University of Heidelberg

Department of Economics

Discussion Paper Series | No. 473

Nonparametric Regression on Latent Covariates with an Application to Semiparametric GARCH-in-Mean Models

Christian Conrad and
Enno Mammen

July 2008

Nonparametric Regression on Latent Covariates with an Application to Semiparametric GARCH-in-Mean Models

Christian Conrad[†] and Enno Mammen[‡]

[†]Faculty of Economics and Social Studies, University of Heidelberg, Grabengasse 14, 69117 Heidelberg, Germany. E-mail: christian.conrad@awi.uni-heidelberg.de

[‡]Department of Economics, University of Mannheim, L7,3-5, 68229 Mannheim, Germany. E-mail: emammen@rumms.uni-mannheim.de

July 2008

Abstract

We consider time series models in which the conditional mean of the response variable given the past depends on latent covariates. We assume that the covariates can be estimated consistently and use an iterative nonparametric kernel smoothing procedure for estimating the conditional mean function. The covariates are assumed to depend (non)parametrically on past values of the covariates and of the observations. Our procedure is based on iterative fits of the covariates and nonparametric kernel smoothing of the conditional mean function. An asymptotic theory for the resulting kernel estimator is developed and the estimator is used for testing parametric specifications of the mean function. Our leading example is a semiparametric class of GARCH-in-Mean models. In this set-up our procedure provides a formal framework for testing economic theories that postulate functional relations between macroeconomic or financial variables and their conditional second moments. We illustrate the usefulness of the methodology by testing the linear risk-return relation predicted by the ICAPM.

Keywords: Specification test, GARCH-M, semiparametric regression, risk premium, ICAPM.

JEL Classification: C12, C14, C22, C52, G12

We would like to thank Yacine Ait-Sahalia, Robert F. Engle, Berthold R. Haag, Menelaos Karanasos, Oliver Linton, Kyusang Yu, the participants of the annual meeting of the Verein für Socialpolitik (Bayreuth, 2006), of the annual meeting of the Swiss Society of Economics and Statistics (St. Gallen, 2007), of the North American Summer Meeting of the Econometric Society (Durham, 2007), and of the Conference on Likelihood Methods in Finance (Princeton, 2007) for their helpful comments and suggestions. A previous version of this paper was circulated under the title "A Specification Test for a Class of GARCH-in-Mean Models".

1 Introduction

Economic theory often predicts a relationship between an unobserved covariate and an observed response variable. Standard examples from finance and macroeconomics are the relation between risk and expected return or nominal uncertainty and inflation. Throughout the article we consider the exemplary situation in which a relationship between the level of a variable and an unobserved covariate that depends on the past is modeled. A prominent example for such a covariate is the conditional variance of the variable given the past. More specifically, we consider an econometric specification for a random variable Y_t of the form

$$\mathbf{E}[Y_t|\mathcal{F}_{t-1}] = m(h_t), \tag{1}$$

where $\mathcal{F}_{t-1}$ represents the information set available at $t-1$ and where h_t is an unobserved covariate that is measurable with respect to $\mathcal{F}_{t-1}$. We assume that h_t depends on its own past values and on the past values of Y_t, i.e. $h_t = f_{m,\psi}(Y_{t-1}, Y_{t-2}, ..., h_{t-1}, h_{t-2}, ...)$ with a function $f_{m,\psi}$ parametrized by the mean function m and a finite- or infinite-dimensional parameter ψ. We propose to estimate the covariate process h_t, the parameter ψ and the regression function m by an iterative procedure. In each cycle of the procedure, m is estimated by regressing Y_t nonparametrically on the fitted values of h_t, then the estimate of ψ is updated by using the new fit of m, and finally, a new estimate of h_t is given by applying the function $f_{\widehat{m},\widehat{\psi}}$ to the actual fits $\widehat{m}, \widehat{\psi}$ of m and ψ. The iteration is repeated until convergence of the estimated mean function is achieved.

We develop an asymptotic theory for the resulting estimator of m and propose a test for parametric specifications of m. For the estimator of m we show the following "oracle property". Asymptotically the nonparametric mean function can be fitted as well as if the fit would have been based on the true unobservable covariate. Our test for parametric specifications of m is based on a comparison of a parametric estimator of m with our nonparametric estimator. The idea of comparing parametric and nonparametric regression fits for testing the appropriateness of a particular parametric model goes back to e.g. Härdle and Mammen (1993) who concentrated on regressions involving independently and identically distributed observations. The problem of testing for linearity in autoregressive time series models has been considered by e.g. Hjellvik and Tjøstheim (1995), while Kreiss et al. (2002) test for linearity in a more general times series setting which is not necessarily autoregressive. In all previous studies the test statistic is based on the difference between a nonparametric and a parametric regression fit, but in contrast to our study the dependent and independent variables are observed directly. The main contribution of this article is to deal with a situation in which the regressor is unobservable and replaced by an appropriate estimate.[1]

[1]The problem considered in this article is closely related to the treatment of nonparametric regressions on generated regressors. See, e.g., Sperlich (2007) for a discussion of the situation where the unobserved variable is i.i.d.

As for estimation, we show that under certain regularity conditions the asymptotic results for the test statistic based on the iteratively fitted values of h_t are the same as if the process h_t had been observed. Since the asymptotic distribution of the test statistic is approached quite slowly as the sample size goes to infinity, we suggest a bootstrap algorithm from which the critical values can be computed. Monte-Carlo simulations show that the bootstrap distribution approximates the distribution of the test statistic under the null hypothesis reasonably well in finite samples. Under the alternative, the test statistic reveals good power properties.

The leading example for our general theory is the situation in which h_t represents the conditional variance of Y_t. In particular, we think of the case where h_t is given by some GARCH-type equation. Then the parameter vector ψ contains the GARCH-parameters and we have a semiparametric GARCH-in-Mean (GARCH-M) model with nonparametric specification of the risk premium $m(\cdot)$. In this model, our nonparametric estimator of m is similar to that proposed by Linton and Perron (2003). In certain cases economic theory directly implies a particular parametric specification $m = m_\gamma$ with γ being a parameter vector. One of the workhorses in financial econometrics, the GARCH-M model introduced by Engle et al. (1987) is a primary example of such a specification where m_γ is typically assumed to be linear or logarithmic in the conditional variance. Our test can be applied for checking such parametric specifications of the risk premium.

We employ the suggested procedure in an empirical application for testing Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM) which suggests that the conditional expected excess return on the market, say $\mathbf{E}[Y_t|\mathcal{F}_{t-1}]$, is proportional to the conditional market variance, i.e. $m_\gamma(h_t) = \lambda h_t$. For monthly as well as daily excess return data on the CRSP value-weighted index we estimate GARCH$(1,1)$-M models. In line with previous studies, we find a positive but insignificant relation between the market excess return and its conditional variance when using monthly data, while we find a highly significant and positive relation using daily data. Under the alternative we estimate a semiparametric model which only assumes the risk premium to be some smooth function. For the daily data we find some evidence against the linear relationship when volatility is extremely high.

Recently, Christensen et al. (2008) followed our approach and proposed another variant of a semiparametric GARCH-M model. In contrast to the general specification considered in this paper, they analyze the simplified model with $h_t = f_\psi(Y_{t-1}, Y_{t-2}, ..., h_{t-1}, h_{t-2}, ...)$, i.e. where the conditional variance does not depend on m. Christensen et al. (2008) provide detailed Monte-Carlo simulations for comparing their estimator with ours. Applying their procedure to the same CRSP data set, they find evidence which supports our empirical results. However, Christensen et al. (2008) do not consider the problem of testing for parametric specifications of m.

The remainder of the article is organized as follows. Section 2 reviews the empirical literature on testing the risk-return relationship by GARCH-M models. Section 3 introduces our general semiparametric framework and discusses the estimation of the nonparametric mean function. In Section 4 we then motivate the test statistic, derive its asymptotic distribution and explain the bootstrap procedure. The empirical properties of our procedure are evaluated in a Monte-Carlo simulation study in Section 5. Section 6 illustrates the method by an application to CRSP excess return data. Finally, we summarize the main conclusions in Section 7 and discuss several directions in which our approach can be naturally extended. All proofs are deferred to the appendix.

2 Modelling the Risk-Return Relation

Merton's (1973) ICAPM predicts that the conditional expected excess return on the market is linear in two components: the conditional market variance (the risk component) and the conditional market covariance with the investment opportunities (the hedge component). Under certain conditions, the equilibrium expected excess return on the market can be approximated as

$$\mathbf{E}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1}) \approx \lambda \cdot \mathbf{Var}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1}), \tag{2}$$

where $r_{M,t}$ denotes the return on the market portfolio, $r_{f,t}$ the return on the risk-free asset and λ is a positive constant equal to the representative agent's Arrow-Pratt measure of relative risk aversion.[2] Equation (2) is often referred to as a conditional single-factor model, while equations which include the covariances with the state variables are labelled conditional multi-factor models. Empirical researchers testing equation (2) have to make an assumption concerning the intertemporal nature of the conditional variance of the market. The class of GARCH-M models provides a natural workhorse in which $h_t \triangleq \mathbf{Var}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1})$ is modelled as some GARCH-type equation and $Y_t \triangleq \mathbf{E}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1}) = m_\gamma(h_t) = \lambda h_t$.

Many attempts have been undertaken to test Merton's (1973) prediction by using various formulations of the GARCH-M model.[3] The somewhat disappointing result, however, is that most empirical studies on the risk-return relation led to controversial findings, some of which indicate a positive relationship such as French et al. (1987) or Lundblad (2007), some indicate a negative relationship such as Glosten et al. (1993), while others do not find a significant relationship at all such as Bodurtha and Mark (1991).

A potential explanation for the controversial findings was rationalized by Backus and Gregory (1993).

[2]The approximation holds either if the partial derivative of the representative agent's utility with respect to wealth is much larger than the partial derivative with respect to the state variables or if the variance of the change in wealth is much larger than the variance of the change in the state variables (see Merton, 1980, p. 329).

[3]It is common to specify the mean as $m(h_t) = \mu + \lambda g(h_t)$ where $g(h_t)$ is either the conditional variance itself, the conditional standard deviation or the log of the conditional variance.

Using Mehra and Prescott's (1985) dynamic exchange economy model they show that the relation between the excess return and its conditional variance can have virtually any shape: increasing, decreasing, flat, U-shaped, inverse U-shaped or non-monotonic depending on both the preferences of the representative agent and the probability structure across states. Similarly, Genotte and Marsh (1993) constructed a general equilibrium model in which the relationship $m_\gamma(h_t) = \lambda h_t + k(h_t)$ holds, with $k(\cdot)$ depending on preferences and on the parameters of the distribution of asset returns. The Merton (1973) relationship with $k(\cdot) = 0$ is obtained only as a very special case, namely if the representative agent has logarithmic utility. Similarly, Whitelaw (2000) investigates the relation between risk and excess return in a general equilibrium exchange economy characterized by a regime-switching consumption process. While a single-regime model generates a positive and essentially linear relation between expected returns and volatility, a two-regime model leads to a complex, nonlinear relation. At the market level this relation will be negative in the long-run.

The theoretical considerations of Backus and Gregory (1993), Genotte and Marsh (1993) and Whitelaw (2000) suggest that a misspecified, i.e. too inflexible, mean function might have caused the controversial empirical results in the above mentioned studies. In the following we review two recent studies which allow for more flexible specifications of the conditional mean.[4]

Das and Sarkar (2000) suggest the ARCH-in-Nonlinear-Mean (ARCH-NM) model which defines the risk premium as a Box-Cox power transformation of the conditional variance. Obviously, this model nests the simpler parametric specifications mentioned above under certain constraints on the power transformation parameter. Although the ARCH-NM specification is favored compared to the standard specification when applied to stock return data, Das and Sarkar (2000) conclude that the model fit is not entirely satisfactory. They conjecture that the ARCH-NM is still not nonlinear enough. Going a step ahead, Linton and Perron (2003) suggest an algorithm for estimating a semiparametric (E)GARCH-M model which does not assume a functional form for the shape of the risk premium a-priori. The model is semiparametric in the sense that the conditional variance equation is modelled parametrically as GARCH or EGARCH, while the shape of the conditional mean is estimated nonparametrically.[5] Although no asymptotic theory is provided for their estimator, Monte-Carlo simulations show that the procedure works reasonably well. An application of the semiparametric EGARCH-M to excess returns on the CRSP

[4]Some studies such as Scruggs (1998) argue that the controversial results are due to an omitted variable bias: if the true relationship is a multi-factor model then single-factor models are misspecified and their estimates of λ are subject to an omitted variable bias. However, Guo and Whitelaw (2006) find that this argument should not apply when using daily data. This is because investment opportunities change slowly at the business cycle frequency and can be treated as being constant at higher frequencies.

[5]Masry and Tjøstheim (1995) investigate the problem of nonparametrically estimating both the mean and the conditional variance function. However, their procedure does not allow for a risk premium.

value-weighted index reveals a hump-shaped pattern of the risk premium which could not be detected by the parametric EGARCH-M model.

Several studies employ nonparametric techniques to estimate the conditional variance. Pagan and Ullah (1988) and Pagan and Hong (1990) argue that the conditional variance is a highly nonlinear function of the past whose form is not adequately captured by parametric GARCH-M models. They firstly estimate the conditional variance nonparametrically and then regress the excess return on the estimated conditional variance by least squares methods. Using this procedure they find a negative but insignificant in-mean coefficient. Pagan and Hong (1990) restrict h_t to be a function of the last p observations $\{Y_{t-1}, \ldots, Y_{t-p}\}$ for some fixed p in order to avoid the well known "curse of dimensionality": the optimal rate of convergence decreases with dimensionality p. This restriction however is problematic since – as has been shown in many other studies – the conditional variance is a highly persistent process and so it is unlikely that its dynamics can be adequately captured by such an estimator. Linton and Mammen (2005) suggested an alternative approach based on kernel smoothing and profiled likelihood circumventing the curse of dimensionality and nevertheless allowing the conditional variance to depend on the whole past of the process Y_t. They specify the conditional variance as additive in Y_{t-j} with the restriction that the different additive functions are proportional to each other. This implies that only one univariate function needs to be estimated. Hence their semiparametric ARCH(∞) model is capable of taking into account both nonlinearity and high persistence in the conditional variance. A similar approach is used by Li et al. (2005) who propose a test for the existence of an in-mean effect. Recently, Chen and Ghysels (2008) have extended the Linton and Mammen (2005) approach by introducing mixed data sampling (MIDAS) in the variance equation. This extension allows to recover the link between returns over short horizons and future volatility over longer horizons. MIDAS specifications for the conditional variance have been proven as useful tools for testing the risk return trade-off (see, Ghysels, 2005, and Anderson et al., 2007).

Several potential explanations (misspecification of the conditional variance, omitted variables bias, ect.) for the controversial empirical findings on the risk-return relation were addressed in the literature, but without convincing success. In this paper we focus on the obvious possibility of misspecification of the mean function. Since the parametric specification of the risk premium implied by the Merton (1973) ICAPM results from very specific assumptions, it seems natural to ask for the appropriateness of the commonly imposed functional form. Our framework allows to consider a general class of in-mean models which nest the standard GARCH-M as a special case. For such a model we address the problem of estimating nonparametrically the conditional mean function and testing for the correct choice of a particular parametric specification. The recent paper by Christensen et al. (2008) can be seen as complementary to our work. The issue of estimation is considered for a special case of our set-up, namely the situation in which h_t does not depend on m.

Recently, nonparametric approaches have received considerable attention in the financial econometrics literature. The articles by Chen and Ghysels (2008), Christensen et al. (2008), Connor et al. (2007) and Linton and Sancetta (2007) are only a few examples for this development. The problem considered in this article further extends this path of research.

3 Estimation Strategy for the General Semiparametric Model

In this section we define the general model and introduce the estimation strategy. The important issue of testing for parametric specifications of m is discussed in the next section.

