The Shape of the Risk Premium: Evidence From a Semiparametric GARCH Model

Oliver Linton
Department of Economics
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom
lintono@lse.ac.uk.

Benoit Perron
Dept. de sc. économiques, CIREQ, CIRANO
Université de Montréal
C.P. 6128, Succursale Centre-ville
Montréal, Québec H3C 3J7
Canada
benoit.perron@umontreal.ca.

August 13, 2002

Abstract

We examine the relationship between the risk premium on the CRSP value-weighted index total return and its conditional variance. We propose a new semiparametric model in which the conditional variance process is parametric, while the conditional mean is an arbitrary function of the conditional variance. For monthly CRSP value-weighted excess returns, the relationship between the two moments that we uncover is nonlinear and non-monotonic.

KEYWORDS: ARCH; Asset Pricing; Backfitting; Fourier Series; Kernel; Risk Premium.

JEL CLASSIFICATION: C13, C14, G12

1. INTRODUCTION

Modern asset pricing theories such as Abel (1987, 1999), Cox, Ingersoll, and Ross (1985), Merton (1973), and Gennotte and Marsh (1993) imply restrictions on the time series properties of expected returns and conditional variances of market aggregates. These restrictions are generally quite complicated, depending on utility functions as well as on the driving process of the stochastic components of the model. However, in an influential paper Merton (1973) obtained very simple restrictions albeit under somewhat drastic assumptions; he showed in the context of a continuous time partial equilibrium model that

$$\mu_t = E[(r_{mt} - r_{ft})|\mathcal{F}_{t-1}] = \gamma \text{var}[(r_{mt} - r_{ft})|\mathcal{F}_{t-1}] = \gamma\sigma_t^2, \tag{1}$$

where r_{mt}, r_{ft} are the returns on the market portfolio and risk-free asset respectively, while $\mathcal{F}_{t-1}$ is the market wide information available at time $t-1$. The constant γ is the Arrow–Pratt measure of relative risk aversion.

The simplicity of the above restrictions and their apparent congruence with the original CAPM restrictions (see Sharpe 1964 and Lintner 1965) has motivated a large number of empirical studies that test some variant of this restriction. A convenient statistical framework for examining the relationship between the quantities μ_t and σ_t^2 in discrete financial time series is the ARCH class of models, see the survey papers of Bollerslev, Chou, and Kroner (1992) and Bollerslev, Engle, and Nelson (1994) for references. Engle, Lilien, and Robins (1987) examined the relationship between government bonds of different maturities using the ARCH–M model in which the errors follow an ARCH(p) process and $\mu_t = \mu(\sigma_t^2)$ for some parametric function $\mu(\cdot)$. They examined $\mu_t = \gamma_0 + \gamma_1\sigma_t$ and $\mu_t = \gamma_0 + \gamma_1 \ln(\sigma_t^2)$, finding that the latter specification provided the better fit. French, Schwert, and Stambaugh (1987) and Nelson (1991) also examine this relationship using GARCH models.

Gennotte and Marsh (1993) argue that the linear relationship (1) should be regarded as a very special case. They construct a general equilibrium model of asset returns and derive the equilibrium relationship

$$\mu_t = \gamma\sigma_t^2 + g(\sigma_t^2), \tag{2}$$

where the form of $g(\cdot)$ depends on preferences and on the parameters of the distribution of asset returns. If the representative agent has logarithmic utility, then $g(\cdot) \equiv 0$ and the simple restrictions of Merton pertain. In addition, Backus, Gregory, and Zin (1989) and Backus and Gregory (1993) provide simulation evidence that $g(\cdot)$, and hence $\mu(\cdot)$, could be of arbitrary functional form in general equilibrium. Whitelaw (2000) develops these empirical findings into an equilibrium asset-pricing model with regime changes in which the relation is linear within each regime but overall nonlinear

due to the presence of the two distinct regimes. Veronesi (2001) also develops a model in which investors receive noisy signals in which the shape of the relation between the risk premium and the conditional variance is ambiguous and depends on investor uncertainty.

Pagan and Hong (1990) argue that the risk premium μ_t and the conditional variance σ_t^2 are highly nonlinear functions of the past whose form is not captured by standard parametric GARCH–M models. They estimate μ_t and σ_t^2 nonparametrically finding evidence of considerable nonlinearity. They then estimated δ from the regression

$$r_{mt} - r_{ft} = \beta' x_t + \delta\sigma_t^2 + \eta_t, \tag{3}$$

by least squares and instrumental variables methods with σ_t^2 substituted by the nonparametric estimate, finding a negative but insignificant δ. Perron (in press) analyses this approach using weak instrument asymptotics and finds similar results.

There are a number of drawbacks with their approach. Firstly, the conditional moments are calculated using a restricted conditioning set - the information set used in defining μ_t, σ_t^2 contained only a finite number of lags, i.e., $\mathcal{F}_{t-1} = \{y_{t-1}, \ldots, y_{t-p}\}$ for some fixed p and data series $y_t = r_{mt} - r_{ft}$. This greatly restricts the dynamics for the variance process. In particular, if the conditional variance is highly persistent, the non-parametric estimator of the conditional variance provides a poor approximation as confirmed by the simulation evidence reported in Perron (1998). Secondly, linearity of the relationship between μ_t and σ_t^2 is imposed, and this seems to be somewhat restrictive in view of earlier findings.

In this paper, we investigate the relationship between the risk premium and the conditional variance of excess returns on the CRSP value-weighted index. We consider a semiparametric specification that differs from previous treatments. In particular, we choose a parametric form for the variance dynamics (in our case EGARCH), while allowing the mean to be an unknown function of σ_t^2. This model takes account of the high level of persistence and leverage effect found in stock index return volatility, while at the same time allowing for an arbitrary functional form to describe the relationship between risk and return at the market level. We develop two estimation methods for this model: a Fourier series method and a method based on kernels. The kernel method is based on iterative one-dimensional smoothing and is similar in this respect to the backfitting method for estimating additive nonparametric regression, see Hastie and Tibshirani (1990). We also suggest a bootstrap algorithm for obtaining confidence intervals. Using these methods, we find evidence of a nonlinear and non-monotonic relationship between the risk premium and the conditional variance.

Other work applying nonparametric methods to this problem can be found in Boudoukh, Richardson, and Whitelaw (1997) and Harvey (2002). Our work differs from these in the parametric specification we choose for the conditional variance. This allows for the joint estimation of the two elements

of interest as will be described below.

In the next section we discuss the specification of our model, while in Sections 3 and 4 we describe how to obtain point and interval estimates respectively. In Section 5, we present our empirical results and the results of a small simulation experiment, while section 6 concludes.

2. A SEMIPARAMETRIC-MEAN EGARCH MODEL

We suppose that the excess returns y_t are generated as follows

$$y_t = \mu(\sigma_t^2) + \varepsilon_t \sigma_t, \ t = 1, 2, \ldots, T, \tag{4}$$

where ε_t is a martingale difference sequence with unit (conditional) variance, while $\mu(\cdot)$ is a smooth function, but of unknown functional form. The restriction that $E[y_t|\mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1} = \{y_{t-j}\}_{j=1}^{\infty}$, only depends on the past through σ_t^2 is quite severe but is a consequence of asset pricing models such as for example Backus and Gregory (1993) and Gennotte and Marsh (1993). In any case, it is possible to generalize this formulation in a number of directions. It is straightforward to incorporate fixed explanatory variables, lagged σ_t^2, or lagged y_t either as linear regressors or inside the unknown function $\mu(\cdot)$. More complicated dynamics for ε_t, such as an ARMA(p, q) model, and a multivariate extension can also be accommodated.

We propose using a parametric function for the conditional variance so as to allow for rich dynamics in the volatility. To be specific we shall consider the Exponential GARCH model introduced by Nelson (1991):

$$h_t \equiv \log(\sigma_t^2) = a + \sum_{j=1}^{p} b_j \log(\sigma_{t-j}^2) + \sum_{k=1}^{q} c_k \left[|\varepsilon_{t-k}| - E\,|\varepsilon_{t-k}| + d\varepsilon_{t-k} \right]. \tag{5}$$

The presence of the lagged dependent variables h_{t-j} allows very rich dynamics for the variance process itself which cannot yet be achieved by nonparametric methods. The above model also allows both the sign and the level of ε_{t-k} to affect σ_t^2 — good news and bad news can have different effects on volatility, hence allowing the possibility of the so-called leverage effect in stock returns. The parameter d controls the relative importance of the symmetric versus asymmetric effects. Evidence of such a leverage effect in returns on stock indices is widespread in the literature and can be found in Nelson (1991) for daily data and in Braun, Nelson, and Sunier (1991) for monthly data.

A number of authors, e.g., Nelson (1991), have found that standardized residuals from estimated GARCH models are leptokurtic relative to the normal, see also Engle and Gonzalez–Rivera (1991). We therefore assume that ε_t has a distribution within the exponential power family

$$f(\varepsilon) = \frac{\nu \exp\left(-\frac{1}{2}|\varepsilon/\lambda|^{\nu}\right)}{\lambda 2^{(1+1/\nu)}\Gamma(1/\nu)}; \ \lambda = [2^{(-2/\nu)}\Gamma(1/\nu)/\Gamma(3/\nu)]^{1/2}, \tag{6}$$

where Γ is the gamma function. The GED family of errors includes the normal ($\nu = 2$), uniform ($\nu = \infty$) and Laplace ($\nu = 1$) as special cases. The distribution is symmetric about zero for all ν, and has finite second moments for $\nu > 1$. With this density, we obtain that $E|\varepsilon_t| = (\lambda 2^{1/\nu}\Gamma(2/\nu))/\Gamma(1/\nu)$ (Hamilton 1994, p. 669).

We assume that the parameter values satisfy the requirements for stationarity given in Nelson (1991). Carrasco and Chen (2002) establish a general result about the dependence properties of a general class of volatility models, which suggests that the process y_t is β-mixing under some conditions.

Newey and Steigerwald (1997) have recently shown that quasi-likelihood estimators in GARCH models based on distributions other than the normal are generally inconsistent. Therefore, we also investigate our EGARCH(p,q) specification for the variance combined with a normal error distribution.

