Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models

Yongmiao Hong

Department of Economics &
Department of Statistical Science
Cornell University
Ithaca, NY 14853-7601
Email: yh20@cornell.edu

and

Tae-Hwy Lee

Department of Economics
University of California, Riverside
Riverside, CA 92521-0427
Email: tae.lee@ucr.edu

Revised, July 2002

ABSTRACT

It is often documented, based on autocorrelation, variance ratio and power spectrum, that exchange rates approximately follow a martingale process. Because autocorrelation, variance ratio and spectrum check serial uncorrelatedness rather than martingale difference, they may deliver misleading conclusions in favor of the martingale hypothesis when the test statistics are insignificant. In this paper, we explore whether there exists a gap between serial uncorrelatedness and martingale difference for exchange rate changes, and if so, whether nonlinear time series models admissible in the gap can outperform the martingale model in out-of-sample forecasts. Applying the generalized spectral tests of Hong (1999) to five major currencies, we find that the changes of exchange rates are often serially uncorrelated, but there exists strong nonlinearity in conditional mean, in addition to the well-known volatility clustering. To forecast the conditional mean, we consider the linear autoregressive, autoregressive polynomial, artificial neural network and functional-coefficient models, as well as their combination. The functional coefficient model allows the autoregressive coefficients to depend on investment positions via an moving average technical trading rule. We evaluate out-of-sample forecasts of these models relative to the martingale model, using four criteria— the mean squared forecast error, the mean absolute forecast error, the mean forecast trading return, and the mean correct forecast direction. White's (2000) reality check method is used to avoid data-snooping bias. It is found that suitable nonlinear models, particularly their combination, do have superior predictive ability over the martingale model for some currencies in terms of certain forecast evaluation criteria.

Key Words: bootstrap, combined forecast, data snooping, directional forecast, functional-coefficient model, generalized spectral test, martingale, moving average technical trading rule, nonlinearity in mean, trading return.

JEL Classification: C2, C5, F3

I. INTRODUCTION

Conventional wisdom holds that exchange rate changes approximately follow an martingale difference sequence (MDS) so that future changes are unpredictable using publicly available information.[1] This hypothesis has been tested using autocorrelation (Box-Pierce-Ljung's portmanteau tests), variance-ratio (Lo and MacKinlay 1988) and spectrum (Durlauf 1991), based on data of various frequencies and sample periods. Unlike the stock markets, where it has been well documented (cf. Lo and MacKinlay 1999 and references therein) that stock price changes are not an MDS, the statistical evidence supporting or refuting the martingale hypothesis for exchange rates seems mixed.[2] Many existing studies have incorporated the martingale hypothesis in modelling exchange rates, focusing on volatility forecasts (e.g., Bollerslev 1990, Brock, Hsieh and LeBaron 1991, Engle, Ito and Lin 1990, West and Cho 1995) or density forecasts (Diebold, Hahn and Tay 1999, Hsieh 1993).

From a nonlinear time series perspective, it is important to distinguish an MDS from a serially uncorrelated (or white noise, WN) process. There exists a nontrivial gap between an MDS and a WN. The former implies the latter, but not vice versa. A nonlinear time series can have zero autocorrelation but a non-zero mean conditional on its past history. Examples are a nonlinear moving average process $Y_t = be_{t-1}e_{t-2} + e_t$ and a bilinear autoregressive process $Y_t = be_{t-1}Y_{t-2} + e_t$, where $\{e_t\}$ is IID. These are WN processes, but there exist predictable nonlinearities in mean. Autocorrelation, variance ratio and power spectrum can easily miss these structures. Misleading conclusions could be reached in favor of MDS when the test statistics based on these measures are insignificant. It is therefore important to explore whether there exists a gap between an MDS and a WN for exchange rate changes, and if so, whether the neglected nonlinearity in mean can be explored to forecast exchange rate changes.

We will first explore serial dependence (i.e., any departure from IID) for exchange rate changes using Hong's (1999) generalized spectrum. The generalized spectrum can capture

---

[1] The terminologies "random walk" and "martingale" have been interchangeably used in the literature. Throughout this paper, we distinguish the random walk hypothesis from the martingale hypothesis. The innovation series is independent and identially distributed (IID) for the former and is an MDS for the latter.

[2] For example, Bekaert and Hodrick (1992), Fong and Ouliaris (1995), LeBaron (1999), Levich and Thomas (1993), Liu and He (1991), McCurdy and Morgan (1988), and Sweeney (1986) find evidence against the martingale hypothesis for nominal or real exchange rates, while Diebold and Nason (1990), Fong, Koh and Ouliaris (1997), Hsieh (1988,1989,1993), McCurdy and Morgan (1987), and Meese and Rogoff (1983a,b) find little evidence against the martingale hypothesis for nominal or real exchange rates.

1

any type of pairwise serial dependence over various lags, including those that could be missed by power spectrum and such higher order spectra as the bispectrum.[3] The generalized spectrum does not require any moment conditions. When proper moments exist, it can be differentiated to capture specific aspects of serial dependence, which are informative in revealing possible types of serial dependence. We find that among other things, most of the currencies we examine are WN in changes, but all of them are not martingales. There exists significant and predictable nonlinearity in the conditional mean of exchange rate changes.

To forecast the nonlinearity in conditional mean, we use linear autoregressive (AR), autoregressive polynomial (PN), artificial neural network (NN), functional-coefficient (FC) models, as well as their combination. The FC model, introduced in Cai, Fan, and Yao (2000, CFY), allows the autoregressive coefficients to be a function of some variables, which we choose according to an moving average technical trading rule (MATTR) so that the autoregressive coefficients vary with investment positions. We evaluate out-of-sample forecasts of these models relative to the martingale model, using four criteria—mean squared forecast error (MSFE), mean absolute forecast error (MAFE), mean forecast trading return (MFTR), and mean correct forecast direction (MCFD). As some complicated models may lead to significant superior predictive ability over the martingale model purely by luck, data-snooping is a serious concern (cf. Lo and MacKinlay 1999, Ch.8). To avoid this, we use White's (2000) test, which accounts for the dependence among the models under comparison. It is found that suitable nonlinear models, particularly their combination, do have significant superior predictive ability over the martingale model in terms of some criteria.

There has been an increasing interest in forecasting exchange rate changes. The nonlinear models used in the literature include bilinear models (Brooks 1997), threshold autoregressive models (Kräger and Kugler 1993, Brooks 1997), nonlinear dynamic systematic filtering models (Lisi and Medio 1997), artificial neural networks (Kuan and Liu 1995, Gençay 1999), and nearest neighbors regression (Diebold and Nason 1990). Recent works have apparently presented conflicting evidence for the types of nonlinearity in exchange rate changes. Hsieh (1989, 1993), Engle *et al.* (1990) and Frances and van Homelen (1998) find that most of nonlinearity in daily exchange rates arises from time-varying volatility, which, however, does not imply predictivity in mean unless there are ARCH-in-mean effects. Using nonparametric

---

[3]One example is an ARCH process with symmetric innovations, which has zero autocorrelations and zero third order cumulants.

regressions, Diebold and Nason (1990) and Meese and Rose (1990) find little improvement in out-of-sample forecasts for many major dollar spot rates in post-1973 float periods. Meese and Rose (1991) examine several structural exchange rate models but find that incorporation of nonlinearities into the structural models does not help forecast the conditional mean of exchange rate changes. In contrast, Kuan and Liu (1995), using the feedforward and recurrent neural network models, find a substantially lower MSFE than the martingale model. Lisi and Medio (1997) also find that a nonlinear filtering model outperforms the martingale model in terms of MSFE. Gençay (1999) finds that buy-sell signals of simple technical trading rules generated from using nearest neighbors and neural networks yield a lower MSFE than the martingale model.

In Section II, we introduce Hong's (1999) generalized spectrum and use it to document serial dependence for five major exchange rates. In Section III, we introduce the FC model, MATTR and other nonlinear time series models, and present the in-sample results on estimation and specification testing. Section IV evaluates the out-of-sample forecasts of these nonlinear models and their combination. Section V concludes.

II. SERIAL DEPENDENCE IN EXCHANGE RATE CHANGES

A. Generalized Spectral Analysis

To explore serial dependence of exchange rate changes, we use Hong's (1999) generalized spectrum, which is proposed as an alternative time series analytic tool to power spectrum and higher order spectrum. The basic idea is to transform a strictly stationary series $\{Y_t\}$ and consider the spectrum of the transformed series. Suppose that $\{Y_t\}$ has an marginal characteristic function $\varphi(u) \equiv Ee^{iuY_t}$ and a pairwise joint characteristic function $\varphi_j(u,v) \equiv Ee^{i(uY_t+vY_{t-|j|})}$, where $i \equiv \sqrt{-1}, u,v \in (-\infty,\infty)$, and $j = 0, \pm 1, \ldots$. Define the covariance function between the transformed variables e^{iuY_t} and $e^{ivY_{t-|j|}}$:

$$\sigma_j(u,v) \equiv \mathrm{cov}(e^{iuY_t}, e^{ivY_{t-|j|}}), \tag{2.1}$$

Straightforward algebra yields $\sigma_j(u,v) = \varphi_j(u,v) - \varphi(u)\varphi(v)$. Because $\varphi_j(u,v) = \varphi(u)\varphi(v)$ for all u,v if and only if Y_t and $Y_{t-|j|}$ are independent, $\sigma_j(u,v)$ can capture any type of pairwise serial dependence over various lags, including those with zero autocorrelation.

When $\sup_{u,v\in(-\infty,\infty)} \sum_{j=-\infty}^{\infty} |\sigma_j(u,v)| < \infty$, the Fourier transform of $\sigma_j(u,v)$ exists:

$$f(\omega,u,v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u,v)e^{-ij\omega}, \quad \omega \in [-\pi,\pi]. \tag{2.2}$$

3

Like $\sigma_j(u,v)$, $f(\omega,u,v)$ can capture all pairwise serial dependencies in $\{Y_t\}$ over various lags. It requires no moment condition. When $\text{var}(Y_t) = \sigma^2$ exists, the power spectrum $H(\omega)$ of $\{Y_t\}$ can be obtained by differentiating $f(\omega,u,v)$ with respect to (u,v) at $(0,0)$:

$$H(\omega) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j)e^{-ij\omega} = -\frac{\partial^2}{\partial u \partial v} f(\omega,u,v) \Big|_{(u,v)=(0,0)},$$

where $\gamma(j) \equiv \text{cov}(Y_t, Y_{t-|j|})$. For this reason, we call $f(\omega,u,v)$ a "generalized spectral density" of $\{Y_t\}$.

Hong (1999, Theorem 1) shows that $f(\omega,u,v)$ can be consistently estimated by

$$\hat{f}_n(\omega,u,v) \equiv \frac{1}{2\pi} \sum_{j=1-n}^{n-1} (1-|j|/n)^{\frac{1}{2}} k\,(j/p)\,\hat{\sigma}_j(u,v)e^{-ij\omega}, \tag{2.3}$$

where $\hat{\sigma}_j(u,v) \equiv \hat{\varphi}_j(u,v) - \hat{\varphi}_j(u,0)\hat{\varphi}_j(0,v)$ is the empirical generalized covariance, $\hat{\varphi}_j(u,v) \equiv (n-|j|)^{-1} \sum_{t=|j|+1}^{n} e^{i(uY_t+vY_{t-|j|})}$ is the empirical pairwise characteristic function, $p \equiv p_n$ is a bandwidth or lag order, and $k(\cdot)$ is a kernel function or "lag window". Commonly used kernels include Bartlett, Daniell, Parzen and Quadratic-Spectral kernels. The factor $(1-|j|/n)^{\frac{1}{2}}$ modifies the variance of $\hat{\sigma}_j(u,v)$. It could be replaced by 1, but it gives better finite sample performance for the tests based on $\hat{f}_n(\omega,u,v)$.

