Queen's Economics Department Working Paper No. 843

A Comparison of Risk-Premium Forecasts implied by Parametric versus Nonparametric Conditional Mean Estimators

Thomas H. McCurdy   Thansis Stengos

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

5-1991

A Comparison of Risk-Premium Forecasts implied by Parametric versus Nonparametric Conditional Mean Estimators

by

Thomas H. McCurdy[1] and Thansis Stengos[2]

Discussion Paper #843

August, 1990

This revision May 1991
(forthcoming, **Journal of Econometrics**)

[1]Department of Economics
Queen's University
Kingston, Ontario,
Canada, K7L 3N6

bitnet MCCURDY@QUCDN
phone (613) 545-2292
fax (613) 545-6668

[2]Department of Economics
University of Guelph
Guelph, Ontario,
Canada, N1G 2W1
and
European University Institute
Firenze, Italy

We would like to thank, without implicating, the participants at the 1990 Canadian Econometrics Study Group and also Anil Bera, Michael Durland, Allan Gregory, Ieuan Morgan, Adrian Pagan, Peter Pauly, Gregory Smith and Tony Wirjanto for helpful comments. Financial support from the Social Sciences and

A Comparison of Risk-Premium Forecasts implied by Parametric versus Nonparametric Conditional Mean Estimators

by

Thomas H. McCurdy[1] and Thansis Stengos[2]

Discussion Paper #843

August, 1990

This revision May 1991
(forthcoming, **Journal of Econometrics**)

[1]Department of Economics
Queen's University
Kingston, Ontario,
Canada, K7L 3N6

bitnet MCCURDY@QUCDN
phone (613) 545-2292
fax (613) 545-6668

[2]Department of Economics
University of Guelph
Guelph, Ontario,
Canada, N1G 2W1
and
European University Institute
Firenze, Italy

We would like to thank, without implicating, the participants at the 1990 Canadian Econometrics Study Group and also Anil Bera, Michael Durland, Allan Gregory, Ieuan Morgan, Adrian Pagan, Peter Pauly, Gregory Smith and Tony Wirjanto for helpful comments. Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

Abstract

This paper computes parametric estimates of a time-varying risk premium model and compares the one-step-ahead forecasts implied by that model with those given by a nonparametric kernel estimator of the conditional mean function. The conditioning information used for the nonparametric analysis is that implied by the theoretical model of time-varying risk. Thus, the kernel estimator is used, in conjunction with a nonparametric diagnostic test for in-sample residual nonlinear structure, to assess the adequacy of the parametric model in capturing any structure in the excess returns.

Our results support the parametric specification of an asset pricing model in which the conditional beta is the ratio of the relevant components of the conditional covariance matrix of returns modelled as a bivariate generalized ARCH process. Although the predictable component of the conditional moments is relatively small, the parametric estimator of the risk premia has somewhat more out-of-sample forecasting ability than does the kernel estimator. Hence, the superior in-sample performance of the latter may be attributed to overfitting.

1. Introduction

The issue of forecastability of out-of-sample values of the conditional mean of asset returns occupies a large literature. Linear unpredictability has long been maintained as an implication of efficient markets, although predictability of returns could also be consistent with the efficient markets hypothesis if it reflects time-varying risk. In practice, the difficulty in finding a model of first moments which out-forecasts a random walk (Meese and Rogoff (1983) provide evidence of this type for the exchange rate case), has frequently led to a martingale process being maintained as the preferred model of asset price dynamics.

Volatility clustering has also been recognized in the literature at least since Mandelbrot (1963). The ARCH model of Engle (1982) presented a parsimonious structure with which to model this time-varying volatility and has led to substantial evidence of conditional heteroskedasticity associated with various financial asset returns (for references, see the recent survey by Bollerslev, Chou and Kroner (1990)). In addition, some part of this time-varying volatility is generally predictable (for example, Engle, Hong and Kane (1990), Pagan and Schwert (1990)).

The above discussion suggests that while the second moment is predictable, at least in part, the first moment may not be. In fact, Diebold and Nason (1990) present evidence that temporal dependence in nominal exchange rates is due to persistence in the conditional variance which is not exploitable for point prediction of the first moment. However, in the case of equity returns, there is some evidence (for example, Fama (1990) and references therein) that there may be a component of returns which is predictable.

While temporal dependence in higher-order moments could be structural,[1] it is well-known that it could also be the result of misspecification of the conditional mean. Indeed both sources could be present. Suppose there is temporal dependence in the second moments and also a time-varying risk premium such that the conditional mean is a nonlinear function of conditional second moments. In this case, modelling the conditional mean as a linear function of the information set will ignore predictable nonlinear components.

[1] For example, Gallant, Hsieh and Tauchen (1989) allow serial dependence in the mixing variable (information arrival) of a mixture model for returns. This generates dependence in higher-order conditional moments of the returns.

Of course, the class of potential nonlinear alternatives is large. Therefore, this paper computes parametric estimates of a time-varying risk premium model and assesses its adequacy in capturing the structure of asset returns, both within-sample and for out-of-sample forecasts. In particular, two types of nonparametric diagnostic testing are pursued. The first involves within-sample testing for the presence of residual nonlinear structure, and the second compares the one-step-ahead forecasts implied by the parametric model with those given by a nonparametric kernel estimator (Robinson (1983)) of the conditional mean function.

The within-sample evaluation includes a nonparametric test developed by Brock, Dechert and Sheinkman (1987), hereafter referred to as the BDS test, which has been applied recently as a diagnostic for the presence of nonlinear structure in asset prices data (for example, Hsieh (1989), Prescott and Stengos (1989)). This test may detect structure which is not captured by more traditional diagnostics which look for deviations from the null of an i.i.d. process in specific directions. This diagnostic is also used to ensure that the kernel specification captures the nonlinearity of the error structure adequately within sample.

The second stage, which compares the forecasts implied by the parametric versus the nonparametric estimators of the conditional mean, investigates whether the identified structure has any forecastable component.[2] The conditioning information used for the nonparametric analysis is that implied by the theoretical model of time-varying risk. Therefore, unlike standard implementations of the kernel estimator for conditional means, we add the squares and cross-products of the conditioning variables in order to reflect the potential contribution of the conditional variance and covariance to the conditional mean through time-varying risk. This transformation of the regression function is introduced as a simple method of bias reduction.