The main difference between our model and previous treatments is that we do not restrict the functional form of $\mu(\cdot)$ *a priori*. This has a number of implications both for estimation and testing. In particular, a simple consistent estimator of $\mu(\cdot)$ is difficult to obtain and would appear to depend on first obtaining consistent estimates of the parameters of the variance process. On the other hand, to estimate these parameters we need to have a good estimate of $\mu(\cdot)$. In the next section we propose a solution to this problem.

3. ESTIMATION

3.1 Parametric Estimation

Estimation of the unknown parameters by maximum likelihood when $\mu(\cdot)$ is known apart from a finite number of parameters, say τ, is considered in Engle, Lilien, and Robins (1987) and Nelson (1991). In this case, let $\theta = (\phi, \tau)$, where $\phi = (a, b_1, \ldots, b_p, c_1, \ldots, c_q, d, \nu)'$, while τ are the vector of unknown mean parameters. Then $\varepsilon_t(\theta)$ and $h_t(\theta)$ can be built up recursively given initial conditions, and the conditional log-likelihood function is

$$\ell_T(\theta) = \sum_{t=1}^{T} \ell_t(\theta) = \sum_{t=1}^{T} \log f(\varepsilon_t(\theta); \nu) - \frac{1}{2}\sum_{t=1}^{T} h_t(\theta), \tag{7}$$

The likelihood function can be maximized with respect to ϕ and τ using the BHHH algorithm, viz.

$$\theta^{[i+1]} = \theta^{[i]} - \lambda^{[i]} \left[\sum_{t=1}^{T} \dot{\ell}_{t\theta}\dot{\ell}'_{t\theta}\right]^{-1} \sum_{t=1}^{T} \dot{\ell}_{t\theta}, \tag{8}$$

where $\lambda^{[i]}$ is a variable step length chosen to maximize the log likelihood function in the given direction, and the score functions $\dot{\ell}_{t\theta}$ are evaluated at $\theta^{[i]}$. Although the likelihood function is not smooth in all parameters [because of the presence of the absolute value of ε], this derivative-based method seems to work well in practice. Some authors have modified the specification by using a smooth substitute for the absolute function for values around zero to avoid this problem. This proved unnecessary in our case.

3.2 Semiparametric Estimation

We propose several methods of constructing estimates of ϕ and $\mu(\cdot)$ in the semiparametric model. We estimate μ using two main approaches: the first one consists of treating the $T \times 1$ vector $\underline{\mu} = (\mu_1, \mu_2, \ldots, \mu_T)'$ as unknown parameters and estimating them through a kernel smoothing method inside the optimization routine. The second approach is to parametrize $\mu(\cdot)$ in a flexible way using series expansion methods. The basis we will use is the Fourier Flexible Form of Gallant (1981), although others could be used. Estimation of ϕ is then achieved by concentrating the likelihood function. We describe the estimation and the construction of confidence intervals for each method in turn.

Kernel Estimation.

The first method estimates μ by a smoothing procedure based on kernels (see Härdle 1990, Härdle and Linton 1994 and Pagan and Ullah 1999 for a discussion of kernel nonparametric regression estimation). Suppose that we could obtain some estimate of $\mu(.)$, then one could easily estimate the parameters of the variance and error distribution using maximum likelihood on the residuals. Unfortunately, it is very difficult to obtain a satisfactory direct estimate of $\mu(.)$. In our time series model, the relevant information set is the entire infinite past, i.e., $\mu(.) = E[y_t|\mathcal{F}_{t-1}]$ depends on the entire past of the series, so it is infeasible to literally compute this expectation empirically. One could argue — as do Pagan and Hong (1990) — that consistent estimates of $E[y_t|\mathcal{F}_{t-1}]$ could be obtained using nonparametric regression with a truncated information set $\mathcal{F}_{t-1}^{P(T)} = \{y_{t-1}, \ldots, y_{t-P}\}$, where $P(T) \Rightarrow \infty$ at a very slow rate. This estimate could then be used to obtain consistent estimates of the parameters of h_t. This is not a particularly appealing procedure from a practical point of view because of the high dimension of the conditioning set. Silverman (1986) dramatically illustrates the curse of dimensionality by showing the effective sample size needed to achieve a certain precision.

In semiparametric problems where one cannot obtain direct estimates of the nonparametric function, one can often instead use a semiparametric profile likelihood method as described in Powell (1994) in which the nonparametric function is estimated for each given parameter value and then the parameters are chosen to minimize some criterion function that would have been the likelihood

if the functions were known rather than estimated. In general, such parametric estimators are root-n consistent and asymptotically normal, and the nonparametric estimators are at least consistent. Unfortunately, in our model, we cannot define the corresponding profiled quantity $\widehat{\mu}_\phi(\sigma_t^2)$ so easily, since σ_t^2 depends, in addition to the parameters, on lagged ε's, which in turn depend on lagged μ's. Therefore, we need to know the entire function $\mu(.)$ [or at least its values at the T sample points] to construct $\widehat{\mu}_\phi(\sigma_t^2)$.

This might at first glance appear to make the estimation procedure hopeless, but this is a false impression. The same sort of issues arise in the estimation of additive nonparametric models and an enormous literature has arisen that proposes estimation algorithms, and, more recently, distribution theory, see for example Breiman and Freedman (1985), Hastie and Tibshirani (1990), Opsomer and Ruppert (1997), and Mammen, Linton, and Nielsen (1998). We borrow from this literature and suggest an estimation procedure based on iterative updating of both the finite dimensional parameters ϕ and the function $\mu(.)$. Our procedure first requires picking starting values for $\underline{\mu}$ and ϕ. We then define a modified version of the BHHH algorithm to update our estimates of ϕ. Finally, we update our estimates of $\underline{\mu}$ using kernel estimates based on the previous iterations filtered log variances. The main advantage of the procedure is that it relies on only one-dimensional smoothing operations at each step, so that the curse of dimensionality does not operate. The main disadvantage is that the procedure is time consuming and may not converge or may converge to local minima.

For convenience we describe our algorithm for the case $p = q = 1$. We smooth on the log of variance h_t instead of the variance itself. Since the logarithm is a monotonic transformation, the two approaches are equivalent, but since log variance has a more symmetric distribution with less effect from outliers, it helps in selecting a bandwidth. Our main algorithm is as follows.

Kernel Estimation Algorithm

1. Choose starting values for $\phi^{[0]}$ and $\{\mu_s^{[0]}\}_{s=1}^T$ which imply $\{h_s^{[0]}\}_{s=1}^T$.

2. Given $\{h_t^{[r-1]}\}_{t=1}^T$, calculate

$$\mu_t^{[r]} = \frac{\sum_s K\left(\frac{h_t^{[r-1]} - h_s^{[r-1]}}{\delta}\right) y_s}{\sum_s K\left(\frac{h_t^{[r-1]} - h_s^{[r-1]}}{\delta}\right)} \tag{9}$$

for $t = 1, 2, \ldots, T$, where $\delta > 0$ is a small bandwidth parameter, while K is a bounded kernel satisfying $\int K(u)du = 1$.

3. Given initial values $h_0^{[r]}(\phi)$ and $\varepsilon_0^{[r]}(\phi)$, define recursively for any parameter value ϕ

$$\begin{aligned} h_t^{[r]} &= a + bh_{t-1}^{[r]} + c_1\left(|\varepsilon_{t-1}^{[r]}| - E|\varepsilon_{t-1}^{[r]}| + d\varepsilon_{t-1}^{[r]}\right), \\ \varepsilon_t^{[r]} &= \frac{y_t - \mu_t^{[r]}}{\exp(h_t^{[r]})}, \end{aligned}$$

for $t = 1, 2, \ldots, T$. Then for any ϕ construct $\ell_t^{[r]}(\phi) = \ell_t(\phi; \underline{\mu}^{[r]})$, the period t contribution to the r^{th} likelihood function, where $\underline{\mu}^{[r]} = (\mu_1^{[r]}, \ldots, \mu_T^{[r]})'$.

4. Calculate

$$\phi^{[r]} = \phi^{[r-1]} - \lambda^{[r]}\left[\sum_{t=1}^{T} \dot{\ell}_{t\phi}^{[r]} \dot{\ell}_{t\phi}^{\prime[r]}\right]^{-1} \sum_{t=1}^{T} \dot{\ell}_{t\phi}^{[r]}, \tag{10}$$

where $\dot{\ell}_{t\phi}^{[r]}$ is the vector of partial derivatives of $\ell_t^{[r]}(\phi)$ with respect to ϕ evaluated at $\phi^{[r]}, \underline{\mu}^{[r]}$.

5. Repeat until convergence. We define convergence in terms of the relative gradient and the change in the nonparametric estimate, i.e.,

$$\max\left\{\max_k \left|\frac{\sum_{t=1}^{T} \dot{\ell}_{t\phi_k} \cdot \phi_k}{\ell(\phi)}\right|, \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\underline{\mu}^{[r+1]} - \underline{\mu}^{[r]}}{\underline{\mu}^{[r]}}\right|\right\} < \varepsilon, \tag{11}$$

for some small prespecified ε (we set $\varepsilon = 10^{-4}$). Denote the resulting estimates by $\hat{\phi}$ and $\underline{\hat{\mu}}$. ■

We are unable to prove convergence of the above algorithm, although in practice it seems to work reasonably well and to give similar answers for a range of starting values. Note that convergence of the backfitting algorithm for separable nonparametric regression has only been shown in some special cases, specifically when the estimator is linear in the dependent variable. However, backfitting has been defined and widely used to estimate more general models than additive nonparametric regression (see Hastie and Tibshirani 1990), and is widely believed to do a good job in such cases. In addition, in recent work, Audrino and Bühlmann (2001) have proposed an iterative algorithm for estimating a nonparametric volatility model. They provided a result on convergence in a special case where a contraction property can be established. See also Dominitz and Sherman (2001) for some related results in parametric cases. Unfortunately, no such contraction property can be guaranteed here.