When $\{Y_t\}$ is IID, $f(\omega,u,v)$ becomes a "flat" generalized spectrum:

$$f_0(\omega,u,v) \equiv \frac{1}{2\pi}\sigma_0(u,v), \quad \omega \in [-\pi,\pi].$$

Any deviation of $f(\omega,u,v)$ from the flat spectrum $f_0(\omega,u,v)$ is evidence of serial dependence. Thus, to detect serial dependence, we can compare $\hat{f}_n(\omega,u,v)$ with the estimator

$$\hat{f}_0(\omega,u,v) \equiv \frac{1}{2\pi}\hat{\sigma}_0(u,v), \quad \omega \in [-\pi,\pi].$$

Once the existence of generic serial dependence is detected, one may like to further explore the nature of serial dependence. For example, is dependence operative primarily through the mean or through higher order moments? If serial dependence exists in mean, is it linear or nonlinear? If dependence exists in variance, does there exist linear or nonlinear and asymmetric ARCH? Different types of serial dependence have different implications for predictability of Y_t. If $\{Y_t\}$ is an MDS, for example, then serial dependence in higher moments will not help predict the conditional mean of Y_t.

4

To explore the nature of serial dependence, one can compare the derivative estimators

$$\hat{f}_n^{(0,m,l)}(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=1-n}^{n-1} (1 - |j|/n)^{\frac{1}{2}} k(j/p)\hat{\sigma}_j^{(m,l)}(u, v)e^{-ij\omega},$$

$$\hat{f}_0^{(0,m,l)}(\omega, u, v) \equiv \frac{1}{2\pi}\hat{\sigma}_0^{(m,l)}(u, v),$$

where $\hat{\sigma}_j^{(m,l)}(u, v) \equiv \partial^{m+l}\hat{\sigma}_j(u, v)/\partial^m u \partial^l v$ for $m, l \geq 0$. Just as the characteristic function can be differentiated to generate various moments, generalized spectral derivatives can capture various specific aspects of serial dependence, thus providing information on possible types of serial dependence.

Hong (1999) proposes a class of tests based on the quadratic norm:

$$Q\left(\hat{f}_n^{(0,m,l)}, \hat{f}_0^{(0,m,l)}\right) \equiv \int\int_{-\pi}^{\pi} |\hat{f}_n^{(0,m,l)}(\omega, u, v) - \hat{f}_0^{(0,m,l)}(\omega, u, v)|^2 d\omega dW_1(u)dW_2(v)$$

$$= \frac{2}{\pi} \int \sum_{j=1}^{n-1} k^2(j/p)(1 - j/n)|\hat{\sigma}_j^{(m,l)}(u, v)|^2 dW_1(u)dW_2(v), \qquad (2.4)$$

where the second equality follows by Parseval's identity, the unspecified integrals are taken over the support of $W_1(\cdot)$ and $W_2(\cdot)$, which are positive nondecreasing weighting functions that set weight about zero equally. An example of $W_1(\cdot)$ and $W_2(\cdot)$ is the N(0,1) CDF, which is commonly used in the characteristic function literature. As we will see below, proper choices of $W_1(\cdot)$ and $W_2(\cdot)$ as well as (m, l) allow us to test various specific aspects of serial dependence. The test statistic is a standardized version of the quadratic form:

$$M(m, l) \equiv \left[\int \sum_{j=1}^{n-1} k^2(j/p)(n - j)|\hat{\sigma}_j^{(m,l)}(u, v)|^2 dW_1(u)dW_2(v) - \hat{C}_0^{(m,l)} \sum_{j=1}^{n-1} k^2(j/p) \right]$$

$$\div \left[\hat{D}_0^{(m,l)} \sum_{j=1}^{n-2} k^4(j/p) \right]^{1/2}, \qquad (2.5)$$

where the centering and standardization factors

$$\hat{C}_0^{(m,l)} \equiv \int \hat{\sigma}_0^{(m,m)}(u, -u)dW_1(u) \int \hat{\sigma}_0^{(l,l)}(v, -v)dW_2(v),$$

$$\hat{D}_0^{(m,l)} \equiv 2 \int |\hat{\sigma}_0^{(m,m)}(u, u')|^2 dW_1(u)dW_1(u') \int |\hat{\sigma}_0^{(l,l)}(v, v')|^2 dW_2(v)dW_2(v').$$

Given (m, l), $M(m, l)$ is asymptotically one-sided N(0,1) under the null hypothesis of serial independence. For a kernel $k(\cdot)$ with unbounded support, $M(m, l)$ employs all $n - 1$ lags in

5

the sample. This is desirable when the alternative has persistent serial dependence. Non-uniform kernels, such as the Daniell kernel $k(z) = \sin(\pi z)/\pi z$, $z \in (-\infty, \infty)$, usually weight down higher order lags. This is expected to enhance good power of the tests in empirical study, because economic agents normally discount past information. This is particularly true of foreign exchange markets, where investors digest information relatively fast. In fact, the Daniell kernel maximizes the power of $M(m, l)$ over a class of kernels that include Parzen and Quadratic-Spectral kernels. The latter is optimal for spectral density estimation, but not necessarily for hypothesis testing (cf. Hong 1999).

In practice, we may first choose $(m, l) = (0, 0)$ to check if there exists any type of serial dependence. Once generic serial dependence is discovered using $M(0, 0)$, we may use various combinations of (m, l) to check specific types of serial dependence. For example, we can set $(m, l) = (1, 0)$ to check whether there exists serial dependence in mean. This checks whether $E(Y_t | Y_{t-j}) = E(Y_t)$ for all $j > 0$, and so it is a more suitable test for MDS than those based on autocorrelation, variance ratio and power spectrum. It can detect a wide range of deviations from MDS, including those with zero autocorrelations. To explore whether there exists linear dependence in mean, we can set $(m, l) = (1, 1)$. If $M(1, 0)$ is significant but $M(1, 1)$ is not, we can speculate that there may exist only nonlinear dependence in mean. We can go further to choose $(m, l) = (1, l)$ for $l = 2, 3, 4$, testing if $\text{cov}(Y_t, Y_{t-j}^l) = 0$ for all $j > 0$. These essentially check whether there exist ARCH-in-mean, skewness-in-mean and kurtosis-in-mean effects, which may arise from the existence of time-varying risk premium, asymmetry and improper account of the concern over large losses respectively. For convenience, we list a variety of spectral derivative tests and the types of dependence they can detect in Table 1.

B. Bootstrapping the Generalized Spectral Tests

To provide implement the generalized spectral tests, we will use bootstrap procedures, which can provide more accurate reference in finite samples than asymptotic theory. Recall M_i, $i = 1, \ldots, 13$, as in Table 1. Let $\hat{M}_i$ be the M_i statistic based on the observed sample $\{Y_t\}_{t=1}^n$. Let $\{Y_t^b\}_{t=1}^n$ be a bootstrap sample of $\{Y_t\}_{t=1}^n$ and let $\hat{M}_i^b$ be the M_i statistic based on $\{Y_t^b\}_{t=1}^n$. Then the bootstrap p-value of $\hat{M}_i$ can be approximated by $p_i^B = B^{-1} \sum_{b=1}^{B} \mathbf{1}(\hat{M}_i^b \geq \hat{M}_i)$, where $\mathbf{1}(\cdot)$ is the indicator function and B is the number of bootstrap replications. We will use naive-bootstrap (Efron, 1979) or wild-bootstrap (Wu 1986, Liu 1988), depending on whether the null hypothesis of interest is IID or MDS.[4]

---

[4]For naive bootstrap, we generate a bootstrap sample by randomly drawing, with replacement, obser-

<div align="center">TABLE 1. Generalized Spectral Tests</div>

Test	Statistic $M(m,l)$	Weights (W_1, W_2)	Test Function $\sigma_j^{(m,l)}(u,v)$	Notation
IID	$M(0,0)$	(W_0, W_0)	$\sigma_j(u,v)$	M_1
MDS	$M(1,0)$	(δ, W_0)	$\mathrm{cov}(Y_t, e^{ivY_{t-j}})$	M_2
correlation	$M(1,1)$	(δ, δ)	$\mathrm{cov}(Y_t, Y_{t-j})$	M_3
ARCH-in-mean	$M(1,2)$	(δ, δ)	$\mathrm{cov}(Y_t, Y_{t-j}^2)$	M_4
skewness-in-mean	$M(1,3)$	(δ, δ)	$\mathrm{cov}(Y_t, Y_{t-j}^3)$	M_5
kurtosis-in-mean	$M(1,4)$	(δ, δ)	$\mathrm{cov}(Y_t, Y_{t-j}^4)$	M_6
nonlinear ARCH	$M(2,0)$	(δ, W_0)	$\mathrm{cov}(Y_t^2, e^{ivY_{t-j}})$	M_7
leverage	$M(2,1)$	(δ, δ)	$\mathrm{cov}(Y_t^2, Y_{t-j})$	M_8
linear ARCH	$M(2,2)$	(δ, δ)	$\mathrm{cov}(Y_t^2, Y_{t-j}^2)$	M_9
conditional skewness	$M(3,0)$	(δ, W_0)	$\mathrm{cov}(Y_t^3, e^{ivY_{t-j}})$	M_{10}
conditional skewness	$M(3,3)$	(δ, δ)	$\mathrm{cov}(Y_t^3, Y_{t-j}^3)$	M_{11}
conditional kurtosis	$M(4,0)$	(δ, W_0)	$\mathrm{cov}(Y_t^4, e^{ivY_{t-j}})$	M_{12}
conditional kurtosis	$M(4,4)$	(δ, δ)	$\mathrm{cov}(Y_t^4, Y_{t-j}^4)$	M_{13}

Notes: When $m = 0$ (or $l = 0$), we put $W_1(\cdot) = W_0(\cdot)$ (or $W_2(\cdot) = W_0(\cdot)$), where $W_0(\cdot)$ is the N(0,1) CDF. When $m > 0$ (or $l > 0$), we put $W_1(\cdot) = \delta(\cdot)$ (or $W_2(\cdot) = \delta(\cdot)$), where $\delta(\cdot)$ is the Dirac delta function.

We first examine the size of the bootstrap tests for IID. All the $M(m,l)$ tests are suitable to test IID, with $M(0,0)$ being an omnibus test for IID. Here, naive bootstrap is appropriate. Table 2 reports the bootstrap sizes at the 10%, 5% and 1% levels for all $M(m,l)$ under an IID-N(0,1) data generation process, with $B = 300$. The results show that the naive bootstrap yields adequate sizes for all $M(m,l)$ tests for IID.

Next, we examine the sizes of the bootstrap tests for MDS. Suitable tests for MDS are $M(1,l)$, $l \geq 0$. We consider a GARCH(1,1)-N(0,1) data generation process, which is an MDS but displays volatility clustering, as is typical for high frequency financial time series. Since the data is conditionally heteroskedastic, we use wild bootstrap. Table 3 reports the bootstrap sizes of these tests for MDS, with $B = 300$. The results show that wild bootstrap yields adequate sizes for the $M(1,l)$ tests of MDS.