The general model is defined as follows:

Assumption 1. *The data are generated by*

$$Y_t = m_0(h_t) + \varepsilon_t, \tag{3}$$

where ε_t *fulfills* $\mathbf{E}[\varepsilon_t|\mathcal{F}_{t-1}] = 0$ *for an increasing* σ*-field* $\mathcal{F}_t$ *with the property that* (ε_t, h_{t+1}) *is* $\mathcal{F}_t$*-measurable.*

The process h_t is an unobserved one-dimensional process. We assume that h_t can be consistently estimated by known functions $\widehat{h}_t$ that depend on parameters ψ and m and on the past observations $Y_{t-1}, Y_{t-2}, \dots, Y_1$. We denote the true parameter values by ψ_0 and m_0, i.e. $h_t = \widehat{h}_t(\psi_0, m_0)$. More generally, we allow this equality to hold only approximately, i.e. that the difference $h_t - \widehat{h}_t(\psi_0, m_0)$ is of asymptotically negligible order, see below. For simplicity, dependence of quantities and functions on $Y_{t-1}, Y_{t-2}, \dots$ is suppressed in our notation. In this section we discuss estimation of the regression function m on a compact interval I. A typical example could be that h_t follows a GARCH$(1,1)$ process or another specification from the GARCH family. Then $\widehat{h}_t(\psi_0, m_0)$ differs from h_t because the starting values of the GARCH autoregression are not known. In the asymptotic treatment, implicitly we assume that the first observations are used in generating the fit of h_t but not in the estimation of m without explicitly mentioning this and indicating this in the notation and theoretical discussion. This allows us to assume that $h_t - \widehat{h}_t(\psi_0, m_0)$ is small for all t and it simplifies the notation. Most importantly, we allow $\widehat{h}_t$ to depend on the function m_0. In particular, this is the case if $\widehat{h}_t$ depends on ψ_0 and on the residuals $\varepsilon_1, ..., \varepsilon_{t-1}$, see also the discussion above. Our central assumption on $\widehat{h}_t$ is that it is measurable with respect to $\mathcal{F}_{t-1}$.

Assumption 2. *The (random) function* $\widehat{h}_t$ *is measurable with respect to* $\mathcal{F}_{t-1}$*. It holds that*

$$|\widehat{h}_t(\psi_0, m_0) - h_t| \le U_T$$

for all values t with $h_t \in I_b$. Here I_b is the set of all points x with distance from I less than b. Furthermore, U_T is a random variable with $U_T = o_P(T^{-\omega})$ with $\omega > \eta$, where b is the bandwidth of our kernel smoothing and $T^{-\eta}$ is the order of b, see Assumption 8 below.

We make the following mixing condition for the covariate process.

Assumption 3. *The process h_t is stationary and β-mixing with mixing coefficients $\beta(j) \leq cv^j$ for constants $c > 0$ and $0 < v < 1$. The density f_h of h_t is Lipschitz continuous and bounded away from 0 on I. The joint density of h_t and h_{t+s} is bounded on $I \times I$, uniformly in s.*

The β-mixing condition in Assumption 3 could be replaced by the assumption that $\beta(j) \leq aj^{-c}$ for a constant $a > 0$ and for a constant c that is large enough. We avoided an exact check of the necessary size of the constant c because we have no examples of ARCH models where Assumption 3 does not hold but where this weaker assumption applies.

For the function m we assume that it is a smooth function on I and that it is parametrically specified by a finite dimensional parameter vector γ outside I. In the framework of testing we consider the hypothesis that m is specified on the whole real line by the parameter γ. We denote the supnorm over I by $\|m\|_\infty = \sup_{x\in I} |m(x)|$. We also write m for the restriction of m to I and write then $\widehat{h}_t(\psi, \gamma, m)$. The parameter vector $\theta = (\psi, \gamma)$ is an element of a normed space endowed with the norm $\| \cdot \|$. When we consider the issue of testing we will restrict the discussion to the parametric case that this normed space is finite dimensional.

We consider an iterative estimation scheme where in each iteration step the estimators of m and $\theta = (\psi, \gamma)$ are updated. We first discuss the asymptotic theory for the case of one iteration step. The general theory then follows by an iterative application of the result. The initial estimators are denoted by $\widetilde{m}$, $\widetilde{\theta} = (\widetilde{\psi}, \widetilde{\gamma})$ and the updated estimators by $\widehat{m}$, $\widehat{\theta} = (\widehat{\psi}, \widehat{\gamma})$. Our theoretical result implies that the updated estimator $\widehat{m}$ fulfills the conditions needed for the starting value of the iteration. Thus, our result can be employed for an iterative application. For estimation, the estimator of Linton and Sancetta (2007) can be used as an initial estimator, for testing, the iteration can start with a parametric estimator, see the discussions in Section 4. We make the following assumptions on the preliminary estimators and on the dependence of $\widehat{h}_t$ on their arguments.

Assumption 4. *The estimators $\widetilde{m}$ and $\widetilde{\theta}$ fulfill*

$$\|\widetilde{\theta} - \theta_0\| = o_P(T^{-\delta_\theta}),$$

$$\|\widetilde{m} - m_0\|_\infty = o_P(T^{-\delta_m}),$$

$$\|D_2\widetilde{m} - D_2 m_0\|_\infty = o_P(T^{\xi_0})$$

for constants $0 < \delta_\theta < 1/2$, $0 < \delta_m$ and real ξ_0. Put also $\xi = \max\{\xi_0, 0\}$. Here, we write $D_j m$ for the j-th derivative of a function m.

Assumption 5. *For $\theta_j = (\psi_j, \gamma_j), m_j$ ($j = 1, 2$) with $\|\theta_j - \theta_0\| \le T^{-\delta_\theta}$, $\|m_j - m_0\|_\infty \le T^{-\delta_m}$, $\|D_2 m_j - D_2 m_0\|_\infty \le T^{\xi_0}$ we assume that*

$$|\widehat{h}_t(\theta_1, m_1) - \widehat{h}_t(\theta_2, m_2)| \le V_T\|\theta_1 - \theta_2\| + W_T\|m_1 - m_2\|_\infty + R_T.$$

Here V_T and W_T are random variables with $V_T = O_P(T^{\rho_\theta})$, $W_T = O_P(T^{\rho_m})$ and $R_T = o_P(T^{-\rho_0})$ with constants $0 \le \rho_m < \delta_m - \eta$, $0 \le \rho_\theta < \delta_\theta - \eta$ and $0 \le \rho_0 < (1 + \eta)/2$.

Assumption 6. *For $\epsilon > 0$ it holds with a constant $C > 0$ that*

$$H(\epsilon, \|\cdot\|, \{\theta : \|\theta - \theta_0\| \le T^{-\delta_\psi}\}) \le C\epsilon^{-1/2}T^{(\xi+\rho_m-\rho_\theta)/2}.$$

Here for a set A, $H(\epsilon, \|\cdot\|, A) = \log N(\epsilon, \|\cdot\|, A)$ is the entropy of A, i.e. $N(\epsilon, \|\cdot\|, A)$ is the number of balls with radius ϵ that are necessary to cover A.

Note that Assumption 6 is fulfilled for the particular case of finite-dimensional θ.

The next assumption is needed because the techniques from empirical process theory that will be used below require subexponential tails.

Assumption 7. *It holds that $\mathbf{E}[\exp(\rho|\varepsilon_t|)|\mathcal{F}_{t-1}] < C$ almost surely for $\rho > 0$ small enough and $h_t \in I$ with a constant $C < \infty$.*

In this assumption we only require conditional subexponential tails of ε_t if h_t lies in the bounded set I. In particular, it is not assumed that ε_t has unconditional subexponential tails. The condition is fullfilled for GARCH-specifications with i.i.d. $\varepsilon_t/\sqrt{h_t}$ that have a subexponential distribution, e.g. Gaussian.

We now introduce our smoothing estimators of m. For estimation of m we use a Nadaraya-Watson smoother $\widehat{m}^{NW}$ and a local linear estimator $\widehat{m}^{LL}$, for testing we only rely on Nadaraya-Watson smoothing. The construction of our test is such that the bias term cancels by substracting an asymptotically equivalent term. Thus for testing, local linear smoothing does not offer advantages over Nadaraya-Watson smoothing. Our main technical tool is to show that the stochastic part of the local linear and of the Nadaraya-Watson estimator is asymptotically equivalent to the "oracle estimator" $\widehat{m}^{*,LL}$ and $\widehat{m}^{*,NW}$ that is based on smoothing Y_t versus h_t.

We decompose $\widehat{m}^{LL} = \widehat{m}^{LL,A} + \widehat{m}^{LL,B}$, ..., $\widehat{m}^{*,NW} = \widehat{m}^{*,NW,A} + \widehat{m}^{*,NW,B}$ into a stochastic component (superindex A) and a mean part (superindex B). Here,

$$\widehat{m}^{NW,A}(x) = \frac{\widehat{r}^{NW,A}(x)}{\widehat{f}(x)},\ \widehat{m}^{NW,B}(x) = \frac{\widehat{r}^{NW,B}(x)}{\widehat{f}(x)},$$

with $\widehat{r}^{NW,A}(x) = \frac{1}{T}\sum_{t=1}^{T} K_b(\widetilde{h}_t - x)[Y_t - m(h_t)]$, $\widehat{r}^{NW,B}(x) = \frac{1}{T}\sum_{t=1}^{T} K_b(\widetilde{h}_t - x)m(h_t)$, $\widehat{f}(x) = \frac{1}{T}\sum_{t=1}^{T} K_b(\widetilde{h}_t - x)$ and $\widetilde{h}_t = \widehat{h}_t(\widetilde{\psi}, \widetilde{\gamma}, \widetilde{m})$. Here, $K_b(\cdot) = b^{-1}K(\cdot/b)$ is a kernel with kernel function K and bandwidth parameter b. Kernel and bandwidth fulfill the following standard smoothing conditions.

Assumption 8. *The kernel K has bounded support ($[-1, 1]$, say) and a continuous derivative. The bandwidth b is of order $T^{-\eta}$, i.e.*

$$0 < \liminf_{T\to\infty} T^{\eta} b \leq \limsup_{T\to\infty} T^{\eta} b < \infty$$

for a constant η with $0 < \eta < \frac{1}{3}$.

The components of the local linear estimator are defined by $\widehat{m}^{LL,A}(x) = \alpha^A$ and $\widehat{m}^{LL,B}(x) = \alpha^B$ where (α^A, β^A) and (α^B, β^B) minimize

$$\sum_{t=1}^{T} K_b(\widetilde{h}_t - x)[Y_t - m(h_t) - \alpha^A - \beta^A(\widetilde{h}_t - x)]^2 = min,$$
$$\sum_{t=1}^{T} K_b(\widetilde{h}_t - x)[m(h_t) - \alpha^B - \beta^B(\widetilde{h}_t - x)]^2 = min.$$

The oracle estimators $\widehat{m}^{*,LL}$, $\widehat{m}^{*,NW}$ and their components $\widehat{m}^{*,LL,A}$, ..., $\widehat{m}^{*,NW,B}$ are defined as $\widehat{m}^{LL}$, $\widehat{m}^{NW}$, ... with $\widetilde{h}_t$ replaced by h_t. Our first theorem compares the stochastic parts of the local linear and of the Nadaraya-Watson estimators with their oracle counterparts. It states that the differences are asymptotically negligible. The reason for providing a separate theorem for the stochastic parts of the estimators is that for our testing procedures no results on the mean parts are needed.

Theorem 1. *Assume that Assumptions 1 – 8 apply. For κ with $\kappa < \min\{\delta_m - \rho_m, \delta_\theta - \rho_\theta\} - \eta - \xi/4$, $\kappa < \rho_0 - (1 + \eta)/2$ and $\kappa < \omega - \eta$ it holds that*

$$\sup_{x\in I} \left|\widehat{m}^{LL,A}(x) - \widehat{m}^{*,LL,A}(x)\right| = o_P(T^{-(1/2)+(\eta/2)-\kappa}), \tag{4}$$
$$\sup_{x\in I} \left|\widehat{m}^{NW,A}(x) - \widehat{m}^{*,NW,A}(x)\right| = o_P(T^{-(1/2)+(\eta/2)-\kappa}). \tag{5}$$

Under the additional assumption that K is three times continuously differentiable it follows that

$$\sup_{x\in I} \left|D_2\widehat{m}^{LL,A}(x) - D_2\widehat{m}^{*,LL,A}(x)\right| = o_P(T^{-(1/2)+(5\eta/2)-\kappa}), \tag{6}$$
$$\sup_{x\in I} \left|D_2\widehat{m}^{NW,A}(x) - D_2\widehat{m}^{*,NW,A}(x)\right| = o_P(T^{-(1/2)+(5\eta/2)-\kappa}). \tag{7}$$

The essential assumption of the theorem is that the rate of convergence of the preliminary estimator $T^{-\delta_m}$ is faster than the rate of the bandwidth $T^{-\eta}$. If the second derivative of the preliminary estimator does not grow too fast to infinity the constant κ can be chosen as $\kappa > 0$. Then the difference between the stochastic parts of the estimators and their oracle counterparts is of lower order as the rate $T^{-(1/2)+(\eta/2)}$. This is the rate of the oracle estimator. Thus, the differences are asymptotically negligible. For slightly more rapid growth of the second derivative we do not get asymptotic equivalence but it still holds that the rate of convergence of the updated estimators is faster than that of the preliminary estimators. Repeated application of Theorem 1 can be used to show asymptotic equivalence after a finite number of iterations. For such an application we have added in Theorem 1 results on rates for the second derivatives of the estimators.

We now discuss the bias terms of Nadaraya-Watson and local linear smoothing. For the asymptotic treatment we need the following additional assumptions.

Assumption 9. *There exist $\delta_1, \delta_2, \delta_3 > 0$ such that for θ_j and m_j $(j = 1, 2)$ with $\|\theta_j - \theta_0\| < \delta_1, \|m_j - m_0\|_\infty < \delta_2, \|D_1 m_j - D_1 m_0\|_\infty < \delta_3$ it holds that the (multivariate) process $(h_t, \widehat{h}_t(\theta_1, m_1), \widehat{h}_t(\theta_2, m_2))$ is β-mixing with mixing coefficients $\beta(j) < cv^j$ for constants $c > 0$, $0 < v < 1$.*

Assumption 10. *We assume that*

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{E}\left[(\widehat{h}_t(\theta, m) - h_t)K_b(\widehat{h}_t(\theta, m) - x)\right] = o(T^{-2\eta})$$

uniformly for $(\theta, m, x) \in \mathcal{G}_T$, where $\mathcal{G}_T$ is the set of tuples (θ, m, x) with $\|\theta - \theta_0\| \le T^{-\delta_\theta}, \|m - m_0\|_\infty \le T^{-\delta_m}, \|D_2 m - D_2 m_0\|_\infty \le T^{\xi_0}$ and $x \in I$.

The following additional assumption is needed for Nadaraya-Watson smoothing.

Assumption 11. *We assume that*

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{E}\left[(\widehat{h}_t(\theta, m) - x)K_b(\widehat{h}_t(\theta, m) - x) - (h_t - x)K_b(h_t - x)\right] = o(T^{-2\eta})$$

uniformly for $(\theta, m, x) \in \mathcal{G}_T$.

The next theorem discusses local linear and Nadaraya-Watson smoothing for bandwidth of order $T^{-\eta}$ with $\eta = 1/5$. For twice differentiable regression functions the optimal rate is than of order $T^{-2/5}$. The theorem states that the difference between the estimators and their oracle counterparts is asymptotically negligible. In particular, this implies pointwise asymptotic normality of the estimators.

Theorem 2. *Assume that Assumptions 9 – 10 and the assumptions of Theorem 1 apply with $\kappa > 0$, $\eta = 1/5$ and $min(\delta_m - \rho_m, \delta_\theta - \rho_\theta, \rho_0, \omega) > \xi + \rho_m$. Then it holds that*

$$\sup_{x \in I} \left|\widehat{m}^{LL}(x) - \widehat{m}^{*,LL}(x)\right| = o_P(T^{-2/5}). \tag{8}$$

Denote the conditional variance of ε_t by $\sigma^2(x) = \mathbf{E}[\varepsilon_t^2 | h_t = x]$. Suppose that σ^2 does not depend on t and that for an x in the interior of I, σ^2 and f_h is continuous at x and m is twice continuously differentiable at the point x. Then

$$\sqrt{nb}\left[\widehat{m}^{LL}(x) - m(x) - \frac{1}{2}b^2 m''(x)\int u^2 K(u)du\right]$$

converges in distribution to $\mathcal{N}(0, \sigma^2(x) f_h(x) \int K^2(u)du)$. *Under the additional assumption that* f_h *is continuously differentiable and that Assumption 11 holds, one obtains for the Nadaraya-Watson smoother that*

$$\sup_{x \in I} \left| \widehat{m}^{NW}(x) - \widehat{m}^{*,NW}(x) \right| = o_P(T^{-2/5}) \tag{9}$$

and that the same limit result holds for

$$\sqrt{nb} \left[\widehat{m}^{NW}(x) - m(x) - \frac{1}{2} b^2 \left(2 \frac{f_h'(x) m'(x)}{f_h(x)} + m''(x) \right) \int u^2 K(u) du \right].$$

4 Testing for Parametric Mean Specifications

In this section we suggest a procedure for testing parametric specifications of m. The test procedure makes use of the nonparametric estimator of m of the last section. Nonparametric estimation in the context of testing is simpler for two reasons. First, one can use the parametric fit of m as the starting value in the iterative procedure for estimating m. Second, our test statistic is constructed such that bias terms of the nonparametric estimator cancel out. This is achieved by comparing in the test statistic two smoothers that have the same asymptotic bias. The first smoother is based on regressing Y_t on the fit of h_t. The second smoother regresses the parametric fit for the conditional mean of Y_t onto h_t.