Given its flexibility within the sample, the kernel estimator of the conditional mean constitutes a benchmark for the parametric model. On the other hand, any tendency for the nonparametric structure to over-fit will be evaluated by a comparison of the out-of-sample forecasts. Commonly kernel estimates are used in two-step estimation procedures (for example, Pagan and Hong (1989)). In our case, the two methods are self-contained and provide

---

[2] In this paper we are concerned with point forecasts and not with the implications of the time-varying conditional variances for forecast confidence intervals, etc., as discussed in Baillie and Bollerslev (1990) and references therein.

competing specifications.

Our empirical application is to a sample of monthly equity returns from 1970 to 1988 inclusive. The conditional-beta capital asset pricing model (for example, Bollerslev, Engle and Wooldridge (1988), Harvey (1991), Mark (1988) and McCurdy and Morgan (1991a)) is used to price time-varying systematic risk for an equity portfolio with respect to an international benchmark portfolio.

Section 2 reviews the asset pricing paradigm used to evaluate equilibrium returns. Section 3.1 outlines the methods used for in-sample estimation, including the test equation system for the maximum likelihood estimation in subsection 3.1.1 and that for the nonparametric kernel estimation in 3.1.2. Subsection 3.1.3 reviews the BDS nonparametric test for residual nonlinear dependence. The results for the parametric and nonparametric estimators are presented in 3.2.1 and 3.2.2 respectively, while the out-of-sample forecast comparisons are summarized in section 4. Concluding comments are offered in section 5.

2. A Model of Time-Varying Risk Premia

2.1 Notation and data sources

We assume perfect markets in a single good, pure exchange model with a representative consumer. As in Lucas (1982) and Hansen and Hodrick (1983), all prices are expressed in units of domestic currency (U.S. dollars) and the interest rates are nominal. Let

C_t = the number of units of the good consumed at t,

p_t = the price per unit of the consumption good at t,

$M_{t-1,t}$ = the intertemporal marginal rate of substitution of domestic currency between time t-1 and time t,

$R_{B,t}$ = one plus the monthly rate of return on a benchmark portfolio which is conditionally mean-variance efficient,

$R_{j,t}$ = one plus the monthly rate of return on the Morgan Stanley Capital International (MSCI) Japanese equity index in U.S.$ with net dividends reinvested (source: MSCI),

$R_{w,t}$ = one plus the monthly rate of return on the MSCI World equity index in U.S.$ with net dividends reinvested (source: MSCI),

R_{t-1} = one plus the rate of return on U.S. Treasury Bills for the month t computed from the average of bid and ask prices for the TB with maturity closest to 30 days on the last trading day of month t-1 (source: CRSP Risk Free Rate File).

2.2 Evaluation of equity portfolio returns

Subtracting the equilibrium condition for a one dollar investment in a one-period (nominally riskfree) bond from that associated with a position in the Japanese equity portfolio, and expanding using the definition of covariance, yields

$$E_{t-1}[R_{j,t}] - R_{t-1} = - R_{t-1}\ cov_{t-1}[M_{t-1,t}\ ,\ R_{j,t}] \quad , \qquad (1)$$

in which E_{t-1} refers to expectations conditional on information at time t-1.

Note that the expected nominal return on the equity position in excess of the (nominally) riskfree rate, will be zero under risk neutrality and a deterministic price level.[3] Alternatively, this ex ante risk premium is a function of the conditional covariance of the (nominal) intertemporal marginal rate of substitution and the equity return.

For empirical implementation, we use a conditional-beta representation of the consumption-based asset pricing relation (1). Following Hansen and Richard (1987), our single-beta asset pricing relation will be expressed in terms of a benchmark portfolio which is hypothesized to be on the *conditional* mean-variance frontier. For example, if there exists an asset or portfolio return R_m which is perfectly conditionally correlated with the intertemporal marginal rate of substitution, then portfolios which yield returns R_B which are a linear combination of R_m and the riskfree return will be conditionally mean-variance efficient. This implies that the equilibrium expected return on any asset will be a function of its conditional beta with that benchmark portfolio and we can re-express (1) as the conditional single-beta asset pricing relation,

$$E_{t-1}[R_{j,t}] - R_{t-1} = \frac{cov_{t-1}[R_{j,t}\ ,\ R_{B,t}]}{var_{t-1}[R_{B,t}]} (E_{t-1}[R_{B,t}] - R_{t-1}) \ . \qquad (2)$$

Several approaches to measuring such a benchmark portfolio have been proposed. For example, Campbell (1987), Engle, Ng and Rothschild (1990), Hansen and Hodrick (1983), and Giovannini and Jorion (1987), treat the benchmark portfolio as unobservable and use either a latent variable approach or factor representing portfolios to estimate the benchmark portfolio returns. Breeden, Gibbons and Litzenberger (1989) proceed by constructing a portfolio which has returns which are maximally correlated with the growth rate of consumption. McCurdy and Morgan (1991b) use benchmark portfolios for both consumption (a maximum correlation portfolio) and wealth (a world equity portfolio) in an empirical implementation motivated by non-expected utility explanations of asset prices.

In this paper, we use the return on the MSCI world equity index as the benchmark portfolio return, replacing R_B in (2) by R_w . Choosing an

---

[3] A stochastic price level will affect the nominal equity premium even under risk neutrality. For more details see, for example, Engel (1990) and Labadie (1989).

equity-based index, as in Mark (1988) and Harvey (1991), is clearly open to the Roll (1977) critique. Nevertheless, the MSCI world index represents extensive international diversification. The issue of whether or not this choice for the benchmark portfolio in the single-beta formulation of the conditional pricing relation (2) is adequate to price all the relevant risk will be addressed further during our evaluation of the empirical model.

