Semiparametric Estimation of ARCH Models using Nearest Neighbour Regression: Some Monte Carlo Results*

Russell Davidson Dirk Eddelbüttel

May 1995

Abstract

We propose a semiparametric estimation method for regression models with an unknown pattern of autoregressive conditional heteroskedasticity. The idea is based on the nonparametric nearest neighbour estimator of Robinson (1987), but here the skedastic function, estimated nonparametrically, is supposed to depend on lagged squared error terms rather than exogenous variables. The fitted values of this nonparametric regression are used as estimates of the elements of $\hat{\boldsymbol{\Omega}}$, the diagonal covariance matrix of the regression error terms. This estimated covariance matrix then serves for feasible GLS.

In this paper we report the results of Monte Carlo experiments designed to compare this new method with both (infeasible) GLS using the true conditional variances generated in the experiment, and the maximum likelihood estimator of the correctly specified parametric model. We find that, for suitable choice of the nonparametric procedure and the number of nearest neighbours used, the new method can be practically as efficient as GLS knowing the true conditional variance.

JEL Classification : C14, C15

Keywords : Autoregressive conditional heteroskedasticity, Monte Carlo simulations, nearest neighbour regression, nonparametric regression, semiparametric regression

---

*This paper was prepared for the conference "Smoothing and Resampling in Economics", Humboldt-Universität zu Berlin, October 4 - 7, 1995. The author's affiliations are GREQAM, France, and Queen's University, Canada for R. Davidson, and GREQAM, France, for D. Eddelbüttel.

1 Introduction

In this paper we extend an idea of Robinson (1987), who considered a linear regression model

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + \epsilon_t, \qquad t = 1, \ldots, n, \tag{1}$$

in which the error terms ϵ_t are supposed to be heteroskedastic, with a pattern of heteroskedasticity dependent on a set of exogenous variables, $\boldsymbol{Z}_t$, usually a subset of the regressors $\boldsymbol{X}_t$. Thus the covariance matrix of the error terms, $\boldsymbol{\Omega}$, is diagonal, with diagonal elements σ_t^2 that are arbitrary functions of some or all of the variables $\boldsymbol{Z}_t$.

When the form of the dependence of the σ_t^2 on $\boldsymbol{Z}_t$ is known, $\boldsymbol{\beta}$ can be efficiently estimated by generalised least squares (GLS). If the functional form of the dependence depends on unknown parameters, these can often be estimated consistently by regressing the squared residuals from OLS estimation of (1) on the $\boldsymbol{Z}_t$ or on functions of the $\boldsymbol{Z}_t$. In that case, feasible generalised least squares (FGLS) can be used to obtain an estimate of $\boldsymbol{\beta}$ that is asymptotically as efficient as the (infeasible) GLS estimate that would be available if all the σ_t^2 were known exactly. Robinson showed that one also obtains asymptotic efficiency in estimating $\boldsymbol{\beta}$ by using estimates of the σ_t^2 obtained by nonparametric regression of the squared OLS residuals on the $\boldsymbol{Z}_t$. In fact, Robinson limited his attention to nonparametric estimates provided by the k nearest neighbours (kNN) method.

In recent years, autoregressive conditional heteroskedasticity, as formulated in ARCH and GARCH models, has received at least as much attention as exogenous heteroskedasticity of the sort considered by Robinson. In this paper we propose to investigate to what extent nonparametric methods like Robinson's can be applied to models with conditional heteroskedasticity, in which σ_t^2 will be a function, not of exogenous variables, but rather of the squares of past error terms.

In the next section, we briefly discuss traditional ARCH and GARCH models and describe Robinson's approach to GLS using the method of k nearest neighbours. We then propose a number of ways in which this approach can be extended to models with conditional heteroskedasticity. In the following section, we present and discuss some Monte Carlo evidence on the properties of the various procedures we propose. In samples of reasonable size, the new methods appear to work well, and, in favourable circumstances, are capable of yielding estimators that are, within experimental error, almost as precise as (infeasible) GLS.

2 Heteroskedasticity

Models of AutoRegressive Conditional Heteroskedasticity (ARCH), introduced in a seminal paper by Engle (1982), have been much used in the last decade; see Bollerslev, Chou, and Kroner (1992), Bera and Higgins (1993) and Bollerslev, Engle, and Nelson (1994) for surveys.

In particular, the GARCH model of Bollerslev (1986), a parsimonious generalisation of the original model, has been shown to fit a large number of financial data series reasonably well. Many variants of these models have been proposed, but the simple GARCH(1,1) model has established itself as a benchmark in this literature.

As defined by Bollerslev (1986), the GARCH(p,q) model can be written as

$$
\begin{aligned}
\epsilon_t &= y_t - \boldsymbol{X}_t\boldsymbol{\beta}; \\
\epsilon_t|\Psi_{t-1} &\sim N(0,h_t); \\
h_t &= \alpha_0 + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2 + \sum_{j=1}^{p}\gamma_j h_{t-j} \equiv \boldsymbol{z}_t'\boldsymbol{\omega}; \\
\boldsymbol{z}_t' &= \left[1\ \epsilon_{t-1}^2\ \ldots\ \epsilon_{t-q}^2\ h_{t-1}\ \ldots\ h_{t-p}\right]; \\
\boldsymbol{\omega}' &= [\alpha_0\ \alpha_1\ \ldots\ \alpha_q\ \gamma_1\ \ldots\ \gamma_p]; \\
\boldsymbol{\theta}' &= [\boldsymbol{\beta}'\ \boldsymbol{\omega}'].
\end{aligned}
$$

The contribution of observation t to the log-likelihood is

$$
l_t(\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log h_t - \frac{1}{2}\epsilon_t^2 h_t^{-1}.
$$

Differentiating l_t with respect to the parameters of the variance equation yields

$$
\frac{\partial l_t}{\partial \boldsymbol{\omega}} = -\frac{1}{2}h_t^{-1}\frac{\partial h_t}{\partial \boldsymbol{\omega}} + \frac{1}{2}\epsilon_t^2 h_t^{-2}\frac{\partial h_t}{\partial \boldsymbol{\omega}} = \frac{1}{2}h_t^{-1}\frac{\partial h_t}{\partial \boldsymbol{\omega}}\left(\frac{\epsilon_t^2}{h_t} - 1\right)
$$

where

$$
\frac{\partial h_t}{\partial \boldsymbol{\omega}} = \boldsymbol{z}_t + \sum_{i=1}^{p}\gamma_i\frac{\partial h_{t-1}}{\partial \boldsymbol{\omega}}.
$$

Differentiating with respect to the parameters in the mean equation yields

$$
\begin{aligned}
\frac{\partial l_t}{\partial \boldsymbol{\beta}} &= \epsilon_t\boldsymbol{X}_t h_t^{-1} - \frac{1}{2}h_t^{-1}\frac{\partial h_t}{\partial \boldsymbol{\beta}} + \frac{1}{2}\frac{\epsilon_t^2}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\beta}} \\
&= \frac{\epsilon_t\boldsymbol{X}_t}{h_t} + \frac{1}{2}h_t^{-1}\frac{\partial h_t}{\partial \boldsymbol{\beta}}\left(\frac{\epsilon_t^2}{h_t} - 1\right) \qquad (2)
\end{aligned}
$$

where

$$
\frac{\partial h_t}{\partial \boldsymbol{\beta}} = -2\sum_{j=1}^{q}\alpha_j\boldsymbol{X}_{t-j}\epsilon_{t-j} + \sum_{j=1}^{p}\gamma_j\frac{\partial h_{t-j}}{\partial \boldsymbol{\beta}}.
$$

With a suitable numerical maximisation routine, these first derivatives permit maximum likelihood estimation along the lines of the Berndt, Hall, Hall, and Hausman (1974) algorithm

(BHHH hereafter) as suggested by Bollerslev (1986, p. 317). However, the updating scheme in the BHHH algorithm uses an outer product of the gradient estimate of the information matrix. Given the rather poor performance of this estimator in finite samples (see Davidson and MacKinnon (1993, section 13.7) for more details on OPG based inference), one should be wary of the covariance matrix estimate produced by BHHH.