Under the null hypothesis we consider an in-mean model with a parametric mean function depending on a finite-dimensional parameter γ_0:

$$Y_t = m_{\gamma_0}(h_t) + \varepsilon_t, \tag{10}$$

where, as in the last section, ε_t fulfils $\mathbf{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 0$, where $\mathcal{F}_t$ is an increasing σ-field with the property that (ε_t, h_{t+1}) is $\mathcal{F}_t$-measurable. On the hypothesis, the covariates h_t can be approximated by $\widehat{h}_t(\theta_0)$ for parameters $\theta_0 = (\psi_0, \gamma_0)$ and a measurable function $\widehat{h}_t$, such that the difference $h_t - \widehat{h}_t(\psi_0, \gamma_0)$ is asymptotically negligible, see below. The function $\widehat{h}_t$ depends on the parameters ψ, γ and on the past observations $Y_{t-1}, Y_{t-2}, \ldots, Y_1$. Again for simplicity, dependence of quantities and functions on $Y_{t-1}, Y_{t-2}, \ldots$ is suppressed in the notation. We assume that the true parameter vector $\theta_0 = (\psi_0, \gamma_0)$ is in the interior of Θ, a compact, convex, and finite dimensional parameter space. In particular, in this section we consider only parametric specifications of ψ.

The alternative model is given by a semiparametric version of equation (10) with a smooth mean function $m(\cdot)$, but ε_t and h_t as before. The semiparametric alternative has two distinct advantages over previous specifications: (i) it does not rely on any parametric specification of $m(\cdot)$, and (ii) it allows for persistence in the conditional variance process since it does not restrict $\mathcal{F}_{t-1}$ as e.g. in Pagan and Hong (1990). For the special case of GARCH-M models the specification under the alternative is closely related to the model considered by Linton and Perron (2003).

4.1 Iterative Estimation of Conditional Mean and Variance

For some initial parametric estimators $\widehat{\gamma}$ and $\widehat{\psi}^{(0)}$ we consider the estimate $\widehat{h}_t^{(0)} = \widehat{h}_t(\widehat{\theta}^{(0)})$ of h_t. Here, $\widehat{\theta}^{(0)} = (\widehat{\psi}^{(0)}, \widehat{\gamma})$.

We will use iterative updates of the estimate $\widehat{\psi}^{(0)}$. These updates are denoted by $\widehat{\psi}^{(k)}$ with $k \geq 1$. The estimator of γ_0 will not be updated. This is done for the following reason. Because our semiparametric alternative model contains nonparametric components, updates of the parametric estimators will slow down the rate of convergence to nonparametric rates. Our test for the parametric hypothesis is based on the comparison of estimators of m_{γ_0} on the hypothesis and on the alternative. If the estimate of γ_0 is updated this will introduce an additional bias term that does not cancel out when comparing the estimators on the hypothesis and on the alternative.

The iterative update of the estimators of ψ_0 and h_t and of the nonparametric estimator of $m_0 = m_{\gamma_0}$ works as follows. Given the fit $\widehat{h}_t^{(k-1)}$ of h_t calculated in the $(k-1)$-th cycle, the estimate of m_0 is updated by smoothing Y_t versus $\widehat{h}_t^{(k-1)}$. The resulting smoother is denoted by $\widehat{m}^{(k)}$. Then using the observations and $\widehat{m}^{(k)}$, the estimators of ψ_0 and h_t are updated. The resulting estimators are denoted by $\widehat{\psi}^{(k)}$ and $\widehat{h}_t^{(k)}$. We now describe the iteration steps in more detail.

For x in a bounded closed interval I and $k \geq 1$ the updated estimator of m_{γ_0} is defined as

$$\widehat{m}^{(k)}(x) = \frac{\widehat{r}^{(k)}(x)}{\widehat{f}_h^{(k)}(x)} + m_{\widehat{\gamma}}(x), \tag{11}$$

with $\widehat{r}^{(k)}(x) = \frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k-1)} - x)[Y_t - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})]$ and $\widehat{f}_h^{(k)}(x) = \frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k-1)} - x)$ and where $K_b(\cdot) = b^{-1}K(\cdot/b)$ with K being a kernel function and bandwidth parameter b. In the simulations we also use the update

$$\check{m}^{(k)}(x) = \widehat{f}_h^{(k)}(x)^{-1}\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k-1)} - x)Y_t. \tag{12}$$

However, the theoretical treatment of $\widehat{m}^{(k)}(x)$ is easier because, as mentioned above, bias terms cancel in the asymptotic analysis that otherwise could only be analyzed under rather strong additional assumptions, see also the bias discussions in Section 3. For $x \notin I$ the estimate $\widehat{m}^{(k)}(x)$ is put equal to the old estimate $m_{\widehat{\gamma}}(x)$. Thus for $x \notin I$ the estimate of $m_{\widehat{\gamma}}(x)$ is not updated. Alternatively, an updated parametric fit for $x \notin I$ could also be considered. For simplicity, this not pursued here. Furthermore, it could be considered that the choice of the interval I depends on the sample size T and grows to the positive real line for $T \to \infty$. We also do not discuss this here. In the simulations we have chosen a GARCH-specification with h_t as conditional variance and we have fitted m nonparametrically on the whole real line. We conjecture that under our mixing conditions this makes an asymptotically negligible difference.

In a next step the fit of h_t is updated. We suppose that the update $\widehat{h}_t^{(k)}$ can be written as a function of $\widehat{m}^{(k)}$, $\widehat{\gamma}$ and $\widehat{\psi}^{(k)}$ and the observations $Y_1, ..., Y_{t-1}$. Again, we suppress dependence on $Y_1, ..., Y_{t-1}$ in the notation and we write $\widehat{h}_t^{(k)} = \widehat{h}_t(\widehat{\psi}^{(k)}, \widehat{\gamma}, \widehat{m}^{(k)})$ where in abuse of notation we denote the function by $\widehat{h}_t$, as the related function $\widehat{h}_t$ of step 0. We suppose that the function does not depend on k and that $\widehat{h}_t(\widehat{\psi}^{(0)}, \widehat{\gamma}, m_{\widehat{\gamma}}) = \widehat{h}_t(\widehat{\psi}^{(0)}, \widehat{\gamma})$.

The above procedure can be performed for a finite fixed number of iterations or until a convergence criterium is fulfilled. The asymptotic theory is developed for a fixed number of iterations. In the simulations we use the criterium

$$\delta(k) = \frac{\sum_{j=1}^{J}\left(\widehat{m}^{(k)}(x_j) - \widehat{m}^{(k-1)}(x_j)\right)^2}{\sum_{j=1}^{J}\left(\widehat{m}^{(k-1)}(x_j)\right)^2 + \bar{c}} < \bar{c} \tag{13}$$

for some small prespecified $\bar{c}$, where x_j, $j = 1, \ldots, J$, are equally spaced grid points on I. We choose $\bar{c} = 0.001$.

4.2 The Test Statistic

We now come to the test statistic which will be based on the difference between a smoothed version of the initial parametric estimator and a Naradaya-Watson kernel estimator of the regression function. The null and alternative hypothesis can be written as

$$\begin{aligned} & H_0: && \mathbf{P}(m(\cdot) = m_{\gamma_0}(\cdot)) = 1 \text{ for some } \gamma_0 \in \Theta_\gamma = \{\gamma | (\psi, \gamma) \in \Theta\} \\ \text{and} \quad & H_1: && \mathbf{P}(m(\cdot) = m_{\gamma}(\cdot)) < 1 \text{ for any } \gamma \in \Theta_\gamma = \{\gamma | (\psi, \gamma) \in \Theta\}. \end{aligned}$$

The test statistic utilizes the fact that the null hypothesis is equivalent to the condition that the L_2-distance between the two functions is zero.

We consider the following test statistic

$$\widehat{\Gamma}_T^{(k)} = \int \left\{ \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k)} - x)\left[Y_t - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})\right]}{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k)} - x)} \right\}^2 w(x)dx, \tag{14}$$

where $w(x)$ is some nonnegative and bounded weighting function.

Note, that in the test statistic we subtract $m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})$ from Y_t and not $m_{\widehat{\gamma}}(\widehat{h}_t^{(k)})$. This is done in order to have a parametric rate for $m_{\gamma_0}(h_t) - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})$ on the hypothesis. In the simulations we also experimented with $m_{\widehat{\gamma}}(\widehat{h}_t^{(k)})$. Both choices lead to almost identical results.

Equation (14) can be interpreted as the integrated squared difference between a smoothed version of the initial parametric estimate $m_{\widehat{\gamma}}$ and the Naradaya-Watson kernel estimate $\check{m}^{(k+1)}$ of the regression function $m(x)$ defined in equation (12). The reason for smoothing the parametric estimate is that whereas $m_{\widehat{\gamma}}$ is asymptotically unbiased and converging at rate $\sqrt{T}$, the nonparametric estimate $\check{m}^{(k+1)}$ has a kernel smoothing bias and convergence rate $\sqrt{Tb}$. Replacing $m_{\widehat{\gamma}}$ by its smoothed version introduces an artificial bias. As a result, under the null hypothesis the bias of $\check{m}^{(k+1)}$ cancels with the one of the smoothed version of the parametric estimate $m_{\widehat{\gamma}}$. In the simpler set-up of Härdle and Mammen (1993) it was explained that not smoothing $m_{\widehat{\gamma}}$ would lead to a test that asymptotically behaves like a linear test that only looks for deviations from the null hypothesis in one direction.

For the case of independent and identically distributed observations Härdle and Mammen (1993) have shown that under the null hypothesis the above test statistic with h_t observable (and $k = 0$) has an asymptotic normal distribution. Kreiss et al. (2002) extend the results of Härdle and Mammen (1993) to settings with dependent data.

We start with a discussion of the asymptotic behavior of $\widehat{\Gamma}_T^{(k)}$ for $k = 0$.

Assumption 12. *The function $m_\gamma(x)$ is differentiable with respect to γ at the point $\gamma = \gamma_0$ for all $x \in I$ and for the derivative $\dot{m}_{\gamma_0}$ it holds that*

$$\sup_{x\in I, \|\gamma-\gamma_0\|\leq\delta} \left| m_\gamma(x) - m_{\gamma_0}(x) - (\gamma-\gamma_0)^T \dot{m}_{\gamma_0}(x) \right| = O(\delta^2)$$

for $\delta \to 0$. The derivative $\dot{m}_{\gamma_0}$ fulfills the following Lipschitz condition

$$\sup_{u,v\in I, \|u-v\|\leq\delta} |\dot{m}_{\gamma_0}(u) - \dot{m}_{\gamma_0}(v)| = O(\delta^\kappa)$$

for $\delta \to 0$ with a constant $\kappa > 0$. Furthermore, $m_\gamma(x)$ is continuously differentiable with respect to x for $x \in I$.

Assumption 13. *It holds that $\|\widehat{\theta}^{(0)} - \theta_0\| = O_P(T^{-1/2})$.*

Assumption 14. *There exists a stationary sequence $\dot{h}_t$ such that for $C > 0$*

$$\sup \left| \widehat{h}_t(\theta) - \widehat{h}_t(\theta_0) - (\theta-\theta_0)^T \dot{h}_t \right| = o_P(T^{-1/2}\log(T)^{-1/2}),$$

where the supremum runs over all t and θ with $\|\theta - \theta_0\| \leq CT^{-1/2}$, and with $\widehat{h}_t(\theta)$ or $\widehat{h}_t(\theta_0)$ or h_t in I. The process $(\dot{h}_t, h_t, \varepsilon_t)$ is stationary and β-mixing with $\beta(j) \leq cv^j$ for constants c and v as in Assumption 3. Furthermore $\mathbf{E}|\dot{h}_t|^r$ is finite for an $r > 2$.

Assumption 15. *For $C > 0$, $1 \leq t \leq T, \|\theta - \theta_0\| \leq CT^{-1/2}, \|\theta' - \theta_0\| \leq CT^{-1/2}$ it holds that*

$$|\widehat{h}_t(\theta) - \widehat{h}_t(\theta')| \leq R_T\|\theta' - \theta\| + S_T,$$
$$|\widehat{h}_t(\theta_0) - h_t| \leq S_T$$

for random sequences R_T and S_T with $R_T = O_P(T^\varsigma)$ and $S_T = O_P(T^{-\nu})$ for constants ς and ν with $0 < 2\varsigma < 1 - 3\eta$ and $\nu > 3\eta/2$ for η as in Assumption 8.

Assumption 16. *The weight function w is continuous and the closure of its support lies in the interior of I.*

For testing, we do not assume that the bandwidth is of an order that is optimal for estimation under certain smoothness conditions on m_{γ_0}, e.g. that the bandwidth is of order $T^{-1/5}$. Such an assumption would be too restrictive because tests that look for more global deviations from the hypothesis make also sense. Assumption 12 is a condition on the smoothness of the mean function. Assumptions 13 – 15 state conditions on the accuracy of the estimates of θ_0 and h_t and on the smoothness of $\widehat{h}_t(\theta)$ as a function of θ. Assumptions 13 and 14 are needed because we make no assumptions on the specific form of the estimators of the parameters. We remark that Assumption 15 is very weak because it is allowed that the random variable R_T may grow with rate T^{ς} for a positive constant ς. In Assumptions 15 and 16 it would be more realistic to allow for the case that h_t does not have the required properties for an initial period $1 \leq t < T^{\alpha}$ with $\alpha > 0$ small enough. This could be incorporated into our theory, but as for estimation, it is omitted in order to simplify the analysis. The theory directly applies if observations of the initial period are not used for the estimation of m.

The following theorem states that under the null hypothesis $T\sqrt{b}\,\widehat{\Gamma}_T^{(0)}$ is asymptotically normal.

Theorem 3. *Assume that Assumptions 3,8,12 – 16 apply. It holds that* $\sup_{x\in I}\mathbf{E}[\varepsilon_t^{4+\delta}|h_t = x] < \infty$ *for some* $\delta > 0$. *For* $x \in I$ *the conditional moment* $\mathbf{E}[\varepsilon_t^{4+\delta}|h_t = x]$ *and the conditional variance* $\sigma^2(x) = \mathbf{E}[\varepsilon_t^2|h_t = x]$ *of* ε_t *are Lipschitz continuous on* I. *The density* f_h *is continuous and* m *is twice continuously differentiable in* I. *Then under* H_0 *it holds that*

$$T\sqrt{b}\ \frac{\widehat{\Gamma}_T^{(0)} - b^{-1/2}M}{\sqrt{V}} \tag{15}$$

converges in distribution to a standard normal distribution. Here

$$\begin{aligned} M &= K^{(2)}(0)\int \sigma^2(x)w(x)f_h^{-1}(x)dx, \\ V &= 2K^{(4)}(0)\int \sigma^4(x)w^2(x)f_h^{-2}(x)dx, \end{aligned}$$

and $K^{(k)}$ *denotes the* k*-fold convolution of* K *with itself.*

We now discuss the test statistic $\widehat{\Gamma}_T^{(k)}$ for $k \geq 1$. We will show that replacing $\widehat{h}_t^{(0)}$ by the iterative estimator $\widehat{h}_t^{(k)}$ described above does not effect the asymptotic distribution of the test statistic under the null hypothesis. On the other hand we will argue below that the test statistic $\widehat{\Gamma}_T^{(k)}$ leads to a significant increase of the power on the alternative. The following additional assumptions are needed to obtain our next result on the asymptotic distribution of $\widehat{\Gamma}_T^{(k)}$ for $k \geq 1$.

Our next theorem states that on the hypothesis $\widehat{\Gamma}_T^{(k)}$ has the same asymptotic distribution as $\widehat{\Gamma}_T^{(0)}$.

Theorem 4. *Assume that the assumptions of Theorem 3 hold with* $\frac{1}{9} + \frac{8}{9}\rho_m < \eta < \frac{3}{11} - \frac{8}{11}\rho_m$. *Furthermore assume that Assumptions 5 and 7 apply for some* $\delta_\theta, \delta_m < \min[2\eta, (1-\eta)/2]$. *Then under* H_0 *it* *holds that*

$$T\sqrt{b}\ \frac{\widehat{\Gamma}_T^{(k)} - b^{-1/2}M}{\sqrt{V}}$$

converges in distribution to a standard normal distribution. Here M and V are defined as in Theorem 3.

The advantage of using $\widehat{\Gamma}_T^{(k)}$ with $k \geq 1$ in comparison to $\widehat{\Gamma}_T^{(0)}$ may be explained as follows. The power of the test statistic depends on the accuracy with which the nonparametric estimate of the mean function can approximate the true mean function. Under the alternative, the parametric model for the mean which is initially estimated is misspecified. As a consequence, the nonparametric estimate of the mean function based on the inconsistent estimate $\widehat{h}_t^{(0)}$ will poorly approximate the true mean function. This leads to a low power of the test statistic $\widehat{\Gamma}_T^{(0)}$. The simulations in the next section will show that the iterative estimation procedure overcomes this problem and results in a precise estimate of $m(\cdot)$. The test statistic $\widehat{\Gamma}_T^{(k)}$ which is based on this iterated estimate will dispose of considerably better power properties than $\widehat{\Gamma}_T^{(0)}$.