C. Joint Tests

vations from the original sample $\{Y_t\}_{t=1}^n$. For wild bootstrap, let F_t be the CDF of Y_t, $t = 1, \ldots, n$. We generate a bootstrap sample according to the formula that $Y_t^b = aY_t$ with probability $p = a/\sqrt{5}$ and $Y_t^b = (1-a)Y_t$ with probability $1-p$, where $a = (1+\sqrt{5})/2$. It follows that $E(Y_t^b|F_t) = 0$, $E[(Y_t^b)^2|F_t] = Y_t^2$, and $E[(Y_t^b)^3|F_t] = Y_t^3$. Thus, wild bootstrap can preserve the first three moments of the original data.

As noted earlier, all the $M(m, l)$ tests are suitable to test IID. Of these tests, $M(0, 0)$ is an omnibus test for IID, because it "encompasses" every moment captured by various derivative tests $M(m, l)$ with $m > 0$ and/or $l > 0$. Various derivative tests $M(m, l)$, on the other hand, are informative in revealing types of serial dependence. Similarly, the $M(1, l)$ tests with $l \geq 0$ are suitable to test MDS. Here, $M(1, 0)$ is an omnibus test for MDS, while various derivative tests $M(1, l)$ with $l \geq 1$ are informative in revealing the nature of departure from MDS.

However, the sequential use of the derivative tests requires caution as these test statistics may be mutually dependent. In this case, increasing the number of tests may increase the probability of Type I error, causing overrejection of the correct null hypothesis (cf. Richardson and Stock 1989, Richardson 1993). To avoid this, we consider a joint test formed from a given $k \times 1$ test statistic vector $M \equiv (M_{i_1}, ..., M_{i_k})'$:

$$J \equiv (M - \mu_M)'\Omega_M^{-1}(M - \mu_M),$$

where $\mu_M \equiv E(M)$ and $\Omega_M \equiv E[(M - \mu_M)(M - \mu_M)']$. This statistic J is asymptotically chi-square under the null hypothesis. Since μ_M and Ω_M are unknown, we estimate them via bootstrap: $\mu_M^* \equiv B^{-1} \sum_{b=1}^{B} \hat{M}^b$ and $\Omega_M^* \equiv B^{-1} \sum_{b=1}^{B} (\hat{M}^b - \mu_M^*)(\hat{M}^b - \mu_M^*)'$, where $\hat{M}^b$ is the $k \times 1$ test statistic vector M based on the bootstrap sample $\{Y_t^b\}_{t=1}^n$. Thus, an operational joint test statistic is

$$J^* \equiv (M - \mu_M^*)'(\Omega_M^*)^{-1}(M - \mu_M^*).$$

This test is also asymptotically chi-square under the null hypothesis of interest. The asymptotic approximation, however, may not be accurate in finite samples. Since J^* is asymptotically pivotal and so the bootstrap test may deliver more accurate sizes than the asymptotic test (cf. Hall 1992), we compute the bootstrap joint test statistic

$$\hat{J}_b^* \equiv (\hat{M}^b - \mu_M^*)'(\Omega_M^*)^{-1}(\hat{M}^b - \mu_M^*), \quad b = 1, ..., B,$$

and approximate the bootstrap p-value of J^* by $p_J^B \equiv B^{-1} \sum_{b=1}^{B} 1(\hat{J}_b^* \geq \hat{J}^*)$, where $\hat{J}^* \equiv (\hat{M} - \mu_M^*)'(\Omega_M^*)^{-1}(\hat{M} - \mu_M^*)$ and $\hat{M}$ is the $k \times 1$ test statistic vector M based on the original sample $\{Y_t\}_{t=1}^n$.

When testing IID, we consider two joint tests: $M_{\text{IID}}^{13} \equiv (M_1, ..., M_{13})'$ and $M_{\text{IID}}^{12} \equiv (M_2, ..., M_{13})'$, via naive bootstrap. The latter captures the overall effect of the twelve generalized spectral derivative tests for IID. A comparison between M_{IID}^{12} and the omnibus test

$M(0, 0)$ will indicate whether serial dependence can be explained by the types of serial dependence captured by the twelve derivative tests. When testing MDS, we consider the two joint tests: $M_{\text{MDS}}^5 \equiv (M_2, M_3, M_4, M_5, M_6)'$ and $M_{\text{MDS}}^4 \equiv (M_3, M_4, M_5, M_6)'$, via wild bootstrap. The latter captures the overall effect of the four generalized spectral tests $M(1, l)$, $1 \leq l \leq 4$, for MDS. A comparison between $M(1, 0)$ and the four derivative tests will reveal whether the departure from MDS can be explained by the first four lagged moments. Monte Carlo results in Tables 2 & 3 show that these joint test procedures have excellent sizes.

D. Serial Dependence in Exchange Rate Changes

We now use the generalized spectral tests to explore serial dependence of five major currencies—the nominal exchange rates of Canada (CD), Germany (DM), UK (BP), Japan (JY) and France (FF), per US dollar. The data is a weekly series from 1/1/1975 to 12/31/1998. The daily noon buying rates in New York City certified by the Federal Reserve Bank of New York for customs and cable transfers purposes are obtained from the Chicago Federal Reserve Board (www.frbchi.org). The weekly series is generated by selecting Wednesdays series (if a Wednesday is a holiday then the following Thursday is used), which has 1253 observations. The use of weekly data avoids the so-called weekend effect, as well as other biases associated with nontrading, bid-ask spread, asynchronous rates and so on, which are often present in higher frequency data. We use the scaled logarithmic difference $Y_t = 100 \ln(\xi_t / \xi_{t-1})$, where ξ_t is an exchange rate level.

The statistic $M(m, l)$ involves the choice of a bandwidth p, which may have significant impact on power. Hong (1999) proposes a data-driven method to choose p. This method still involves the choice of a preliminary bandwidth $\bar{p}$. Simulations in Hong (1999) and in Tables 2 & 3 show that the choice of $\bar{p}$ is less important than the choice of p. We consider $\bar{p}$ in the range $6 - 15$ to examine the robustness of $M(m, l)$ with respect to the choice of $\bar{p}$. We use the Daniell kernel, which maximizes the asymptotic power of $M(m, l)$ over a class of kernels that includes Parzen and Quadratic-Spectral kernels.[5]

Table 4 reports the values of $M(m, l)$ together with the bootstrap p-values of the individual tests and joint tests described earlier, for CD, DM, BP, JY and FF, using the medium preliminary lag order $\bar{p} = 10$. The results for other $\bar{p}$ are quite similar and not reported here. We use the 5% significance level here. For comparison, note that $M(m, l)$ has an asymptotic

---

[5]We have also used the Bartlett, Parzen and Quadratic-Spectral kernels. The results are similar to those based on the Daniell kernel.

one-sided N(0,1) distribution, so the asymptotic critical value at the 5% level is 1.65.

For CD, the geometric random walk hypothesis is strongly rejected (see $M(0,0)$). The martingale test $M(1,0)$ also rejects the martingale hypothesis as its bootstrap p-value is 0.026. This implies that the change of Canadian dollar exchange rates has a nonzero mean conditional on its past history and there exists neglected dependence in the martingale model. However, the correlation test $M(1,1)$ has a bootstrap p-value of 0.090, indicating that $\{Y_t\}$ is a WN. Moreover, the insignificant statistics $M(1,2)$, $M(1,3)$ and $M(1,4)$ indicate that the neglected nonlinearity in mean cannot be explained by its higher order conditional moments. The test $M(2,0)$ shows that there exists strong possibly nonlinear time-varying volatility, and the linear ARCH test $M(2,2)$ indicates very strong linear ARCH effects. The leverage effect $(M(2,1))$ is strong but the ARCH-in-mean effect $(M(1,2))$ is insignificant. There also exist some conditional skewness $(M(3,0))$ and conditional kurtosis $(M(4,0))$.

For DM, $M(0,0)$ suggests that the geometric random walk hypothesis is also strongly rejected. In contrast, the correlation test $M(1,1)$ is insignificant, implying that $\{Y_t\}$ is a WN. This, however, does not necessarily imply that $\{Y_t\}$ is an MDS (as most existing studies conclude), because $\{Y_t\}$ may have zero autocorrelation but a nonzero conditional mean. Indeed, the martingale test $M(1,0)$ strongly rejects the martingale hypothesis as its bootstrap p-value is 0.008. This implies that the change of Deutschemark exchange rates, though serially uncorrelated, has a nonzero mean conditional on its past history. Thus, suitable nonlinear time series models may be able to predict DM exchange rate changes. The test $M(2,0)$ shows strong possibly nonlinear time-varying volatility, and the linear ARCH test $M(2,2)$ indicates very strong linear ARCH effects. Both the leverage effect $(M(2,1))$ and ARCH-in-mean effect $(M(1,2))$ are, however, insignificant. There exist significant conditional skewness $(M(3,3))$ and large conditional kurtosis $(M(4,4))$.

For BP, the geometric random walk hypothesis is also strongly rejected $(M(0,0))$. In contrast, the correlation test $M(1,1)$ indicates that $\{Y_t\}$ is a WN (with a bootstrap p-value of 0.382). Nevertheless, the martingale test $M(1,0)$ strongly rejects the martingale hypothesis with a bootstrap p-value of 0.002. Thus, the change of British pound, though serially uncorrelated, has a nonzero mean conditional on its past history and is predictable nonlinearly using its own past history. Again, the linear ARCH test $M(2,2)$ suggests very strong ARCH effects, and the test $M(2,0)$ also shows significant possibly nonlinear time-varying volatility. The leverage effect $(M(2,1))$ is weak and the ARCH-in-mean effect $(M(1,2))$ is

insignificant. There exist very significant conditional skewness and conditional kurtosis.

For JY, the geometric random walk hypothesis is strongly rejected (see $M(0,0)$). The martingale test $M(1,0)$ also strongly rejects the martingale hypothesis. The rejection may be explained by significant moments as the tests $M(1,l)$, $l = 1, 2, 3, 4$, are significant. Unlike CD, DM, BP and FF (below), JY exhibits strong serial correlation in changes ($M(1,1)$). There also exist strong ARCH effect, leverage effect, ARCH-in-mean effect, conditional skewness and conditional heterokurtosis.

Finally, for FF, the results are similar to those for CD. First, the geometric random walk hypothesis is strongly rejected. While $M(1,1)$ shows no serial correlation (with a bootstrap p-value of 0.178), $M(1,0)$ strongly rejects the martingale hypothesis (with a bootstrap p-value of 0.008). The higher order moments are not significant to explain the nonlinearity in mean (see $M(1,l)$ with $l = 2, 3, 4$). The $M(2,2)$ test suggests very strong ARCH effects, and $M(2,0)$ also suggests significant possibly nonlinear time-varying volatility. The leverage effect ($M(2,1)$) is significant but the ARCH-in-mean effect ($M(1,2)$) is insignificant. There exist significant conditional skewness and conditional kurtosis.

To summarize, we observe:

1) There exists strong serial dependence for the changes of all the five exchange rates. The geometric random walk hypothesis (possibly with drift) is strongly rejected for all the five currencies.

2) Although the changes of exchange rates are often serially uncorrelated (as is the case for CD, DM, BP, FF), they are clearly not an MDS for all the five currencies. There exists strong nonlinearity in mean for the changes of all the five exchange rates.