3. In-Sample Estimation

3.1 Methods

3.1.1 Parametric test equation system for Maximum Likelihood Estimation

The conditional asset pricing relation (2) specifies that the excess returns on the Japanese equity portfolio, $R^*_{j,t} \equiv (R_{j,t} - R_{t-1})$, are expected to be proportional to the excess returns on the benchmark portfolio. The latter are represented in our model by excess returns on the world equity portfolio, $R^*_{w,t} \equiv R_{w,t} - R_{t-1}$. The time-varying proportionality factor is the conditional beta which is a function of the conditional second moments of the joint returns process. This specification suggests a bivariate model which jointly estimates the first and second conditional moments of those returns. In this section, we briefly outline the test equation system used for our quasi-maximum likelihood estimation (QMLE) of (2).

Maintaining rational expectations, we replace expected values in (2) by realized values minus forecast errors. The rational expectations assumption implies that forecast errors have conditional means of zero. Using the notation: $h_{j,t}$ for the conditional variance of $R^*_{j,t}$; $h_{w,t}$ for the conditional variance of $R^*_{w,t}$; $h_{jw,t}$ for the conditional covarance between $R^*_{j,t}$ and $R^*_{w,t}$; μ for a parameter that can be set at zero in order to exclude the risk premium term from the model; $x_{j,t-1}$ and $x_{w,t-1}$ for vectors of explanatory variables known at time t-1; the system of test equations is, analogous to that in McCurdy and Morgan (1991a),

$$R^*_{j,t} = \gamma_j' x_{j,t-1} + \mu \frac{h_{jw,t}}{h_{w,t}} [\gamma_w' x_{w,t-1} + \psi_w \varepsilon_{w,t-1}] + \varepsilon_{j,t}, \qquad (3)$$

$$R^*_{w,t} = \gamma_w' x_{w,t-1} + \psi_w \varepsilon_{w,t-1} + \varepsilon_{w,t}, \qquad (4)$$

$$\varepsilon_t' \mid I_{t-1} \sim N(0, H_t),$$

that is, conditional on information at t-1, the errors $\varepsilon_t' \equiv [\varepsilon_{j,t}\ \varepsilon_{w,t}]$ are hypothesized to be normally distributed with covariance matrix

$$\mathbf{H}_t = \mathbf{C}'\mathbf{C} + \mathbf{A}'\ \varepsilon_{t-1}\varepsilon_{t-1}'\ \mathbf{A} + \mathbf{B}'\ \mathbf{H}_{t-1}\ \mathbf{B}\ , \qquad (5)$$

in which **C**, **A** and **B** are symmetric parameter matrices.

The bivariate specification of the conditional means, given by (3) and (4), includes the vectors of potential explanatory variables $x_{j,t-1}$ and

$x_{w,t-1}$. Except for the intercepts, which are included in the estimated model, those variables are primarily used for the omitted variable tests discussed below. For example, in the case of (3), x_j is used test for the potential importance of variables which might have explanatory power under alternative specifications of the time-varying risk premium model. Therefore, the estimated model for R_j^* includes an intercept and the conditional risk premium.

For the case of the world index excess return in (4), in addition to an intercept, x_w includes a domestic dividend yield variable[4] in excess of the U.S. riskfree rate, $DY_{d,t-1}^*$, which has been shown to have predictive value for U.S. equity returns (see, for example, Fama (1990) and references therein) and for the MSCI world equity index return (Harvey (1991)). An MA(1) term has also been included in (4) in order to capture any serial dependence due to nonsynchronous trading of the components of the index (for example, French, Schwert and Stambaugh (1987) and Chou (1988)).

Time-variation in the conditional second moments of financial data has been extensively documented (for references, see the recent survey by Bollerslev, Chou and Kroner (1990)). We parameterize the variance-covariance structure in (5) using the Baba, Engle, Kraft and Kroner (1989) form of the generalized ARCH structure (Engle (1982), Bollerslev (1986)). This structure ensures that the conditional variance-covariance matrix is positive definite and is also relatively parsimonious with respect to the number of parameters. When conducting omitted variable tests, the structure in (5) is augmented by the inclusion of variables from the information set at time t-1.

This empirical specification of the test equation system allows the risk premia to vary as a result of time variation in the expected benchmark portfolio returns and also time variation in both the variance and covariance components of the equity beta. This also means that the price of covariance risk is allowed to vary -- unlike most implementations of vector GARCH-M models of risk premia. QMLE is implemented with standard errors computed to allow robust inference in the presence of potential departures from conditional normality[5] (Bollerslev and Wooldridge (1988), Weiss (1986), White

---

[4] Computed, as in Fama (1990), by summing monthly dividends associated with the value-weighted NYSE portfolio for the twelve months preceding t-1 and dividing by the value of the portfolio at t-1.

[5] Potential loss of efficiency from such QMLE has been evaluated by Engle and Gonzalez-Rivera (1990) who propose a more efficient semiparametric approach.

(1982)).[6]

3.1.2 Nonparametric Kernel Estimation

Silverman (1986) presents a general introduction to nonparametric density estimation while Ullah (1988) focuses on nonparametric estimation of econometric functionals. The conditional mean of a random variable y, given a vector of conditioning variables x, can be written as $E(y|x) = M(x)$. In parametric estimation M(x) is typically assumed to be linear in x but in the nonparametric approach M(x) remains a general functional form.

Consider now the time series process $\{y_t\}$ and in particular the problem of estimating the mean of y_t conditional on $(y_{t-1},...,y_{t-p})$. Robinson (1983) derives the asymptotic distribution of the nonparametric kernel estimator of the joint density of the time series data generating process (DGP) of (y_t,x_t) and of the conditional mean of y_t, given $(y_{t-1},...,y_{t-p},x_t,...,x_{t-p})$. Let f(z) be the multivariate DGP of the p+1 dimensional random vector z, which we write as $z = (y_t,x_t)$, where $x_t = (y_{t-1},...,y_{t-p})$. The kernel estimator of the joint density is

$$\hat{f}(y,x) = n^{-1}h^{-(p+1)}\sum_{t=1}^{n} K\left(\frac{y-y_t}{h},\frac{x-x_t}{h}\right) \tag{6}$$

where the kernel function K satisfies certain conditions, including: $\int K(z^*)dz^* = 1$, $\int |K(z^*)|dz^* < \infty$ and $||z^*||^{p+1}|K(z^*)| \longrightarrow 0$ as $||z^*|| \longrightarrow \infty$ for $z^* = (z - z_t)/h$ and $||z^*||$ as the usual Euclidean norm of z^*. Since it is possible to choose the function K so that it is continuous, the resulting kernel estimator of the density function will also be continuous. In the present paper, the kernel is chosen to be the standard multivariate normal density function.