Note, that we did not distinguish between the bandwidth parameter used for the estimation of the mean function and the one used in the test statistic. In the derivation of the theorems we treat them as identical. In the simulations and in the application we choose the bandwidth parameter in the iterative estimation procedure by cross-validation as was suggested in Linton and Perron (2003) and is discussed in the next subsection. To reduce notation we do not equip the bandwidth parameter with an index k. Additionally, we will report the test statistic for several choices of the bandwidth in order to document the robustness of the outcome with respect to variations in the bandwidth parameter.

The asymptotic power of both tests ($k = 0$ and $k > 0$) can be analysed under additional assumptions on the parametric estimators on the alternative. If the regression function m differs from the parametric specification $m_\gamma(x)$ by a term $T^{-1/2}b^{-1/4}\delta(x)$ (for a fixed function δ) then the limiting distribution of $T\sqrt{b}(\widehat{\Gamma}_T^{(k)} - b^{-1/2}M)$ is equal to $\mathcal{N}(\int w(x)\delta^2(x)dx, V)$. Thus, deviations of order $T^{-1/2}b^{-1/4}$ are detected. One can show that the power is uniform over Sobolev balls of alternatives. This is in contrast to goodness-of-fit tests that detect $n^{-1/2}$ alternatives, but do not achieve power uniformly. The limit does not depend on the number k of iterations. But, as argued above, for noncontiguous alternatives the power may be quite different. This can be seen in the simulations where the one step test has a very poor power compared to the fully iterated version.

4.3 Parametric and Semiparametric GARCH$(1,1)$-M

We now discuss model (10) for the special case of parametric GARCH$(1,1)$-M specification which is the most popular version of such a model. Then we will briefly explain the semiparametric GARCH$(1,1)$-M version of Linton and Perron (2003) and relate their approach to ours.

The GARCH$(1,1)$-M model is given by

$$Y_t = m_{\gamma_0}(h_t(\theta_0)) + \varepsilon_t, \tag{16}$$

$$\varepsilon_t = \sqrt{h_t(\theta_0)} Z_t, \tag{17}$$

$$h_t(\theta_0) = \omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 h_{t-1}(\theta_0) \tag{18}$$

with i.i.d. mean zero variables Z_t with variance equal to one. The conditional expectation of Y_t is parameterized as $m_{\gamma_0}(h_t(\theta_0)) = \mu_0 + \lambda_0 g(h_t(\theta_0))$. The vector θ contains the parameters of the mean and variance functions, i.e. $\theta_0 = (\psi_0, \gamma_0)$, with $\psi_0 = (\omega_0, \alpha_0, \beta_0)$ and $\gamma_0 = (\mu_0, \lambda_0)$. Three parametric specifications for the function g are commonly applied. The original Engle et al. (1987) specification assumes either $g(h_t(\theta_0)) = h_t(\theta_0)$ or $g(h_t(\theta_0)) = \sqrt{h_t(\theta_0)}$, while some authors also use $g(h_t(\theta_0)) = \ln(h_t(\theta_0))$. As noted by Pagan and Hong (1990) this latter specification is possibly unsatisfactory, since as $h_t(\theta_0) \to 0$ the conditional variance in logs takes very large negative values and the relationship between the conditional variance and Y_t may be overstated. Of course, when λ_0 is restricted to being zero the GARCH-M reduces to the Bollerslev (1986) GARCH model.

The GARCH$(1,1)$-M process will be both strictly and covariance stationary if (i) $Z_t \overset{iid}{\sim} \mathcal{N}(0,1)$ and (ii) $\alpha_0 + \beta_0 < 1$. Note, that strict stationarity and ergodicity of the process only require $\mathbf{E}[\ln(\alpha_0 Z_t^2 + \beta_0)] < 1$ which is weaker then the condition implying covariance stationarity. Specifically, for the parameters of the conditional variance equation we assume that $\omega_0 > 0$, $0 < \alpha_0 < 1$, $0 < \beta_0 < 1$. These restrictions also imply the non-negativity of the conditional variance. General results on the moments and autocorrelation structure of the GARCH(p,q)-M can be found in Karanasos (2001).

Lee and Hansen (1994) and Lumsdaine (1996) derived the distribution theory for the quasi-maximum likelihood estimator in the GARCH$(1,1)$ model. To the best of our knowledge sufficient regularity conditions which ensure consistency and asymptotic normality of the quasi-maximum likelihood estimator for the GARCH-M model have not yet been established. As standard in the literature on GARCH-M we will treat our estimates as if the distribution theory for the GARCH estimator could be directly extended. Note, that in contrast to ARMA-GARCH models which do not allow for an in-mean effect, in the GARCH-M model the information matrix is not block diagonal, and thus consistent estimation of the parameters requires that both the conditional mean and variance functions are correctly specified and estimated simultaneously.[6]

Linton and Perron (2003) propose a semiparametric version of the GARCH$(1,1)$-M model described by equations (16) – (18) in which the functional dependence of Y_t on its conditional variance, $m(h_t)$, is estimated by nonparametric kernel smoothing methods. The estimation procedure is very similar to

[6] Christensen et al. (2008) modify the conditional variance equation (18) to $h_t(\theta_0) = \omega + \alpha_0 Y_{t-1}^2 + \beta_0 h_{t-1}(\theta_0)$. Then, by construction, the ARCH(∞) representation of h_t does no longer depend on m.

the one described above, i.e. based on an iterative updating of both the parameters of the conditional variance equation and the function $m(\cdot)$.

For our simulations we adopt two steps from the Linton and Perron (2003) algorithm. First, the initial parameter estimates $(\widehat{\psi}^{(0)}, \widehat{\gamma})$ will be obtained by estimating the parametric specification described in equations (16) – (18) by quasi-maximum likelihood. Second, in each iteration step the bandwidth for the nonparametric estimate $\widehat{m}^{(k)}$ is chosen as $b = b_0 \sigma(\widehat{h}_t^{(k-1)}) T^{-1/5}$, where $\sigma(\widehat{h}_t^{(k-1)})$ is the standard deviation of the fitted conditional variance from the $(k-1)$-th iteration step and the value of b is determined as the one which produces the lowest value of the cross-validation function

$$CV(b) = \frac{1}{T} \sum_{t=1}^{T} \left(Y_t - \widehat{m}_{-t}^{(k)}(\widehat{h}_t^{(k-1)}) \right)^2 ,$$

where $\widehat{m}_{-t}^{(k)}$ is the leave-one-out estimator and b_0 is allowed to vary between 0.5 and 2.5 in increments of 0.1.

In the simulations as well as in the application we will focus on testing for linearity in the GARCH$(1,1)$-M model. Since many properties of the model such as the behavior of the maximum likelihood estimator are largely unexplored we do not verify our assumptions for this specification. However, it is widely believed that the well known properties of the GARCH$(1,1)$ should also hold for the GARCH$(1,1)$-M. Most of the above assumptions can be easily verified for the GARCH$(1,1)$. Assumption 7 is satisfied by e.g. Gaussian Z_t. Note that the interval I is assumed to be bounded. Carrasco and Chen (2002) show that h_t in the GARCH$(1,1)$ is β-mixing with exponentially decaying mixing coefficients as required in Assumption 3. Assumption 12 is naturally satisfied when m_γ does not depend on h_t and Assumption 13 holds by the results of Lee and Hansen (1994) and Lumsdaine (1996). Finally, Assumption 15 follows directly from the ARCH(∞) representation of h_t.

4.4 Parametric Bootstrap

We expect that the theorems can only give a rough idea about the stochastic behavior of our test statistic for small sample sizes. Indeed we will see in the simulations that the normal approximation does not work very well in our setting. Therefore, it seems appropriate not to use the asymptotic critical values but to compute the critical values by resampling (see Härdle and Mammen, 1993).

Suppose one has obtained initial parameter estimates $(\widehat{\psi}^{(0)}, \widehat{\gamma})$ and final estimates of the conditional variance $\widehat{h}_t^{(k)} = \widehat{h}_t(\widehat{\psi}^{(k)}, \widehat{m}^{(k)})$ according to the algorithm described in Section 4.1. Then one can approximate $\widehat{\Gamma}_T^{(k)}$ by numerical integration. The bootstrap procedure makes use of the fact that under the null hypothesis we have a parametric specification of the conditional mean and variance and can be described as follows:

Step 1: Generate a bootstrap series $\{Y_t^\star\}_{t=1}^T$ according to equations (16) – (18) with $m_{\widehat{\gamma}}$ given by the null hypothesis. As a starting value h_0 we use the estimated unconditional variance. Innovations $Z_t^\star$ are drawn from the standard normal distribution.

Step 2: Apply the algorithm described in Section 4.1 to the bootstrap series $\{Y_t^\star\}_{t=1}^T$ and obtain $m_{\widehat{\gamma}^\star}$ and $\widehat{h}_t^{(k)\star}$. Calculate the value of the bootstrap test statistic $\widehat{\Gamma}_T^{(k)\star}$ by numerical integration.

Step 3: Repeat step 1 and 2 for B times. The bootstrap p-value of $\widehat{\Gamma}_T^{(k)}$ is the relative frequency of the event $\{\widehat{\Gamma}_T^{(k)\star} \geq \widehat{\Gamma}_T^{(k)}\}$ in the B bootstrap resamples.

5 Monte-Carlo Simulation

In this section we examine the finite sample properties of the semiparametric estimation procedure and the empirical level and power of the proposed test statistic. We first compare the performance of the parametric GARCH$(1,1)$-M with the semiparametric procedure under the null hypothesis and then under the alternative. Thereafter, we estimate the empirical level and power and demonstrate the robustness of our results with respect to the choice of the bandwidth. We always use an Epanechnikov kernel and weight function $w(\cdot) = \mathbf{1}_{[\underline{h},\overline{h}]}$, where $\underline{h}$ and $\overline{h}$ are chosen such that approximately 90% of the data are covered.[7] For simplicity we will denote the fitted conditional variance and the corresponding test statistic from the last iteration step by $\widehat{h}_t$ and $\widehat{\Gamma}_T$ suppressing the index k. The integral of the test statistic $\widehat{\Gamma}_T$ is numerically approximated on 50 equally spaced grid points on the interval $[\underline{h},\overline{h}]$. The parameters of the conditional variance equation are chosen to be $\omega_0 = 0.01$, $\alpha_0 = 0.1$ and $\beta_0 = 0.85$ which represent typical parameter values in empirical applications. The innovations are drawn from the standard normal distribution. All the simulations are carried out for a sample size of $T = 1000$. The Monte-Carlo experiments are repeated $M = 200$ times and the bootstrap resampling is performed $B = 200$ times for each sample. Initial parameter estimates for the mean and variance equation are obtained by quasi-maximum likelihood. The variance parameters are updated by estimating a parametric GARCH$(1,1)$ on the residuals $Y_t - \widehat{m}_b^{(k)}(\widehat{h}_t^{(k-1)})$. In each iteration step we impose the parameter restrictions described in Section 4.3 implying covariance stationarity and nonnegativity of the conditional variance. The bandwidth parameter b is chosen in each iteration step according to the cross-validation criterion discussed in Section 4.3. Throughout the simulations we set $I = (0, \infty)$.

[7] Alternatively, we used a standard normal kernel and obtained virtually identical results.

5.1 Performance of the Estimation Procedure

We first evaluate the performance of the estimation procedure for three linear specifications which reflect the null hypothesis:

$$\text{(N1)} \quad m(h_t) = 0.05 \cdot h_t \qquad \text{(N2)} \quad m(h_t) = 0.5 \cdot h_t \qquad \text{(N3)} \quad m(h_t) = h_t$$

Table 1 presents in Panel A the median estimates for the mean and variance equation parameters of the parametric GARCH$(1,1)$-M and in Panel B the median estimates of the parameters from the conditional variance equation obtained by the semiparametric procedure.[8] In both panels we also provide the 25% and 75% quantiles for the estimated parameters over the 200 replications. The median parametric parameter estimates presented in Panel A of Table 1 are – as expected under the null – very close to the true parameter values of the model for the different values of λ_0. In particular, the in-mean parameter λ_0 is very well estimated as shown by the 25% and 75% quantiles. However, from the estimates of the quantiles it is evident that the true value λ_0 can be recovered much better for higher values of λ_0 than for smaller ones. From Panel B it becomes clear that the semipametric estimator leads to very precise estimates of the conditional variance equation parameters, although it unnecessarily applies the iterating procedure. Figure 1 shows the true mean function, the pointwise median of the parametric and the nonparametric estimate along with the pointwise 25% and 75% quantiles of the nonparametric estimate for model N3. Under the null hypothesis both estimation procedures perform equally well in recovering the true structure of the model. Similar figures are available for models N1 and N2, but are omitted for space considerations.

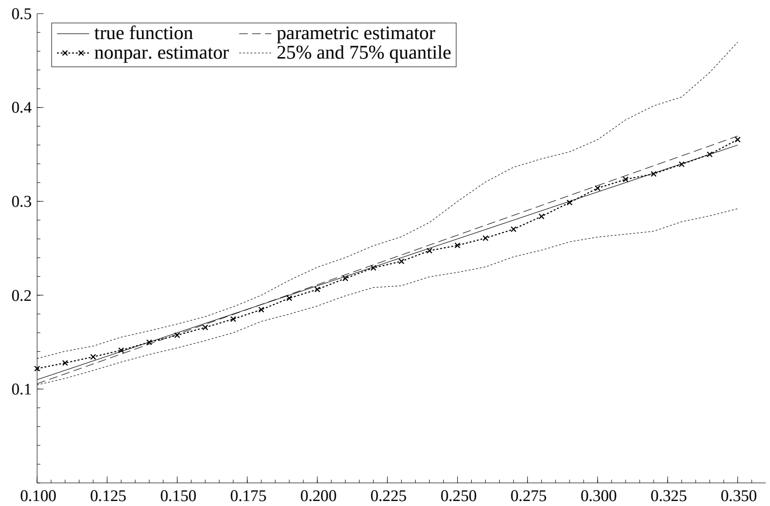

Figure 1: Parametric and nonparametric estimate for model N3.

[8] Similar results were obtained for the square root and log specification and are available upon request.

Table 1: Monte-Carlo estimates of the parametric and semiparametric regression model.

	Panel A: Median parametric estimates				
	$\widehat{\mu}$	$\widehat{\lambda}$	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$
N1	-0.003	0.050	0.011	0.097	0.841
($\lambda_0 = 0.05$)	(-0.024, 0.019)	(-0.078, 0.184)	(0.009, 0.015)	(0.080, 0.117)	(0.811, 0.868)
N2	-0.005	0.508	0.011	0.098	0.841
($\lambda_0 = 0.5$)	(-0.033,0.017)	(0.386,0.689)	(0.009, 0.015)	(0.082, 0.115)	(0.813, 0.866)
N3	-0.007	1.032	0.010	0.097	0.849
($\lambda_0 = 1$)	(-0.037, 0.024)	(0.879, 1.197)	(0.008, 0.013)	(0.084, 0.109)	(0.824, 0.868)
A1	0.746	-0.809	0.013	0.109	0.825
($\zeta_0 = 0.5$)	(0.682, 0.794)	(-1.094, -0.525)	(0.010, 0.017)	(0.090, 0.122)	(0.798, 0.855)
A2	0.012	0.275	0.011	0.099	0.842
($\zeta_0 = 0.1$)	(-0.022, 0.047)	(0.103, 0.435)	(0.008, 0.014)	(0.086, 0.115)	(0.818, 0.869)
A3	0.077	0.760	0.010	0.096	0.852
($\zeta_0 = 0.12$)	(0.054, 0.102)	(0.627, 0.895)	(0.008, 0.014)	(0.081, 0.109)	(0.823, 0.871)

Notes: Entries in Panel A and B are the median of the estimated parameters over the 200 replications. The entries in the parenthesis are the 25% and 75% quantiles over the 200 replications.

Monte-Carlo estimates of the parametric and
semiparametric regression model (continued).