3) For CD, DM, BP and FF, the nonlinearity in mean cannot be explained by the polynomials of lagged exchange rate changes. It is of a complicated and unknown form.

4) There exist strong ARCH effects for all the five currencies. While the leverage effect is significant for CD, JY and FF, the ARCH-in-mean effect is significant only for JY.

5) There are significant conditional skewness and/or conditional kurtosis.

It is important to explore the implications of these stylized facts. In the rest of this paper, we focus on finding 2). The fact that exchange rate changes are not an MDS implies that exchange rate changes are predictable in mean.[6] We will use some nonlinear time series

---

[6] This does not necessarily imply that exchange rate markets are inefficient, or the exchange rates are not rational assessments of "fundamental" values. As Lucas (1978) has shown, rational expectation equilibrium

models in combination with an MATTR to forecast the often neglected nonlinearity in mean for exchange rate changes.

III. NONLINEAR TIME SERIES MODELS

To forecast the changes of exchange rates, we consider models for $E(Y_t|I_{t-1})$, where $I_t \equiv \{Y_t, Y_{t-1}, ...\}$ is the information set available at time t. The evidence in Section II suggests that $E(Y_t|I_{t-1})$ is time-varying but is of a complicated form. In particular, it cannot be modeled simply by autoregressive polynomials in lagged exchange rate changes for all the currencies except JY. Various parametric and nonparametric models can be used. Examples of parametric models are autoregressive bilinear and threshold models. Examples of nonparametric models are artificial neural network, kernel and nearest neighbor regression models. As noted earlier, these models have been used in the literature, with apparently mixed results on the predictivity of exchange rate changes in mean.

We consider the following models for $E(Y_t|I_{t-1})$: linear autoregressive, AR(d); autoregressive polynomial, PN(d, m); artificial neural network, NN(d, q); functional coefficient, FC(d, L); and their combination. These models are described in Table 5. Among them, NN(d, q) is a popular nonlinear model (see, e.g., Campbell, Lo and MacKinlay 1997, Ch.12). FC(d, L) is a new nonlinear time series model introduced by CFY (2000), with time-varying and state-dependent coefficients. It can be viewed as a special case of Priestley's (1980) state-dependent model, but it includes the models of Tong (1990), Chen and Tsay (1993) and regime-switching models as special cases. In our application, the autoregressive coefficients depend on an MATTR.

In addition to forecasting $E(Y_t|I_{t-1})$, we will also forecast the direction of changes. For directional forecasts, we consider two additional models—MATTR(L) and Buy&Hold, as described in Table 5. For MATTR(L), one has $\text{sign}(\hat{Y}_{t+1}) = \text{sign}(U_{t+1})$, where $\text{sign}(x) \equiv 1(x > 0) - 1(x < 0), 1(\cdot)$ is the indicator function, and U_t is as in Table 5. It generates a "buy" signal (i.e., to purchase the foreign currency using U.S. dollars) when the current exchange rate level ξ_t is above the moving average $L^{-1} \sum_{j=1}^{L} \xi_{t-j}$, and a "sell" signal (i.e., to purchase U.S. dollars using the foreign currency) when it is below. For Buy&Hold, one has $\text{sign}(\hat{Y}_{t+1}) = 1$ for all t. No estimation is required for these two models.

asset prices need not follow a martingale sequence.

To estimate $AR(d)$ and $PN(d,m)$, we use the ordinary least squares method. $NN(d,q)$ is estimated by the nonlinear least squares method with the Newton-Raphson algorithm. See (e.g.) Kuan and Liu (1995) for more discussion. Because the FC model is relatively new, we now describe this model, its estimation and testing in some detail.

TABLE 5. Models

Name	Models for $E(Y_t|I_{t-1})$ and $\text{sign}[E(Y_t|I_{t-1})]$
Benchmark	$E(Y_t|I_{t-1}) = \mu.$
$AR(d)$	$E(Y_t|I_{t-1}) = \beta_0 + \sum_{j=1}^{d}\beta_j Y_{t-j}.$
$PN(d,m)$	$E(Y_t|I_{t-1}) = \alpha_0 + \sum_{j=1}^{d}\sum_{i=1}^{m}\alpha_{ij}Y_{t-j}^{i}.$
$NN(d,q)$	$E(Y_t|I_{t-1}) = \beta_0 + \sum_{j=1}^{d}\beta_j Y_{t-j} + \sum_{i=1}^{q}\delta_i G(\gamma_{0i} + \sum_{j=1}^{d}\gamma_{ji}Y_{t-j}),$
	where $G(z) = (1 + e^{-z})^{-1}.$
$FC(d,L)$	$E(Y_t|I_{t-1}) = a_0(U_t) + \sum_{j=1}^{d}a_j(U_t)Y_{t-j},$ where $U_t = \xi_{t-1} - L^{-1}\sum_{j=1}^{L}\xi_{t-j-1}.$
Combined(1-4)	Combined forecast of $AR(2)$, $PN(2,4)$, $NN(2,5)$ and $FC(2,26)$,
	which is to be defined in equation (4.6) below.
$MATTR(L)$	$\text{sign}[E(Y_t|I_{t-1})] = \text{sign}(U_t),$ where $U_t = \xi_{t-1} - L^{-1}\sum_{j=1}^{L}\xi_{t-j-1}.$
Buy&Hold	$\text{sign}[E(Y_t|I_{t-1})] = 1.$

A. Functional-Coefficient Model

Let $\{(Y_t, U_t)'\}_{t=1}^{n}$ be a stationary process, where Y_t and U_t are scalar variables. Also let $X_t \equiv (1, Y_{t-1}, \ldots, Y_{t-d})'$, a $(d+1) \times 1$ vector. We assume

$$E(Y_t|I_{t-1}) = a_0(U_t) + \sum_{j=1}^{d} a_j(U_t)Y_{t-j}, \tag{3.1}$$

where the $\{a_j(U_t)\}$ are the autoregressive coefficients depending on U_t, which may be chosen as a function of X_t or something else. Intuitively, the FC model in (3.1) is an AR process with time-varying autoregressive coefficients.

The coefficient functions $\{a_j(U_t)\}$ can be estimated by local linear regression. At each point u, we approximate $a_j(U_t)$ locally by a linear function $a_j(U_t) \approx a_j + b_j(U_t - u)$, $j = 0, 1, ..., d$, for U_t near u, where a_j and b_j are constants. The local linear estimator at point u is then given by $\hat{a}_j(u) = \hat{a}_j$, where $\{(\hat{a}_j, \hat{b}_j)\}_{j=0}^{d}$ minimizes the sum of local weighted squares $\sum_{t=1}^{n}[Y_t - E(Y_t|I_{t-1})]^2 K_h(U_t - u)$, with $K_h(\cdot) \equiv K(\cdot/h)/h$ for a given kernel function $K(\cdot)$ and bandwidth $h \equiv h_n \to 0$ as $n \to \infty$. We select h using a modified multi-fold "leave-one-out-type" cross-validation based on MSFE (cf. CFY 2000, p. 944).

B. Choosing U_t Based on MATTR

It is important to choose an appropriate smooth variable U_t. Knowledge on data or economic theory may be helpful. When no prior information is available, U_t may be chosen as a function of explanatory vector X_t or using such data-driven methods as AIC and cross-validation. See Fan, Yao and Cai (2002) for further discussion on the choice of U_t. For exchange rate changes, we choose U_t as the difference between the exchange rate at time $t-1$ and the moving average of the most recent L periods of exchange rates at time $t-1$:

$$U_t \equiv \xi_{t-1} - L^{-1} \sum_{j=1}^{L} \xi_{t-j}. \tag{3.2}$$

The moving average $L^{-1} \sum_{j=1}^{L} \xi_{t-j}$ is a proxy for the trend at time $t-1$. We choose $L = 26$ (half a year). Intuitively, U_t is expected to reveal useful information on the direction of changes. For more discussion on the rationales of using an MATTR, see (e.g.) Levich and Thomas (1993) and Brock, Lakonishock and LeBaron (1999).

C. Testing for and Estimation of Functional-Coefficients

To justify the use of the FC model, we apply CFY's (2000) goodness-of-fit test for an $AR(d)$ model against a $FC(d, L)$ alternative. In the framework of (3.1), the null hypothesis of $AR(d)$ can be stated as

$$\mathsf{H}_0 : a_j(U_t) = \beta_j, \quad j = 0, 1, \ldots d, \tag{3.3}$$

where β_j is the autoregressive coefficient in $AR(d)$. Under H_0 in (3.3), $\{Y_t\}$ is linear in mean conditional on X_t. Under the alternative to H_0 in (3.3), the autoregressive coefficients depend on U_t and the $AR(d)$ model suffers from "neglected nonlinearity".

To test H_0 in (3.3), CFY compares the residual sum of squares (RSS) under H_0

$$RSS_0 \equiv \sum_{t=1}^{n} \hat{\varepsilon}_t^2 = \sum_{t=1}^{n} [Y_t - \hat{\beta}_0 - \sum_{j=1}^{d} \hat{\beta}_j Y_{t-j}]^2$$

with the RSS under the alternative

$$RSS_1 \equiv \sum_{t=1}^{n} \tilde{\varepsilon}_t^2 = \sum_{t=1}^{n} [Y_t - \hat{a}_0(U_t) - \sum_{j=1}^{d} \hat{a}_j(U_t) Y_{t-j}]^2.$$

The test statistic is $T_n = (RSS_0 - RSS_1)/RSS_1$. We reject H_0 in (3.3) for large values of T_n. CFY suggest the following bootstrap method to obtain the p-value of T_n: (i) generate the bootstrap residuals $\{\varepsilon_t^b\}_{t=1}^{n}$ from the centered residuals $\tilde{\varepsilon}_t - \bar{\varepsilon}$ where $\bar{\varepsilon} \equiv n^{-1} \sum_{t=1}^{n} \tilde{\varepsilon}_t$ and

14

define $Y_t^b \equiv X_t'\hat{\beta} + \varepsilon_t^b$, where $\hat{\beta}$ is the OLS estimator for AR(d); (ii) calculate the bootstrap statistic T_n^b using the bootstrap sample $\{Y_t^b, X_t', U_t\}_{t=1}^n$; (iii) repeat steps (i) and (ii) B times ($b = 1, ..., B$) and approximate the bootstrap p-value of T_n by $B^{-1} \sum_{b=1}^B \mathbf{1}(T_n^b \geq T_n)$. The bootstrap p-values of T_n for five currencies, based on $B = 1000$, are 0.031 (CD), 0.000 (DM), 0.025 (BP), 0.022 (JY) and 0.063 (FF), which indicate strong rejection of AR(2) in favor of FC(2, 26).

IV. PREDICTIVE ABILITY OF NONLINEAR MODELS

So far, we have explored nonlinearity in mean for exchange rate changes via in-sample analysis. We now examine whether the above nonlinear models can outperform the martingale model in out-of-sample forecasts. As the martingale test $M(1,0)$ checks the null hypothesis that $E(Y_t|Y_{t-j}) = E(Y_t)$ for all $j > 0$, we choose the martingale model $Y_t = \mu + \varepsilon_t$ as the benchmark. When several forecast models using the same data are compared, it is crucial to take into account the dependence among the models. Otherwise, the resulting inference is commonly referred to as "data-snooping" and can be misleading (cf. Lo and MacKinlay 1999, Ch.8). White (2000) develops a novel test for out-of-sample multiple model comparison that accounts for data-snooping biases. We use this method here.