In practice, the estimated parameters of h_t appear to be quite robust to different parametric specifications of the mean equation. The filtered estimate of h_t based on $\mu_t^{[0]} = T^{-1}\sum_{s=1}^{T} y_s$ should be close to the true h_t and should provide good starting values. We also use the fitted values from am

EGARCH-M model as starting values to check for robustness. As in the parametric case, additional iterations should improve the performance of the estimated parameters and function.

The stopping rule (11) was arrived at after some experimentation. It is desirable to ensure that the entire parameter vector $(\phi, \underline{\mu})$ is convergent.

Fourier Series Estimation.

The second approach we consider is to parametrize the mean equation using a flexible functional form. By letting the number of terms grow with sample size and with a suitable choice of basis functions, this method can approximate arbitrary functions. This is an example of sieve estimation, but for a given sample size, it reduces to a parametric method with a finite number of parameters, and the estimation algorithm is just the standard BHHH algorithm given above.

The basis we will use is a modification of the flexible Fourier form of Gallant (1981) by adding sine and cosine terms to a linear function. Because it uses trigonometric terms, it is convenient for the data to lie in the $[0, 2\pi]$ interval. To do so, we recenter and rescale the estimates of h_t and define a new variable

$$h_t^* = (h_t - \underline{h}) \frac{2\pi}{(\overline{h} - \underline{h})}, \tag{12}$$

where $\underline{h}$ and $\overline{h}$ are scalars such that $\underline{h}$ is less than $\min(h_t)$ and $\overline{h}$ is greater than $\max(h_t)$. Then the Fourier approximation is

$$\mu(h_t^*) = \gamma_0 + \gamma_1 h_t^* + \sum_{j=1}^{M} \psi_j \sin(j h_t^*) + \sum_{j=1}^{M} \varphi_j \cos(j h_t^*). \tag{13}$$

The number of parameters to estimate is $p + q + 2M + 5$.

4. INFERENCE

There is a general theory of inference for maximum likelihood and quasi-maximum likelihood estimators in time series, see Wooldridge (1994) for a state of the art survey. Specifically, Bollerslev and Wooldridge (1992) showed, under high level conditions, that quasi-maximum likelihood estimators in a parametric GARCH model are consistent and asymptotically normal provided only that the mean and the variance equations are correctly specified. However, their theory is based on high-level conditions which are rather difficult to verify even in the simplest cases. Papers that have derived an asymptotic theory for these models from primitive conditions are: Weiss (1986) for ARCH models, and Lumsdaine (1996) and Lee and Hansen (1994) for the GARCH(1,1) model. For other specifications in the GARCH class, the asymptotic theory that is used in practice is not known

to be valid. Similarly, the distribution theory for the EGARCH model of Nelson even in the special case with no mean effects and normal errors has not yet been established rigorously. However, there is much simulation evidence to support the normal approximation in this general class of models, and the results of Bollerslev and Wooldridge (1992) are widely believed to hold more generally, and are frequently used in practice. Gonzalez-Rivera and Drost (1999) have investigated the efficiency of various different estimation criteria under different specifications.

Given the complicated structure of our semiparametric model, it is not surprising that we cannot provide rigorous asymptotic theory for our estimators. However, if h_t were observed, a kernel estimate of $\mu(\cdot)$ as in (9) would be consistent and asymptotically normal under appropriate conditions, since the process h_t is weakly dependent. Therefore, the results of Robinson (1983) can be applied to establish consistency, provided $\delta(T) \to 0$ at an appropriate rate; this argument can be extended to the case where h_t is replaced by a consistent parametric estimate. Indeed, the asymptotic distribution of nonparametric estimates is usually independent of any preliminary parametric estimation [Powell (1994)]. We therefore expect $\hat{\mu}_t$ to be consistent at the usual nonparametric rate. As regards $\hat{\phi}$, we expect it to be $\sqrt{T}$ consistent and to have a limiting normal distribution with the variance including some component arising from the estimation of μ.

We now turn to the construction of standard errors for the parameter estimates and the risk premium. In the former case, we report analytical and bootstrap standard errors. The analytical standard errors are obtained by taking the outer product of the gradient with respect to the estimated parameters when the GED distribution is used. When the conditional distribution is Gaussian, we use the Bollerslev-Wooldridge (1992) QMLE standard errors. For the kernel estimator, the estimated parameters are just ϕ, the parameters of the error distribution and the variance process, while for the series estimator we are estimating these parameters jointly with the pseudo parameters τ of the mean function. For the series estimator we therefore compute standard errors from the matrix $[\sum_{t=1}^{T} \dot{\ell}_{t\theta}\dot{\ell}'_{t\theta}(\widehat{\theta})]^{-1}$, while for the kernel estimators we compute them from the smaller matrix $[\sum_{t=1}^{T} \dot{\ell}_{t\phi}\dot{\ell}'_{t\phi}(\widehat{\phi},\widehat{\mu})]^{-1}$. The kernel standard errors asymptotically understate the true uncertainty associated with the parameter estimates, since they neglect the loss of efficiency associated with the non-parametric estimation of $\mu(\cdot)$.

The second method of obtaining standard errors is through the bootstrap. There are now many methods for time series models including some that make very weak assumptions regarding the dependence structure, like the block bootstrap and the sieve bootstrap. In practice, however, their performance depends a lot on the implementation and the model structure. We instead prefer a bootstrap procedure that uses some of our model structure. We give an algorithm for calculating such confidence intervals for $p = q = 1$ in the case of the kernel procedure. We use a modified version of the wild bootstrap (see Härdle 1990, p 247) because we do not wish to rule out higher

order conditional heterogeneity, as this is relevant for the sampling variability of our estimators.

WILD BOOTSTRAP ALGORITHM

1. Given estimates $\underline{\mu}$, $\hat{\phi}$, $h_t(\widehat{\phi}, \underline{\mu})$, and $\hat{\varepsilon}_t = \varepsilon_t(\hat{\phi}, \underline{\mu})$, calculate the recentered residuals $\hat{\varepsilon}_t^c = (\hat{\varepsilon}_t - T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t)$.

2. Let z_t be a random variable with $E(z_t^j) = 0$ for $j = 1, 3$ and $E(z_t^j) = 1$ for $j = 2, 4$. Draw a random sample $\{z_1, \ldots, z_T\}$ from this distribution and let $\varepsilon_t^* = \hat{\varepsilon}_t^c \cdot z_t$. The variable ε_t^* will satisfy $E(\varepsilon_t^*) = 0$, $E(\varepsilon_t^{*2}) = \widehat{\varepsilon}_t^{c2}$, $E(\varepsilon_t^{*3}) = 0$, and $E(\varepsilon_t^{*4}) = \widehat{\varepsilon}_t^{c4}$. We choose z_t be a discrete variable which takes values -1 and 1 with equal probability.

3. Given starting values $\hat{h}_0$ and ε_0^*, define recursively

$$\hat{h}_t = \hat{a} + \hat{b}\hat{h}_{t-1} + \widehat{c}_1\left(\left[|\varepsilon_{t-1}^*| - E|\varepsilon_{t-1}^*|\right] + \widehat{d}\varepsilon_{t-1}^*\right).$$

   and

$$y_t^* = \hat{\mu}(\hat{h}_t; \{h_s\}_{s=1}^T) + \varepsilon_t^*\sigma_t^*,$$

   with the corresponding choice of $\mu(\cdot)$. In the case of the kernel estimator, some auxiliary bandwidth parameter $\tilde{\delta}$ that oversmooths the data should be chosen, where

$$\hat{\mu}(x; \{h_s\}_{s=1}^T, \delta) = \frac{\sum_{s\neq t} K\left(\frac{x-h_s}{\delta}\right) y_s}{\sum_{s\neq t} K\left(\frac{x-h_s}{\delta}\right)}.$$

   whereas for the Fourier series

$$\hat{\mu}(\hat{h}_t) = \hat{\gamma}_0 + \hat{\gamma}_1\hat{h}_t^* + \hat{\psi}_1 \sin\left(\hat{h}_t^*\right) + \hat{\varphi}_1 \cos\left(\hat{h}_t^*\right)$$

   with $\hat{h}_t^* = \left(\hat{h}_t - \underline{h}\right)\frac{2\pi}{\left(\overline{h}-\underline{h}\right)}$.

4. Given $\{y_t^*\}_{t=1}^T$ calculate parameter estimates $\hat{\phi}^*$ using the above quasi-Newton procedure.

5. Repeat steps 2–4 m times. The standard errors are estimated from the sample standard deviation of the bootstrap parameter estimates $\hat{\phi}^*$. ■

This method of obtaining standard errors is time-consuming for large datasets since it relies on simulation. However, it should reflect fully the loss of precision associated with estimating $\mu(\cdot)$. We impose a condition of symmetry on the errors for simplicity. However, we do not impose the restriction $E(\varepsilon_t^{*2}) = 1$ because this would require $E(z_t^2) = 1/\hat{\varepsilon}_t^{c2}$, which is numerically unstable

and generates paths with very large outliers. Our chosen distribution for z_t^j is the Rademacher distribution advocated by Davidson and Flachaire (2001) based on Edgeworth expansions.

The second problem, the construction of confidence intervals for $\widehat{\mu}$ can be approached in two ways: we can think of standard errors that are conditional on a value of h_t [and therefore allows us to look at the issue of the shape of the risk premium], and those that are conditional on all observables and thus allow us to run real-time experiments, and would be of interest to a decision maker. The second type is more difficult to construct as h_t depends on the infinite past, hence these standard errors have to be built up recursively.

On the other hand, computing standard errors conditional on the value of h_t is rather simple. For the kernel method, the variance of $\widehat{\mu}_t$ is given by Härdle (1990):

$$\frac{1}{n\delta}\frac{\sigma_t^2 \int k\left(u\right)^2 du}{f\left(h_t\right)}, \tag{14}$$

where $f\left(h_t\right)$ is the ergodic density of h_t evaluated at h_t. This quantity can be estimated by replacing σ_t^2 and $f(h_t)$ by estimates $\widehat{\sigma}_t^2$ and $\widehat{f}(\widehat{h}_t)$ respectively.