An important consideration in the literature is the choice of the bandwidth parameter h. Too large a value of h induces bias and too small a value induces imprecise estimates. Robinson (1983) and Ullah (1988), among others, summarize the conditions that the kernel function and the bandwidth parameter h will have to satisfy to obtain the asymptotic properties of the regression function estimator.

One of the ways to ensure bias-reduction for particular choices of h, are the so called "higher-order" kernels, proposed by Bartlett (1963) and introduced in the econometric literature by Robinson (1988) in a semiparametric context. The higher order kernels are of the form

[6] Software used for this QMLE was developed by I.G. Morgan and T.H. McCurdy with the underlying optimization routines of Numerical Algorithms Group.

$K(z) = \sum_{j=0}^{1/2(m-2)} c_j z^2 K(z)$, where K(z) is the Gaussian kernel. The constants c_j satisfy a system of linear equations. However in a pure nonparametric context the higher order kernels require a very large data set for a noticeable bias improvement and at the same time they can introduce a lot of additional variance in the estimates. Robinson (1988) discusses alternative bias-reduction methods that have been proposed in the statistics literature.

In this paper, using a priori information from the economic model, we include the squares and cross products of the regressors in an attempt to linearize the regression function. The reason behind the inclusion of these terms is that if the true regression function is linear and the regressors are uniformly distributed, the kernel estimates will be nearly unbiased.

The estimator of the regression function can de derived to be

$$\hat{E}(y|x) = \sum_{t=1}^{n} y_t r_t \tag{7}$$

where

$$r_t = K_2\left(\frac{x - x_t}{h}\right) \Big/ \sum_{t=1}^{n} K_2\left(\frac{x - x_t}{h}\right),$$

with $K_2(x^*) = \int K(y^*, x^*)dy$, $x^* = (x - x_t)/h$ and $y^* = (y - y_t)/h$. The above expression can be evaluated at any value of x to yield the nonparametric estimator of the regression function. Clearly, out-of-sample forecasts conditional on a set of known x values can be calculated, see for example Moschini, Prescott and Stengos (1988).

For the conditional variance of y given x the kernel estimator is derived to be

$$\hat{V}(y|x) = \sum_{t=1}^{n} y_t^2 r_t(x) [\hat{E}(y|x)]^2 \tag{8}$$

where $\hat{E}(y|x)$ and $r_t(x)$ are defined in (7).

The response or regression coefficient of y with respect to changes in a regressor, say x_j is defined to be $\beta(x) = \partial E(y|x)/\partial x_j$. The kernel estimator of $\beta(x)$, say $\hat{\beta}(x)$, is defined as

$$\hat{\beta}(x) = \sum_{t=1}^{n} y_t (r_{1t} - r_{2t}) \tag{9}$$

where $r_{1t} = K'(x_t^*)/\sum_{t=1}^{n} K(x_t^*)$, $r_{2t} = K(x_t^*)\sum_{t=1}^{n} K'(x_t^*)(\sum_{t=1}^{n} K(x_t^*))^{-2}$ and $K'(x_t^*) = \partial K(x_t^*)/\partial x_j$.

In this paper the fixed response or fixed regression coefficient estimates were obtained by estimating $\hat{\beta}(\bar{x})$, where $\bar{x}$ is the sample mean. However, the speed for adjustment of $\hat{\beta}(\bar{x})$ is $(nh^{p+2})^{1/2}$, which is slower than the usual $n^{1/2}$ rate of adjustment that one obtains with parametric estimation, since $h \longrightarrow 0$ as $n \longrightarrow \infty$. This slower rate of convergence implies that the standard errors of the nonparametric estimates typically turn out to be larger than their corresponding parametric counterparts and test statistics based on them are less efficient. One could obtain $n^{1/2}$ consistency (that is, the parametric rate of convergence) by averaging over all $\hat{\beta}(x)$. However, Ullah (1988) argues that, although $\hat{\beta}(\bar{x})$ may be less efficient than the average derivative estimate, it may be more robust.

In applications, the investigator must choose the window width h as well as the kernel function. The choice of the window width is important since bias is an increasing function of h while variance is a decreasing function of h. Ullah (1988) suggests setting the window width h_i in the following way:

$$h_i = s_i n^{-1/(4+p)}, \text{ for } i=1,\dots,p, \tag{10}$$

where s_i denotes the standard deviation of x_i.

3.1.3 Testing for Nonlinearities

The BDS statistic is designed to test the null hypothesis that a time series is i.i.d. against a variety of alternative nonlinear hypotheses. Below we will discuss briefly its structure and the intuition behind it.

Let $\{y_t$: t=1,2,...,T) be a sequence of observations that are i.i.d.. From this series, construct the m-dimensional vector, or "m-history"

$$y_t^m = \{y_t, y_{t+1}, \dots, y_{t+m-1}\}.$$

Using these m-histories we can compute the following quantity, known as the correlation integral:

$$C_m(\varepsilon) = \lim_{T\to\infty} \frac{2}{T_m(T_m-1)} \sum_{t<s} I_\varepsilon(y_t^m, y_s^m) \tag{11}$$

where $T_m = (T - m + 1)$ and $I_\varepsilon(y_t^m, y_s^m)$ is an indicator function that equals unity if $|y_t - y_s| < \varepsilon$. Here $|.|$ is the supnorm. The correlation integral measures the proportion of the m-dimensional points that are "close" to each other, where "close" is defined in terms of the supnorm criterion.

Given a sample of size T, the following sample correlation dimension statistic can be computed

$$C_m(\varepsilon, T) = \frac{2}{T_m(T_m-1)} \sum_{t<s} I_\varepsilon(y_t^m, y_s^m). \tag{12}$$

Brock, Dechert and Scheinkman (1987) show that if $\{x_t\}$ is i.i.d. with a non degenerate density f(.) then, for fixed m and ε,

$$C_m(\varepsilon, T) \longrightarrow [C_1(\varepsilon)]^m \text{ with probability 1, as } T \longrightarrow \infty.$$

Furthermore,

$$\sqrt{T}(C_m(\varepsilon, T)) - C_1(\varepsilon, T)^m) \longrightarrow N(0, V_m(\varepsilon)).$$

The standardized form of the above is the BDS statistic and it is given by

$$W_m(\varepsilon, T) = \sqrt{T}(C_m(\varepsilon, T) - C_1(\varepsilon, T)^m)/\sqrt{V_m(\varepsilon)}. \tag{13}$$

The derivation of the asymptotic distribution is based on results from the theory of U-statistics, see Serfling (1980).