A. Forecast Evaluation Criteria

Suppose there are $n + 1$ $(\equiv R + P)$ observations. We use the most recent R observations available at time t, $R \leq t < n + 1$, to generate P forecasts using each model. For each time t in the prediction period, we use a rolling sample $\{Y_{t-R+1}, ..., Y_t\}$ of size R to estimate model parameters. We can then generate a sequence of 1-step-ahead forecasts $\{\hat{Y}_{t+1}\}_{t=R}^n$, which is used to evaluate each model. For more discussion on rolling estimation, see West (1996) and McCracken (2000).

Our main aim is to investigate out-of-sample forecasts of the models described in Table 5, relative to the martingale model. We compare them in terms of MSFE, MAFE, MFTR, and MCFD respectively:

$$MSFE \equiv P^{-1} \sum_{t=R}^n (Y_{t+1} - \hat{Y}_{t+1})^2,$$

$$MAFE \equiv P^{-1} \sum_{t=R}^n |Y_{t+1} - \hat{Y}_{t+1}|,$$

15

$$MFTR \equiv P^{-1} \sum_{t=R}^{n} \text{sign}(\hat{Y}_{t+1})Y_{t+1},$$

$$MCFD \equiv P^{-1} \sum_{t=R}^{n} \mathbf{1}(\text{sign}(\hat{Y}_{t+1})\text{sign}(Y_{t+1}) > 0).$$

We ignore issues such as interest differentials, transaction costs and market depth, and assume that there exists no budget constraint. Because the investors are ultimately trying to maximize profits rather than minimize forecast errors, MSFE and MAFE may not be the most appropriate evaluation criteria.[7] Granger (1999) emphasizes the importance of model evaluation using economic measures such as MFTR rather than statistical criteria such as MSFE and MAFE. MCFD is closely associated with an economic measure as it relates to market timing. Mutual fund managers, for example, can adjust investment portfolios in a timely manner if they can predict the directions of changes, thus earning a return higher than the market average. Note that $MFTR^{\text{Buy\&Hold}} = P^{-1} \sum_{t=R}^{n} Y_{t+1} \to \mu$ in probability as $P \to \infty$, where $\mu \equiv E(Y_t)$.

B. Comparing Forecasting Models

Multiple model comparison can be conveniently formulated as hypothesis testing of suitable moment conditions. Consider an $l \times 1$ vector of moments, $E(\boldsymbol{\psi}^*)$, where $\boldsymbol{\psi}^* \equiv \boldsymbol{\psi}(Z, \beta^*)$ is an $l \times 1$ vector with elements $\psi_k^* \equiv \psi_k(Z, \beta^*)$, $1 \leq k \leq l$, $Z \equiv (Y, X')'$ and $\beta^* \equiv \text{plim} \ \hat{\beta}_t$. These moments have incorporated the information of the l models and the benchmark model. Define the out-of-sample $l \times 1$ moment vector

$$\bar{\boldsymbol{\psi}} \equiv P^{-1} \sum_{t=R}^{n} \boldsymbol{\psi}(Z_{t+1}, \hat{\beta}_t).$$

Noting that MSFE and MAFE are to be minimized while MFTR and MCFD are to be maximized, we compare model k ($k = 1, ..., l$) with the martingale model (model 0) via

$$\bar{\psi}_k \equiv -(MSFE^k - MSFE^0),$$
$$\bar{\psi}_k \equiv -(MAFE^k - MAFE^0),$$
$$\bar{\psi}_k \equiv MFTR^k - MFTR^0,$$
$$\bar{\psi}_k \equiv MCFD^k - MCFD^0.$$

---

[7]Clements and Hendry (1993) show that MSFE may be an inadequate and potentially misleading basis for model selection because it is not invariant to data transformations.

By a suitable central limit theorem, we have $\sqrt{P}[\bar{\boldsymbol{\psi}} - E(\boldsymbol{\psi}^*)] \to N(0, \Omega)$ in distribution as $P \equiv P_n \to \infty$ when $n \to \infty$,[8] where Ω is a $l \times l$ variance-covariance matrix

$$\Omega \equiv \lim_{n\to\infty} \text{var} \left[P^{-\frac{1}{2}} \sum_{t=R}^{n} \boldsymbol{\psi}(Z_{t+1}, \hat{\beta}_t) \right].$$

In general, this matrix is rather complicated because it depends on a component due to parameter estimation uncertainty in the estimated parameter $\hat{\beta}_t$ (cf. West 1996). However, when $E(\partial \boldsymbol{\psi}^*/\partial\beta) = 0$ or $P/R \to 0$ as $n \to \infty$, West (1996, Thm. 4.1) and McCracken (2000, Thm. 2.3.1) shows that, under proper regularity conditions, Ω does not depend on parameter estimation uncertainty and can be simplified as:

$$\Omega = \lim_{n\to\infty} \text{var} \left[P^{-\frac{1}{2}} \sum_{t=R}^{n} \boldsymbol{\psi}(Z_{t+1}, \beta^*) \right], \tag{4.1}$$

as is the case of West's (1996, Thm. 4.1(a)) and Diebold and Mariano (1995).[9] Here, the effect of using $\hat{\beta}_t$ rather than β^* is asymptotically negligible.

When we compare an individual model, say model k, with a benchmark, the null hypothesis is that model k is no better than the benchmark:

$$\mathsf{H}_1 : E(\psi_k^*) \leq 0 \ \text{ for each } k = 1, ..., l. \tag{4.2}$$

We can use the tests of Diebold and Mariano (1995), West (1996), McCracken (2000) or Pesaran and Timmermann (1992) with an appropriate estimator of Ω.

When we compare l models against a benchmark jointly, the null hypothesis of interest is that the best model is no better than the benchmark:

$$\mathsf{H}_2 : \max_{1 \leq k \leq l} E(\psi_k^*) \leq 0. \tag{4.3}$$

Under the alternative to H_2, the best model is superior to the benchmark. To test H_2, sequential use of an individual test may result in a data-snooping bias since the test statistics are mutually dependent. To account for possible data snooping bias, we use White's (2000) method. White (2000) proposes the following test statistic for H_2:

$$\bar{V}_P \equiv \max_{1 \leq k \leq l} \sqrt{P}[\bar{\psi}_k - E(\psi_k^*)], \tag{4.4}$$

---

[8] One referee points out that tests of equal MSFE for two nested models may have a nonstandard limit distribution. Fortunately, under the condition that $(P/R) \log \log R \to 0$, $\sqrt{P}[\bar{\psi} - E\psi^*]$ is still asymptotically normal for the MSFE criterion (cf. Clark and McCracken 2001, Thm. 3.1(b); 2002, Thm. 3.3(b)). Thus, White's bootstrap procedure is valid and applicable.

[9] Only McCracken (2000) considers nondifferentiable ψ for the case of MAFE in this literature.

whose limit distribution is however unknown, due to unknown Ω. To obtain the p-value for $\bar{V}_P$, White (2000) suggests and justifies using the stationary bootstrap of Politis and Romano (1994): (i) obtain a bootstrap sample $\{Z_{t+1}^b\}_{t=R}^n$; (ii) estimate $\{\hat{\beta}_t^b\}_{t=R}^n$ using $\{Z_{t+1}^b\}_{t=R}^n$; (iii) compute the stationary bootstrap statistic

$$\bar{V}_P^b \equiv \max_{1 \leq k \leq l} \sqrt{P}(\bar{\psi}_k^b - \bar{\psi}_k), \tag{4.5}$$

where $\bar{\psi}_k^b \equiv P^{-1} \sum_{t=R}^n \psi_k(Z_{t+1}^b, \hat{\beta}_t^b)$; (iv) repeat steps (i)–(iii) B times ($b = 1, \ldots, B$), and approximate the bootstrap p-value of $\bar{V}_P$ by $B^{-1} \sum_{b=1}^B 1(\bar{V}_P^b \geq \bar{V}_P)$. This bootstrap p-value for testing H_2 is called the "reality check p-value" for data snooping.

White (2000, Thm. 2.3 & Cor. 2.4) shows that a sufficient condition for the validity of the stationary bootstrap is $(P/R) \log \log R \to 0$ as $n \to \infty$, no matter whether $E(\partial \psi^*/\partial \beta) = 0$.[10] White (2000) only considers differentiable ψ (e.g., MSFE). As noted in White (2000, p.1100), it is possible to extend his procedure to nondifferentiable ψ (e.g., MAFE, MFTR, MCFD). Checking White's (2000) proof, we see that when no parameter estimation is involved, White's (2000) procedure is applicable to nondifferentiable ψ. We expect that when parameter estimation is involved, the impact of parameter estimation uncertainty is asymptotically negligible when P grows at a suitably slower rate than R. In this case, we conjecture that White's (2000) procedure continues to hold for nondifferentiable ψ no matter whether $\partial E(\boldsymbol{\psi}^*)/\partial \beta = 0$. However, the proof is much involved and has to be pursued in further work. McCracken's (2000) approach may be useful here.

C. Combined Forecast

In practice it is not uncommon that some forecast models perform well in certain periods while other forecast models perform well in other periods. It is difficult to find a forecast model that outperforms all the other models in all prediction periods. To improve forecasts over individual models, combined forecasts have been suggested. Bates and Granger (1969), Stock and Watson (1999), Knox, Stock and Watson (2000) and Yang (2002) show that forecast combinations can improve forecast accuracy over a single model. Granger (2001) emphasizes that combined forecasts may provide an insurance to diversify risk of forecast errors, analogous to investing on portfolios rather than on individual securities. We will combine AR(2), PN(2, 4), NN(2, 5) and FC(2, 26) to forecast the conditional mean of exchange

---

[10] In our applications, we use $(R, P) = (613, 612)$ and $(817, 408)$. For these choices, $(P/R) \log \log R = 0.444$ and 0.232 respectively.

rate changes. Let $\hat{Y}_{kt}$ be the forecast for $E(Y_t|I_{t-1})$ by model k, where $k = 1, 2, 3, 4$ denotes AR(2), PN(2, 4), NN(2, 5) and FC(2, 26) respectively. We consider the combined forecast:

$$\hat{Y}_t^* \equiv \sum_{k=1}^{4} w_{kt}\hat{Y}_{kt}, \tag{4.6}$$

where the weight

$$w_{kt} \equiv \frac{\exp[-\lambda_t \sum_{s=1}^{t-1}(Y_s - \hat{Y}_{ks})^2]}{\sum_{k'=1}^{4} \exp[-\lambda_t \sum_{s=1}^{t-1}(Y_s - \hat{Y}_{k's})^2]},$$

$\lambda_t = 1/(2S_t^2)$, and S_t^2 is the sample variance of $\{Y_s\}_{s=1}^{t-1}$; that is, $S_t^2 = (t-2)^{-1} \sum_{s=1}^{t-1}(Y_s - \mu_t)^2$ and $\mu_t = (t-1)^{-1} \sum_{s=1}^{t-1} Y_s$. The combined model in (4.6) will be denoted as Combined(1-4).The weighting scheme w_{kt} is proposed in Yang (2002). Intuitively, w_{kt} gives a large weight to model k at period t when it forecasted well at period $t-1$, and gives a small weight to model k at period t when it forecasted poorly at period $t-1$. Yang and Zhou (2002) have applied this method to choose ARIMA models and find that it has a clear stability advantage in forecasting over some existing popular model selection criteria.