For the series approximation, we define $\widehat{\tau}$ as the estimated mean parameters and H_t be the vector of slopes, i.e., $\partial\mu/\partial\tau|_{\widehat{\tau}}$. For instance, for the Fourier series

$$H_t = (1, h_t^*, \sin\left(h_t^*\right), \cos(h_t^*), \ldots, \sin(Mh_t^*), \cos(Mh_t^*))'. \tag{15}$$

Then,

$$\text{var}\left[\mu\left(h_t\right)|h_t\right] = H_t' \text{var}\left(\widehat{\tau}\right) H_t, \tag{16}$$

where $\text{var}(\widehat{\tau})$ is the appropriate submatrix of the covariance matrix of $\widehat{\theta}$ obtained by the bootstrap method as described above.

Finally, choice of bandwidth is a nontrivial problem here. It is necessary to undersmooth our estimate of $\mu(\cdot)$ to obtain good estimates of ϕ as has been pointed out by Robinson (1988) for example. We adopt a cross-validation approach in which we maximize the likelihood function for each point on a grid of δ and choose the value that maximizes the (leave-one-out) likelihood function. However, to obtain a reasonable choice of bandwidth, it was necessary to remove the outliers when doing this and we removed 25% at each end of the data.

5. NUMERICAL RESULTS

5.1 EMPIRICAL RESULTS

Data.

We examine the monthly excess returns on the most comprehensive CRSP value-weighted index (including dividends) — the monthly continuously-compounded return on the index minus the monthly return on the 30-day T-bills— over the period January 1926 to December 2001. The data is obtained from the Center for Research on Security Prices (CRSP), which includes the NYSE, AMEX, and Nasdaq and is perhaps the best readily available proxy for 'the market'. We also conducted an analysis on the S&P500 series and obtained similar results. The data are plotted in the top panel of Figure 1. In Table 1 below we report sample moments for the data over the whole sample and two subsamples, each containing approximately half of the data: I (1926–1961) and II (1962–2000).

*** Figure 1 Here ***

*** Table 1 Here ***

There is strong evidence of leptokurtosis and negative skewness in the full sample and in both subsamples. The table reveals some differences in moments across subsamples. In particular, the first sub-period has much higher mean and variance, more pronounced negative skewness, and fatter tails than the rest of the sample. The standard deviation is approximately ten times the size of the mean, and this appears to support the widely held view that it is fundamentally difficult to estimate any mean effect in the presence of such large volatility [making the association that global mean corresponds to signal and global standard deviation corresponds to noise+signal]. However, from the nonparametric point of view this evidence is not by itself convincing since the global moments are one end of the smoothing spectrum where bandwidth is infinite; the other end of the smoothing spectrum is where bandwidth is zero and corresponds to the point mean being equal to the observation itself and the point standard deviation being the same quantity. To illustrate this point we computed a running mean and running standard deviation with 7 observations and equal weighting. The results are in he bottom two panels of figure 1 and show the time-varying nature of the mean and volatility. At this frequency, the mean and standard deviation are much closer in magnitude. Note also that this approach to estimating volatility provides similar estimates to those obtained from the dynamic models that we propose. Estimated volatility is high around well-known events: the Depression years, World War II, the oil shock and the 1987 crash in both cases.

Estimation.

We first discuss some model selection choices that had to be made. For the series estimator, values of the tuning parameters of up to 3 were considered with the models selected by the Akaike criterion (AIC) which maximizes $2\ln L(\omega) - 2k$ where k is the number of parameters in the model

and the Bayesian criterion (BIC) which maximizes $2\ln L(\omega) - k\ln T$. The criteria gave somewhat conflicting results, but the model with $p = 2$, $q = 1$, and $M = 1$ is well liked by both criteria. For the EGARCH-M model, AIC chooses $p = 1$ and $q = 3$ while BIC chooses $p = 1$ and $q = 1$. The selected model is the second choice for both criteria. For the model with Fourier terms, AIC chooses $p = 2$, $q = 1$ and $M = 2$ while BIC chooses $p = 1$, $q = 1$,and $M = 0$. The selected model is only marginally worse than these preferred ones. The values $p = 2$ and $q = 1$ were also chosen by Nelson (1991).

We chose the same values of p and q when estimating the model using the kernel approach. Results for other choices of p and q are available from the authors upon request. It is difficult to compare the fit of the model estimated with the kernel for various values of p and q as the models are then non-nested. As explained above, the bandwidth was selected by cross-validation over a grid of potential bandwidths. The bandwidth has the form:

$$\delta = k\sigma(h_t)\, T^{-\frac{1}{5}}, \tag{17}$$

where $\sigma(h_t)$ is the standard deviation of h_t, updated at each iteration to reflect the new estimates of h_t. The bandwidth constant k is allowed to vary between 0.5 and 2.5 in increments of 0.1, and the estimated value of k is the one that produces the highest value of the cross-validated likelihood. We set the values of $\underline{h}$ and $\overline{h}$ at -10 and -2 respectively based on the results from the kernel estimation which does not impose such restrictions. We also check to ensure that there is no value of h_t outside of these values in the course of optimization.

We now turn to the estimation results. The results from the estimation using the two methods considered here and their associated standard errors ($se_\phi(\phi)$) are presented in Table 2.

*** Table 2 Here ***

Our parameter estimates appear quite robust to the method chosen to do the estimation. They are also consistent with many other studies in the area such as Nelson (1991), Glosten, Jagannathan, and Runkle (1993) or Bollerslev, Engle, and Nelson (1994). In particular, volatility persistence is quite high (the sum of the estimates of b_1 and b_2 is well over 0.9), and the estimate of the leverage effect d is negative. However, this parameter is not precisely estimated with the kernel procedure and is not significantly different from 0. Finally, the estimated value of ν is around 1.4 which is again consistent with previous findings. The distribution we find has fatter tails than the normal which is a special case with $\nu = 2$. Note that the bootstrap standard errors tend to be larger than the analytic standard errors, sometimes dramatically so. The QMLE standard errors are not appreciably different from those obtained from the estimation with the GED.

The last row of Table 2 provides results of a likelihood ratio test for the significance of the coefficients on the nonlinear terms in the Fourier series. The results clearly show that linearity is strongly rejected at usual significance levels.

*** Figure 2 Here ***

The risk premium estimated using the kernel method is graphed in the top left corner of Figure 2 as a function of h_t. Confidence intervals at the 95% level constructed using the pointwise kernel confidence intervals are also provided. The figure clearly reveals a non-monotonic relation between h_t and $E\left[y_t \left| \mathcal{F}_{t-1} \right.\right]$. This is consistent with the findings of Backus and Gregory (1993), Whitelaw (2000), and Veronesi (2001) that in general equilibrium, the risk premium may have virtually any shape. Although the estimated risk premium is not significantly different from a constant at this level for some part of its range, the evidence is stronger in the middle range $h_t \in [-7.5, -5.5]$, which is where most of the data lie [the bottom left corner of Figure 2 plots the marginal density]. The evolution of the estimated risk premium, conditional standard deviation, and Sharpe ratio (in monthly terms) are presented in Figure 3. The episodes of high volatility revealed by this figure coincide closely with those obtained by a simple running average as done in Figure 1. Note that stocks were a great deal in the 1990s according to the Sharpe ratio but that they have become much less so in recent years.

*** Figure 3 Here ***

The top right panel of Figure 2 provides the shape of the risk premium estimated using the Fourier series. The graph also includes the analytical 95% confidence intervals conditional on h_t. Again, the estimated shape is nonlinear.

The two smoothing methods both have advantages and disadvantages. The kernel estimate appears rather wiggly in the end points where there is not much data. The Fourier series method on the other hand is very smooth and gives the appearance of being precisely estimated. However, there is a pronounced upward slope at the high end, which seems at odds with the kernel method finding. This end-trend is quite symptomatic of these polynomial-based methods; we view it with some skepticism. Notice also the difference in the standard errors for the two methods. The Fourier series method has a confidence band whose width is almost the same throughout the shown range, while the confidence band for the kernel is very wide at the end points, which reflects the relative paucity of the data in this region. Thus the Fourier series confidence band gives the appearance of being very precisely estimated in a region where we have little data. This is because it is a global

fitting method that draws its estimates from all the data. We thus redraw the two estimates on the same plot in the bottom right corner of Figure 2. The methods agree quite closely - there is a hump shape, which is first concave and then convex.

Finally, we provide some diagnostics on the standardized residuals $\hat{\varepsilon}_t = (y_t - \hat{\mu}_t)/\widehat{\sigma}_t$. We just report the results for the kernel, but similar results have been obtained for the series approach. The plots of the autocorrelogram of both the residuals and their squares indicates that they are close to white noise: there are 4 significant autocorrelation coefficients at the 5% level among the first 100 lags in the levels and 5 significant autocorrelations in the squares.

*** Figure 4 Here ***

Subsample Estimation.

In order to see how robust our estimates are, we re-estimated the model over two sub-samples: 1926-1961 and 1962-2001 using the kernel method. The results are presented in Table 3 below (with analytical standard errors in parentheses).

*** Table 3 Here ***

The results show quite a bit of instability in the point estimates. Figure 5 shows the estimated risk premium using the same scale as in the other figures. Because the last subsample is characterized by lower volatility than the beginning of the sample, the estimated log-volatility is concentrated towards the left of the graph for that period. The risk premium we estimate in the second period is much flatter than that of the first period because of the much larger bandwidth constant chosen, though the point estimate suggests a similar non-monotonic shape as for the full sample and the first subsample.

*** Figure 5 Here ***

5.2 MONTE CARLO

In order to appreciate the performance of our kernel procedure in estimating the risk premium in financial data, we carried out four simulation experiments. Each experiment is repeated 5000 times on samples of size 500. To make the experiments as realistic as possible, the parameters of each experiment are set to values estimated from our dataset. The data generating processes used for the experiments are presented in Table 4.