	Panel B: Median semiparametric estimates		
	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$
N1	0.0102	0.0916	0.8505
($\lambda_0 = 0.05$)	(0.0082, 0.0138)	(0.0768, 0.1100)	(0.8242, 0.8724)
N2	0.0101	0.0911	0.8507
($\lambda_0 = 0.5$)	(0.0082,0.0141)	(0.0765,0.1103)	(0.8251, 0.8735)
N3	0.0101	0.0923	0.8554
($\lambda_0 = 1$)	(0.0076, 0.0131)	(0.0782, 0.1074)	(0.8318, 0.8777)
A1	0.0102	0.0913	0.8541
($\zeta_0 = 0.5$)	(0.0077, 0.0131)	(0.0793, 0.1066)	(0.8323, 0.8762)
A2	0.0101	0.0925	0.8551
($\zeta_0 = 0.1$)	(0.0077, 0.0128)	(0.0784, 0.1061)	(0.8330, 0.8777)
A3	0.0101	0.0910	0.858
($\zeta_0 = 0.12$)	(0.0078, 0.0134)	(0.0774, 0.1024)	(0.8320, 0.8778)

Notes: As in Table 1.

Next, we investigate the accuracy of the iterative estimation algorithm under the alternative. We employ the following mean functions:

$$\begin{array}{llcl} \text{(A1)} & m(h_t) & = & h_t + \zeta_0 \cdot \sin(10 \cdot h_t) \\ \text{(A2)} & m(h_t) & = & 0.5 \cdot h_t + \zeta_0 \cdot \sin(0.5 + 20 \cdot h_t) \\ \text{(A3)} & m(h_t) & = & h_t + \zeta_0 \cdot \sin(3 + 30 \cdot h_t). \end{array}$$

These alternatives represent shapes of the risk premium which are not covered by the standard specification but can be viewed as motivated by the results of Backus and Gregory (1993), Genotte and Marsh (1993) and the empirical findings of Linton and Perron (2003). Alternative A1 and A2 are inverse U-shaped and U-shaped while A3 is a hump-shaped alternative. The parameter ζ_0 can be regarded as a measure for the distance between the linear null hypothesis and the alternative.

The lower part of Table 1 presents the results of the Monte-Carlo simulations performed for models A1 – A3 with specific values for ζ_0. Again, Panel A reports the mean and variance parameter estimates from the parametric GARCH(1, 1)-M with $m(h_t) = \mu + \lambda h_t$ while Panel B reports the estimates for the conditional variance equation obtained by the semiparametric procedure. Figures 2 and 3 show the pointwise median parametric and nonparametric estimate along with the 25% and 75% pointwise quantiles of the latter and the true mean function for alternatives A1 and A3. Additionally, we plot the pointwise median estimate of the semiparametric procedure that is obtained after the first iteration step.

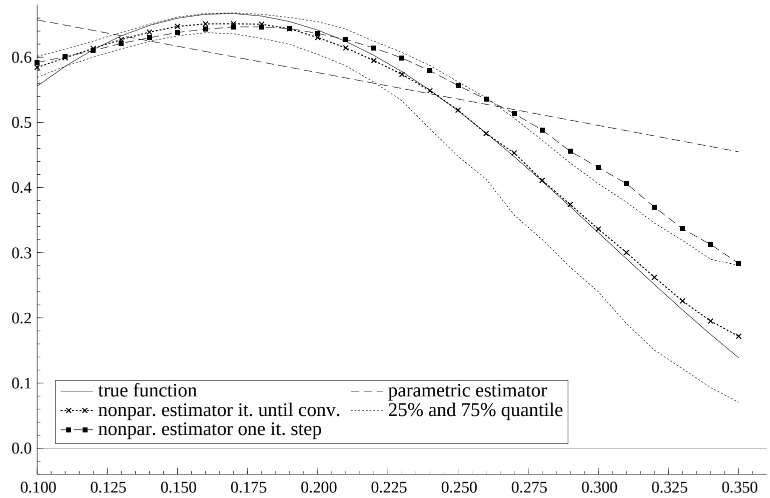

Figure 2: Parametric and nonparametric estimate for model A1 ($\zeta_0 = 0.5$).

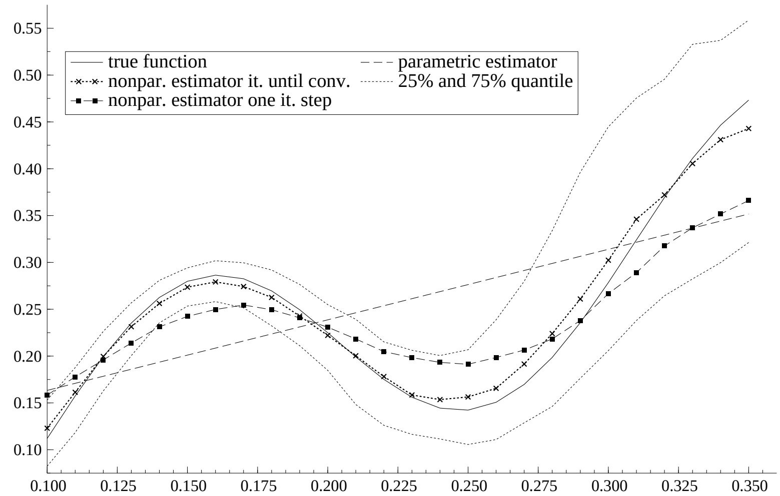

Figure 3: Parametric and nonparametric estimate for model A3 ($\zeta_0 = 0.12$).

The figures reveal that the nonparametric estimate of the mean function does again perform very well in uncovering the true mean function. The parametric estimate – which is restricted to being linear – fails to do so. In particular, in model A1 the true mean function is increasing for values of the conditional variance up to 0.175 while it is decreasing from 0.175 onwards. The parametric estimate of the mean function either over or underestimates the true risk premium. This example shows that one can easily find a negative relationship by applying the parametric model to a non-linear risk premium. A curve similar to A1 is presented by Whitelaw (2000, Figure 3) as a reasonable relationship between the expected return and its volatility in his two regime model when the economy is in a contractionary regime. Merely, the application of the semiparametric procedure makes it possible to obtain the true relationship, i.e. the risk premium is increasing until volatility exceeds a critical value, and then it becomes decreasing. A similar interpretation holds for A2.[9] Finally, A3 is a hump-shaped alternative as suggested by the findings of Linton and Perron (2003). Although, the parametric model captures the overall increasing tendency, it would predict very misleading values for the risk premium. The nonparametric fit on the other hand follows closely the true risk premium. These examples clearly illustrate the superiority of the semiparametric approach. Moreover, it is possible to construct non-monotonic shapes of the risk premium which lead to insignificant estimates of the parameter λ_0 and hence would suggest that there is no relationship between h_t and Y_t, while the semiparametric procedure recovers the true relationship. This failure of the parametric estimator may explain the finding of an insignificant $\widehat{\lambda}$ in many studies using the parametric GARCH$(1,1)$-M specification. The graphical intuitions are supported by the estimation results reported in Table 1. It is clear that now – as the parametric model is misspecified – the estimates of λ_0 are completely misleading. Nevertheless, the parameters in the conditional variance equation are still surprisingly well estimated using the parametric model. Finally, the semiparametric estimation procedure results in very accurate estimates of the conditional variance parameters ω_0, α_0 and β_0.

Figures 2 and 3 also help to illustrate the gains that are obtained by iterating in the semiparametric estimation procedure. It is evident that the one step iteration estimator cannot capture the nonlinearities by the same degree of accuracy as the iteration until convergence estimator. While this seems to be the case for A1 only for large values of h_t, it is generally true for A3 where the one step iteration estimator simply leads to a regression function which is too smooth. It appears that by doing only one iteration step it is not possible to move sufficiently far away from the parametric estimate to be close to the true mean function. This requires further iterations. We will see in the next subsection that this directly effects the power properties of our test statistic.[10]

[9] The corresponding figure is omitted for reasons of brevity.

[10] Christensen et al. (2008) replicate our simulations and compare our estimation approach with theirs. For the alternatives considered here both procedures provide almost the same results.

5.2 Monte-Carlo Estimates of Level and Power

This subsection evaluates the performance of the test statistic. In Table 2 we check for models N1, N2 and N3 and for different choices of the bandwidth parameter b whether the estimated level of the test reflects the nominal level. We report the estimated levels in comparison to the nominal 5% and 10% levels. In general, the estimated levels are very stable around the nominal levels of 5% and 10% for a wide range of bandwidths. The lowest bandwidth $b = 0.015$ produces too conservative results, i.e. we

Table 2: Monte-Carlo estimates of the level.

	b	0.015	0.020	0.025	0.030	0.035	0.040	0.045
N1	5%	0.030	0.050	0.055	0.060	0.055	0.050	0.055
	10%	0.075	0.070	0.095	0.100	0.105	0.110	0.105
N2	5%	0.025	0.045	0.045	0.050	0.050	0.060	0.070
	10%	0.080	0.090	0.090	0.105	0.110	0.115	0.105
N3	5%	0.025	0.045	0.040	0.040	0.060	0.060	0.070
	10%	0.065	0.080	0.075	0.085	0.085	0.095	0.100

Notes: Entries are rejection rates over the 200 replications at the 5% and 10% nominal level.

observe underrejection. A bandwidth of $b = 0.02$ produces estimates of the level which are in most cases slightly below 5% and 10% respectively, while a bandwidth of $b = 0.045$ leads to estimates slightly above 5% and 10%. Overall, the the bootstrap procedure seems to do a very good job in estimating the 5% and 10% levels close to the nominal ones. The optimal bandwidth as chosen by cross-validation in the last iteration step of the semiparametric procedure is in the neighborhood of $b = 0.02$. For model N3 we plot the density of $T\sqrt{b}\,\widehat{\Gamma}_T$ and six bootstrap approximations in Figure 4 (upper). The figure shows that the bootstrap approximations estimate the distribution of $T\sqrt{b}\,\widehat{\Gamma}_T$ very well when the underlying model reflects the null hypothesis. Moreover, it is evident that the test statistic is not normally distributed and therefore one should not rely on the asymptotic critical values. Figure 4 (lower) shows the simulated density of $T\sqrt{b}\,\widehat{\Gamma}_T$ and six bootstrap replications for model A1. Under A1 the simulated density of $T\sqrt{b}\,\widehat{\Gamma}_T$ and the six bootstrap densities are very different, suggesting that the test statistic may have good power properties. Figure 5 displays the empirical power of the test for alternatives A1, A2 and A3 and two choices of bandwidths. The mean functions under the alternative are constructed such that the models move further away from the null hypothesis as ζ_0 increases. For all three alternatives we find the desired property that the power is monotonically increasing in the value of ζ_0. Moreover, the power is very similar across the two choices for the bandwidth parameter. The overall performance of the test

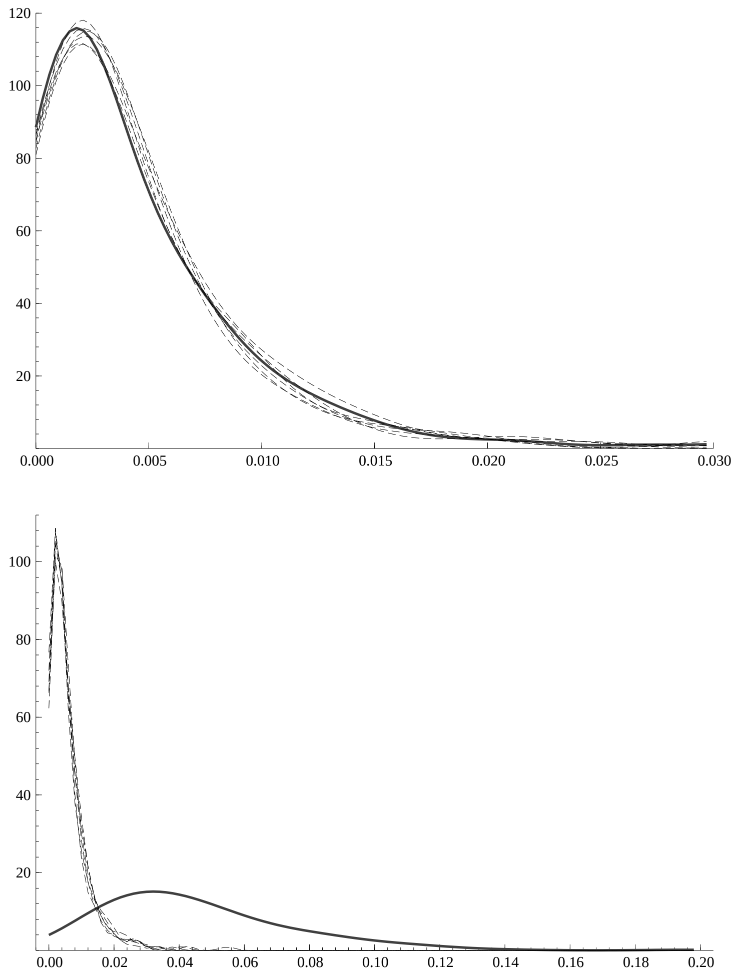

Figure 4: Simulated density of test statistic (solid) and six bootstrap approximations (dashed) for model N3 (upper) and A1 (lower).

applied under the alternative is very satisfactory. We conclude that the bootstrap procedure works well in our setting.

We also examined the power properties of the one step iteration estimator in comparison to the fully iterated estimator. For all three alternatives the tests based on the full iteration estimator lead to higher power than the corresponding test statistics based on the one step estimator. For instance, for A1 the fully iterated estimator produces empirical powers at the 5% and 10% nominal level of $(0.615, 0.750)$, $(0.875, 0.950)$ and $(0.945, 0.975)$ for $\zeta_0 \in \{0.3, 0.5, 0.7\}$ and $b = 0.03$, respectively. The corresponding figures for the one step estimator are $(0.400, 0.595)$, $(0.750, 0.870)$ and $(0.890, 0.945)$. Thus, the difference in the power of the test based on the fully iterated and the one step estimator can be striking.[11] In the

[11]Similarly, for A2 we obtain empirical powers at the 5% and 10% nominal level of $(0.470, 0.615)$, $(0.730, 0.820)$ and $(0.975, 0.995)$ for $\zeta_0 \in \{0.075, 0.1, 0.15\}$ for the fully iterated estimator and $(0.200, 0.335)$, $(0.350, 0.540)$ and $(0.720, 0.860)$

light of Figures 2 and 3 this is not surprising, since the one step estimator is almost everywhere closer to the parametric estimator than the full iteration estimator.

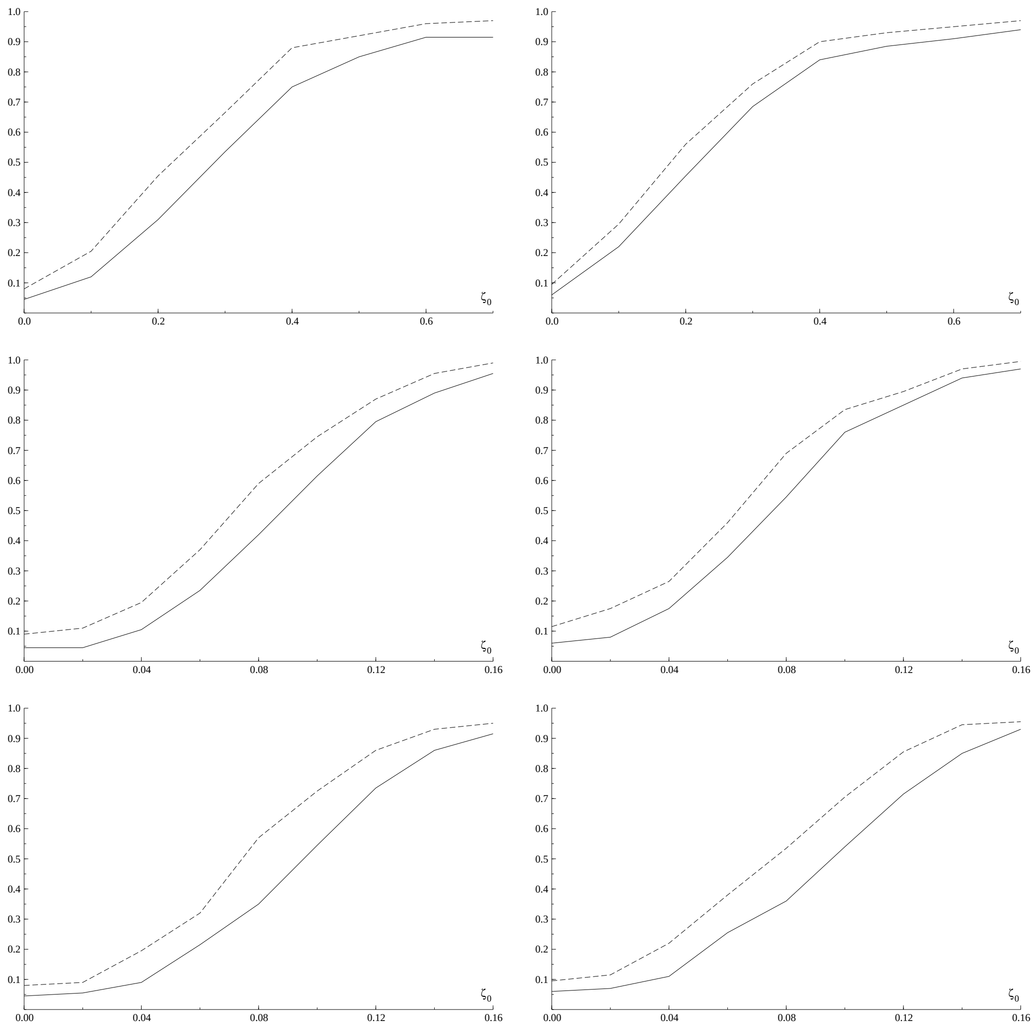

Figure 5: Simulated power for model A1 (upper), A2 (middle) and A3 (lower) with $b = 0.02$ (left) and $b = 0.04$ (right). Levels are given by 5% (solid) and 10% (dashed).

for the one step estimator. Finally, for A3 the empirical powers at the 5% and 10% nominal level are $(0.590, 0.735)$, $(0.770, 0.885)$ and $(0.940, 0.955)$ for $\zeta_0 \in \{0.1, 0.12, 0.15\}$ for the fully iterated estimator and $(0.235, 0.450)$, $(0.405, 0.575)$ and $(0.540, 0.725)$ for the one step estimator.