D. Predictivity of Exchange Rate Changes

We now evaluate the out-of-sample forecasts of the models described in Table 5. Table 6 reports the results on White's (2000) test, where P_{RC}^1 is the bootstrap p-value for comparing a single model with the martingale model, and P_{RC}^2 is the bootstrap reality check p-value for comparing l models with the martingale model. The l-th number for P_{RC}^2 is the bootstrap reality check p-value for the null hypothesis that the best of the first l models has no superior predictive power over the martingale model. The last number for P_{RC}^2 checks if the best of all the models under comparison has superior predictive ability over the martingale model. The difference between each P_{RC}^1 and the last P_{RC}^2 gives an estimate of data-snooping bias.

For CD, none of AR(2), PN(2, 4), NN(2, 5) and FC(2, 26) outperform the martingale model in terms of all the four criteria, but the combined forecast of these models does outperform the martingale model in terms of MSFE and MAFE (Panel A).

For DM, in terms of MSFE and MAFE, none of AR(2), PN(2, 4), NN(2, 5) and FC(2, 26) outperform the martingale model, while Combined(1-4) outperforms the martingale model. In terms of MFTR and MCFD in Panel B, PN(2, 4) performs best and better than the martingale model. It generates a 9.2% profit and correctly predicts 54.2% of the directions of changes. It has marginally significant bootstrap p-values $P_{RC}^1 = 0.089$ and 0.051 in terms of MFTR and MCFD respectively for the null hypothesis that PN(2, 4) is no better than the

martingale model. So, comparing only PN$(2,4)$ with the martingale model would suggest that PN$(2,4)$ may improve the forecast over the martingale model. However, P_{RC}^2, the reality check p-value for testing the null hypothesis that the best of the seven models is no better than the martingale model, tells a different story. The last (7th) numbers for P_{RC}^2 are 0.363 and 0.245 in terms of MFTR and MCFD respectively, indicating that the best forecast model (PN$(2,4)$) among the seven models is no better than the martingale model.[11]

For BP, in terms of MSFE (Panel B) and MAFE (Panels A & B), Combined(1-4) yields significant P_{RC}^1. In terms of MCFD, NN$(2,5)$ and FC$(2,26)$ give significant P_{RC}^1 in Panel B, which remains significant even using the reality check p-value with P_{RC}^2.

For JY, AR(2) outperforms the martingale model in terms of MAFE (Panel B) and MCFD (Panels A and B). This is consistent with the in-sample evidence of serial correlation for JY (see $M(1,1)$ in Table 4). The superior predictive ability is most clear in terms of MCFD with significant reality check p-values P_{RC}^2. All the other models also outperform the martingale model in terms of MCFD. In terms of MAFE, Combined(1-4) yields significant P_{RC}^1 in Panels A and B.

For FF, Combined(1-4) outperforms the martingale model in terms of many criteria in Panels A and B. AR(2) and PN$(2,4)$ outperform the martingale model in terms of MFTR and MCFD (Panels A and B). FC$(2,26)$ has MFTR $= 9.8\%$ with $P_{RC}^1 = 0.052$ (Panel A).

To sum up, we observe:

1) The combined forecast model performs best in most cases. It does provide some insurance to diversify risk of forecast errors.

2) It is hardest to predict CD and easiest to predict JY. This is consistent with the in-sample evidence that the martingale test $M(1,0)$ reported in Table 4 is least significant for CD and most significant for JY. Although the significant in-sample evidence of nonlinearity in mean suggested from $M(1,0)$ does not carry over to significant out-of-sample forecasts for some currencies in terms of certain criteria (after taking data-snooping bias into account), the results of in-sample and out-of-sample analysis match each other in terms of the degree of significance. In particular, the degree of significance of the martingale test $M(1,0)$ decreases in the order of JY, BP, DM, FF and CD. The degree of significance of the superior predictive ability of nonlinear models over the martingale model more or less follows the same order.[12]

---

[11] Of course, it is possible that the insignificant reality check p-values are due to low power of the reality check test procedure in finite samples. This, however, remains to be investigated.

[12] The existence of some degree of correspondence between the strength of the rejection using $M(1,0)$ and

3) For (and only for) JY, the linear AR(2) model has some superior predictive ability over the martingale model in out-of-sample forecasts. This is consistent with the in-sample evidence in Table 4, where the correlation test $M(1,1)$ is significant only for JY.

4) The choice of loss function affects the forecast evaluation results. It appears that it is more difficult to beat the martingale model in terms of statistical criteria (MSFE and MAFE), while it is easier in terms of the economic criteria (MFTR and MCFD).

V. CONCLUSIONS

Using Hong's (1999) generalized spectral tests, we have documented that there exists strong nonlinear serial dependence for exchange rate changes, which cannot be solely attributed to the well-known volatility clustering. The generalized spectrum also provides useful tools to learn about the nature of nonlinear serial dependence in the exchange rate changes. In particular, we find that there exists significant nonlinearity in mean for exchange rate changes, although most exchange rate changes are serially uncorrelated.

To forecast the neglected nonlinearity in mean, we consider some nonlinear time series models and their combination. After filtering out possible data-snooping bias via White's (2000) test, we find that some of these models have superior predictive ability for JY and FF, to lesser degree for CD, DM, and BP, over the martingale model, particularly in terms of trading returns and/or directional forecasts. The choice of loss function generally affects the forecast evaluation results. The combined forecast model performs best in most cases.

Although the in-sample inference and out-of-sample forecasts share similar patterns in the degree of significance, the in-sample significant nonlinear serial dependence in mean does not carry over to significant out-of-sample forecasts in terms of some criteria, after accounting for data-snooping. Perhaps the nonlinear models used are not the most suitable ones. Alternatively, parameter estimation uncertainty may dominate the nonlinearity in mean that may not be strong enough to be exploited for forecasting. Also, nonlinearities may be exogenous, arising from outliers, structural shifts and government intervention, which can render various nonlinearity tests to reject the null hypothesis of linearity while not being useful for out-of-sample forecasts.

the degree of out-of-sample predictability in mean implies that to certain extent, the in-sample test $M(1,0)$ is indicative of out-of-sample predictability for exchange rate changes. It seems to suggest that there exists some systematic predictable component in foreign exchange markets that prevails throughout the whole sample period.

In this paper, we have only exploited the predictability of exchange rate changes in mean. The generalized spectrum also reveals significant linear and nonlinear dependence in higher moments, which have important implications on the predictability for higher order moments and the entire density, which is important for correctly assessing exchange rate risk. We leave this for future work.

REFERENCES

Bates, J.M. and C.W.J. Granger (1969), "The Combination of Forecasts", *Operations Research Quarterly*, 20, 451-468.

Bekaert, G. and R. Hodrick (1992), "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets", *Journal of Finance*, 467-509.

Bollerslev, T. (1990), "Modelling the Coherence in Short-run Nominal Exchange rates: A Multivariate Generalized ARCH Model", *Review of Economics and Statistics*, 72, 498-505.

Brock, W., D. Hsieh and B. LeBaron (1991), *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence.* MIT Press: Cambridge, MA.

Brock, W., J. Lakonishock and B. LeBaron (1992), "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns", *Journal of Finance*, 47, 1731-64.

Brook, C. (1997), "Linear and Nonlinear (Non-)Forecastability of High-frequency Exchange Rates", *Journal of Forecasting*, 16, 125-145.

Cai, Z., J. Fan, and Q. Yao (2000), "Functional-coefficient Regression Models for Nonlinear Time Series", *Journal of American Statistical Association,* 95(451), 941-956.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay (1997), *The Econometrics of Financial Markets,* Princeton University Press.

Chen, R. and R.S. Tsay (1993), "Functional-coefficient Autoregressive Models", *Journal of the American Statistical Association*, 88, 298-308.

Clark, T.E. and M.W. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models", *Journal of Econometrics,* 105, 85-110.

Clark, T.E. and M.W. McCracken (2002), "Evaluating Long-Horizon Forecasts," Department of Economics, University of Missouri, Columbia.

Clements, M.P. and D.F. Hendry (1993), "On the Limitations of Comparing Mean Square Forecast Errors", *Journal of Forecasting,* 12, 617-637.

Diebold, F.X., J. Hahn, and A.S. Tay (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange", *Review of Economics and Statistics,* 81(4), 661-673.

Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X. and J.A. Nason (1990),"Nonparametric Exchange Rate Prediction?" *Journal of International Economics*, 28, 315-332.

Durlauf, S. (1991), "Spectral Based Tests for the Martingale Hypothesis", *Journal of Econometrics,* 50, 1-19.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife", *Annals of Statistics*, 7, 1-26.

Engle, R., T. Ito, and W. Lin (1990), "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market", *Econometrica,* 58, 525-542.

Fan, J., Q. Yao, and Z. Cai (2002), "Adaptive Varying-Coefficient Linear Models", *Journal of the Royal Statistical Society Series B,* forthcoming.

Fong, W.M. and Ouliaris, S. (1995), "Spectral Tests of the Martingale Hypothesis for Exchange Rates", *Journal of Applied Econometrics*, 10, 255-271.

Fong, W.M., S.K. Koh and S. Ouliaris (1997), "Joint Variance Ratio Tests of the Martingale Hypothesis for Exchange Rates", *Journal of Business and Economic Statistics,* 15, 51-59.

Franses, P.H. and P. van Homelen (1998), "On Forecasting Exchange Rates Using Neural Networks", *Applied Financial Economics*, 8, 589-596.

Gençay, R. (1999), "Linear, Nonlinear and Essential Foreign Exchange Rate Prediction with Simple Technical Trading Rules", *Journal of International Economics*, 47, 91-107.

Granger, C.W.J. (1999), *Empirical Modeling in Economics: Specification and Evaluation,* Cambridge University Press: London.

Granger, C.W.J. (2001), "Thick Models", Department of Economics, UC San Diego.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion,* Springer-Verlag: New York.

Hong, Y. (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach", *Journal of the American Statistical Association*, 84, 1201-1220.

Hsieh, D.A. (1988), "The Statistical Properties of Daily Foreign Exchange Rates: 1974-1983", *Journal of International Economics*, 24, 129-145.

Hsieh, D.A. (1989), "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates", *Journal of Business*, 62, 339-368.

Hsieh, D.A. (1993), "Implications of Nonlinear Dynamics for Financial Risk Management," *Journal of Financial and Quantitative Analysis*, 28, 41-64.

Knox, T., J.H. Stock, and M.W. Watson (2000), "Empirical Bayes Forecasts of One Time Series Using Many Predictors," Working Paper, Department of Economics, Harvard University.

Kräger, H. and P. Kugler (1993), "Nonlinearity in Foreign Exchange Markets: A Different Perspective", *Journal of International Money and Finance*, 12, 195-208.

Kuan, C.-M. and T. Liu (1995), "Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks", *Journal of Applied Econometrics*, 10, 347-364.

LeBaron, B. (1999), "Technical Trading Rule Profitability and Foreign Exchange Intervention", *Journal of International Economics*, 49(1), 125-143.

Lee, T.-H., H. White and C.W.J. Granger (1993), "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests", *Journal of Econometrics*, 56, 269-290.

Levich, R. and L. Thomas (1993), "The Significance of Technical Trading Rule Profits in the Foreign Exchange Market: A Bootstrap Approach", *Journal of International Money and Finance*, 12, 451-474.