*** Table 4 Here ***

The first simulation experiment involves generating a risk premium from a linear model. We thus estimated an EGARCH-M model with GED errors from the data (these are the results presented in the last column of Table 2) and used it to generate 5000 samples. We then applied our non-parametric procedure to these simulated samples. The results are presented in the upper left panel of Figure 6. The solid line represents the true risk premium which is linear. The line with the long dashes is the median estimated function at each point on our equispaced grid. The short dashes represent the 25th and 75th percentile respectively. The method appears to do quite well as the median estimate deviates from the true function marginally for all values of the log conditional variance. The interquartile range of the estimates is relatively narrow in the middle of the distribution, but it increases dramatically for large or small volatility as there is less data.

*** Figure 6 Here ***

The second experiment used the model estimated by the Fourier series and GED errors presented in the previous section to generate the data. The results for this experiment are in the upper right panel of Figure 6. The kernel procedure unveils the nonlinear mean function well except for small log conditional variance.

The third experiment is a GARCH-M model with normal errors and linear mean. This experiment is designed to check the robustness of our results to misspecification in the conditional variance process and the innovation density (the parametric components of our model). The results are in the lower left panel of the figure. The kernel procedure discovers the linear mean very well. However, the confidence bands are very wide reflecting the additional uncertainty caused by misspecification.

Finally, the last experiment consists of a GARCH model with normal errors and mean function estimated with Fourier series. Once again, the mean function is well estimated where most data lies, but the uncertainty is once again large due to misspecification.

Table 5 presents the median and interquartile ranges for the estimated parameters over the 5000 replications. Some of these parameters are difficult to interpret since the estimated model is misspecified in experiments 3 and 4. For experiments 1 and 2 in which the model is correctly specified, we see that the procedure estimates most parameters well. It has a tendency to underestimate c_1, the effect of past innovation on the log conditional variance. Also, it does not distinguish well between the effect of h_{t-1} and h_{t-2} individually in experiment 2, although the overall persistence is well estimated. For the two misspecified models where data is generated from a GARCH(1,1) model,

the parameter values appear reasonable as they suggest a single lag of h_{t-1} and no leverage effect. Moreover, the normality of the innovations is well discovered.

*** Table 5 Here ***

Overall, these results suggest that our kernel procedure performs well in uncovering possible nonlinearities in the data. Yet, if the model were truly linear, the procedure would not mislead us. It is thus a useful tool for looking at the shape of the risk premium.

6. CONCLUSIONS

We have found a highly nonlinear relationship between the first two moments of index returns as suggested by Backus and Gregory (1993) and Gennotte and Marsh (1993). In particular, the risk premium appears to be non-monotonic and indeed hump-shaped. This result appears to be quite robust to the estimation method and the tuning parameters selected. However, the estimated risk premia are subject to quite a bit of variability and are not uniformly significantly different from zero at the 95% level. This and some instability over time must temper our interpretations to some degree.

Acknowledgement

We are grateful to Adrian Pagan, participants at the 1999 EC^2 conference in Madrid and at seminars at Montréal, Queen's, and UC, Santa Barbara, an associate editor, and two anonymous referees for comments and discussion. Perron acknowledges financial assistance from the Fonds pour la Formation des chercheurs et l'aide à la recherche (FCAR) and the Mathematics of Information Technology and Complex Systems (MITACS) network. Linton would like to thank the ESRC and STICERD for financial support.

References

[1] Abel, A. B. (1987), "Stock Prices Under Time-Varying Dividend Risk: An Exact Solution in an Infinite-Horizon General Equilibrium Model," *Journal of Monetary Economics*, 22, 375-393.

[2] ———— (1999), "Risk Premia and Term Premia in General Equilibrium," *Journal of Monetary Economics*, 43, 3-33.

[3] Audrino, F., and Bühlmann, P. (2001), "Tree-structured GARCH models," *Journal of The Royal Statistical Society,* 63, 727-744.

[4] Backus, D. K., and Gregory, A. W. (1993), "Theoretical Relations Between Risk Premiums and Conditional Variances," *Journal of Business and Economic Statistics*, 11, 177-185.

[5] Backus, D. K., Gregory, A. W., and Zin, S. E. (1989), "Risk Premiums in the Term Structure: Evidence from Artificial Economies," *Journal of Monetary Economics*, 24, 371-399.

[6] Black, F. (1976), "Studies in Stock Price Volatility Changes," Proceedings of the 1976 Business and Economic Statistics Section, American Statistical Association.

[7] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

[8] Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992), "ARCH Modelling in Finance," *Journal of Econometrics*, 52, 5-59.

[9] Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994), "ARCH Models," in *Handbook of Econometrics, volume IV*, eds. R. F. Engle and D. L. McFadden, Elsevier Science, 2959-3038.

[10] Bollerslev, T. and Wooldridge, J. M. (1992). "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models With Time Varying Covariances," *Econometric Reviews,* 11, 143-172.

[11] Boudoukh, J., Richardson, M. and Whitelaw, R. F. (1997), "Nonlineairties in the Relation Between the Equity Premium and the Term Structure," *Management Science*, 43, 371-385.

[12] Braun, P. A., Nelson, D. B., and Sunier, A. M. (1995), "Good News, Bad News, Volatility and Betas," *Journal of Finance,* 50, 1575-1604.

[13] Breiman, L. and Friedman, J. H. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association*, 80, 580-598.

[14] Carrasco, M. and Chen, X. (2002), "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18, 17-39.

[15] Cox, J., Ingersoll, J., and Ross, S. (1985), "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363-384.

[16] Davidson, R. and Flachaire, E. (2001): "The Wild Bootstrap, Tamed at Last", Unpublished manuscript, Queen's University at Kingston.

[17] Dominitz, J., and Sherman R. (2001). "Convergence Theory for Stochastic Iterative Procedures With an Application to Semiparametric Estimation," Unpublished manuscript, California Institute of Technology.

[18] Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.

[19] Engle, R. F., Lilien, D. M., and Robins, R. P. (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH–M Model," *Econometrica*, 55, 391-407.

[20] Engle, R. F., and Gonzalez-Rivera, G. (1991), "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9, 345-359.

[21] French, K. R., Schwert, G. W., and Stambaugh, R. B. (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3-29.

[22] Gallant, A. R. (1981), "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics,* 15, 211-245.

[23] Gennotte, G., and Marsh, T. (1993), "Valuations in Economic Uncertainty and Risk Premiums on Capital Assets," *European Economic Review*, 37, 1021-1041.

[24] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Returns on Stocks," *Journal of Finance*, 48, 1779-1801.

[25] Gonzalez-Rivera, G., and Drost, F. C. (1999), "Efficiency Comparisons of Maximum-Likelihood Based Estimators in GARCH Models," *Journal of Econometrics,* 93, 93-111.

[26] Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

[27] Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press.

[28] Härdle, W., and Linton, O. (1994), "Applied Nonparametric Methods, " in *Handbook of Econometrics, volume IV,* eds. R. F. Engle and D. L. McFadden, Elsevier Science, 2295-2339.

[29] Hastie, T., and Tibshirani, R. (1990), *Generalized Additive Models*, Chapman and Hall.

[30] Lintner, J. (1965), "The Valuation of Risky Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13-37.

[31] Lee, S., and Hansen, B. (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29-52.

[32] Lumsdaine, R. L. (1996), "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, 64, 575-596.

[33] Mammen, E., Linton, O., and Nielsen, J. P. (1999), "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm Under Weak Conditions," *The Annals of Statistics,* 27, 1443-1490.

[34] Merton, R. C. (1973), "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867-887.

[35] Nelson, D. B. (1990), "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318-334.

[36] ——— (1991), "Conditional Heteroscedasticity in Asset Returns: A New Approach," *Econometrica*, Vol. 59, 347-370.

[37] Newey, W. K., and Steigerwald, D. G. (1997), "Asymptotic Bias for Quasi-Maximum Likelihood Estimators in Conditional Heteroskedasticity Models," *Econometrica*, 65, 587-599.

[38] Opsomer, J. D., and Ruppert, D. (1997), "Fitting a Bivariate Additive Model by Local Polynomial Regression," *The Annals of Statistics*, 25, 186 - 211.

[39] Pagan, A. R., and Hong, Y. S. (1990), "Non-parametric Estimation and the Risk Premium," In *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, eds. W. A. Barnett, J. Powell, and G. Tauchen, Cambridge University Press, 51-75.

[40] Pagan, A. R., and Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press.

[41] Perron, B. (in press), "Semi-parametric Weak Instrument Regressions with an Application to the Risk-Return Trade-off," *Review of Economics and Statistics.*

[42] ——— (1998), "A Monte Carlo Comparison of Non-parametric Estimators of the Conditional Variance," Unpublished manuscript, Université de Montréal.

[43] Powell, J. (1994), "Estimation of Semiparametric Models," in *Handbook of Econometrics, volume IV,* eds. R. F. Engle and D. L. McFadden, Elsevier Science, 2443-2521.

[44] Robinson, P. M. (1983), "Nonparametric Estimators for Time Series." *Journal of Time Series Analysis,* 4, 185-207.

[45] ———— (1988), "Root-N-Consistent Semiparametric Regression." *Econometrica*, 56, 931-954.

[46] Scruggs, J. T. (1998), "Resolving the Puzzling Intertemporal Relation between the Market Risk Premium and Conditional Market Variance: A Two-Factor Approach," *Journal of Finance*, 53, 575-603.

[47] Schwert, G. W. (1989), "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance*, 44, 1115-1153.

[48] Sharpe, W. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance,* 19, 567-575.

[49] Tadikamalla, P. R. (1980), "Random Sampling from the Exponential Power Distribution," *Journal of the American Statistical Association*, 75, 683-686.

[50] Veronesi, P. (2001), "How Does Information Quality Affect Stock Returns?" *Journal of Finance*, 55, 807-837.

[51] Weiss, A. (1986), "Asymptotic Theory for ARCH models: Estimation and Testing," *Econometric Theory,* 2, 107-131.

[52] Whistler, D. (1988), "Semiparametric ARCH Estimation of Intra-Daily Exchange Rate Volatility," unpublished manuscript.

[53] Whitelaw, R. F. (2000), "Stock Market Risk and Return: An Equilibrium Approach," *Review of Financial Studies*, 13, 521-547.