6 Application: The Shape of the Risk Premium

6.1 Data

The usefulness of the specification test will now be assessed in an application to test for linearity in the risk-return relation. For this we employ monthly and daily excess return data on the CRSP value-weighted index, which includes the NYSE, the AMEX and the NASDAQ and can be considered as the best available proxy for "the market". Monthly excess returns (including dividends) are calculated as the continuously compounded return on the CRSP minus the yield on a one month Treasury bill (from Ibbotson Associates), $Y_t = r_{M,t} - r_{f,t}$. Daily excess returns are calculated analogously, whereby daily yields are calculated by dividing the monthly yield by the number of trading days in the month and, hence, assuming constant yields for each calender day. The monthly data ranges from July 1963 to December 2001 (462 observations).[12] Daily return data was obtained from the Kenneth R. French data library for the period July 1963 to July 2005 (10593 observations). We only briefly summarize the descriptive statistics. The average excess return for the monthly (daily) data is about 0.37% (0.02%) with a standard deviation of 4.50% (0.89%). The distributions of the monthly as well as the daily excess returns are characterized by excess kurtosis (6.08 and 21.16). Moreover, the 12-th and 24-th order Ljung-Box statistics in combination with the results of the Engle LM-test for ARCH effects (both not reported) indicate serial correlation in the squared return series and highlight the importance of an appropriate modelling of the conditional variance of the excess returns.

The motivation for investigating both monthly and daily excess return data is to see whether there is any systematic difference in the analysis of the two. First, as argued by Andersen and Bollerslev (1998) more precise estimates of conditional volatility may be obtained by employing daily data in comparison with monthly data, and thus a better estimate of the true risk-return relation. Second, as shown by Scruggs (1998) a hedge demand which is not included as an explanatory variable may lead to an omitted variable bias in estimating the risk-return relation. However, since Guo and Whitelaw (2006) find that the investment opportunities change slowly at the business cycle frequency, these changes can be regarded as approximately constant at a daily frequency. Thus, it should be possible to precisely estimate the risk-return relation at a daily frequency even without explicitly incorporating the hedge demand in the regression equation.

[12] The monthly data was kindly provided by Oliver Linton and is analyzed by Linton and Perron (2003). Although their full data set goes back to January 1926 we decided to use only the observations from July 1963 onwards. A preliminary analysis of the complete data set revealed that the GARCH parameter estimates were very unreliable. This is because the Great Depression was characterized by extremely high volatility compared to the period thereafter. Hence, fitting a single GARCH model without allowing for changes in the volatility regime appeared to be questionable. Details on this are available from the authors upon request.

6.2 Parametric GARCH$(1,1)$-M Estimates

Next, we estimate parametric GARCH$(1,1)$-M models with $m(h_t) = \mu + \lambda h_t$ for the two data sets. In both regressions we include a constant μ to account for market imperfections such as taxes or transaction costs. Parameter estimates are provided in Table 3. The constant turned out to be significant for the daily data only. For the monthly as well as the daily data the GARCH parameter estimates $\widehat{\alpha}$ and $\widehat{\beta}$ are highly significant, satisfy the condition for covariance stationarity and imply a high degree of persistence in the conditional variance ($\hat{\alpha} + \hat{\beta} = 0.949$ for the monthly data and $\widehat{\alpha} + \widehat{\beta} = 0.995$ for the daily data). The finding of a high degree of persistence is an important result, since Poterba and Summers (1986) show that only persistent increases in volatility will effect the discount factors applied to future cash flows and thereby current prices. Therefore, they argue that persistence in the volatility is a necessary condition for fluctuations in volatility to have a significant impact on explaining risk premia. Similarly, Bekaert and Wu (2000, p. 2) reason that the predicted positive effect of volatility on excess returns relies "first of all on the fact that volatility is persistent". In line with the previous literature the estimate for λ is positive but insignificant when monthly data is used.

Table 3: GARCH-M estimates for CRSP data.

	$\widehat{\mu}$	$\widehat{\lambda}$	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$	Q^2_{12}
monthly data	-0.003	3.870	0.0001	0.074	0.875	4.61
	(-0.579)	(1.130)	(1.798)	(3.105)	(16.561)	[0.97]
daily	0.0003	3.844	$6.66 \cdot 10^{-7}$	0.089	0.906	12.76
	(2.976)	(2.714)	(3.938)	(8.187)	(103.919)	[0.39]

Notes: Bollerslev and Wooldridge (1992) robust t-statistics are reported in parenthesis $(\cdot)$. Q^2_{12} are the Ljung-Box statistics at the 12-th lag for the squared standardized residuals. Numbers in brackets $[\cdot]$ are p-values.

In sharp contrast, we estimate a positive and highly significant in-mean effect for the daily data. In particular, the estimate of λ is significant at the 1% level. Moreover, the value estimated for λ based on the daily data is almost identical to the one estimated for the monthly data. This is reasonable since both the risk premium and the conditional variance should be approximately proportional to the length of the measurement interval. If – as argued in Guo and Whitelaw (2006) – the omitted hedge term does not effect the estimation of the risk-return relation when daily data is employed, the finding of similar $\widehat{\lambda}$'s for monthly and daily data suggests that the omitted variable bias argument of Scruggs (1998) does also not hold at a monthly frequency. This is because in the presence of such an effect the estimate of λ based on monthly data should be considerably different from the one on daily data. Therefore, our results are much more in favor of the argument by Andersen and Bollerslev (1998), namely that the estimates based on daily data provide a more accurate measure of the conditional volatility and hence allow for a more precise estimation of the risk-return relation. As a result of this we find a significant in-mean effect using the daily data. Following French et al. (1987) $\widehat{\lambda}$ can be interpreted as an estimate for the parameter of relative risk aversion. The value we estimate is plausible for the coefficient of relative risk aversion. We conclude that the parametric GARCH$(1, 1)$-M models deliver convincing evidence for a positive and at a daily frequency significant relation between risk and excess returns.[13]

According to the Ljung-Box statistics the null hypothesis of uncorrelated squared standardized residuals is accepted for both models. Finally, the GARCH$(1, 1)$-M models were preferred by the AIC and BIC information criteria to models of higher order.

6.3 Testing the Linear Hypothesis

Next, we will apply our specification test to the CRSP excess return data to check whether the functional relationship between excess returns and risk can be confirmed to be linear as assumed by the parametric GARCH$(1, 1)$-M. Recall from Section 2 that Linton and Perron (2003) found support for a hump-shaped pattern of the risk premium.

The application of the test procedure requires the choice of an appropriate bandwidth b and of an interval $[\underline{h}, \overline{h}]$ on which the test statistic is evaluated.[14] For the two data sets we evaluate the test statistic on two different intervals. The larger one is chosen such that it covers 90% of the data, the smaller one covers only 70%. In both situations $\underline{h}$ corresponds to the 5% quantile $(q_{0.05}(\widehat{h}_t))$ of the distribution of the estimated conditional variances from the last iteration step. Accordingly, we choose $\overline{h}$ approximately as the 75% or 95% quantile $(q_{0.75}(\widehat{h}_t)$ and $q_{0.95}(\widehat{h}_t))$. As a guide for choosing the bandwidth we use $b = \sigma(\widehat{h}_t) \cdot T^{-1/5}$, where $\sigma(\widehat{h}_t)$ and T refer only to the observations in $[\underline{h}, \overline{h}]$. This choice of the bandwidth usually results in values slightly above the cross-validated bandwidth from the last iteration step. We additionally report the test statistic and the corresponding p-values for two larger choices of b, whereby the largest bandwidth is always based on the full distribution of $\widehat{h}_t$. Such choices of b can be considered as oversmoothing in comparison to the optimal bandwidth for estimation.

The test results are presented in Table 4. We begin by discussing the results for the monthly data. Several interesting findings emerge. Besides the estimated 95% quantile of the fitted conditional variances $q_{0.95}(\widehat{h}_t)$, we report the median of the 95% quantiles of the fitted conditional variances over the 200

[13]Of course, the simple GARCH$(1, 1)$-M model could be augmented in several directions. For example, we could incorporate a volatility feedback effect in the conditional variance equation (see, e.g., Smith, 2006).

[14]As in the simulation section, we will denote the fitted conditional variance and the corresponding test statistic from the last iteration step by $\widehat{h}_t$ and $\widehat{\Gamma}_T$ suppressing the index k.

bootstrap replications denoted by $q_{0.95}(h_t^\star)$. We observe that $q_{0.95}(\widehat{h}_t)$ and $q_{0.95}(h_t^\star)$ are very close to each other reflecting the fact that the fitted conditional variances from the bootstrap procedure mimic very well the distribution of the fitted conditional variances from the observed data. As can be seen from the table we cannot reject the null hypothesis that the risk premium is linear in the conditional variance at any reasonable significance level.

Table 4: Testing for linearity in the risk-return relation.

	monthly data				daily data			
	$q_{0.95}(\widehat{h}_t) = 29.41$, $q_{0.95}(h_t^\star) = 27.76$				$q_{0.95}(\widehat{h}_t) = 2.34$, $q_{0.95}(h_t^\star) = 2.38$			
b	0.99	1.25	1.50	1.71	0.05	0.07	0.10	0.14
	$[\underline{h}, \overline{h}] = [12, 25]$				$[\underline{h}, \overline{h}] = [0.2, 1.5]$			
$T\sqrt{b}\,\widehat{\Gamma}_T$	0.984	0.787	0.662	0.596	8.062	5.913	4.393	4.403
p-value	0.610	0.545	0.481	0.422	0.086	0.136	0.161	0.100
	$[\underline{h}, \overline{h}] = [12, 30]$				$[\underline{h}, \overline{h}] = [0.2, 2.34]$			
$T\sqrt{b}\,\widehat{\Gamma}_T$	1.922	1.561	1.278	1.096	62.303	56.079	48.484	39.459
p-value	0.797	0.754	0.711	0.690	0.015	0.025	0.075	0.075

Notes: The smallest bandwidth always corresponds to the smaller interval, while the second smallest bandwidth is chosen according to the larger interval. The two largest bandwidths can be regarded as oversmoothing.

Figure 6 shows the parametric and nonparametric estimate of the risk premium for the monthly data.[15] The shape of the nonparametric estimate reveals some non-linearity which could be called hump-shaped as in Linton and Perron (2003). Nevertheless, the nonparametric estimate trends very closely with the linear parametric estimate making the test result plausible.

For the daily data, we again find that the 95% quantiles of the fitted and bootstrap conditional variances are very close to each other. However, the test results are less uniform. While the results for the smaller interval are in line with the linear hypothesis, the results for the broader interval suggest that the hypothesis of linearity should be rejected.

To check for the robustness of our results we also tested the hypothesis of no in-mean effect, i.e. $H_0 : m_\gamma(h_t) = \mu$. This hypothesis was rejected in the overwhelming majority of cases. In summary, we find that there is convincing evidence for the existence of an in-mean effect. While for the monthly data we

[15] Pointwise 95% asymptotic standard errors for the nonparametric estimate are given by

$$\widehat{m}_b^{(k)}(x) \pm 1.96 \cdot \sqrt{\frac{1}{Tb}\frac{x\int K(u)^2du}{\widehat{f}_h(x)}},$$

where we use the fact that for the GARCH-M model it holds that $\mathbf{Var}(Y_t|h_t = x) = x$.

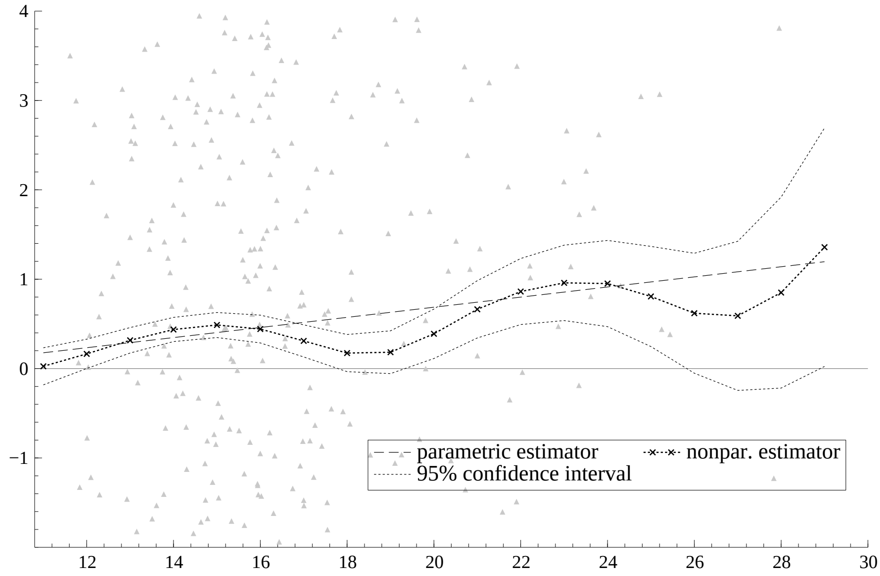

Figure 6: Parametric and nonparametric fit for monthly data. The monthly returns are expressed in %.

cannot reject the hypothesis of the effect being linear, the daily data points to the existence of some non-linearities in the risk-return relationship for large values of the conditional variance. Such non-linearities may be related to the existence of leverage or volatility feedback effects (see, e.g., Smith, 2006). It could also be necessary to distinguish between short- and long-run volatility components (see, e.g., Engle et al., 2007, and Adrian and Rosenberg, 2008).

7 Conclusions

This article deals with the asymptotic behavior of nonparametric regressions with unobserved covariates. First, we use iterative procedures to fit the unobservable regressors and propose nonparametric smoothing estimators based on the fitted covariates. Second, we study tests for parametric specifications that are based on the comparison of a parametric estimator with our nonparametric fit. Exploiting tools from empirical process theory we show oracle efficiency of our nonparametric procedure, i.e. the nonparametric procedure behaves as if the regressor were observable. This property is shown for both estimation and testing.

Our general model nests a specification which has received considerable attention in the financial econometrics literature, the class of parametric GARCH-M models. Those models are heavily used in the analysis of the risk-return relationship as well as to investigate the causal relationship between the level and the uncertainty of macroeconomic variables such as inflation and output growth. The parametric functional form of the risk premium assumed in the GARCH-M is mainly motivated by the ICAPM or imposed simply for convenience.

We apply our test procedure empirically to daily as well as monthly return data on the CRSP. While the results for the monthly data are in line the prediction made by the ICAPM, the results for the daily data question the appropriateness of the linear specification for the risk premium when volatility is very high. This finding is line with the results reported in Christensen et al. (2008) for the same data and period and might explain some of the controversial results presented in the previous literature.

Finally, we would like to point to natural extensions of the model studied in this article. First, we could allow for higher dimensional explanatory variables in the mean function. The explanatory variable could have several unobserved and observed components. In particular, the covariate could contain lagged values of an unobservable covariate and/or observed macroeconomic variables. In case of high dimensional covariates one could consider structured nonparametric models like an additive model, see e.g. Mammen, Linton and Nielsen (1999). Moreover, when the conditional variance process is more complicated, the unobserved covariate could be a specific volatility component only. E.g. it would be natural to extend the class of GARCH-MIDAS models considered in Engle et al. (2007) by allowing one or both of the volatility components to effect the conditional mean in a nonparametric fashion. Similarly, the semiparametric MIDAS model of Chen and Ghysels (2008) could be augmented by an in-mean term.

Appendix

In the proofs of the theorems we make use of the following lemmas. The first lemma contains an exponential inequality for martingales. This inequality is a modification of e.g. Lemma 8.9 in van de Geer (2000).