It should be noted that ML estimation of a correctly specified GARCH model is in general more efficient than any form of GLS, even asymptotically. This is because, although GLS takes account of the conditional heteroskedasticity, it does not take account of the fact that the parameter $\boldsymbol{\beta}$ of the regression function influences the skedastic function through the squared residuals. In other words, GLS ignores the dependence of h_t on $\boldsymbol{\beta}$. In the first-order conditions given above for GARCH ML estimates, one sees two terms in the derivative of the loglikelihood with respect to $\boldsymbol{\beta}$, the second involving the factor $\partial h_t/\partial\boldsymbol{\beta}$. GLS is implicitly solving first-order conditions that do not have this second term. Robinson's problem, in which the heteroskedasticity depends on exogenous variables only, does not have this feature. Thus, in the case of autoregressive conditional heteroskedasticity, we cannot expect our nonparametric procedure to achieve the efficiency of GARCH maximum likelihood estimation, always assuming, of course, that the GARCH model is properly specified. All feasible GLS procedures suffer from the same drawback, since asymptotically the best they can achieve is the efficiency of infeasible GLS.

Another way to see this point is to note that the two terms in the loglikelihood derivative with respect to $\boldsymbol{\beta}$ are uncorrelated under standard regularity conditions. The variance of the sum of the two terms is thus equal to the sum of the variances, and must therefore be greater than the variance of the first term by itself. It follows that the information about $\boldsymbol{\beta}$ is greater in the GARCH ML case than in the case of GLS.

Critics of GARCH models argue that their asymptotics are fairly complicated. Another complication attending these models is that estimation must normally be undertaken with non-negativity restrictions on the parameters α_i and γ_i, in order to ensure that the conditional variance is positive for all t. More recently, studies of high-frequency data on exchange rates have shown that the standard GARCH(1,1) does not work at all well at these low levels of time aggregation (see for example Guillaume, Dacorogna, Davé, Müller, Olsen, and Pictet (1994) and Guillaume, Pictet, and Dacorogna (1995)).

Yet another criticism can be made. (G)ARCH models are *ad hoc* by their very nature. Such models try to explain the second moment of the error terms by a linear combination of squared lagged residuals. Although many variants have been proposed that introduce asymmetric effects, threshold effects, and non-linearities of different sorts, none of these variants has been able to establish itself as general enough to work well on the full variety of available data sets.

In the face of all these criticisms, in this paper we follow Robinson (1987) in modelling the variance by means of an auxiliary nonparametric regression. The square roots of the fitted values from this regression are then used as weights in a feasible GLS estimation of the original model. As did Robinson, we use the k nearest neighbours procedure. In this procedure, the fitted value for observation t is computed as a weighted average of those k values of the dependent variable (here the squared OLS residuals) for which the explanatory variables $\boldsymbol{Z}_i$, exogenous in Robinson's case, but not necessarily so in ours, are closest to $\boldsymbol{Z}_t$ in an appropriate Euclidean metric. This method is described in Robinson (1987), Härdle (1990) and Härdle and Linton (1994). An efficient algorithm for the procedure is due to Friedman, Baskett, and Shustek (1975); an implementation of this algorithm is provided by Delgado (1993). We are grateful to Miguel Delgado for providing us with his Fortran program; we actually used a C++ translation of it for our Monte Carlo work.

One of our proposed methods is directly in the spirit of the GARCH literature. First, the underlying model (1) is estimated by OLS, or, if the underlying model is nonlinear, by nonlinear least squares. Next, the squared residuals from this regression are used as the dependent variable in a kNN nonparametric regression, in which the explanatory variables are lags of these same squared residuals. Finally the square roots of the fitted values of this nonparametric regression, which are necessarily positive, are used as the weights in an FGLS procedure. In the next section, we use only the first lags of the squared residuals, but the method generalises in an obvious way to an arbitrary number of lags. Essentially, what we have is a nonparametric version of an ARCH(1) estimation procedure, which should be capable of capturing richer nonlinear dependences than can the basic ARCH(1) model. An alternative procedure is also considered, in which, perfectly agnostically, we regress the squared residuals simply on the index t of the observation, that is, on a grid of equidistant values. This is the "standard" nonparametric smoothing that can be performed by a wide variety of methods, ranging from orthogonal series estimation and kernel estimation to wavelets.

The next section describes our experiments on some of these procedures.

3 Monte Carlo Results

We use the graphical methods proposed by Davidson and MacKinnon (1994) for the presentation of simulation results. In their paper, they propose three forms of graphs to present the outcome of a Monte Carlo study. All of them suppose that N independent drawings of a test statistic are generated in a Monte Carlo experiment. The test statistic is further supposed to follow approximately some nominal distribution, which can be used to compute the probability values (P values) corresponding to the N realisations of the statistic. The graphs are based on the empirical distribution function $\hat{F}$ of the P values. Davidson and MacKinnon propose three kinds of graphs: "P value plots", "P value discrepancy plots" and "size-power curves". In the first, the values of $\hat{F}(x_i)$ are plotted against a set of points x_i chosen in advance. If the actual distribution of the statistic is close to nominal, the resulting graph may well be almost indistinguishable from the 45° line. In that event, "P value discrepancy plots" can reveal more detail. Here $\hat{F}(x_i)-x_i$ is plotted against x_i. In a "size-power curve" two empirical distribution functions are plotted against each other: the first, $\hat{F}(x)$, is generated by a data-generating process (DGP) which satisfies the null hypothesis, whereas the second, $\hat{F}^{\star}(x)$ is computed using a data generating process that does not. Plotting the pairs $(\hat{F}(x_i), \hat{F}^{\star}(x_i))$ generates a size-power curve for the DGP used to generate $\hat{F}^{\star}(x)$ based on *corrected* size; see Davidson and MacKinnon (1994) for more details and examples. Such size-power curves can be used in our setting to compare estimators as well as test statistics: we calculate the asymptotic t-statistics based on different estimates of the parameter(s) of interest and of the covariance matrix.

Two different DGPs were used in our experiments. In the first, we used a GARCH(1,1) process with standard normal errors defined by

$$\begin{aligned}
y_t &= x_t\beta + \epsilon_t \\
\epsilon_t &= \zeta_t h_t \\
h_t &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \gamma_1 h_{t-1} \\
\zeta_t &\sim N(0,1) \\
(\beta, \alpha_0, \alpha_1, \gamma_1) &= (1.0, 1.0, 0.2, 0.75).
\end{aligned}$$

Using the results in Bollerslev (1986), we easily verify that the second and fourth moment of this GARCH process exist.

We also use a simpler ARCH(1) DGP that is defined by

$$\begin{aligned}
y_t &= x_t\beta + \epsilon_t \\
\epsilon_t &= \zeta_t h_t \\
h_t &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 \\
\zeta_t &\sim N(0,1) \\
(\beta, \alpha_0, \alpha_1) &= (1.0, 1.0, 0.5).
\end{aligned}$$

We estimate the parameter β using data generated by these two DGPs, using four different estimators. The first one is the correctly specified maximum likelihood estimator (MLE) for either a GARCH(1,1) or an ARCH(1) model. The numerical maximisation is undertaken with analytical gradients and the expected Hessian. This estimator attains the Cramér-Rao bound: no other estimator can be more efficient the MLE when the model is correctly specified. Figure 1 shows one realisation of each of the two DGPs along with the maximum likelihood fit for the skedastic function.