The parameter ε plays a similar role as the bandwidth does in the case of kernel estimation. There are no results however suggesting an optimal choice of ε. Additionally m is a choice parameter as well. For a given value m, ε should not be too small, otherwise the sample correlation integral will capture too few points. Similarly, ε should not be chosen to be too large. Since there is no unique choice for these two parameters, users report a number of statistics. Although these statistics are not independent, a battery of significant BDS statistics does provide strong evidence against the null hypothesis.

Monte carlo simulations by Brock, Dechert and Scheinkman (1987) provide evidence that the BDS statistic has good power against a variety of nonlinear alternatives. More recently, extensive simulations by Brock, Hsieh and LeBaron (1991) indicate that the BDS statistic has good size and power characteristics even in moderately sized samples. Moreover, the statistic has good power against a wide variety of nonlinear alternatives, including tent map chaotic processes and stochastic processes such as autoregressive, threshold autoregressive, nonlinear moving average and ARCH.

There is a note of caution when one applies the BDS in practice. In empirical work, the BDS is usually applied to residuals from some preliminary estimation of the regression function. The "nuisance-parameter" problem affects the behaviour of the BDS statistic in finite samples and leads in general to an actual size of the test that is greater than the nominal one. The problem persists in larger samples when the residuals come from an ARCH/GARCH model. The BDS in this case lacks power to reject the false model.

3.2 In-Sample Results

3.2.1. Parametric model estimates and evaluation

The first 16.5 years of the sample (2nd month of 1970 to the 6th month of 1986) are used for in-sample estimation. The remaining 2.5 years of data are used to evaluate out-of-sample forecasts. The in-sample quasi-maximum likelihood results are summarized in tables 1 to 3. Table 1 presents the coefficient and robust standard error estimates for the test equation system (3) to (5). Table 2 summarizes the in-sample fit and tests associated with the importance of the time-varying risk term while table 3 reports the results of our statistical evaluation of the model.

The conditional risk premium model maintained in (2) implied that $\gamma_{0j} = 0$

and $\mu = 1$ in (3). The first panel of estimates in table 1 indicate that neither of these restrictions can be rejected on the basis of robust t-tests. The second panel of results, for the excess return on the world portfolio, R_w^*, shows that the estimated coefficients for the intercept γ_{ow} and for the variable DY_d^*, the domestic dividend yield in excess of the riskfree rate, are both significantly positive. The MA(1) term is positive in sign but is insignificantly different from zero.

The final panel of table 1 reports the estimates for the conditional variance-covariance structure. Although the cross-equation restrictions associated with the quadratic structure make it difficult to associate persistence with a particular parameter, there is clearly significant conditional heteroskedasticity for both returns. In addition, the conditional covariance between the two returns is statistically significant. For the in-sample estimates reported in table 1, a likelihood ratio test does not reject a restriction that the off-diagonal elements of A and B, a_{jw} and b_{jw} respectively, are zero. However, the model with those two parameters included did somewhat better out-of-sample, perhaps due to the outliers associated with the market crash in October 1987. For this reason, we did not restrict the coefficient matrices A and B to be diagonal.

Table 2 summarizes evidence concerning the statistical importance of the estimated risk premium in explaining the Japanese equity excess return as well as some summary statistics relating to the in-sample fit of the model reported in table 1. The likelihood ratio test associated with restricting the conditional beta risk premium to be zero ($\mu = 0$ versus μ unrestricted) has a very low p-value indicating that we can convincingly reject the zero risk premium hypothesis. This result, together with the estimates for γ_{oj} and μ in table 1, lend support to the conditional beta formulation of the time-varying risk premia for this sample. Note that the sample average of the ex ante risk premia is about one half as large as the average ex post excess return over the sample.[7] However, the standard deviation of R_j^* is over three times larger than the sample standard deviation of the risk premia. Figure 1 plots the estimated conditional beta for the in-sample period. The beta is clearly time-varying. The estimated price of covariance risk, $E_{t-1}(R_{wt}^*)/h_{w,t}$, is also time-varying. Therefore, it appears to be important to allow each of the

---

[7] If the estimated intercept, $\hat{\gamma}_{oj}$, is added to the average conditional risk premium, this sample mean of the predicted R_j^* is similar in size to that of the ex post R_j^*.

components of the conditional risk premia to vary.

The specification tests reported in the first panel of table 3 do not indicate any statistical problems with the model reported in table 1, except for skewness associated with the residuals from (3). These tests include: a nonparametric runs test; portmanteau statistics: for autocorrelation in the first ten lags of the *standardized* residuals[8]; for remaining heteroskedasticity in the same number of lags of the squared *standardized* residuals; and for neglected heterogeneity using the cross products of the *standardized* residuals of the two series. Although these portmanteau tests may be affected by the presence of predetermined regressors and any remaining time variation in higher-order moments (see, for example, Cumby and Huizinga (1988)), the associated p-values are sufficiently large to suggest that the persistence in the first two conditional moments has been adequately captured. In particular, note that the $Q_{jw}(10)$ result indicates that the parameterization (5) adequately captures any persistence in the conditional covariances. Further evidence that the model specified in (3) to (5) captures the persistence in the data is given in the second and third columns of table 5 where the BDS test rejects the i.i.d. hypothesis for R_j^* but fails to so for the *standardized* residuals from the parametric model.