Lemma 1. *For random variables* $..., e_{-1}, e_0, e_1, ..., e_T$ *suppose that* e_t *is* $\mathcal{F}_t$*-measurable for an increasing* σ*-field* $\mathcal{F}_t$*, that* $\mathbf{E}[e_t|\mathcal{F}_{t-1}] = 0$ *and that* $\sup_t \mathbf{E}[\exp(c|e_t|)|\mathcal{F}_{t-1}] < \infty$ *(a.s.) for a constant* $c > 0$ *small enough. Consider a sequence of random variables* $r_1, r_2, ...$ *where* r_t *is measurable with respect to the* σ*-field generated by* $\mathcal{F}_{t-1}$*. Assume that* $\max_{1\le t\le T} |r_t| \le c/2$ *(a.s.). Then it holds that*

$$\mathbf{E}\left[\exp\left(\sum_{t=1}^{T} r_t e_t\right)\right] \le \left\{\mathbf{E}\left[\exp(C\sum_{t=1}^{T} r_t^2)\right]\right\}^{1/2},$$

where C *is a deterministic a.s. bound of* $\mathbf{E}\left[e_t^2 \exp\left(\frac{c}{2}|e_t|\right)|\mathcal{F}_{t-1}\right]$.

The next lemma contains bounds on the nominator and denominator of the Nadaraya-Watson estimator which is applied to the covariates h_t and $\widehat{h}_t(\psi_0, m_0)$.

Lemma 2. *Under the assumptions of Theorem 1 it holds that*

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T}K_b(h_t-x)-f_h(x)\right|=O_P(b^2+\sqrt{\log(T)}(Tb)^{-1/2}), \tag{19}$$

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T}K_b(h_t-x)\varepsilon_t\right|=O_P(\sqrt{\log(T)}(Tb)^{-1/2}), \tag{20}$$

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T}[K_b(h_t-x)-K_b(\widehat{h}_t(\psi_0,m_0)-x)]\right|=O_P(T^{-\omega}b^{-1}), \tag{21}$$

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T}[K_b(h_t-x)-K_b(\widehat{h}_t(\psi_0,m_0)-x)]\varepsilon_t\right|=O_P(T^{-\omega}b^{-3/2}T^{-1/2}\sqrt{\log(T)}). \tag{22}$$

The constant ω was introduced in Assumption 8.

Proof of Lemma 2. For a proof of (19) see Masry (1996). The proof of (20) also follows classical lines. Because of Assumption 1 it suffices to show

$$\sup_{x\in I_T}\left|\frac{1}{T}\sum_{t=1}^{T}K_b(h_t-x)\varepsilon_t\right|=O_P(\sqrt{\log(T)}(Tb)^{-1/2}). \tag{23}$$

Here I_T is a grid of points of I with cardinality growing polynomially in T. Equality (23) can be proved by application of the exponential bound in Lemma 1 and by use of the Markov inequality $\mathbf{P}[\sum_{t=1}^{T} r_t e_t \geq c] \leq \exp(-sc)\mathbf{E}[\exp(s\sum_{t=1}^{T} r_t e_t)]$. We apply these bounds with $e_t=\varepsilon_t$ and with $r_t=K_b(h_t-x)\sqrt{h_t}$ if $T^{-1}\sum_{s=1}^{t}\mathbf{1}(|h_t-x|\leq 2b)\leq Cb$ and $r_t=0$ else. Here C is a constant that is chosen large enough. Note that for such a choice

$$T^{-1}\sum_{s=1}^{t}\mathbf{1}(|h_t-x|\leq 2b)\leq Cb \tag{24}$$

for all $x\in I$ with probability tending to one.

Equation (21) follows by a direct bound. For a proof of (22) one proceeds similarly as in the proof of (20). □

Lemma 3. *Under the assumptions of Theorem 1 it holds that*

$$\sup_{x\in I,(\theta_1,m_1),(\theta_2,m_2)\in\mathcal{M}_T}\left|\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta_2,m_2)-x)\varepsilon_t -\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta_1,m_1)-x)\varepsilon_t\right|=O_P(T^{-1/2+\eta/2-\kappa}), \tag{25}$$

$$\sup_{x\in I,(\theta_1,m_1),(\theta_2,m_2)\in\mathcal{M}_T}\left|\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta_2,m_2)-x) -\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta_1,m_1)-x)\right|=O_P(T^{-\delta_m+\rho_m+\eta}+T^{-\delta_\theta+\rho_\theta+\eta}), \tag{26}$$

where $\mathcal{M}_T=\{(\theta,m):\|\theta-\theta_0\|\leq T^{-\delta_\theta}, m\in\mathcal{M}_T^*\}$ and $\mathcal{M}_T^*=\{m:\|m-m_0\|\leq T^{-\delta_m},\|D_2m-D_2m_0\|\leq T^{\xi}\}$.

Proof of Lemma 3. Claim (26) follows by a direct bound. We now show claim (25). For simplicity we neglect the discussion of the θ-component and we show

$$\sup_{x\in I, m_1, m_2 \in \mathcal{M}_T^*} \left| \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(m_2)-x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(m_1)-x)\varepsilon_t \right| = O_P(T^{-1/2+\eta/2-\kappa}), \quad (27)$$

where $\widehat{h}_t(m) = \widehat{h}_t(\theta_0, m)$. For a proof of (27) we use a chaining argument, compare e.g. the proof of Lemma 3.2 in van de Geer (2000). Put $\delta = T^{-\delta_m}$ and for $s \geq 1$ consider $2^{-s}\delta$ covering sets $\mathcal{M}_s^*$ of $\mathcal{M}_T^*$, i.e. for each $m \in \mathcal{M}_T^*$ there exists $m^* \in \mathcal{M}_s^*$ with $\|m^* - m\|_\infty \leq 2^{-s}\delta$. The covering sets can be chosen such that their cardinality $\sharp\mathcal{M}_s^*$ does not exceed $C^* \exp[(2^{-s}\delta)^{-1/2}T^{\xi/2}]$ for a constant $C^* > 0$. This is a standard bound for coverings of Sobolev balls, see van de Geer (2000). We now write $\Delta_t(m, m^*) = T^{-1}\{K_b(\widehat{h}_t(m)-x) - K_b(\widehat{h}_t(m^*)-x)\}\varepsilon_t^*$ with $\varepsilon_t^* = \varepsilon_t \mathbf{1}[|\varepsilon_t| \leq C^{**}\log T] - \mathbf{E}\{\varepsilon_t \mathbf{1}[|\varepsilon_t| \leq C^{**}\log T]\}$ for a constant $C^{**} > 0$ that is large enough. Now for $C^{**} > 0$ large enough

$$\sup_{x\in I, m, m^* \in \mathcal{M}^C} \left| \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(m^*)-x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(m)-x)\varepsilon_t - \sum_{t=1}^T \Delta_t(m, m^*) \right| = O_P(T^{-1/2+\eta/2-\kappa}).$$

For $m_1, m_2 \in \mathcal{M}^C$ we choose now $m_1^s, m_2^s \in \mathcal{M}_s^C$ with $\|m_1^s - m_1\|_\infty \leq 2^{-s}\delta$, $\|m_2^s - m_2\|_\infty \leq 2^{-s}\delta$ and we consider the chain

$$\begin{aligned} \sum_{t=1}^T \Delta_t(m_1, m_2) &= \sum_{t=1}^T \Delta_t(m_1^0, m_2^0) - \sum_{s=1}^{G_T}\sum_{t=1}^T \Delta_t(m_1^{s-1}, m_1^s) + \sum_{s=1}^{G_T}\sum_{t=1}^T \Delta_t(m_2^{s-1}, m_2^s) \\ &\quad - \sum_{t=1}^T \Delta_t(m_1^{G_T}, m_1) + \sum_{t=1}^T \Delta_t(m_2^{G_T}, m_2) \end{aligned}$$

where G_T is the largest integer with $2^{G_T}T^{-1/2-\eta/2-\kappa+\delta_m-\rho_m}\log(T) < c^*$ for a constant c^* that is large enough. We now give a bound on $\mathbf{P}[\sup_{m_1\in\mathcal{M}^C}\sum_{s=1}^{G_T}\sum_{t=1}^T \Delta_t(m_1^{s-1}, m_1^s) > T^{-1/2+\eta/2-\kappa}]$. Similar bounds can be proved for the other terms and for $\mathbf{P}[\inf_{m_1\in\mathcal{M}^C}\sum_{s=1}^{G_T}\sum_{t=1}^T \Delta_t(m_1^{s-1}, m_1^s) < -T^{-1/2+\eta/2-\kappa}]$. We get the following inequality with $\eta_s = c2^{-3s/4}$ where c is chosen such that $\sum_{s=1}^\infty \eta_s \leq 1$. With constants $c_1, c_2, ... > 0$ we get with $r = c_0 T^{1/2+5\eta/2+2\delta_m-2\rho_m-\kappa} 2^{2s}\eta_s$ for $c_0 > 0$ small enough:

$$\begin{aligned}
&\mathbf{P}\left[\sup_{m_1\in\mathcal{M}_T^*}\sum_{s=1}^{G_T}\sum_{t=1}^{T}\Delta_t(m_1^{s-1}, m_1^s) > T^{-1/2+\eta/2-\kappa}\right]\\
&\quad\le \sum_{s=1}^{G_T}\mathbf{P}\left[\sup_{m_1\in\mathcal{M}_T^*}\sum_{t=1}^{T}\Delta_t(m_1^{s-1}, m_1^s) > \eta_s T^{-1/2+\eta/2-\kappa}\right]\\
&\quad\le \sum_{s=1}^{G_T}\sharp\mathcal{M}_{s-1}^*\sharp\mathcal{M}_s^*\mathbf{P}\left[\sum_{t=1}^{T}\Delta_t(m_1^{s-1}, m_1^s) > \eta_s T^{-1/2+\eta/2-\kappa}\right]\\
&\quad\le \sum_{s=1}^{G_T} c_1\exp[2(2^{-s}\delta)^{-1/2}T^{\xi/2}]\mathbf{P}\left[\sum_{t=1}^{T}\Delta_t(m_1^{s-1}, m_1^s) > \eta_s T^{-1/2+\eta/2-\kappa}\right]\\
&\quad\le \sum_{s=1}^{G_T} c_1\exp[2T^{\delta_m/2+\xi/2}2^{s/2}]\left\{\mathbf{E}\left[\exp[c_2 r^2\sum_{t=1}^{T}\Delta_t^2(m_1^{s-1}, m_1^s)]\right]\right\}^{1/2}\exp[-r\eta_s T^{1/2+\eta/2-\kappa}]\\
&\quad\le \sum_{s=1}^{G_T} c_1\exp[2T^{\delta_m/2+\xi/2}2^{s/2} - c_3 2^{s/2}T^{2\delta_m-2\rho_m-2\eta-2\kappa-2\xi}b^4(\log T)^{-1}]
\end{aligned}$$

with a constant $c_* > 0$. The last inequality follows by application of the exponential inequality of Lemma 2. At this point it is also used that $2^{s/4}T^{-1/2-\eta/2-\kappa+\delta_m-\rho_m}\log(T)$ is small enough by appropriate choice of constant c^* for $s \le G_T$. Thus, we can apply the bound of Assumption 7 for $s \le G_T$. We now apply that the argument of the exponential function in the upper bound of the last inequalities converges to $-\infty$. This concludes the proof of the lemma. □

Proof of Theorem 1. We only prove (4). Equations (5)-(7) follow by similar arguments. For the proof of (4) we apply Lemma 3 with $m_1 = \widetilde{m}$, $\theta_1 = \widetilde{\theta}$, $m_2 = m_0$ and $\theta_2 = \theta_0$. Then, (4) follows by application of Lemma 2. □

Proof of Theorem 2. We only prove (9). Claim (8) follows by similar arguments. The asymptotic normality results follow from (8), (9) and classical results on nonparametric regression under dependence, see e.g. Masry (1996). Because of (5), for the proof of (9) it remains to show

$$\sup_{x\in I}\left|\widehat{m}^{NW,B}(x) - \widehat{m}^{*,NW,B}(x)\right| = o_P(T^{-2/5}). \tag{28}$$

Note that uniformly for $x \in I$ with $\widetilde{h}_t = \widehat{h}_t(\widetilde{\theta}, \widetilde{m})$

$$\begin{aligned}
\widehat{m}^{NW}(x) &= m(x) + \widehat{f}_h(x)^{-1}\frac{1}{T}\sum_{t=1}^{T}(h_t - x)K_b(\widetilde{h}_t - x)m'(x)\\
&\quad + \widehat{f}_h(x)^{-1}\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}(h_t - x)^2 K_b(\widetilde{h}_t - x)m''(x) + o_P(T^{-2/5})\\
&= m(x) + \widehat{f}_h(x)^{-1}\frac{1}{T}\sum_{t=1}^{T}[(h_t - \widetilde{h}_t) + (\widetilde{h}_t - x)]K_b(\widetilde{h}_t - x)m'(x)\\
&\quad + \widehat{f}_h^*(x)^{-1}\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}(h_t - x)^2 K_b(h_t - x)m''(x) + o_P(T^{-2/5})
\end{aligned}$$

with $\widehat{f}_h^*(x) = T^{-1}\sum_{t=1}^T K_b(h_t - x)$. Thus for (28) it suffices to show that uniformly for $(\theta, m, x) \in \mathcal{G}_T$

$$\frac{1}{T}\sum_{t=1}^T \left[(\widehat{h}_t(\theta,m) - h_t)K_b(\widehat{h}_t(\theta,m) - x)\right] = o_P(T^{-2\eta}), \tag{29}$$

$$\frac{1}{T}\sum_{t=1}^T \left[(\widehat{h}_t(\theta,m) - x)K_b(\widehat{h}_t(\theta,m) - x) - (h_t - x)K_b(h_t - x)\right] = o_P(T^{-2\eta}). \tag{30}$$

We only prove (29). Claim (30) follows similarly. Because of Assumption 10 it suffices to show that

$$\sup_{(\theta,m,x)\in\mathcal{G}_T} \left|\frac{1}{T}\sum_{t=1}^T a_t(\theta,m,x)\right| = o_P(T^{-2\eta}), \tag{31}$$

where $a_t(\theta,m,x) = (\widehat{h}_t(\theta,m) - h_t)K_b(\widehat{h}_t(\theta,m) - x) - \mathbf{E}\left[(\widehat{h}_t(\theta,m) - h_t)K_b(\widehat{h}_t(\theta,m) - x)\right]$. Choose $\rho, \delta > 0$. We will show that for constants $c_1, ..., c_4 > 0$ (depending only on ρ and δ) it holds that

$$\sup_{(\theta,m,x)\in\mathcal{G}_T} \mathbf{E}\left[\left|\frac{1}{T}\sum_{t=1}^T a_t(\theta,m,x)\right| > \rho T^{-2\eta}\right] \leq c_1\exp(-c_2T^{1/2}) + c_3\exp(-c_4T^{1+2\delta_{min}}b^{5+\delta}), \tag{32}$$

where δ_{min} is the minimum of $\delta_m - \rho_m$, $\delta_\theta - \rho_\theta$, ρ_0 and ω. We assume for simplicity that $|V_T| \leq c_0T^{\rho_\theta}$, $|W_T| \leq c_0T^{\rho_m}$ and $|R_T| \leq c_0T^{-\rho_0}$ for a constant $c_0 > 0$. A more detailed discussion that does not require this additional assumption would need an additional standard truncation argument. Using our entropy bounds one can show that it suffices for (31) to show that

$$\sup_{(\theta,m,x)\in\mathcal{G}_T^*} \left|\frac{1}{T}\sum_{t=1}^T a_t(\theta,m,x)\right| = o_P(T^{-2\eta}), \tag{33}$$

where $\mathcal{G}_T^*$ is a suitable chosen finite set of $O(\exp(c_5T^{\xi/2+\rho_m/2+\eta})T^{c_6})$ points with $c_5, c_6 > 0$. Claim (29) follows from (32) and (33) because of $\delta_{min} > \xi + \rho_m$ and $\delta > 0$ small enough. Thus for (29) it remains to show (32).