[54] Wooldridge, J. M. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics, volume IV,* eds. R. F. Engle and D. L. McFadden, Elsevier Science, 2659-2738.

Tables and Figures

Table 1. Raw Data by Sub Period

	Full sample	1926:1-1961:12	1962:1-2001:12
Mean ($\times 100$)	0.4987	0.6686	0.3457
Variance ($\times 100$)	0.3031	0.4132	0.2043
Skewness	-0.5037	-0.4224	-0.7422
Excess Kurtosis	6.8015	3.5760	2.8460

Table 2. Full Sample Estimates

$$r_t - r_t^f = \mu_t + \sigma_t \varepsilon_t$$
$$h_t = \ln(\sigma_t^2) = a + b_1 h_{t-1} + b_2 h_{t-2} + c_1 (|\varepsilon_{t-1}| - E|\varepsilon_{t-1}| - d_{t-1})$$
$$\text{Fourier: } \mu_t = \gamma_0 + \gamma_1 h_t^* + \psi_1 \sin(h_t^*) + \varphi_1 \cos(h_t^*)$$
$$\varepsilon_t \sim GED(\nu) \text{ or } N(0,1)$$

	Kernel-GED	Kernel-QMLE	Fourier-GED	Fourier-QMLE	EGARCH-M
a	-0.311 (0.112) (0.209)	-0.222 (0.089) (0.196)	-0.407 (0.085) (0.177)	-0.399 (0..136) (0.163)	-0.340 (0.109)
b_1	0.780 (0.333) (0.417)	0.978 (0.097) (0.432)	0.452 (0.150) (0.146)	0.379 (0..181) (0.116)	0.353 (0.138)
b_2	0.169 (0.325) (0.402)	-0.015 (0.086) (0.419)	0.484 (0.147) (0.154)	0.555 (0.179) (0.117)	0.593 (0.136)
c_1	0.293 (0.091) (0.088)	0.260 (0.053) (0.099)	0.241 (0.029) (0.117)	0.254 (0.039) (0.053)	0.267 (0.048)
d	-0.142 (0.122) (0.222)	-0.080 (0.122) (0.211)	-0.763 (0.131) (0.355)	-0.721 (0..190) (0.175)	-0.536 (0.182)
ν	1.444 (0.078) (0.121)	-	1.425 (0.088) (0.192)	-	1.419 (0.844)
γ_0	-	-	-0.370 (0.022) (0.157)	-0.363 (0.131) (0.102)	-0.003 (0.003)
γ_1	-	-	0.117 (0.006) (0.054)	0.115 (0.039) (0.034)	0.002 (0.000)
ψ_1	-	-	0.137 (0.009) (0.050)	0.133 (0.047) (0.036)	-
φ_1	-	-	-0.009 (0.010) (0.021)	-0.007 (0.014) (0.024)	-
Bandwidth constant	0.9	0.9	-	-	-
Likelihood	1502.3	1489.5	1510.8	1493.5	1507.2
Linearity test $H_0 : \psi_i = \varphi_i = 0, i > 1$ $(p-value)$	-	-	7.369 (0.025)	8.956 (0.011)	-

Note: The numbers in parentheses are analytical and wild bootstrap standard errors respectively. For the GED, the analytical standard errors are from the outer-product of gradient (OPG), while for the QMLE, the analytical standard errors are those of Bollerslev and Wooldridge (1992).

Table 3. Sub-period Estimates

	1926-1961	1962-2001
a	-0.135 (0.088) (0.216)	-0.754 (0.473) (0.718)
b_1	1.137 (0.468) (0.465)	0.469 (0.334) (0.488)
b_2	-0.159 (0.460) (0.454)	0.411 (0.333) (0.467)
c_1	0.223 (0.118) (0.116)	0.341 (0.137) (0.108)
d	-0.098 (0.165) (0.779)	-0.244 (0.229) (0.476)
ν	1.444 (0.125) (0.177)	1.487 (0.107) (0.194)
Bandwidth constant	0.7	2.5
Likelihood	680.7	829.9

Note: See table 2.

Table 4. Data Generating Processes Used for the Simulation Experiments

Experiment 1: Linear mean, EGARCH conditional variance, and GED errors
$\mu_t = -0.003 - 0.002h_t$
$h_t = -0.340 + 0.353h_{t-1} + 0.593h_{t-2} + 0.267\left(\lvert\varepsilon_{t-1}\rvert - E\lvert\varepsilon_{t-1}\rvert - 0.536\varepsilon_{t-1}\right)$
$\varepsilon_t \sim GED\,(1.419)$
Experiment 2: Fourier mean, EGARCH conditional variance, and GED errors
$\mu_t = -0.370 + 0.117h_t^* + 0.137\sin(h_t^*) - 0.009\cos(h_t^*)$
$h_t = -0.407 + 0.452h_{t-1} + 0.484h_{t-2} + 0.241\left(\lvert\varepsilon_{t-1}\rvert - E\lvert\varepsilon_{t-1}\rvert - 0.763\varepsilon_{t-1}\right)$
$\varepsilon_t \sim GED\,(1.425)$
Experiment 3: Linear mean, GARCH conditional variance, and normal errors
$\mu_t = 0.013 + 0.001h_t$
$\sigma_t^2 = 7.402 \times 10^{-5} + 0.867\sigma_{t-1}^2 + 0.109u_{t-1}^2$
$\varepsilon_t \sim N\,(0,1)$
Experiment 4: Fourier mean, GARCH conditional variance, and normal errors
$\mu_t = -0.229 + 0.073h_t^* + 0.082\sin(h_t^*) - 0.006\cos(h_t^*)$
$\sigma_t^2 = 7.163 \times 10^{-5} + 0.867\sigma_{t-1}^2 + 0.117u_{t-1}^2$
$\varepsilon_t \sim N\,(0,1)$

Table 5. Median Estimated Parameters in Simulation Experiments

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
a	-0.340 (-0.489,-0.267)	-0.382 (-0.598,-0.242)	-0.273 (-0.417,-0.146)	-0.181 (-0.326,-0.086)
b_1	0.375 (0.352,0.766)	0.715 (0.415,1.155)	0.890 (0.423,1.351)	0.978 (0.596,1.391)
b_2	0.545 (0.148,0.593)	0.190 (0.219,0.494)	0.045 (-0.402,0.498)	-0.027 (-0.429,0.348)
c_1	0.216 (0.129,0.267)	0.136 (0.081,0.199)	0.199 (0.122,0.267)	0.196 (0.120,0.237)
d	-0.536 (-0.823,-0.444)	-0.838 (-1.335,-0.531)	-0.029 (-0.211,0.123)	-0.010 (-0.160,0.137)
ν	1.419 (1.346,1.460)	1.411 (1.325,1.506)	1.969 (1.813,2.136)	1.981 (1.833,2.143)

Note: Entries are the median of the estimated parameters over the 5000 replications. The entries in parentheses are the 25th and 75th percentile over the 5000 replications respectively.

Figure captions

Figure 1. Data. The top panel is a time plot of the continuously compounded returns on the CRSP value-weighted index, 1926-2001. The middle panel is a rolling mean estimate of the risk premium obtained as a moving average using a window width of seven and equal weighting. The bottom panel is a rolling estimate of the standard deviation of the excess returns also using a window width of seven observations and equal weighting.

Figure 2. Empirical Results for the Full Sample. The top left panel is the kernel estimate of the risk premium as a function of the log conditional variance. The solid line represents the point estimate and the dashed lines are the limits of a 95% confidence interval computed using (14). The top right panel is similar but uses the Fourier series estimator. The bottom left panel plots the marginal density of the log conditional variance. Finally, the bottom right panel superposes the point estimates from the kernel and Fourier series. Note that the horizontal scale is the same for all panels and that the vertical scale is the same for all panels except the lower left one.

Figure 3. Time Plots of Kernel-Estimated Risk Premium, Standard Deviation, and Sharpe Ratio. The top and middle panels plot the estimated risk premium and standard deviation over time. The bottom panel plots the estimated Sharpe ratio as the ratio of the estimated risk premium to the estimated standard deviation. The results are in monthly terms and have not been annualized.

Figure 4. Autocorrelation of Standardized Residuals and Their Squares. The top panel plots the autocorrelation function of the standardized residuals obtained from the kernel estimator for the first 100 lags along with the asymptotic 95% confidence bands under independence. The bottom panel plots the same information for the squared standardized residuals.

Figure 5. Subsample Results. The figure provides kernel estimates of the risk premium for the two subsamples along with the 95% confidence bands. The scales are identical to those in figure 2. The top panel provides results for the first subsample, 1926:1-1961:12, while the bottom panel gives the same information for the second subsample, 1962:1-2001:12.

Figure 6. Monte Carlo Results. This figures reports the results on the estimated risk premium from the four simulation experiments. In each panel, the solid line is the true relationship, the long-dashed line is the median among the 5000 replications and the short-dashed lines are the 25th and 75th percentile respectively. The scales are the same as those in figures 2 and 5.