The bottom panel of table 3 reports results of some OPG LM tests for the potential importance of variables which might have explanatory power under alternative specifications of the time-varying risk premium model or, of course, some other alternative model. While the model passes these tests, it is possible that a more flexible functional form and/or a more general parameterization of the first two conditional moments of the joint returns series might improve the fit of the conditional risk premia.[9] For example, more extensive tests might indicate that the MSCI world equity portfolio excess return is an inadequate measure for R_B^*, or that there are additional risk factors which are not priced by the conditional-beta model specified by (2) -- for example, McCurdy and Morgan (1991c). The nonparametric specification reported in the next section has been designed to alleviate

---

[8] The vector of raw residuals ε_t is standardized as $u_t = H_t^{-1/2}\ \varepsilon_t$ where $H_t^{-1/2}$ is obtained from orthonormal transformation of the conditional covariance matrix H_t. We thank Michael Durland for providing code for this transformation.

[9] Pagan and Schwert (1990) provide evidence that the EGARCH parameterization proposed by Nelson (1991) outperforms the GARCH specification for conditional variances of monthly U.S. stock returns for the sample 1834-1925.

Table 1

Parametric model estimates*

Japanese equity excess returns:

$$R^*_{j,t} = \gamma_{0j} + \mu \frac{h_{jw,t}}{h_{w,t}} \left(\gamma_{0w} + \gamma_{1w}\, DY^*_{d,t-1} + \psi_w\, \varepsilon_{w,t-1}\right) + \varepsilon_{j,t}$$

$\hat{\gamma}_{0j}$	$\hat{\mu}$
0.0068	1.7869
(0.0047)	(0.5186)

Benchmark portfolio excess returns:

$$R^*_{w,t} = \gamma_{0w} + \gamma_{1w}\, DY^*_{d,t-1} + \psi_w\, \varepsilon_{w,t-1} + \varepsilon_{w,t}$$

$\hat{\gamma}_{0w}$	$\hat{\gamma}_{1w}$	$\hat{\psi}_w$
0.0175	0.0534	0.1141
(0.0044)	(0.0184)	(0.0962)

Conditional Variance-Covariance Matrix:

$$\begin{bmatrix} h_{j,t} & h_{jw,t} \\ h_{jw,t} & h_{w,t} \end{bmatrix} = \begin{bmatrix} c_j & c_{jw} \\ c_{jw} & c_w \end{bmatrix} \begin{bmatrix} c_j & c_{jw} \\ c_{jw} & c_w \end{bmatrix} + \begin{bmatrix} a_j & a_{jw} \\ a_{jw} & a_w \end{bmatrix} \begin{bmatrix} \varepsilon^2_{j,t-1} & \varepsilon_{j,t-1}\varepsilon_{w,t-1} \\ \varepsilon_{j,t-1}\varepsilon_{w,t-1} & \varepsilon^2_{w,t-1} \end{bmatrix} \begin{bmatrix} a_j & a_{jw} \\ a_{jw} & a_w \end{bmatrix}$$

$$+ \begin{bmatrix} b_j & b_{jw} \\ b_{jw} & b_w \end{bmatrix} \begin{bmatrix} h_{j,t-1} & h_{jw,t-1} \\ h_{jw,t-1} & h_{w,t-1} \end{bmatrix} \begin{bmatrix} b_j & b_{jw} \\ b_{jw} & b_j \end{bmatrix}$$

$\hat{c}_j$	$\hat{c}_{jw}$	$\hat{c}_w$	$\hat{a}_j$	$\hat{a}_{jw}$	$\hat{a}_w$
0.0325	0.0235	0.0040	0.1029	-0.0701	0.3549
(0.0170)	(0.0039)	(0.0116)	(0.1246)	(0.1017)	(0.0640)

$\hat{b}_j$	$\hat{b}_{jw}$	$\hat{b}_w$
0.7356	-0.0940	0.7975
(0.2535)	(1.6550)	(0.0950)

*Robust standard errors are in parentheses below the coefficient estimates.

Table 2

Evidence concerning fit and risk premia for the model in Table 1

	risk premium$_j$	risk premium$_j$ + intercept	R_j^*	R_w^*
average	0.00606	0.01286	0.01189	0.00367
standard deviation	0.01701	0.01701	0.05891	0.04038
total sum of squares			0.70809	0.32221
residual sum of squares			0.63146	0.29251
$\mu = 0$	15.94			
p-value	(0.000)			

Sample means are expressed as monthly returns. The last row reports the likelihood ratio test statistic associated with restricting the conditional beta risk premium in (5) to be zero ($\mu = 0$ versus unrestricted).

Table 3

Statistical evaluation of the parametric model reported in Table 1

1. Specification checks

	j	jw	w
R	-0.06 (0.95)		-1.17 (0.24)
Q(10)	6.60 (0.76)		11.11 (0.35)
$Q^2(10)$	4.48 (0.92)		5.55 (0.85)
$Q_{jw}(10)$		5.92 (0.82)	
SK	9.19 (0.00)		0.02 (0.89)
KU	0.99 (0.32)		2.52 (0.11)

2. OPG LM tests for variables omitted from $R^*_{j,t}$

	mean	variance
$R^*_{j,t-1}$	0.96 (0.33)	0.26 (0.61)
$R^*_{w,t-1}$	3.20 (0.07)	2.33 (0.15)
R_{t-1}	0.10 (0.75)	0.86 (0.35)
January	1.34 (0.25)	0.09 (0.76)

R is the test statistic for runs above the mean, SK and KU are conditional moment tests for skewness and excess kurtosis, Q(10) the Ljung-Box form of the portmanteau statistic for autocorrelation in the first 10 lags of the standardized residuals, $Q^2(10)$ the same for the squared standardized residuals, and $Q_{jw}(10)$ is the portmanteau statistic for the autocorrelation function for the cross products of the standardized residuals of the two series. The square of the variable listed is used for the OPG tests associated with the variance. The p-values, shown in parenthesis, are for the chi-square distribution except for R where they are for the unit normal two-tailed test.

some restrictions implied by the parametric formulation.

3.2.2. Nonparametric estimates

In this section we evaluate the nonparametric regression and present the in-sample estimates. The dependent variable is R^*_j . The independent variables that enter the kernel regression include the lagged excess return variables and their squares and cross products. In other words, the information sets for the parametric and nonparametric specifications are comparable since they contain similar information. Table 4 presents the estimates of the derivatives of the regression function evaluated at the sample means. Also the standard errors of these estimates are reported. Since these estimates only incorporate information at one point of the sample space they are less efficient than the parametric estimates which incorporate information from the whole sample. Alternatively, one could compute the mean of all the partial derivative estimates, hence using information from the whole sample. Ullah (1988) argues that such mean estimates, although more efficient, are nevertheless less robust than the estimates of these derivatives at the sample means of the regressors. The nonparametric estimates are indeed quite inefficient. Only the lagged dependent variable appears to be significant at the 10% level. The choice of h was proportional to $n^{-1/9}$. Different choices of h, slightly larger or smaller than the above, led to qualitatively similar results.