The second estimator that we use is infeasible GLS. In situations other than Monte Carlo studies, the econometrician does not know the precise form of heteroskedasticity. But because this is a simulation study, we are able to estimate *à la* Aitken by premultiplying regressors and regressand with $h_t^{-1/2}$ where we use the true values h_t that were used to construct the data. The third estimator is semiparametric: we regress $\boldsymbol{y}$ on $\boldsymbol{X}$, square the residuals, "treat" them nonparametrically and use the fitted values in a weighted regression

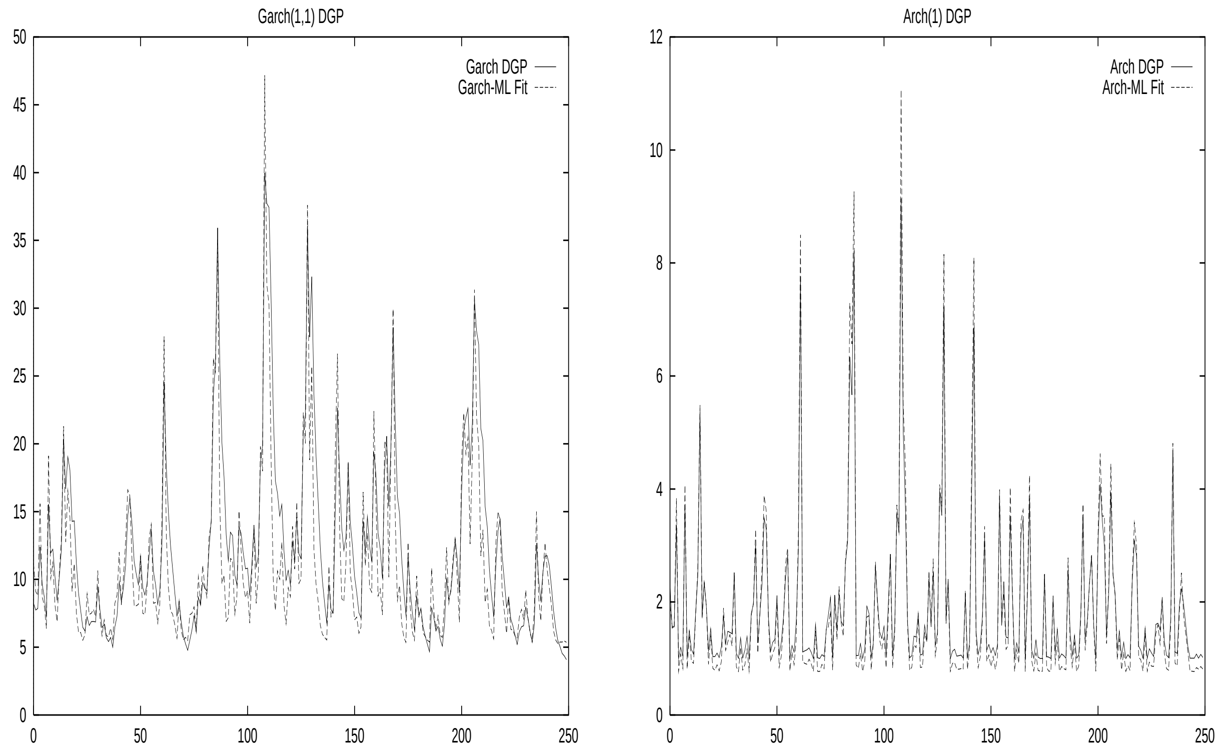

Figure 1: Maximum likelihood fitted values for the skedastic function

of $\boldsymbol{y}$ on $\boldsymbol{X}$. Two different nonparametric regressions are examined: a regression of squared residuals on their first lags, and a regression of squared residuals on an equidistant grid.

It is interesting to see graphically how well the nonparametric regression captures the true skedastic function. Figure 2 therefore repeats for the nonparametric fits what we saw in Figure 1 for the ML fits. The upper two graphs are the results from the nonparametric regression on the first lag, and the lower two graphs are from the grid-based nonparametric nearest-neighbour regression. The data-generating processes are, as in figure 1, a GARCH(1,1) on the left and an ARCH(1) on the right. The same data as in figure 1 are used.

In figure 1, we see that the correctly specified maximum likelihood estimator achieves a remarkable fit; but this is hardly surprising, as this estimator attains the Cramér-Rao bound. The nonparametric estimates, also unsurprisingly, track the true skedastic function less well – in every graph, we find some peaks that are not well captured. It is interesting to note that the nonparametric fit based on the lagged squared residuals is better for the ARCH(1) DGP than for the GARCH(1,1). This corresponds to the fact that, in some sense, the nonparametric regression is correctly specified for ARCH(1) – the conditional variance is indeed a function only of the squared error term lagged once – but not for GARCH(1,1), where the conditional variances depend on all lagged squared errors. Nevertheless, both nonparametric estimators provide a reasonable fit for GARCH, whereas ARCH is clearly better fit by the regression on the first lag than by the one on the grid. The grid regressions