Lisi, F. and A. Medio (1997), "Is a Random Walk the Best Exchange Rate Predictor?" *International Journal of Forecasting*, 13, 255-267.

Liu, C.Y. and J. He (1991), "A Variance Ratio Test of Random Walks in Foreign Exchange Rates", *Journal of Finance*, 46, 773-785.

Liu, R.Y. (1988), "Bootstrap Procedures under Some Non-iid Models", *Annals of Statistics*, 16, 1697-1708.

Lo, A.W. and A.C. MacKinlay (1988), "Stock Market Prices Do not Follow Random Walks: Evidence from a Simple Specification Test", *Review of Financial Studies*, 1, 41-66.

Lo, A.W. and A.C. MacKinlay (1999), *A Non-Random Walk Down Wall Street*, Princeton University Press: Princeton.

Lucas, R.E. (1978), "Asset Prices in an Exchange Economy", *Econometrica,* 46, 1429-1446.

McCracken, M.W. (2000), "Robust Out-of-sample Inference", *Journal of Econometrics,* 99(2), 195-223.

McCurdy, T.H. and I.G. Morgan (1987), "Tests of the Martingale Hypothesis for Foreign Currency Futures with Time-Varying Volatility", *International Journal of Forecasting*, 3, 131-148.

McCurdy, T.H. and I.G. Morgan (1988), "Testing the Martingale Hypothesis in Deutsche Mark Futures with Models Specifying the Form of Heteroskedasticity", *Journal of Applied Econometrics*, 3, 187-202.

Meese, R.A. and K. Rogoff (1983a), "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample", *Journal of International Economics*, 14, 3-24.

Meese, R.A. and K. Rogoff (1983b), "The Out of Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?" in: J. Frenkel, ed., *Exchange Rates and International Economics*, University of Chicago Press: Chicago.

Meese, R.A. and A.K. Rose (1990), "Non-linear, Non-parametric, Non-essential exchange rate estimation", *American Economic Review*, 80, 192-196.

Meese, R.A. and A.K. Rose (1991), "An Empirical Assessment of Nonlinearities in Models of Exchange Rate Determinations", *Review of Economic Studies*, 58, 603-619.

Pesaran, M.H. and A. Timmermann (1992), "A Simple Nonparametric Test of Predictive Performance," *Journal of Business and Economic Statistics*, 10, 461-465.

Politis, D. N. and J. P. Romano (1994), "The Stationary Bootstrap", *Journal of the American Statistical Association,* 89, 1303-1313.

Priestley, M.B. (1980), "State-dependent Models: A General Approach to Nonlinear Time Series Analysis", *Journal of Time Series Analysis,* 1, 47-71.

Richardson, M. (1993), "Temporary Components of Stock Prices: A Skeptic's View", *Journal of Business and Economic Statistics,* 11(2), 199-207.

Richardson, M. and J.H. Stock (1989), "Drawing Inferences from Statistics Based on Multiyear Asset Returns", *Journal of Financial Economics*, 25, 323-348.

Stock, J.H. and M.W. Watson (1999), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series", in *Cointegration, Causality, and Forecasting, A Festschrift in Honor of C.W.J. Granger*, (eds.) R.F. Engle and H. White, Oxford University Press: London, pp.1-44.

Sweeney, R.J. (1986), "Beating the Foreign Exchange Market", *Journal of Finance,* XLI, 163-182.

Tong, H. (1990), *Nonlinear Time Series: A Dynamic System Approach.* Clarendon Press: Oxford.

West, K.D. (1996), "Asymptotic Inference about Predictive Ability", *Econometrica,* 64 1067-1084.

West, K.D. and D. Cho (1995), "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics,* 69, 367-391.

White, H. (2000), "A Reality Check for Data Snooping", *Econometrica*, 68(5), 1097-1126.

Wu, C.F.J. (1986), "Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis", *Annals of Statistics*, 14, 1261-1350.

Yang, Y. (2002), "Combining Forecasting Procedures: Some Theoretical Results", Working Paper, Department of Statistics, Iowa State University.

Yang, Y. and H. Zou (2002), "Combining Time Series Models for Forecasting", Working Paper, Department of Statistics, Iowa State University.

**TABLE 2. Sizes of Generalized Spectral Tests
for the Null Hypothesis of IID Using Naive Bootstrap**

DGP: $Y_t = e_t$ where e_t is IID N(0, 1)

	$n = 100$									$n = 200$								
α	0.10			0.05			0.01			0.10			0.05			0.01		
$\bar{p}$	6	10	15	6	10	15	6	10	15	6	10	15	6	10	15	6	10	15
$M(0,0)$	.086	.080	.086	.038	.038	.034	.004	.004	.004	.094	.092	.094	.046	.044	.048	.016	.018	.018
$M(1,0)$	.080	.088	.088	.040	.042	.042	.006	.006	.008	.102	.102	.110	.064	.064	.062	.004	.010	.014
$M(1,1)$	.100	.092	.088	.042	.044	.044	.008	.008	.014	.102	.094	.088	.042	.042	.052	.006	.006	.006
$M(1,2)$	.080	.076	.072	.030	.036	.026	.006	.006	.010	.102	.104	.110	.042	.050	.056	.006	.004	.004
$M(1,3)$	.086	.092	.086	.044	.048	.046	.004	.006	.008	.110	.110	.106	.040	.046	.038	.006	.006	.004
$M(1,4)$	.076	.078	.076	.028	.028	.030	.004	.006	.006	.108	.112	.120	.050	.052	.058	.006	.004	.004
$M(2,0)$	.096	.084	.096	.044	.050	.044	.006	.006	.006	.102	.108	.114	.052	.046	.054	.016	.016	.010
$M(2,1)$	.074	.072	.076	.038	.042	.042	.008	.006	.008	.114	.110	.106	.054	.064	.056	.008	.004	.004
$M(2,2)$	.112	.104	.102	.048	.048	.042	.016	.014	.012	.094	.084	.084	.046	.040	.040	.012	.012	.010
$M(3,0)$	.108	.106	.114	.050	.060	.058	.008	.008	.006	.082	.086	.092	.056	.056	.060	.012	.010	.008
$M(3,3)$	.104	.096	.088	.054	.044	.056	.012	.014	.018	.106	.098	.086	.050	.054	.052	.006	.010	.006
$M(4,0)$	.092	.082	.092	.046	.048	.044	.014	.018	.016	.096	.098	.096	.048	.050	.044	.016	.010	.012
$M(4,4)$	.086	.090	.086	.046	.046	.044	.010	.010	.010	.082	.084	.084	.044	.048	.048	.006	.006	.006
Joint12	.096	.102	.104	.050	.054	.060	.016	.014	.014	.114	.124	.118	.058	.056	.062	.018	.014	.010
Joint13	.100	.102	.098	.054	.058	.060	.018	.016	.016	.116	.120	.116	.054	.062	.058	.014	.012	.012

Notes: We compute bootstrap p-values using 300 bootstrap replications, and the empirical rejection rates are based on 500 Monte Carlo replications. We have computed the bootstrap p-values for preliminary lag order $\bar{p} = 6, \ldots, 15$, but report only for three values of $\bar{p} = 6, 10, 15$. The data $\{Y_t\}_{t=1}^{n}$ are generated from an IID N(0,1) distribution via the GAUSS Window version 3.2.38 pseudo random number generator RNDN. Joint12 and Joint13 are the joint tests for $M_{\text{IID}}^{12} = (M_2, \ldots, M_{13})'$ and $M_{\text{IID}}^{13} = (M_1, \ldots, M_{13})'$ respectively, as described in Section II.C.

**TABLE 3. Sizes of Generalized Spectral Tests
for the Null Hypothesis of MDS Using Wild Bootstrap**

DGP: $Y_t = \sigma_t e_t$ where $\sigma_t^2 = 0.1 + 0.2e_{t-1}^2 + 0.7\sigma_{t-1}^2$ and e_t is IID N(0, 1)

	$n = 100$									$n = 200$								
α	0.10			0.05			0.01			0.10			0.05			0.01		
$\bar{p}$	6	10	15	6	10	15	6	10	15	6	10	15	6	10	15	6	10	15
$M(1,0)$	.088	.096	.096	.044	.048	.046	.012	.012	.012	.074	.080	.086	.054	.050	.052	.012	.012	.008
$M(1,1)$	.084	.090	.090	.044	.044	.042	.012	.014	.014	.124	.108	.108	.040	.050	.054	.010	.006	.004
$M(1,2)$	.130	.134	.138	.070	.074	.074	.014	.014	.014	.132	.130	.132	.062	.064	.064	.010	.010	.012
$M(1,3)$	.110	.104	.098	.056	.052	.052	.006	.006	.004	.162	.156	.156	.046	.046	.052	.000	.000	.000
$M(1,4)$	.108	.106	.100	.036	.042	.042	.004	.006	.008	.136	.132	.126	.048	.052	.056	.002	.002	.004
Joint4	.130	.116	.112	.060	.052	.054	.016	.016	.014	.116	.116	.126	.044	.046	.044	.006	.006	.008
Joint5	.106	.100	.102	.054	.050	.050	.020	.016	.016	.098	.106	.114	.036	.034	.040	.004	.006	.006

Notes: We compute bootstrap p-values using 300 bootstrap replications, and the empirical rejection rates are based on 500 Monte Carlo replications. We have computed the bootstrap p-values for preliminary lag order $\bar{p} = 6, \ldots, 15$, but report only for three values of $\bar{p} = 6, 10, 15$. The data $\{Y_t\}_{t=1}^n$ are generated from a GARCH(1,1) process $Y_t = \sigma_t e_t$, where $\sigma_t^2 = 0.1 + 0.2e_{t-1}^2 + 0.7\sigma_{t-1}^2$ and $\{e_t\}$ is an IID N(0,1) innovation generated via the GAUSS pseudo random number generator RNDN. We generate $n + 1000$ observations (with $n = 100$ or 200), and then discard the first 1000 ones to alleviate the impact of using some initial values. Joint4 and Joint5 are the joint tests for $M_{\text{MDS}}^4 \equiv (M_3, M_4, M_5, M_6)'$ and $M_{\text{MDS}}^5 \equiv (M_2, M_3, M_4, M_5, M_6)'$ respectively, as described in Section II.C.