The BDS statistics of the kernel residuals, reported in table 5, suggest that there is no linear or nonlinear dependence present and that the model is adequately specified. An additional point that needs some emphasis is that the total fit of the kernel regression is quite good, with a considerable reduction of the sum of squared residuals when compared with the parametric regression. However, the in-sample superiority of the nonparametric fit should be viewed with caution, since it does not lead to superior forecasts as will be seen in the next section. Hence, part of the in-sample performance of the kernel regression should be attributed to "over-fitting", a problem that is encountered often in nonparametric regression.

4. Out-of-Sample Forecasts

4.1. Evaluation of parametric versus nonparametric forecasts

To examine the predictive ability of the parametric and nonparametric specifications we generated a sample of 30 out-of-sample, one-step-ahead forecasts. Both the parametric and nonparametric formulations were fitted to a subset of the data (the last 30 observations were deleted) and a single

Table 4

Kernel Estimates at Sample Means

Independent Variables	Coefficient Estimates
$R^*_{w,t-1}$	-0.0012
	(0.0035)
$R^*_{j,t-1}$	0.0060
	(0.0034)
$(R^*_{w,t-1})^2$	-0.0001
	(0.0019)
$(R^*_{j,t-1})^2$	0.0001
	(0.0002)
$(R^*_{w,t-1})(R^*_{j,t-1})$	-0.0017
	(0.0026)
Mean of Dependent Variable	0.01189
total sum of squares	0.70809
residual sum of squares	0.53034

*The numbers below the estimates are standard errors.

Table 5

BDS Statistics from the Parametric and Nonparametric Specifications

ε = normalized standard deviation

Embedding Dimension	Raw Series	Parametric Residuals	Kernel Residuals
3	1.4208	-0.8359	-0.9476
5	2.0378	0.2290	-0.1170
7	2.6651	1.1762	0.2056
9	2.3362	1.6864	0.0529

ε = normalized standard deviation scaled by 1.25

Embedding Dimension	Raw Series	Parametric Residuals	Kernel Residuals
3	1.2306	-0.6748	-0.1122
5	1.7131	-0.1479	-0.6933
7	1.9790	0.4502	0.0021
9	1.8987	0.9165	-0.1709

The BDS statistics are distributed as standard normal variates. The residuals from the parametric formulation are the standardized residuals. The ε is chosen to be proportional to the standard deviation of each of the series divided by its range.

Table 6

Regression of actual values on one-step-ahead forecasts*

	Kernel	Parametric
Constant	0.0215	-0.0089
	(0.0148)	(0.0311)
Forecast	1.2493	1.5370
	(1.1160)	(1.1614)
Mean of Dep. Variable	0.0284	0.0284
total sum of squares	0.1824	0.1824
residual sum of squares	0.1514	0.1489
R^2	0.0428	0.0589
MSE	0.0056	0.0050
LogL	36.7658	37.0191

In-sample fit for the same subsample

residual sum of squares	0.1093	0.1503
R^2	0.3088	0.0500

* Standard errors are shown in parentheses below the coefficient estimates.

one-period-ahead forecast was computed and stored. The estimation sample was then increased by one observation and the models were re-estimated and used to compute a forecast for a single period. In this way we generated the sequence of 30 out-of-sample one-step-ahead forecasts.

The results from regressing the actual values on the forecasts appear in table 6. Both the R^2s from these regressions and the mean squared errors suggest that the parametric forecasts are somewhat superior to the nonparametric ones. This constrasts to the in-sample fit for which the kernel estimator did considerably better. The second panel of table 6 shows the in-sample fit achieved by the two estimators for the same subsample over which the one-step-ahead forecasts were computed. This demonstrates that the poorer performance of the kernel estimator out-of-sample was not due to the characteristics of that particular time period. These comparisons suggest that the kernel estimator is overfitting within sample for this application.

Both sets of forecasts are unbiased, although numerically the nonparametric estimate is closer to unity than the parametric one. Residual-based diagnostics from these regressions did not uncover any remaining structure. The nonparametric forecasts serve as a benchmark in evaluating the ability of the parametric model to detect any structure exploitable for out-of-sample forecasts. The kernel forecasts were no better than the parametric ones which seems to suggest that the parametric model has adequately captured the underlying structure of the DGP.

5. Concluding Comments

This paper presents results that support the parametric formulation of a time-varying risk premium for the excess returns of a Japanese equity portfolio. A battery of diagnostics, including the BDS nonparametric test for residual structure, support this claim.

In addition, a nonparametric model was estimated using kernel regression. This nonparametric specification allows for flexible functional form. Given its flexibility within sample, the kernel estimator of the conditional mean constituted a benchmark for the parametric model. A tendency for the nonparametric structure to overfit was revealed by a comparison of the out-of-sample forecasts.

The out-of-sample forecasts suggest that the parametric specification is well-specified and produces unbiased forecasts with a lower MSE than the nonparametric forecasts. The out-of-sample performance of the parametric model acts as a validation for its good performance in-sample.

References

Baba, Y., R.F. Engle, D.F. Kraft and K.F. Kroner, 1989, **Multivariate** simultaneous generalized ARCH, **manuscript**.

Baillie, R.T. and T. Bollerslev, 1990, Conditional forecast densities from dynamic models with GARCH innovations, **manuscript**, J.L. Kellog Graduate School of Management, Northwestern University.

Bartlett, M.S., 1963, Statistical estimation of density functions, **Sankhya**, Series A, 25, 145-154.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, **Journal of Econometrics** 31, 307-327.

Bollerslev, T., R.Y. Chou and K.F. Kroner, 1990, ARCH modeling in finance: a review of the theory and empirical evidence, **Journal of Econometrics**, forthcoming.