Fig. 1. Monthly excess returns
CRSP value-weighted index
1926-2001

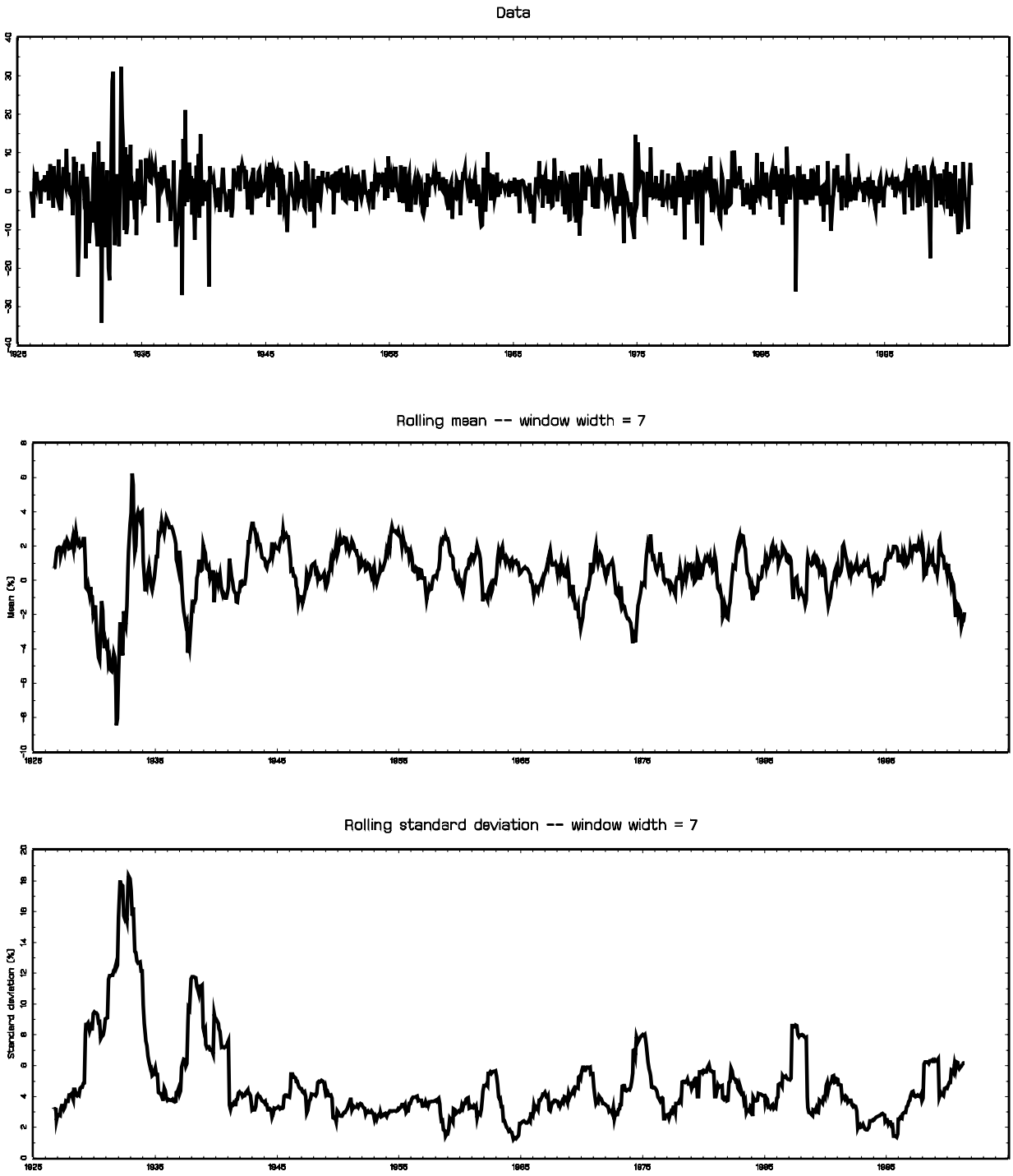


Figure 2. Estimation results

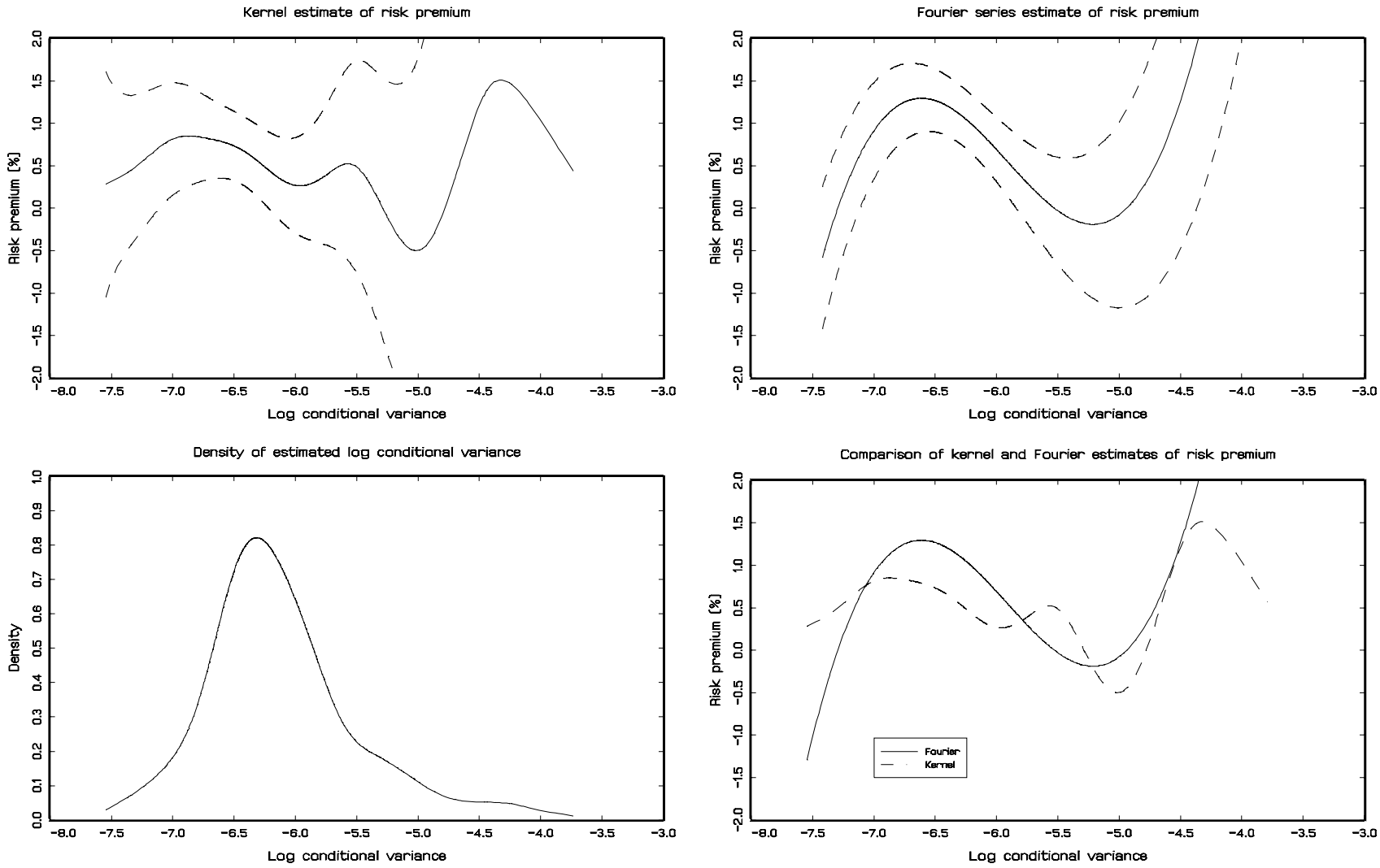


Figure 3. Time plots of estimated quantities
Kernel

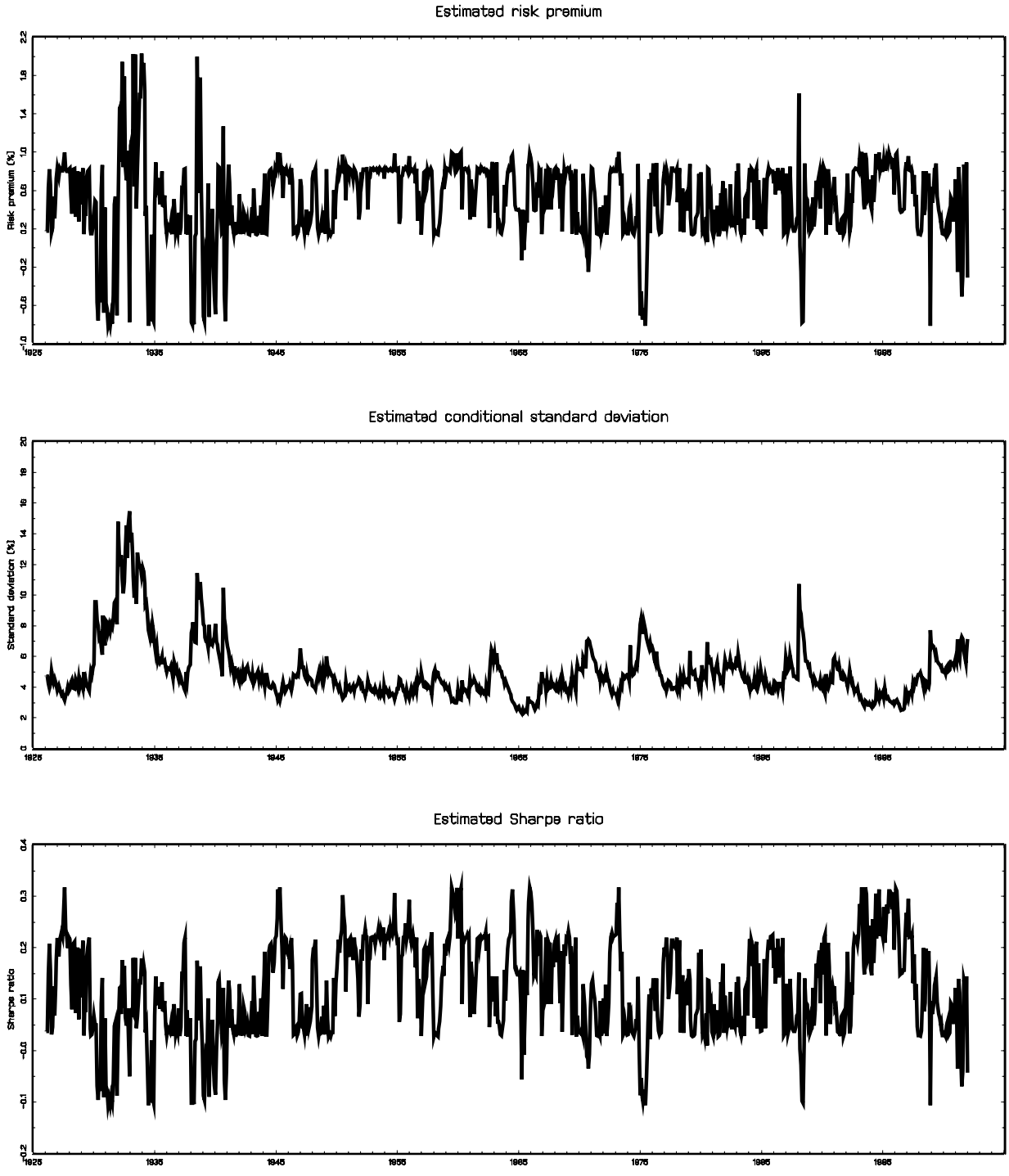
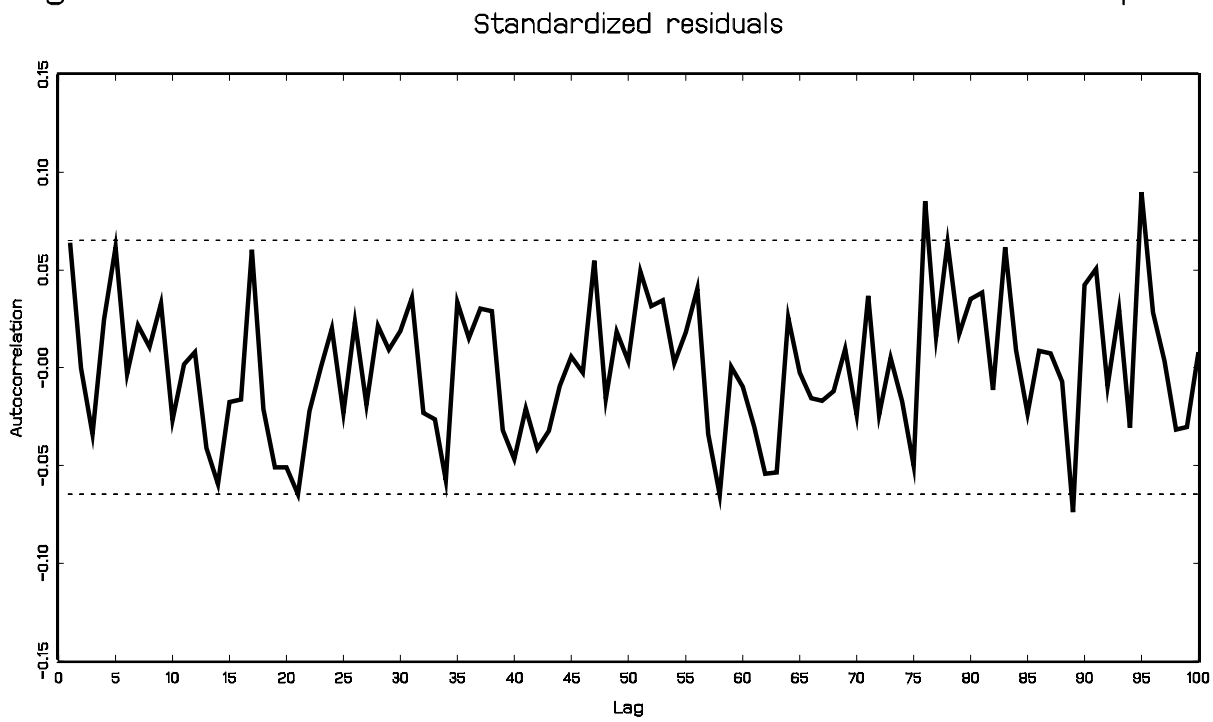