TABLE 4. Generalized Spectral Tests: Bootstrap p-Values for Five Currencies

	CD			DM			BP			JY			FF		
	statistic	P_B	P_W	statistic	P_B	P_W	statistic	P_B	P_W	statistic	P_B	P_W	statistic	P_B	P_W
$M(0,0)$	12.283	.000		10.717	.000		6.121	.002		20.243	.000		10.601	.000	
$M(1,0)$	3.187	.012	.026	4.798	.000	.008	5.720	.002	.002	10.175	.000	.000	4.572	.000	.008
$M(1,1)$	2.396	.028	.090	1.235	.146	.180	0.769	.158	.382	8.152	.000	.002	1.306	.116	.178
$M(1,2)$	-0.099	.410	.674	-0.356	.444	.704	-0.496	.608	.872	11.351	.000	.016	0.978	.132	.302
$M(1,3)$	-0.012	.340	.686	-0.187	.420	.616	2.463	.050	.414	5.695	.004	.100	-0.170	.424	.610
$M(1,4)$	1.730	.060	.294	0.829	.158	.346	2.340	.048	.466	9.076	.000	.068	1.352	.094	.282
$M(2,0)$	10.283	.000		7.488	.000		5.023	.000		4.074	.010		7.685	.000	
$M(2,1)$	5.264	.004		0.539	.302		1.964	.058		3.582	.020		3.201	.028	
$M(2,2)$	15.369	.002		18.942	.000		35.280	.000		13.752	.000		14.148	.002	
$M(3,0)$	2.311	.034		1.748	.098		4.086	.006		-0.482	.664		2.041	.062	
$M(3,3)$	-0.651	.332		5.585	.028		14.932	.012		4.150	.020		7.006	.018	
$M(4,0)$	2.800	.020		2.691	.050		1.210	.104		-0.699	.760		2.414	.030	
$M(4,4)$	-0.710	.250		4.467	.030		17.971	.014		-0.050	.080		7.370	.022	
Joint4			.168			.310			.482			.004			.432
Joint5			.096			.060			.012			.000			.086
Joint12		.000			.000			.000			.000			.002	
Joint13		.000			.000			.000			.000			.002	

Notes: P_B denotes the naive bootstrap p-values and P_W denotes the wild bootstrap p-values, both are based on 500 bootstrap replications. We have computed the bootstrap p-values for $\bar{p} = 6, \ldots, 15$, but reported only for $\bar{p} = 10$. The joint tests are defined in Tables 2 & 3. For comparison, all M(m,l) statistics are asymptotically one-sided $N(0,1)$ and thus upper-tailed asymptotic critical values are 1.65 and 2.33 at the 5% and 1% levels, respectively.

TABLE 6. Predictive Ability Tests for Five Currencies
Panel A. $(R, P) = (613, 612)$
Canadian Dollar (CD)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	0.366			0.469			0.000			0.492		
1	AR(2)	0.368	.775	.775	0.468	.340	.340	-0.029	.838	.838	0.480	.678	.678
2	PN(2, 4)	0.374	.948	.850	0.470	.716	.542	-0.042	.899	.937	0.462	.858	.831
3	NN(2, 5)	0.365	.412	.751	0.469	.555	.738	0.014	.343	.584	0.510	.255	.455
4	FC(2, 26)	0.544	.946	.867	0.500	.974	.860	0.023	.213	.522	0.516	.173	.406
5	Combined(1-4)	0.365	.056	.841	0.467	.005	.723	0.009	.410	.546	0.510	.234	.430
6	MATTR(26)							-0.010	.586	.579	0.497	.390	.470
7	Buy&Hold							0.028	.042	.509	0.479	.774	.486

Deutschemark (DM)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	2.232			1.142			0.016			0.493		
1	AR(2)	2.238	.669	.669	1.141	.425	.425	0.002	.581	.581	0.490	.540	.540
2	PN(2, 4)	2.267	.826	.864	1.150	.799	.625	0.077	.193	.284	0.518	.150	.199
3	NN(2, 5)	2.312	.951	.936	1.173	1.000	.780	0.004	.522	.375	0.477	.699	.293
4	FC(2, 26)	2.374	.996	.959	1.178	.995	.859	0.042	.335	.425	0.482	.684	.336
5	Combined(1-4)	2.219	.044	.827	1.137	.026	.687	0.038	.376	.444	0.497	.436	.350
6	MATTR(26)							0.021	.485	.485	0.487	.592	.387
7	Buy&Hold							-0.015	.635	.527	0.502	.400	.449

British Pound (BP)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	2.054			1.048			-0.024			0.466		
1	AR(2)	2.071	.939	.939	1.054	.983	.983	-0.092	.846	.846	0.456	.628	.628
2	PN(2, 4)	2.470	.893	.984	1.074	.895	.990	0.019	.290	.432	0.489	.203	.271
3	NN(2, 5)	2.087	.784	.989	1.069	.972	.996	0.057	.121	.292	0.487	.233	.375
4	FC(2, 26)	2.166	.988	.995	1.086	1.000	1.000	0.001	.385	.345	0.490	.171	.409
5	Combined(1-4)	2.053	.392	.951	1.046	.085	.851	0.041	.199	.368	0.479	.295	.422
6	MATTR(26)							-0.037	.560	.419	0.475	.350	.471
7	Buy&Hold							0.005	.364	.472	0.544	.003	.015

Japanese Yen (JY)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	2.479			1.154			0.040			0.482		
1	AR(2)	2.476	.433	.433	1.149	.285	.285	0.059	.412	.412	0.520	.049	.049
2	PN(2, 4)	3.203	.917	.754	1.199	.911	.640	0.100	.233	.310	0.528	.019	.047
3	NN(2, 5)	2.662	.998	.855	1.198	.996	.758	0.100	.237	.413	0.515	.116	.083
4	FC(2, 26)	2.528	.808	.901	1.160	.687	.816	0.104	.257	.432	0.529	.043	.095
5	Combined(1-4)	2.467	.178	.876	1.146	.040	.706	0.136	.105	.279	0.539	.011	.043
6	MATTR(26)							0.108	.226	.326	0.539	.025	.055
7	Buy&Hold							-0.040	.733	.389	0.515	.205	.104

French Franc (FF)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	2.094			1.103			-0.029			0.469		
1	AR(2)	2.102	.694	.694	1.101	.338	.338	0.093	.031	.031	0.525	.004	.004
2	PN(2, 4)	2.074	.176	.296	1.101	.367	.454	0.075	.082	.066	0.515	.038	.012
3	NN(2, 5)	2.186	.991	.481	1.138	.998	.645	0.027	.274	.117	0.487	.264	.042
4	FC(2, 26)	2.127	.768	.597	1.117	.890	.754	0.098	.052	.123	0.500	.113	.051
5	Combined(1-4)	2.082	.008	.597	1.097	.002	.586	0.076	.103	.132	0.500	.134	.059
6	MATTR(26)							0.037	.203	.154	0.500	.126	.080
7	Buy&Hold							-0.014	.396	.168	0.495	.141	.093

Panel B. $(R, P) = (817, 408)$

Canadian Dollar (CD)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	0.377			0.483			0.071			0.532		
1	AR(2)	0.381	.906	.906	0.484	.633	.633	-0.011	.995	.995	0.498	.916	.916
2	PN(2, 4)	0.387	.934	.926	0.484	.656	.747	0.008	.974	.996	0.512	.734	.879
3	NN(2, 5)	0.385	.920	.988	0.486	.791	.891	-0.024	.995	.997	0.483	.934	.934
4	FC(2, 26)	0.464	.956	.994	0.506	.954	.937	0.012	.959	.997	0.505	.822	.960
5	Combined(1-4)	0.380	.969	1.000	0.485	.921	.987	-0.025	.998	.997	0.483	.970	.970
6	MATTR(26)							0.019	.947	.999	0.500	.855	.979
7	Buy&Hold							0.071	1.000	.881	0.532	1.000	.848

Deutschemark (DM)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	2.273			1.127			-0.021			0.493		
1	AR(2)	2.276	.544	.544	1.126	.385	.385	0.040	.236	.236	0.512	.224	.224
2	PN(2, 4)	2.286	.612	.824	1.133	.672	.633	0.092	.089	.149	0.542	.051	.088
3	NN(2, 5)	2.290	.673	.904	1.136	.807	.772	0.004	.401	.236	0.502	.377	.139
4	FC(2, 26)	2.472	.943	.934	1.158	.932	.838	0.001	.406	.263	0.498	.432	.156
5	Combined(1-4)	2.262	.087	.840	1.126	.197	.839	0.068	.155	.279	0.525	.150	.170
6	MATTR(26)							-0.002	.426	.315	0.488	.530	.205
7	Buy&Hold							0.021	.380	.363	0.502	.383	.245

British Pound (BP)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	1.925			0.973			-0.080			0.449		
1	AR(2)	1.932	.666	.666	0.976	.867	.867	0.054	.369	.369	0.429	.763	.763
2	PN(2, 4)	2.683	.886	.924	1.024	.922	.943	0.044	.094	.127	0.456	.389	.517
3	NN(2, 5)	2.213	.918	.935	1.018	.970	.960	-0.051	.365	.186	0.510	.032	.052
4	FC(2, 26)	1.956	.734	.961	0.986	.901	.977	0.047	.130	.205	0.510	.060	.091
5	Combined(1-4)	1.918	.080	.899	0.968	.022	.812	0.057	.098	.194	0.488	.107	.100
6	MATTR(26)							-0.084	.494	.230	0.480	.160	.115
7	Buy&Hold							-0.031	.348	.278	0.556	.004	.008

Japanese Yen (JY)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	2.392			1.110			0.042			0.468		
1	AR(2)	2.368	.231	.231	1.097	.057	.057	0.021	.577	.577	0.520	.028	.028
2	PN(2, 4)	3.829	.898	.660	1.176	.882	.496	0.052	.452	.522	0.544	.005	.010
3	NN(2, 5)	2.415	.758	.704	1.116	.707	.552	0.161	.097	.196	0.542	.007	.015
4	FC(2, 26)	2.394	.550	.739	1.098	.172	.594	0.160	.126	.221	0.574	.001	.006
5	Combined(1-4)	2.389	.412	.746	1.104	.080	.595	0.065	.432	.231	0.544	.011	.006
6	MATTR(26)							0.134	.182	.273	0.547	.013	.006
7	Buy&Hold							-0.042	.680	.347	0.529	.083	.025

French Franc (FF)

k	Model	MSFE	P_{RC}^1	P_{RC}^2	MAFE	P_{RC}^1	P_{RC}^2	MFTR	P_{RC}^1	P_{RC}^2	MCFD	P_{RC}^1	P_{RC}^2
0	Benchmark	2.158			1.095			-0.059			0.485		
1	AR(2)	2.168	.673	.673	1.093	.338	.338	0.094	.028	.028	0.527	.077	.077
2	PN(2, 4)	2.118	.122	.139	1.088	.209	.218	0.172	.007	.008	0.544	.028	.027
3	NN(2, 5)	2.189	.842	.185	1.103	.828	.349	-0.009	.318	.014	0.495	.396	.046
4	FC(2, 26)	2.218	.886	.300	1.117	.955	.487	0.044	.178	.021	0.507	.269	.079
5	Combined(1-4)	2.151	.096	.300	1.093	.128	.487	0.154	.014	.024	0.532	.066	.086
6	MATTR(26)							0.012	.215	.027	0.500	.303	.099
7	Buy&Hold							0.016	.151	.028	0.505	.184	.107

Notes: (1) The data are weekly series from January 1, 1975 to December 31, 1998, with 1253 observations. After forming the MATTR(26) and lagged variables we have $n + 1 = 1225$ observations, from which R and P observations are used for regressions and predictions, respectively. The rolling scheme is used: we use the fixed window of the past R observations for estimation, and as we move from R to n, older observations are not used in estimation. (2) P_{RC}^1 and P_{RC}^2 denote the stationary bootstrap p-values of White (2000) test, with 1000 bootstrap rereplications and a bootstrap smoothing parameter $q = 0.75$. The other values of q (e.g., $q = 0.25, 0.50$) give similar p-values (not reported but available). P_{RC}^1 is to compare each model k with the martingale model; P_{RC}^2 is to compare the best of the first I models with the martingale model. The I-th number for P_{RC}^2 is the bootstrap reality check p-value for the null hypothesis that the best of the first I models has no superior predictive power over the martingale model. The last number for P_{RC}^2 checks if the best of all the models under comparison has superior predictive ability over the martingale model. (3) The smaller MSFE and MAFE or the larger MFTR and MCFD, the better the predictive ability of a model.