Bollerslev, T. and J.M. Wooldridge, 1988, Quasi maximum likelihood estimation of dynamic models with time varying covariances, **manuscript**, Department of Economics, MIT.

Bollerslev, T., R.F. Engle and J.M. Wooldridge, 1988, A capital asset pricing model with time varying covariances, **Journal of Political Economy** 96, 116-131.

Breeden, D.T., M.R. Gibbons and R.H. Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, **Journal of Finance** 44, 231-262.

Brock, W., W.D. Dechert, and J. Sheinkman 1987 "A test for independence based on the correlation dimension" **SSRI Working Paper # 8702**, Economics, University of Wisconsin.

Brock, W., D.A. Hsieh and B. Lebaron, 1991, **Nonlinear Dynamics, Chaos and Instability**, (M.I.T. Press, Cambridge MA), forthcoming.

Campbell, J.Y., 1987, Stock returns and the term structure, **Journal of Financial Economics** 18, 373-399.

Chou, R.Y., 1988, Volatility persistence and stock valuations: Some empirical evidence using GARCH, **Journal of Applied Econometrics** 3, 279-294.

Cumby, R.E. and J. Huizinga, 1988, Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variable regressions, **manuscript**, University of Chicago.

Diebold, F.X. and J.M. Nason, 1990, Nonparametric exchange rate prediction, **Journal of International Economics**, 28, 315-332.

Engel, C., 1990, On the foreign exchange risk premium in a general equilibrium model, **manuscript**, University of Virginia.

Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, **Econometrica** 50, 987-1008.

Engle, R.F. and G. Gonzalez-Rivera, 1990, Semiparametric ARCH models, **Discussion Paper 89-17R**, University of California, San Diego.

Engle, R.F., T. Hong and A. Kane, 1990, Valuation of variance forecasts with simulated options markets, **manuscript**, Department of Economics, UC San Diego.

Engle, R.F., V. Ng and M. Rothschild, 1990, Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills, **Journal**

of Econometrics, 45, 213-237.

Fama, E.F., 1990, Stock returns, expected returns, and real activity, **Journal of Finance** 45, 1089-1108.

French, K.R., G. William Schwert and R.F. Stambaugh, 1987, Expected stock returns and volatility, **Journal of Financial Economics** 19, 3-30.

Gallant, A.R., D.A. Hsieh and G. Tauchen, 1989, On fitting a recalcitrant series: the pound/dollar exchange rate 1974-83, **manuscript**, Department of Economics, Duke University.

Giovannini, A. and P. Jorion, 1987, Interest rates and risk premia in the stock market and in the foreign exchange market, **Journal of International Money and Finance**, 107-124.

Hansen, L.P. and R.J. Hodrick, 1983, Risk averse speculation in the forward foreign exchange market: An economic analysis of linear models, in **Exchange Rates and International Macroeconomics**, edited by J.A. Frenkel, (University of Chicago Prress for the NBER).

Hansen, L.P. and S.F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, **Econometrica** 55, 587-614.

Harvey, Campbell R., 1991, The world price of covariance risk, **Journal of Finance**, 46, 111-157.

Hsieh, D.A. ,1989, Testing for nonlinear dependence in foreign exchange rates: 1974-1983", **Journal of Business**, 62, 339-368.

Labadie, Pamela, 1989, Stochastic inflation and the equity premium, **Journal of Monetary Economics**, 24, 277-298.

Lucas, R.E., Jr., 1982, Interest rates and currency prices in a two-country world, **Journal of Monetary Economics** 10, 335-360.

Mandelbrot, B., 1963, The variation of certain speculative prices, **Journal of Business** 36, 147-165.

Mark, N., 1988, Time varying betas and risk premia in the pricing of forward foreign exchange contracts, **Journal of Financial Economics** 22, 335-354.

McCurdy, T.H. and I.G. Morgan, 1991a, Tests for a systematic risk component in deviations from uncovered interest rate parity, **Review of Economic Studies**, 58, 587-602.

McCurdy, T.H. and I.G. Morgan, 1991b, Evidence of risk premia in foreign currency futures markets, **Review of Financial Studies**, forthcoming.

McCurdy, T.H. and I.G. Morgan, 1991c, Single beta models and currency futures prices, **Economic Record**, forthcoming.

Meese, R., and K. Rogoff, 1983, Empirical exchange rate models of the seventies: do they fit out-of-sample? **Journal of International Economics** 14, 3-24.

Moschini, G., D.M. Prescott and T. Stengos, 1988, Nonparametric forecasting: an application to the US hog supply, **Empirical Economics** 13, 141-154.

Nelson, Daniel B., 1991, Conditional heteroskedasticity in asset returns: a new approach, **Econometrica** 59, 347-370.

Pagan, A.R. and Y.S. Hong, 1989, Non-parametric estimation and the risk premium, in: W. Barnett, J. Powell and G. Tauchen, eds., **Semiparametric and nonparametric methods in econometrics and statistics**, (Cambridge University Press).

Pagan, A.R. and G.W. Schwert, 1990, Alternative models for conditional stock volatility, **Journal of Econometrics**, 45, 267-290.

Prescott, David M. and T. Stengos, 1989, Do asset markets overlook exploitable nonlinearities? The case of gold, **manuscript**, University of Guelph.

Robinson, P.M., 1983, Nonparametric estimators for time series, **Journal of Time Series Analysis** 4, 185-208.

Robinson, P.M., 1988, Root-n consistent semiparametric regression, **Econometrica**, 56, 931-954.

Roll, R., 1977, A critique of asset pricing theory tests, Part I, **Journal of Financial Economics** 4, 129-176.

Serfling, R.J., 1980, **Approximation theorems of mathematical statistics**, (John Wiley, New York).

Silverman, B.W., 1986, **Density Estimation for Statistics and Data Analysis**, (Chapman and Hall, London and New York).

Ullah, A., 1988, Non-parametric estimation of econometric functionals, **Canadian Journal of Economics** XXI.3, 625-658.

Weiss, A.A., 1986, Asymptotic theory for ARCH models: Estimation and Testing, **Econometric Theory** 2, 107-131.

White, H., 1982, Maximum likelihood estimation of misspecified models, **Econometrica**, 50, 1-26.