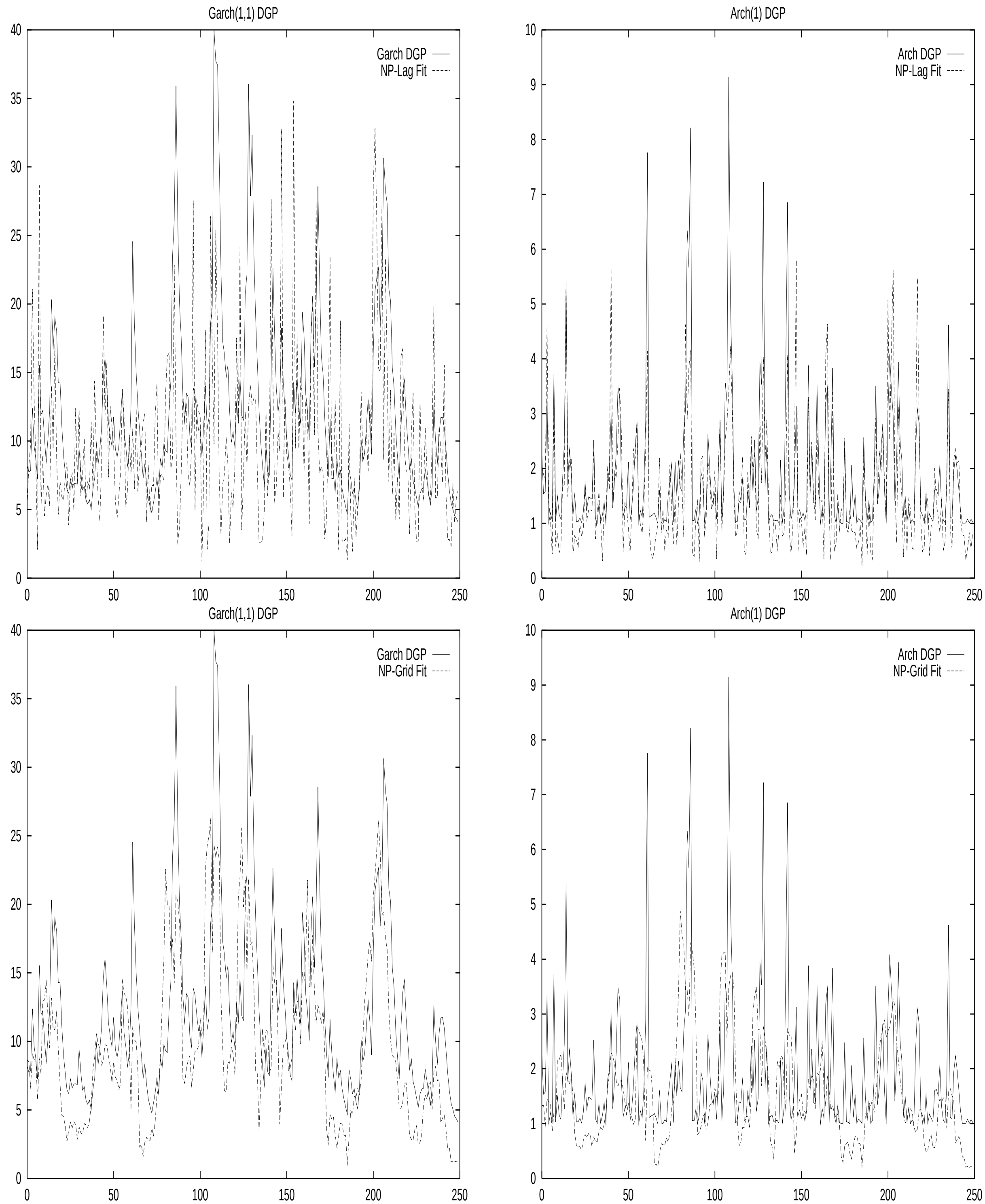

Figure 2: Nonparametrically estimated fitted values for the skedastic function

also seem to provide less variable, smoother, fits than the regressions on the lagged squared errors. This is presumably simply a consequence of the fact that the regression on the grid is primarily a smoothing procedure. But one should be cautious in interpreting these graphs, as they come from one single realisation of each DGP.

A fundamental problem with nonparametric estimation is that it requires the selection of a parameter that determines the "degree of localness" for the smoothing. For kernel estimation, the appropriate parameter is the window width; for nearest-neighbour estimation it is the number k of neighbours to include in the weighting. Few theoretical results are available concerning the best choice of this parameter. Indeed, asymptotic theory can tell us only that, in order to minimise the mean squared error of the kNN estimator, k should asymptotically be of the order of $n^{4/5}$; see Härdle (1990). Under this assumption, Robinson (1987) obtains results on the asymptotic properties of the kNN estimator, but little is said about what happens when k is chosen badly.

Cross-validation is commonly used as a data-driven mechanism for finding "good" parameter values. For many problems, there exist proofs that minimising the cross-validation criterion function is an asymptotically optimal search strategy; see for instance Härdle (1990, section 5.1). For our Monte Carlo experiments, it would have been prohibitive to calculate and minimise the cross-validation function for each Monte Carlo replication. Instead, we performed a series of experiments in each of which k was fixed, for various values of k.

This could of course be a very bad strategy if different realisations of the same DGP lead to cross-validation functions that are minimised by dramatically different values of k. In order to check on this, we explicitly calculated cross-validation curves for a small number of independent realisations of each DGP. The results are shown in figure 3, where we plot the value of the cross-validation function $\mathrm{CV}(k)$ for six different realisations of each DGP. There are four groups of results, one for each of the two nonparametric regressions under consideration, and each of the two DGPs, distributed on the page as in Figure 2. The cross-validation functions are scaled appropriately so as to facilitate comparison.

Figure 3 shows that the cross-validation functions may indeed vary considerably for different realisations of one data-generating process. On the other hand, almost all of the functions have a global minimum for a quite small value of k, even if, for larger values, their behaviour is highly variable. The first graph in the upper-left hand corner shows results for GARCH(1,1) data and a nonparametric regression on lagged squared errors. After the global minimum for a small k, the functions are not generally monotonic, although there is a visible trend in the upwards direction. For the GARCH(1,1) data and a grid-based regression much the same can be said, although the curves seem to vary over a smaller range. The curves for the regression on lagged squared errors with the ARCH(1) data all seem to flatten out for values of k above 30. For lower values of k, their behaviour is much more variable, and, in two cases, there seems not even to be a minimum at a small value of k; this differs from the GARCH results. The grid-based regression on the ARCH data gives similar results, but with still more variability. It is difficult to draw any clear conclusions, without minimising

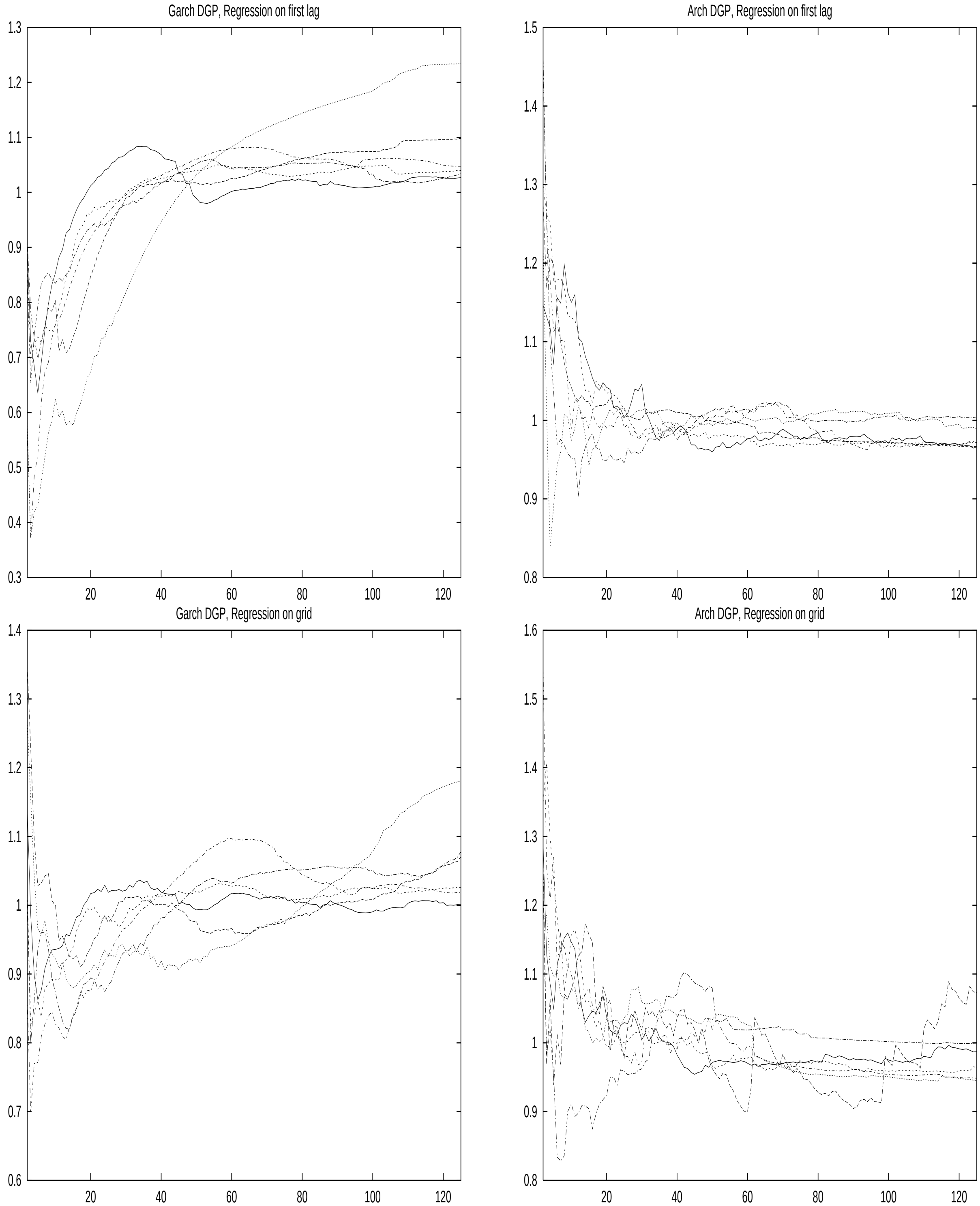

Figure 3: Cross-validation functions

the cross-validation function on each replication, about how well cross-validation works on average.