For the proof of (32) we apply the exponential inequality for mixing processes stated in Theorem 1.3. (2) in Bosq (1998) and use Davydov‘s inequality (see Corollary 1.1 in Bosq, 1998) to bound the variance of sums of blocks of summands. We apply the exponential inequality with blocks of $T^{1/2+3\eta/2-\delta_{min}/2}$ summands. This shows (32). $\square$

Proof of Theorem 3. The test statistic has the following representation: $\widehat{\Gamma}_T^{(0)} = \widehat{\Gamma}_{T,1}^{(0)} + \widehat{\Gamma}_{T,2}^{(0)} + \widehat{\Gamma}_{T,3}^{(0)}$, where

$$\begin{aligned}
\widehat{\Gamma}_{T,1}^{(0)} &= \int \left\{\frac{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(0)} - x)}\right\}^2 w(x)dx, \\
\widehat{\Gamma}_{T,2}^{(0)} &= -2\int \left\{\frac{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(0)} - x)}\right\} \\
&\qquad \times \left\{\frac{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(0)} - x)\left[m_{\widehat{\gamma}}(\widehat{h}_t^{(0)}) - m_{\gamma_0}(h_t)\right]}{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(0)} - x)}\right\} w(x)dx, \\
\widehat{\Gamma}_{T,3}^{(0)} &= \int \left\{\frac{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(0)} - x)\left[m_{\widehat{\gamma}}(\widehat{h}_t^{(0)}) - m_{\gamma_0}(h_t)\right]}{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(0)} - x)}\right\}^2 w(x)dx.
\end{aligned}$$

We show that

$$\widehat{\Gamma}^{(0)}_{T,1} = \widetilde{\Gamma}_T + o_P(T^{-1}b^{-1/2}), \quad (34)$$

$$\widehat{\Gamma}^{(0)}_{T,2} = o_P(T^{-1}b^{-1/2}), \quad (35)$$

$$\widehat{\Gamma}^{(0)}_{T,3} = o_P(T^{-1}b^{-1/2}), \quad (36)$$

where

$$\widetilde{\Gamma}_T = \frac{1}{T^2}\sum_{s,t=1}^{T} \frac{K^{(2)}(h_t - h_s)}{f_h(h_t)f_h(h_s)} w(h_s)\varepsilon_s\varepsilon_t.$$

For the proof of claim (34) one applies first Lemma 5, that is stated below. This shows that

$$\widehat{\Gamma}^{(0)}_{T,1} = \int \left\{ \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)} \right\}^2 w(x)dx + o_P(T^{-1}b^{-1/2}).$$

Claim (34) now follows from continuity of w and f_h and (19). It can be easily checked that (19) holds under the assumptions of Theorem 3, see Masry (1996).

For a proof of claim (35) one first applies Assumption 12, Assumption 14, $\widehat{\theta} - \theta_0 = O_P(T^{-1/2})$, Lemmas 4 and 5 and (19)to show that

$$\widehat{\Gamma}^{(0)}_{T,2} = (\widehat{\gamma} - \gamma_0)\frac{1}{T^2}\sum_{1\le s,t\le T} w_{s,t}\varepsilon_t + (\widehat{\theta} - \theta_0)\frac{1}{T^2}\sum_{1\le s,t\le T} w^*_{s,t}\varepsilon_t + o_P(T^{-1}b^{-1/2}), \quad (37)$$

with $w_{s,t} = \int_I K_b(h_t - x)K_b(x - h_s)\frac{\dot{m}_{\gamma_0}(x)}{f^2(x)}dx$ and $w^*_{s,t} = \int_I K_b(h_t - x)K_b(x - h_s)\frac{m'_{\gamma_0}(x)}{f^2(x)}dx\dot{h}_s$. We now use $\widehat{\theta} - \theta_0 = O_P(T^{-1/2})$, $b|w_{s,t}| \le C$, $b|w^*_{s,t}| \le C$ for a constant C and Davydov's inequality (see Corollary 1.1 in Bosq, 1998). This implies that the right hand side of (37) is of order $o_P(T^{-1}b^{-1/2})$ which shows claim (35).

Claim (36) follows directly from Assumption 12.

For the proof of the theorem it remains to show that $T\sqrt{b}(\widetilde{\Gamma}_T - b^{-1/2}M)/V$ converges in distribution to a standard normal distribution. This can be done by the same arguments as in Fan and Li (1999). □

Lemma 4. *Under the assumptions of Theorem 3 it holds that*

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}^{(0)}_t - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t\right| = O_P(T^{-1/2-\delta}),$$

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}^{(0)}_t - x) - \frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\right| = O_P(T^{-\delta}\sqrt{b})$$

for a $\delta > 0$.

Proof of Lemma 4. For the first claim it suffices to show for $C > 0$ that

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t(\theta_0) - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t\right| = O_P(T^{-1/2-\delta}),$$

$$\sup_{x\in I, \|\theta_1-\theta_2\|\le CT^{-1/2}}\left|\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t(\theta_1) - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t(\theta_2) - x)\varepsilon_t\right| = O_P(T^{-1/2-\delta}).$$

These statement can be shown by similar arguments as in the proofs of the statements of Lemma 2. The second statement of the lemma follows similarly. □

Lemma 5. *Under the assumptions of Theorem 3 it holds that*

$$\sup_{x\in I}\left|\frac{\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(0)}-x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(0)}-x)}-\frac{\frac{1}{T}\sum_{t=1}^{T}K_b(h_t-x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T}K_b(h_t-x)}\right|=O_P(T^{-1/2-\delta})$$

for a $\delta>0$.

Proof of Lemma 5. The statement of Lemma 5 follows directly from Lemma 4. □

Proof of Theorem 4.

For functions m we define

$$\widehat{\Gamma}_T(\theta,m)=\int\left\{\frac{\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta,m)-x)\left[Y_t-m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})\right]}{\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta,m)-x)}\right\}^2 w(x)dx.$$

Note that $\widehat{\Gamma}_T^{(k)}=\widehat{\Gamma}_T(\widehat{\theta}^{(k)},\widehat{m}^{(k)})$ for $k\geq 1$ and $\widehat{\Gamma}_T^{(0)}=\widehat{\Gamma}_T(\widehat{\theta}^{(0)},m_{\widehat{\gamma}})$ with $\widehat{\theta}^{(k)}=(\widehat{\psi}^{(k)},\widehat{\gamma})$. The statement of Theorem 4 follows from the following two claims. For $C>0$ it holds that

$$\sup_{(\theta_1,m_1),(\theta_2,m_2)\in\mathcal{M}^{C,*}}\left|\widehat{\Gamma}_T(\theta_1,m_1)-\widehat{\Gamma}_T(\theta_2,m_2)\right| = o_P(T^{-1}b^{-1/2}), \tag{38}$$

$$\left(\widehat{\theta}^{(k)},\widehat{m}^{(k)}\right) \in \mathcal{M}^{C,*}. \tag{39}$$

Here $\mathcal{M}^{C,*}$ denotes the set of all tuples (θ,m) with $m\in\mathcal{M}^C$ and where $\theta=(\psi,\gamma)$ fulfills $\|\psi-\psi_0\|\leq b^2T^{\iota}+(Tb)^{-1/2}T^{\iota}$, $\|\gamma-\gamma_0\|\leq (T)^{-1/2}T^{\iota}$ for some $\iota>0$ small enough. The set $\mathcal{M}^C$ is the class of all functions m whose second derivative is absolutely bounded by $C(Tb^5)^{-1/2}\sqrt{\log(T)}$, which coincide outside of I with $m_{\widehat{\gamma}}$ and which fulfil:

$$\sup_{x\in I}|m(x)-m_{b,0}(x)|\leq C(Tb)^{-1/2}\sqrt{\log(T)},$$

where

$$m_{b,0}(x)=\frac{\mathbf{E}[K_b(h_t-x)m_0(h_t)]}{\mathbf{E}[K_b(h_t-x)]}.$$

For a proof of (38) it suffices to show that for all $C>0$ for $\delta>0$ small enough

$$\begin{aligned}\sup_{x\in I,(\theta_1,m_1),(\theta_2,m_2)\in\mathcal{M}^{C,*}}&\left|\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta_2,m_2)-x)\varepsilon_t\right.\\&\left.-\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta_1,m_1)-x)\varepsilon_t\right|=O_P(T^{-1/2-\delta}),\end{aligned}\tag{40}$$

$$\begin{aligned}\sup_{x\in I,(\theta_1,m_1),(\theta_2,m_2)\in\mathcal{M}^{C,*}}&\left|\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta_2,m_2)-x)\right.\\&\left.-\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t(\theta_1,m_1)-x)\right|=O_P(\sqrt{b}\,T^{-\delta}).\end{aligned}\tag{41}$$

Using these two bounds claim (38) follows by similar arguments as in the proof of Theorem 1. Claims (40) and (41) follow by Lemma 3. Note that our assumptions allow the choice $\kappa > \eta/2$.

For the proof of (39) we will argue that for $l \leq k$

$$\sup_{x\in I}\left|\widehat{m}^{(l)}(x) - m_{b,0}(x)\right| \leq C(Tb)^{-1/2}\sqrt{\log(T)}, \quad (42)$$

$$\sup_{x\in I}\left|D_2\widehat{m}^{(l)}(x)\right| \leq C(Tb^5)^{-1/2}\sqrt{\log(T)} + C \quad (43)$$

almost shurely for $C > 0$ large enough. For a proof of (42) note that from (40) and (41) it follows that for $\delta > 0$ small enough

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(k)} - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t\right| = O_P(T^{-1/2-\delta}),$$

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(k)} - x) - \frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(0)} - x)\right| = O_P(\sqrt{b}\,T^{-\delta}).$$

Thus (42) follows from our results on $\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t$ and $\frac{1}{T}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(0)} - x)$ in the proof of Theorem 1.

For a proof of (43) we write

$$\widehat{m}^{(k)}(x) = \frac{\widehat{r}^A(x) + \widehat{r}^B(x)}{\widehat{f}_h^{(k)}(x)} + m_{\widehat{\gamma}}(x),$$

where $\widehat{r}^A(x) = T^{-1}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(k-1)} - x)\varepsilon_t$, $\widehat{r}^B(x) = T^{-1}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(k-1)} - x)[m_{\gamma_0}(h_t) - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})]$, and $\widehat{f}_h^{(k)}(x) = T^{-1}\sum_{t=1}^{T}K_b(\widehat{h}_t^{(k-1)} - x)$. For the proof of (39) it suffices to show for $0 \leq j \leq 2$ that $\sup_{x\in I}\left|D_j\widehat{r}^A(x)\right| + \sup_{x\in I}\left|D_j\widehat{r}^B(x)\right| + \sup_{x\in I}\left|D_j\widehat{f}_h^{(k)}(x)\right| \leq C(Tb^{2j+1})^{-1/2}\sqrt{\log(T)} + C$ and $\sup_{x\in I}\left|\widehat{f}_h^{(k)}(x)^{-1}\right| \leq C$, almost shurely for $C > 0$ large enough. This can be done by similar arguments as in the proof of (40) and (41). □

References

[1] Adrian, T., and J. Rosenberg (2008). "Stock returns and volatility: pricing the short-run and long-run components of market risk." *Journal of Finance*, forthcoming.

[2] Andersen, T., and T. Bollerslev (1998). "Answering the skeptics: yes, standard volatility models do provide accurate forecasts." *International Economic Review* 39, 885–905.

[3] Anderson, E. W., E. Ghysels, and J. Juergens (2007). "The impact of risk and uncertainty on expected returns." *Mimeo*, University of North Carolina at Chapel Hill.

[4] Backus, D. K., and A. W. Gregory (1993). "Theoretical relations between risk premiums and conditional variances." *Journal of Business & Economic Statistics* 11, 177–185.

[5] Beckaert, G. and G. Wu (2000). "Asymmetric volatility and risk in equity markets". *The Review of Financial Studies* 13, 1–42.

[6] Bodurtha, J. N., and N. C. Mark (1991). "Testing the CAPM with time-varying risks and returns." *The Journal of Finance* XLVI, 1485–1505.

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

[8] Bosq, D. (1998). *Nonparametric statistics for stochastic processes: estimation and prediction (2nd edition)*, Lecture Notes in Statistics, Vol. 110, Springer-Verlag, New York.

[9] Carrasco, M., and X. Chen (2002). "Mixing and moment properties of various GARCH and stochastic volatility models." *Econometric Theory* 18, 17–39.

[10] Chen, X., and E. Ghysels (2008). "News - good or bad - and its impact on volatility predictions over multiple horizons." *Mimeo*, University of North Carolina at Chapel Hill.

[11] Christensen, B. J., C. M. Dahl, and E. M. Iglesias (2008). "Semiparametric inference in a GARCH-in-Mean model." *Mimeo*, Michigan State University.

[12] Das, S., and N. Sarkar (2000). "An ARCH in the nonlinear mean model." *The Indian Journal of Statistics* 62, 327–344.

[13] Engle, R. F., E. Ghysels, and B. Sohn (2007). "On the economic sources of stock market volatility." *Mimeo*, New York University.

[14] Engle, R. F., D. M. Lilien, and R. P. Robins (1987). "Estimating time varying risk premia in the term structure." *Journal of Business and Economic Statistics* 9, 345–359.

[15] Fan, Y. and Q. Li (1999). "Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing." *Nonparametric Statistics* 10, 245–271.

[16] French, K. R., G. W. Schwert, and R. F. Stambaugh (1987). "Expected stock returns and volatility." *Journal of Financial Economics* 19, 3–29.

[17] Genotte, G., and T. Marsh (1993). "Valuations in economic uncertainty and risk premiums on capital assets." *European Economic Review* 37, 1021–1041.

[18] Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). "There is a risk-return trade-off after all." *Journal of Financial Economics* 76, 509–548.

[19] Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). "On the relationship between the expected value and the volatility of the nominal excess return on stocks." *The Journal of Finance* XLVIII, 1779–1801.

[20] Guo, H., and R. Whitelaw (2006). "Uncovering the risk-return relation in the stock market." *Journal of Finance* LXI, 1433–1463.

[21] Härdle W., and E. Mammen (1993). "Comparing nonparametric versus parametric regression fits." *The Annals of Statistics* 21, 1926–1947.

[22] Hjellvik, V., and D. Tjøstheim (1995). "Nonparametric tests of linearity for time series." *Biometrika* 82, 351–368.

[23] Karanasos, M. (2001). "Prediction in ARMA models with GARCH in mean effects." *Journal of Time Series Analysis* 22, 555–576.

[24] Kreiss, J.-P., M. H. Neumann, and Q. Yao (2002). "Bootstrap tests for simple structures in nonparametric time series regression." Preprint, TU Braunschweig.

[25] Lee, S.-W., and B. E. Hansen (1994). "Aymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator." *Econometric Theory* 10, 29–52.

[26] Li, Q., J. Yang, C. Hsiao, and Y.-J. Chang (2005). "The relationship between stock returns and volatility in international stock markets." *Journal of Empirical Finance* 12, 650–665.

[27] Linton, O., and E. Mammen (2005). "Estimating semiparametric ARCH(∞) models by kernel smoothing methods." *Econometrica* 73, 771–836.

[28] Linton, O., and B. Perron (2003). "The shape of the risk premium: evidence from a semiparametric generalized autoregressive conditional heteroscedasticity model." *Journal of Business & Economic Statistics* 21, 354–367.

[29] Linton, O., and A. Sancetta (2007). "Consistent estimation of a general nonparametric regression function in time series." *Mimeo*, The London School of Economics.

[30] Lundblad, C. (2007). "The risk return tradeoff in the long run: 1836 - 2003." *Journal of Financial Economics* 85, 123-150.

[31] 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.

[32] Mammen, E., O. Linton, and J. P. Nielson (1999). "The existence and asymptotic properties of a backfitting algorithm under weak conditions." *The Annals of Statistics* 27, 1443–1490.

[33] Masry, E. (1996). "Multivariate local polynomial regression for time series: uniform strong consistency and rates." *Journal of Time Series Analysis* 17, 571–599.

[34] Masry, E., and D. Tjøstheim (1995). "Nonparametric estimation and identification of nonlinear ARCH times series." *Econometric Theory* 11, 258–289.

[35] Mehra, R., and E. Prescott (1985). "The equity premium: a puzzle." *Journal of Monetary Economics* 15, 145–161.

[36] Merton, R. C. (1980). "On estimating the expected return on the market - an exploratory investigation." *Journal of Financial Economics* 8, 323–361.

[37] Merton, R. C. (1973). "An intertemporal capital asset pricing model." *Econometrica* 41, 867–886.

[38] Pagan, A. R., and A. Ullah (1988). "The econometric analysis of models with risk terms." *Journal of Applied Econometrics* 3, 87–105.

[39] Pagan, A. R., and Y. S. Hong (1990). "Nonparametric estimation and the risk premium." In Barnett, W. A., J. Powell, and G. E. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, 51–75.

[40] Poterba, J. M., and L. H. Summers (1986). "The persistence of volatility and stock market fluctuations". *The American Economic Review* 76, 1142–1151.

[41] Scruggs, J. T. (1998). "Resolving the puzzling intertemporal relation between the market risk premium and the conditional market variance: a two-factor approach." *The Journal of Finance* LIII, 575–603.

[42] Smith, D. R. (2006). "Risk and return in stochastic volatility models: volatility feedback matters!" *Mimeo*, Simon Fraser University.

[43] Sperlich, S. (2007). "A note on nonparametric estimation with constructed variables and generated regressors." *Mimeo*, University of Göttingen.

[44] Van de Geer, S. (2000). *Empirical processes in M-estimation.* Cambridge University Press, Cambridge.

[45] Whitelaw, R. F. (2000). "Stock market risk and return: an equilibrium approach." *The Review of Financial Studies* 13, 521–547.