Figure 4. Autocorrelation of standardized residuals and squares



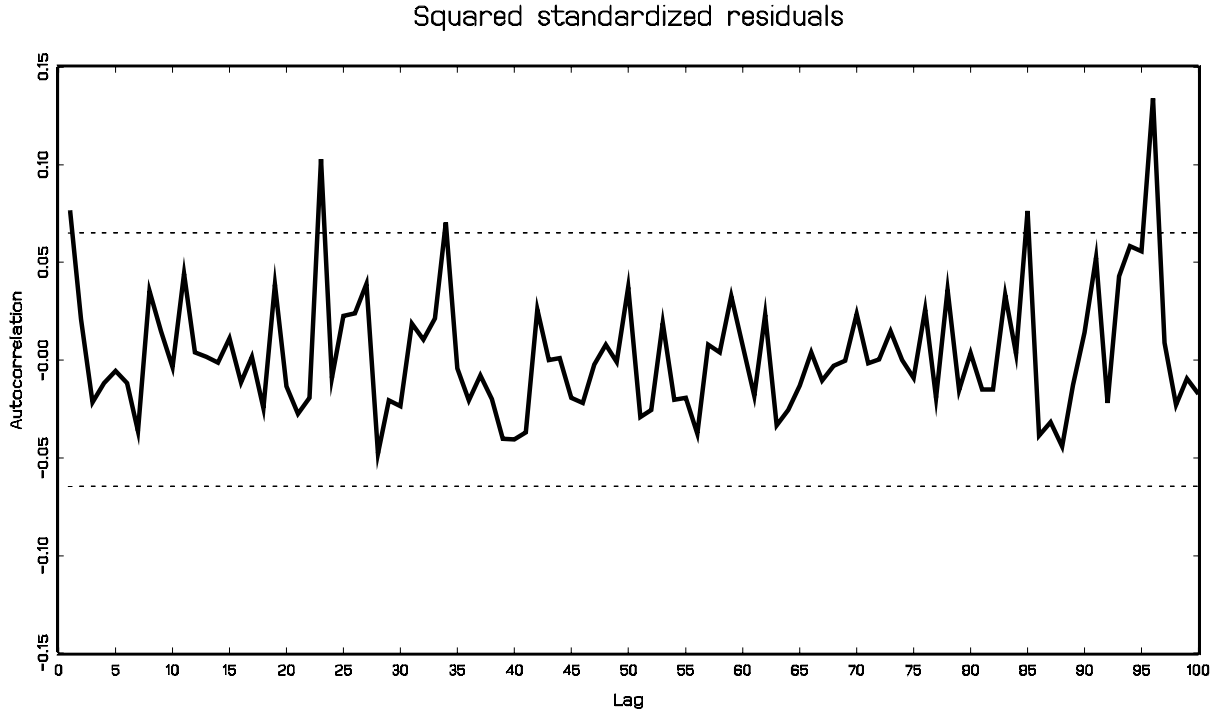


Figure 5. Kernel Estimate of the risk premium

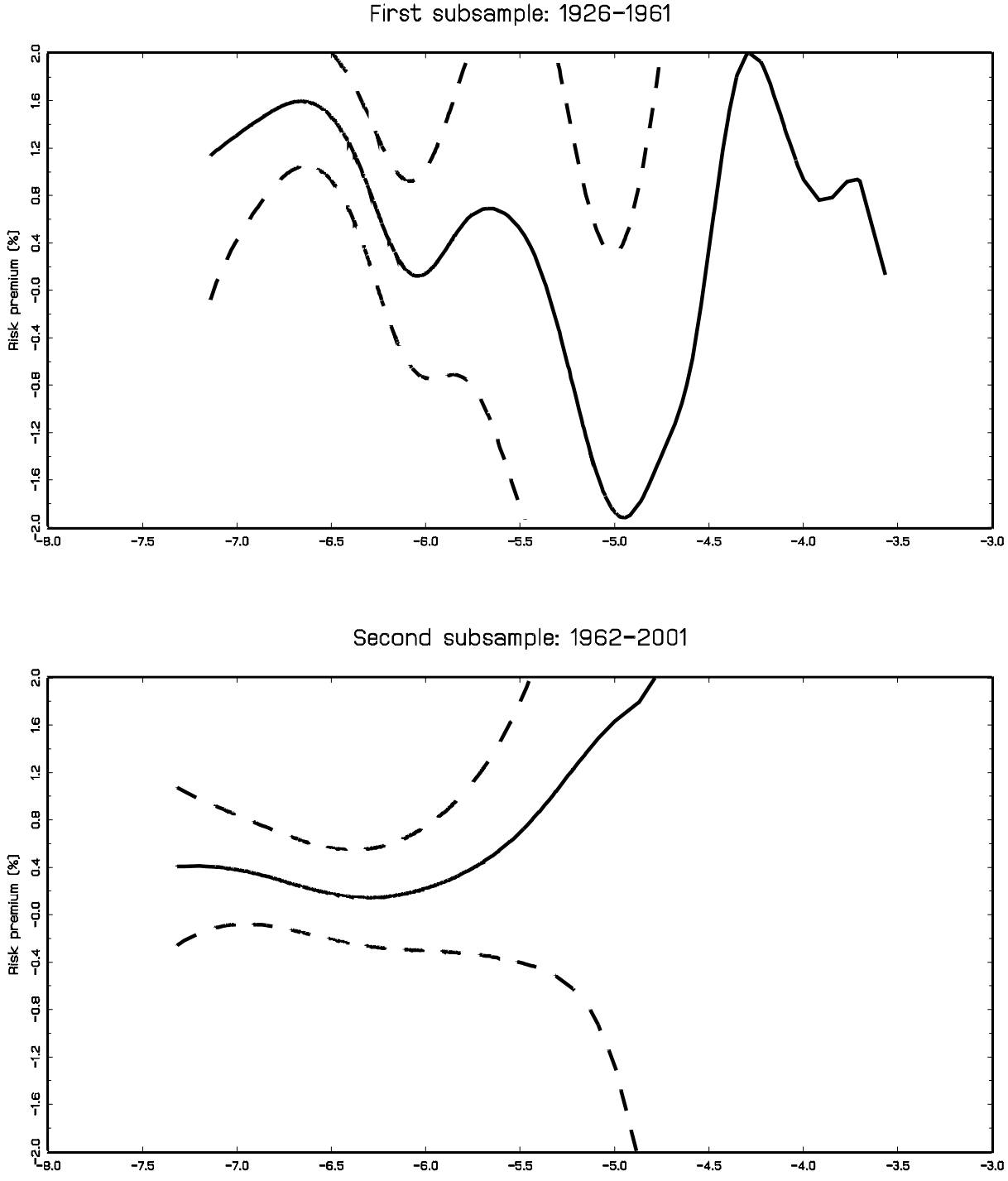


Figure 6. Monte Carlo results