With this reserve, then, we now pass on to the discussion of the main Monte Carlo experiments. The quantities of interest on each replication will be the t-statistics for hypotheses that the single parameter in the mean equation takes on various values. We prefer to focus on t-statistics rather than raw parameter estimates because, in practice, only the former are of use in statistical inference. The consequence of our choice is that we do not study the distribution of the estimator of the parameter β, but rather of this estimator divided by the estimator of its standard error. Although these distributions are different, we argue that it is the second one that is more interesting: a parameter estimate does not allow an investigator to come to any conclusion unless it is accompanied by an estimate of its variance or standard error.

By computing t-statistics for a number of different null hypotheses, we can use the same experiment to investigate both the size and power of the t-test. This procedure differs somewhat from standard practice, where, if one were interested in power, one would normally generate data under both the null and an alternative hypothesis. This is unnecessary for the present experiment, since, for linear regression models, the residuals are invariant with respect to the values of the parameters of the regression function. Thus it is enough to generate data for only one value of β, and use these data to investigate both the size of the test, by considering the hypothesis that the value of β is what it really is, and also the power of the test, by considering a number of false hypotheses. In all cases, P-values are computed for the Student-t distribution with the appropriate number of degrees of freedom, using the routines from Moshier (1989). Size considerations allow us to see whether inference based on nominal critical values or P-values is reliable; power considerations allow us to measure the efficiency of the different estimators we consider.

In the experiments we now report, we set the number of observations $n = 250$ and performed $N = 5000$ Monte Carlo replications. In order to study the impact of sample size, we also conducted a second, smaller, set of experiments with $n = 500$ and $N = 2000$. In all, we have eight different experiments, using two data-generating processes, two nonparametric regressions and two sets of values for n and N. In addition, we consider three reference cases against which we compare the results of the nonparametric procedures. The first is just OLS itself. The OLS t-statistic should have behaviour close to nominal, since both the ARCH and GARCH DGPs yield *unconditionally* homoskedastic error terms. Next is infeasible GLS. By comparing it with OLS, we can see how much of an efficiency gain is theoretically possible by use of a feasible GLS procedure. By comparing it with the nonparametric procedures, we can see how well these do relative to the theoretical limit. Lastly, we consider the maximum likelihood estimator itself, which, according to the argument made in the last section, should be more efficient even than infeasible GLS.

GARCH data; lagged squared errors. Figure 4 shows the results of the first experiment. For $n = 250$ and $N = 5000$, data generated by the GARCH(1,1) DGP were used to estimate β by OLS, infeasible GLS, maximum likelihood using the true GARCH(1,1) model, and the semiparametric estimator based on a kNN fit of the conditional variance obtained from regressing the squared OLS residuals on their first lags. For k, we used values between 20 and 120. Because of space constraints, we show results for only a selection of the values of k considered.

The two left-hand panels display the P-value discrepancy plots for the t-statistics computed using the different estimators. The upper plots show the results for $k = 20, 40, 60$, and the lower plots for $k = 80, 100, 120$. Things seem to move continuously as k varies. Since these plots do not have a natural scale on the vertical axis, it is important to check, when studying the graphs, on how quantitatively important the P-value discrepancies actually are. In the present case, however, it is clear that in only one case, namely for the semiparametric test with lagged squared errors and $k = 20$, is there a P-value discrepancy that emerges from the experimental noise. Our experiments are not sufficiently precise to determine whether any other tests have significant size distorsions; in any event, if they exist, they clearly cannot be large.

The right-hand panels in figure 4 show the size-power curves for the MLE, GLS, OLS, and the semiparametric estimators corresponding to different values of k. The data were generated for $\beta = 1$, and the hypothesis tested was $\beta = 0$. Note that we have not shown the full extent of these curves. It seemed preferable to show a "zoom" of the principal area of action, in the top left hand corner of the graphs.

The results are reasonably clear. Maximum likelihood gives the greatest efficiency, as expected, with infeasible GLS not far behind. The semiparametric estimators are clearly less efficient than either GLS or ML, but are always more efficient than OLS. In fact, all values of k give virtually the same power. It is probably this feature that led to the rather flat cross-validation functions seen in figure 3. On the other hand, $k = 20$ gave significant size distorsion.

GARCH data; grid. Figure 5 is set up like figure 4, and shows the results of the next experiment, in which the nonparametric regression was on an independent variable consisting of an equidistant grid on the unit interval.

Looking at the P-value discrepancy plots, we see that, with one exception, the discrepancies for all the semiparametric tests are almost completely contained between those for the MLE and GLS. The exception is once again for $k = 20$. However, although the discrepancy curve for this value of k seems significantly different from the other curves drawn in the figure, a closer inspection reveals that the absolute size of the size distorsion is no bigger than that for infeasible GLS, around 0.01 at its greatest point.

The size-power curves, in the right-hand panels of Figure 5, appear to show that the semiparametric estimate for $k = 20$ performs better than infeasible GLS! This is presumably just a consequence of experimental noise. It nevertheless seemed interesting to look more closely at small values of k. We thus performed a series of experiments for values of k between

