nonparametric prediction.[5] This is achieved by (explicit or implicit) substitution of nonparametric estimates of the underlying joint and marginal densities into (1). In our dynamic models, the stochastic conditioning vector x is composed of lagged dependent variables:

$$x_t=\{y_{t-1},\ldots,y_{t-p}\}. \tag{2}$$

[3]In other words, *significance* of nonlinearity does not necessarily imply its *economic importance*.

[4]For surveys of various aspects of nonparametric and semiparametric estimation, see Ullah (1988) and Robinson (1988).

[5]Because the meaning is obvious from the context, we use lower-case letters for both random variables and their realizations, and we use f to denote all probability density functions.

We shall work with the very general nonlinear autoregressive structure:

$$y_t = g(y_{t-1}, \ldots, y_{t-p}) + \varepsilon_t,$$
$$\mathrm{E}(\varepsilon_t | y_{t-1}, \ldots, y_{t-p}) = 0, \tag{3}$$

$t = 1, \ldots, T$, so that (1) may be rewritten more specifically as:

$$\mathrm{E}(y_{t+1} | y_t, \ldots, y_{t-p+1}) = \int y_{t+1} f(y_{t+1} | y_t, \ldots, y_{t-p+1}) \, \mathrm{d}y_{t+1}$$
$$= \int y_{t+1} [f(y_{t+1}, y_t, \ldots, y_{t-p+1}) / f(y_t, \ldots, y_{t-p+1})] \, \mathrm{d}y_{t+1}. \tag{4}$$

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.

In this study we make use of a 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:

$$\hat{g}(x^*) = \sum_{t=1}^{T} w_{k_T}(x_t) y_t, \tag{5}$$

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.[6] 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 $0 < \xi \leqq 1$, and let $k_T = \mathrm{int}(\xi \cdot T)$, where $\mathrm{int}(\cdot)$ rounds down to the nearest integer. Then rank the x_t's by Euclidean distance from x^*; call

[6]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.

these $x_1^*, x_2^*, \ldots, x_T^*$. 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_Tth closest neighbor:

$$\lambda(x^*, x_{k_T}^*) = \left[\sum_{j=1}^{p} (x_{k_T j}^* - x_j^*)^2 \right]^{1/2}. \tag{6}$$

Form the neighborhood weight function:

$$v_t(x_t, x^*, x_{k_T}^*) = C[\lambda(x_t, x^*)/\lambda(x^*, x_{k_T}^*)], \tag{7}$$

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

$$C(u) = \begin{cases} (1-u^3)^3, & \text{for } u < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

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

$$\hat{y}^* = \hat{g}(x^*) = x^{*\prime}\hat{\beta}, \tag{9}$$

where

$$\hat{\beta} = \operatorname{argmin} \left[\sum_{t=1}^{T} v_t (y_t - x_t'\beta)^2 \right]. \tag{10}$$

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 the 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.[7]

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.

$$\lim_{T\to\infty} k_T = \infty, \qquad \lim_{T\to\infty} (k_T/T) = 0. \tag{11}$$

[7] See Cleveland, Devlin and Grosse (1988) for further discussion.