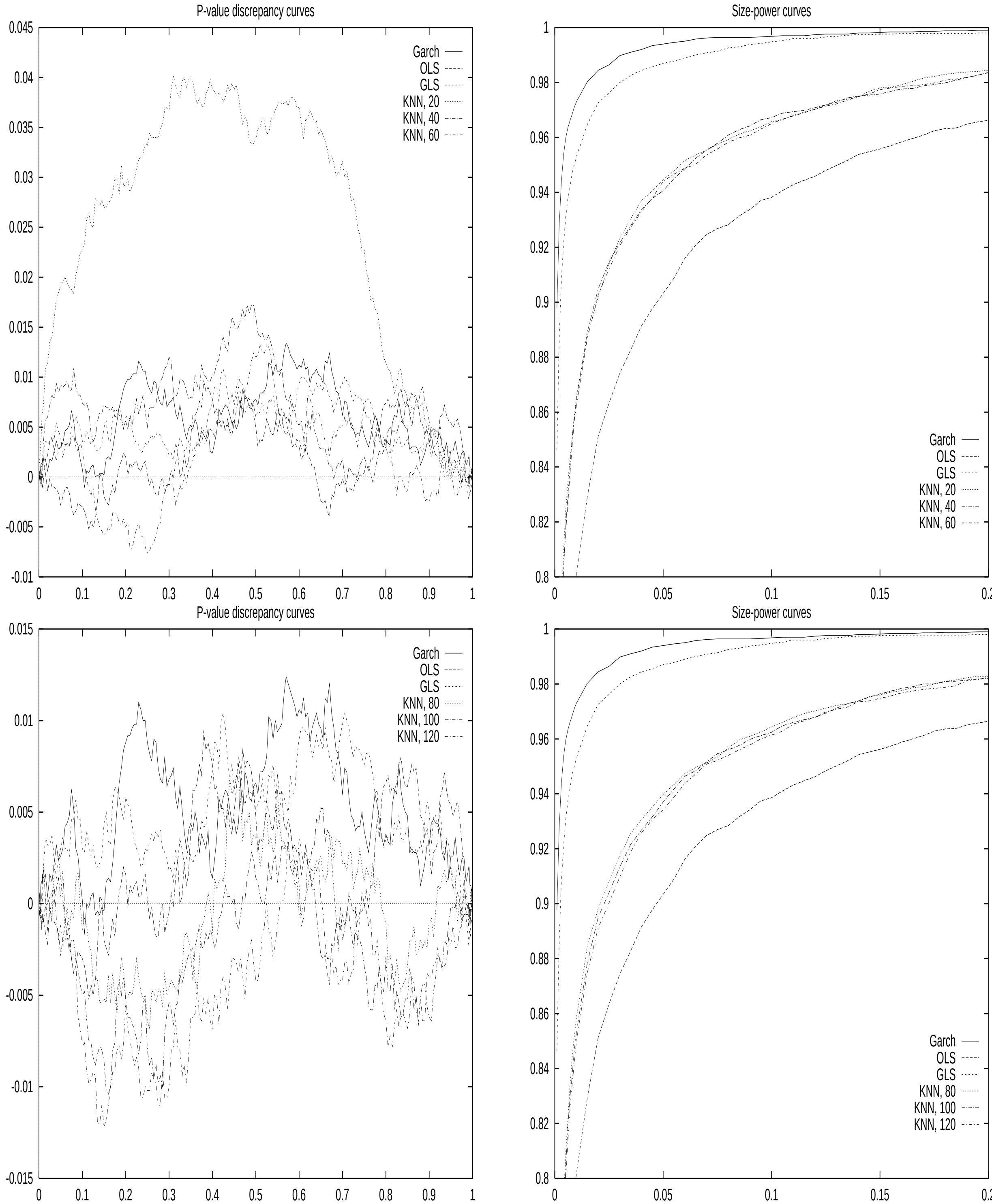

Figure 4: Regression on first lag, GARCH data, $n = 250$, $N = 5000$, $H1 : \beta = 0.0$

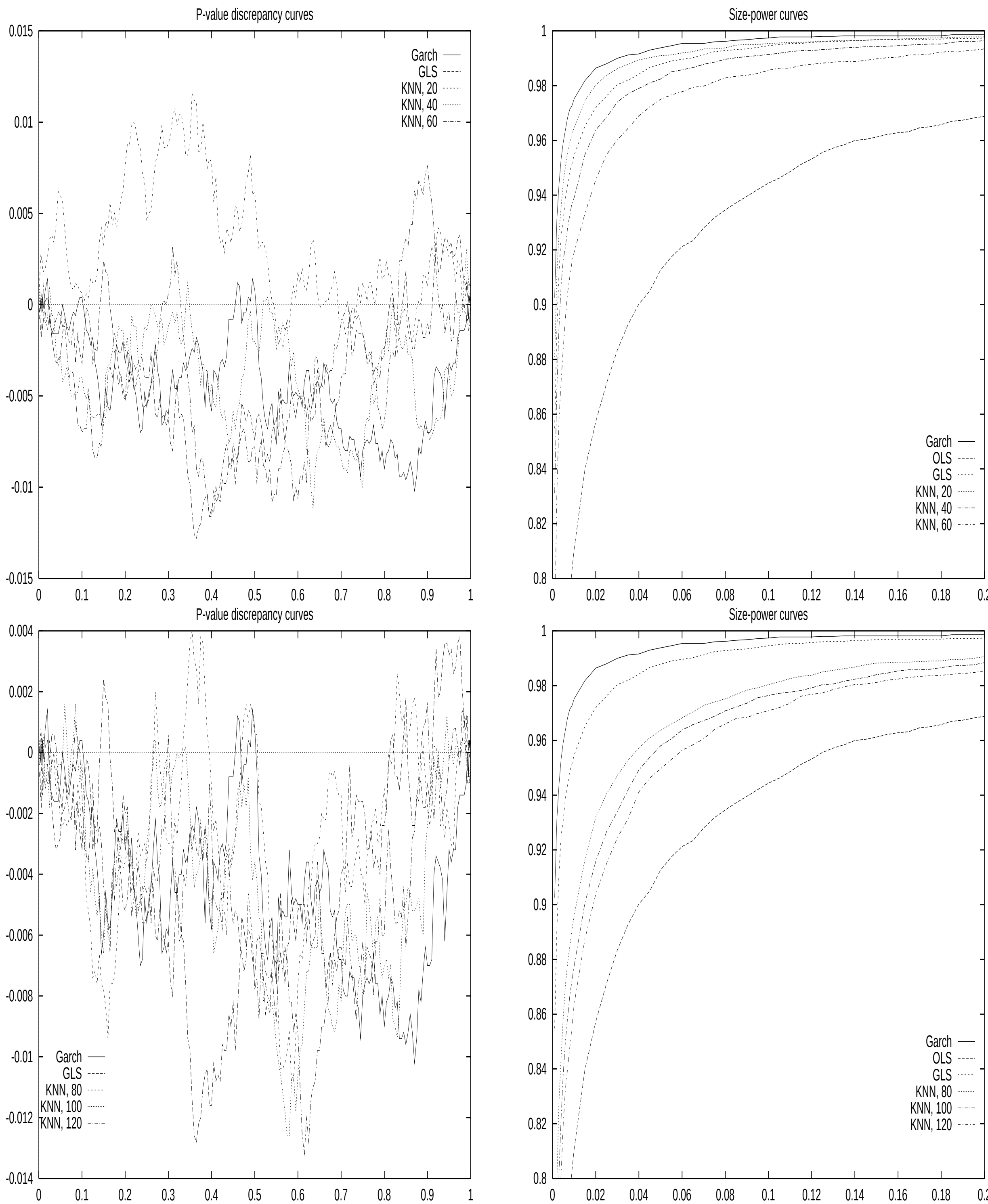

Figure 5: Regression on grid, GARCH data, $n = 250$, $N = 5000$, $H1 : \beta = 0.0$

2 and 30 for only 1000 replications each. For all values between 10 and 30, the size-power curves were almost indistinguishable, at this higher level of experimental noise at least, from the size-power curve for infeasible GLS, and those for k around 20 appeared to have best power.

For $k = 40$ and for $k = 60$, however, the tests based on the semiparametric estimator are clearly less powerful than infeasible GLS. For still higher values of k, as seen in the lower right-hand panel of the figure, we see that the power decreases considerably as k increases. This result is not unexpected, considering that the regression on the grid is just a smoothing procedure, with greater smoothing for larger k, and the true skedastic function is not particularly smooth.

Comparing the size-power curves from figures 4 and 5, we conclude that the semiparametric estimator based on the equidistant grid clearly works better with a GARCH(1,1) DGP than does the variant where we regress on the lagged squared errors. Moreover, the size distortion for the estimator that has best power in the lagged squared errors case is greater than that in the grid case, in which the worst size distortion is no worse than that for infeasible GLS. This feature of our results is most likely due to the fact that, in a GARCH model, the conditional variance depends on more lags that just the first.

ARCH data; lagged squared errors. We now examine the performance of the various tests with data generated by an ARCH(1) DGP. The ML estimator used as a benchmark is also obtained using a correctly specified ARCH(1) model. The principal difference here with respect to the experiments on GARCH-generated data is that the nonparametric regression on lagged squared errors is now correctly specified: the conditional variances are indeed just functions of the first lagged squared error. Another difference is that all the estimators we consider are considerably more efficient with ARCH data. Consequently, in computing size-power curves, the hypothesis tested was $\beta = 0.5$ rather than $\beta = 0$.

The upper left-hand panel of figure 6 shows that significant size distorsions affect even the ARCH MLE, where the greatest estimated distorsion is around 0.021. For $k = 60$, the semiparametric estimator does about as well in this regard as does the ARCH MLE. The best size performance is achieved for $k = 40$. The lower left-hand panel shows the three semiparametric P-value discrepancy plots for higher values of k lying close together, at a greater distance from the horizontal axis than the GLS curve, and at about the same distance as the MLE curve. However, all the P-value discrepancies shown in this lower panel are not very large, even if they may be significant, something which is not clear without a formal test.

The two size-power graphs of figure 6 show a very regular behaviour of the semiparametric estimator for the ARCH(1) DGP. The overall variation among the size-power curves is very low; again this is no doubt the same phenomenon as that which led to the flat cross-validation curves. Once more we performed experiments with a small number of replications for different values of k, and obtained results very similar to those shown. Overall, the semiparametric

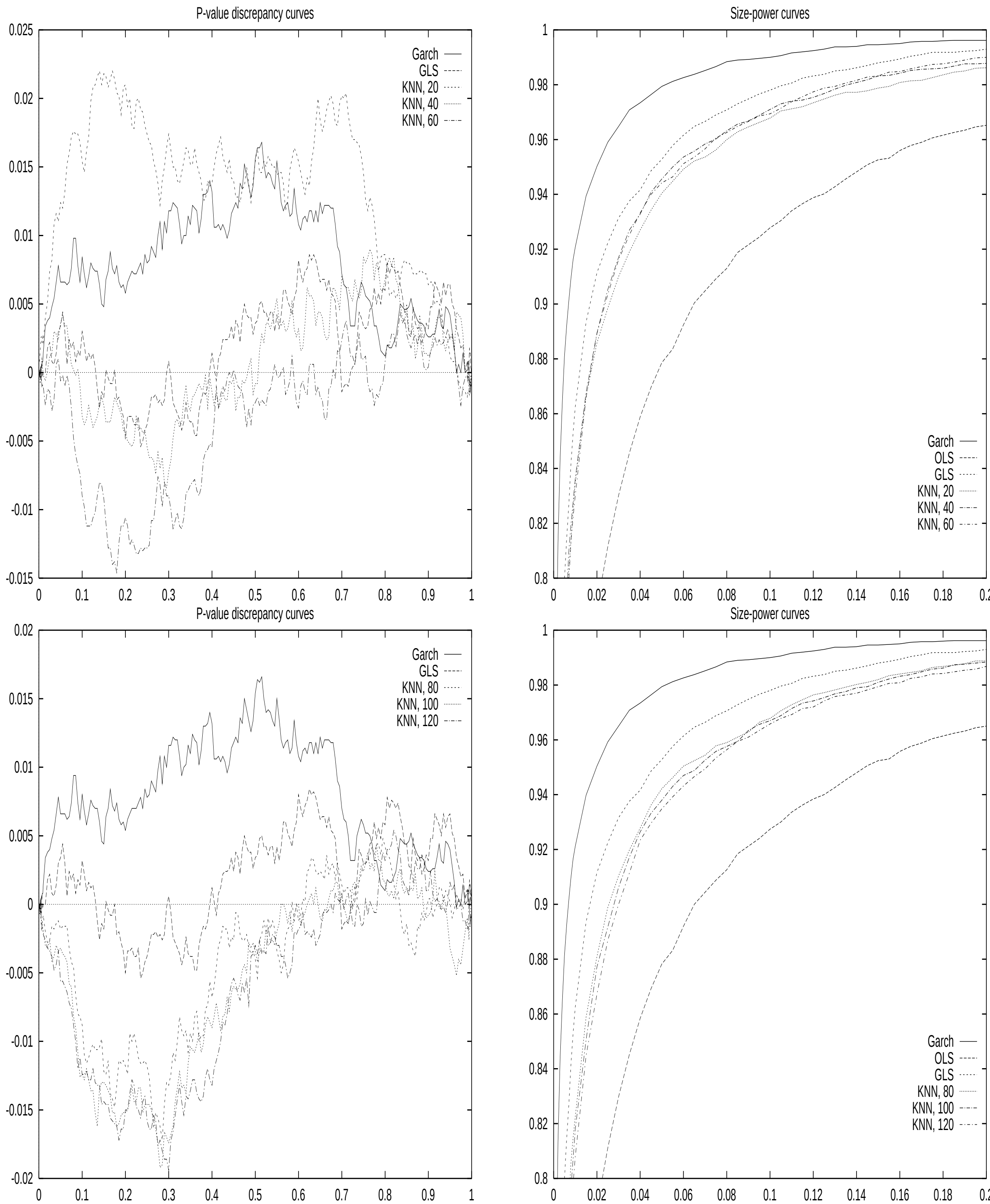

Figure 6: Regression on first lag, ARCH(1) data, $n = 250$, $N = 5000$, $H1 : \beta = 0.5$.

estimator in this case shows the same lack of sensitivity to k as does the grid-based estimator with the GARCH data.

All the size-power curves in figure 6 are closer to GLS than to OLS, whereas in figure 4 the contrary is true. We confirm the greater efficiency of the procedure based on the lagged square errors when we use ARCH data rather than GARCH data.

ARCH data, grid. Next, we consider the nonparametric regression on an equidistant grid applied to data from the ARCH(1) DGP. As above, the value of β under the alternative hypothesis is set to 0.5.

The results shown in figure 7 are clearly the worst of the four main figures we have seen so far. Although for k equal to 40 or greater the P-value discrepancies are small and no worse than those of ML or infeasible GLS, there is the same problem as the one we observed before for $k = 20$, where size distorsion is significant.

But it is in terms of power that the tests considered here do worst. The size-power curves show for the first time that the power of the test based on the semiparametric estimator can be about as low as for the test based on the OLS estimator. The best power is halfway between OLS and GLS; but this is for $k = 20$, for which, as we have seen, there is size distorsion.

If we compare the graphs in figure 5, where the grid-based semiparametric estimator is applied to the GARCH data, with those in figure 7, we see that the grid-based method works disappointingly badly on the ARCH data, despite its very good performance on the GARCH data. Since with ARCH(1) the conditional heteroskedasticity is autoregressive only to first order, the smoothing process implicit in the grid regression evidently gives a poor account of a conditional variance that is evidently not a very smooth function of the time.

Larger sample size. We now present results from a series of experiments in which the sample size was doubled to $n = 500$. In order to keep computing time reasonable, there were only $N = 2000$ replications in these experiments. The increased precision that comes with the greater sample size made it necessary to change the values of β for the alternative hypotheses used when looking at power; the hypothesis that $\beta = 0$ was rejected almost all of the time. Thus, for the GARCH DGP, the hypothesis tested was $\beta = 0.25$, and for the ARCH DGP, $\beta = 0.7$.

The results are set out in figures 8, 9, 10, and 11. These should be compared with figures 4, 5, 6, and 7 respectively. Such a comparison shows at once that the overall pattern of results is hardly changed at all by doubling the sample size. The advantages and the defects of the various procedures remain the same. Even in the size-power curves, the change in the hypothesis tested is sufficient to leave power roughly the same as with the smaller sample size. These experiments, although far from complete, suggest quite strongly that the results obtained so far will not change greatly if one considers still larger sample sizes; we seem to be already close enough to the asymptotic regime.

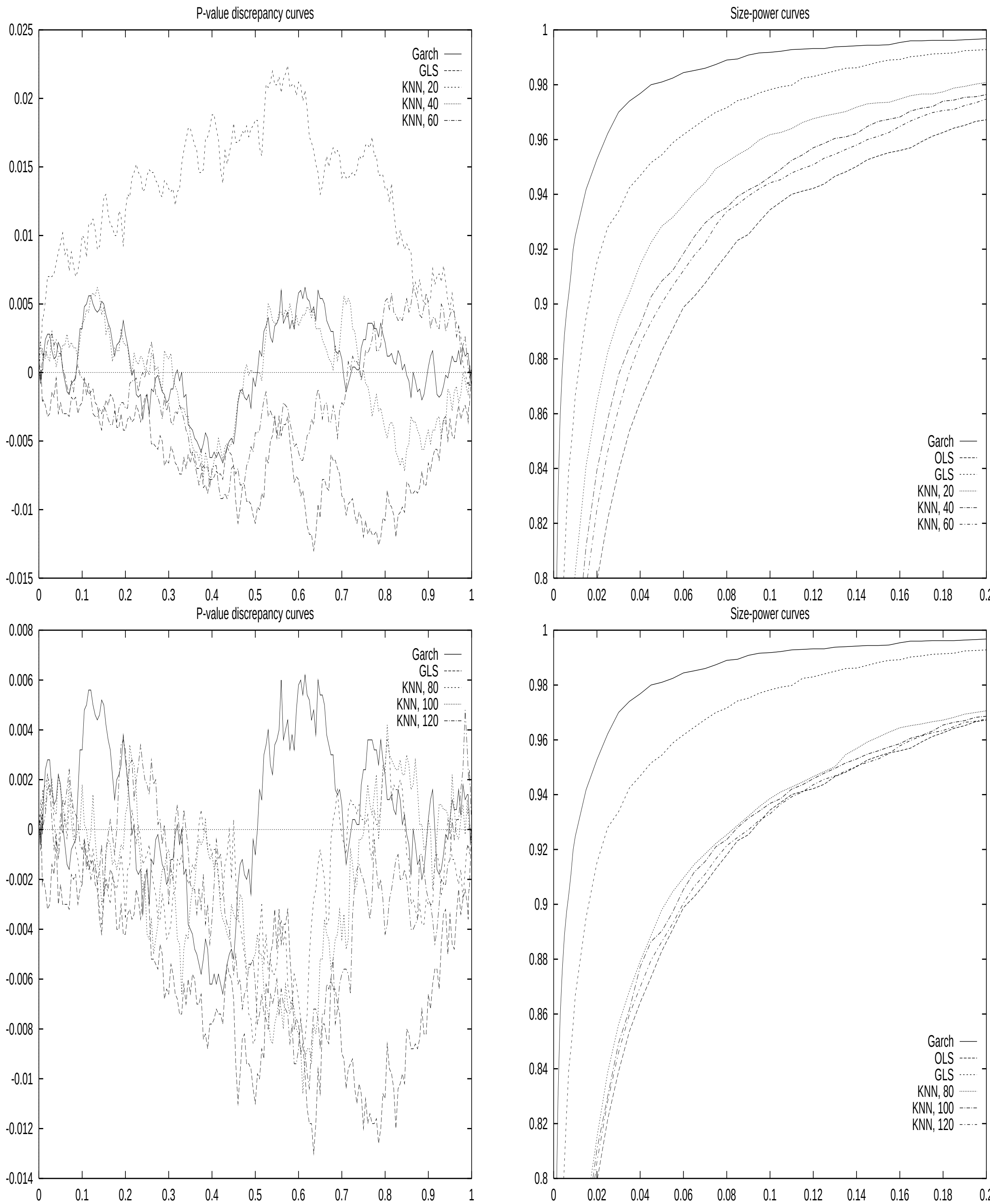

Figure 7: Regression on grid, ARCH(1) data, $n = 250$, $N = 1000$, $H1 : \beta = 0.5$.

4 Conclusion

In this paper, we have proposed a new semiparametric estimator for regression models with autoregressive conditional heteroskedasticity. The method was inspired by the result of Robinson (1987) for models with heteroskedasticity determined by exogenous variables. However, when the conditional skedastic function depends on lagged squared error terms, the parameters of the regression function affect the skedastic function as well, through the squared residuals. Thus we are unable to obtain a result as strong as that of Robinson, who shows that his semiparametric procedure is fully asymptotically efficient. Our benchmark in the Monte Carlo studies has therefore been, not maximum likelihood with a correctly specified model, but rather the somewhat less efficient infeasible GLS. At least for the cases considered in the Monte Carlo, namely GARCH(1,1) and ARCH(1), infeasible GLS was seen to be significantly, but not greatly, less efficient than ML.

Two distinctly different approaches to the nonparametric modelling of the conditional skedastic function are proposed. In the first, the conditional variance is regressed nonparametrically on lagged squared OLS residuals, as a proxy for the unobservable lagged squared errors. In the second, the method used is basically a smoothing procedure: the conditional variance is modelled by a smooth of the squared OLS residuals treated simply as a function of time.

The two methods give interestingly different results. When used with data generated by a GARCH(1,1) process, the regression on lagged squared errors gives a very mediocre fit of the true skedastic function, and consequently a relatively inefficient estimator of the parameter of the regression function. On the other hand, the smoothing procedure – our regression "on a grid" – gives excellent results, to the extent that in our experiments the efficiency achieved appeared to exceed that of infeasible GLS very slightly. The regression on lagged squared errors worked very well on data generated by an ARCH(1) process, in sharp contrast with what was observed with the GARCH data. This is no doubt due to the fact that the conditional skedastic function for an ARCH(1) process depends only on the first lags of the squared errors, so that the nonparametric regression is correctly specified, as it is not with GARCH. But the regression on a grid gives very disappointing results with ARCH data, presumably because smoothing, which involves averaging over a number of observations, is inappropriate for a conditional variance process which is autoregressive of order only one.

Despite the great differences in efficiency among the various semiparametric estimators considered, one finding remains constant. It is that all of these estimators are always at least as good as ordinary least squares. Thus there seems to be nothing to lose by using them, unless one has specific information on the nature of the conditional variance process that would, for instance, make it possible to formulate a correctly specified maximum likelihood model.

The results presented in this paper are clearly incomplete. The advantage of nonparametric methods is that they can adequately take account of nonlinear functional relations. But here we have looked only at ARCH and GARCH DGPs, in which the conditional variance depends linearly on lagged squared errors. It would be interesting to see if the semiparametric estimators perform as well if the conditional skedastic function is nonlinear.

There are many nonparametric regression techniques other than the kNN procedure. Robinson favoured this procedure because it simplified his proofs. Since our approach has been resolutely experimental, we have not (yet) provided any analytical results at all, and are thus free to investigate other nonparametric methods, such as kernel estimation or even smoothing by wavelets.

The disappointing performance of the lagged squared errors regression when used on GARCH data might well be substantially alleviated if the nonparametric regression used more than one lag of the squared residuals. Our use of only one was completely arbitrary, and it is entirely possible that, in practice, data-driven methods could be developed to select an approximately optimal number of lags. What seems clear already is that the nonparametric regression works much better when it is correctly specified.

Lastly, we have left open the question of how much cross-validation can help in choosing the parameter k used in the kNN procedure, or, more generally the window width used in the nonparametric regression. Our results serve only to illustrate that this problem merits further study; they are otherwise quite inconclusive. Despite the daunting amount of computer time it promises to use up, a study in which the cross-validation function is minimised on each Monte Carlo replication could be extremely illuminating.

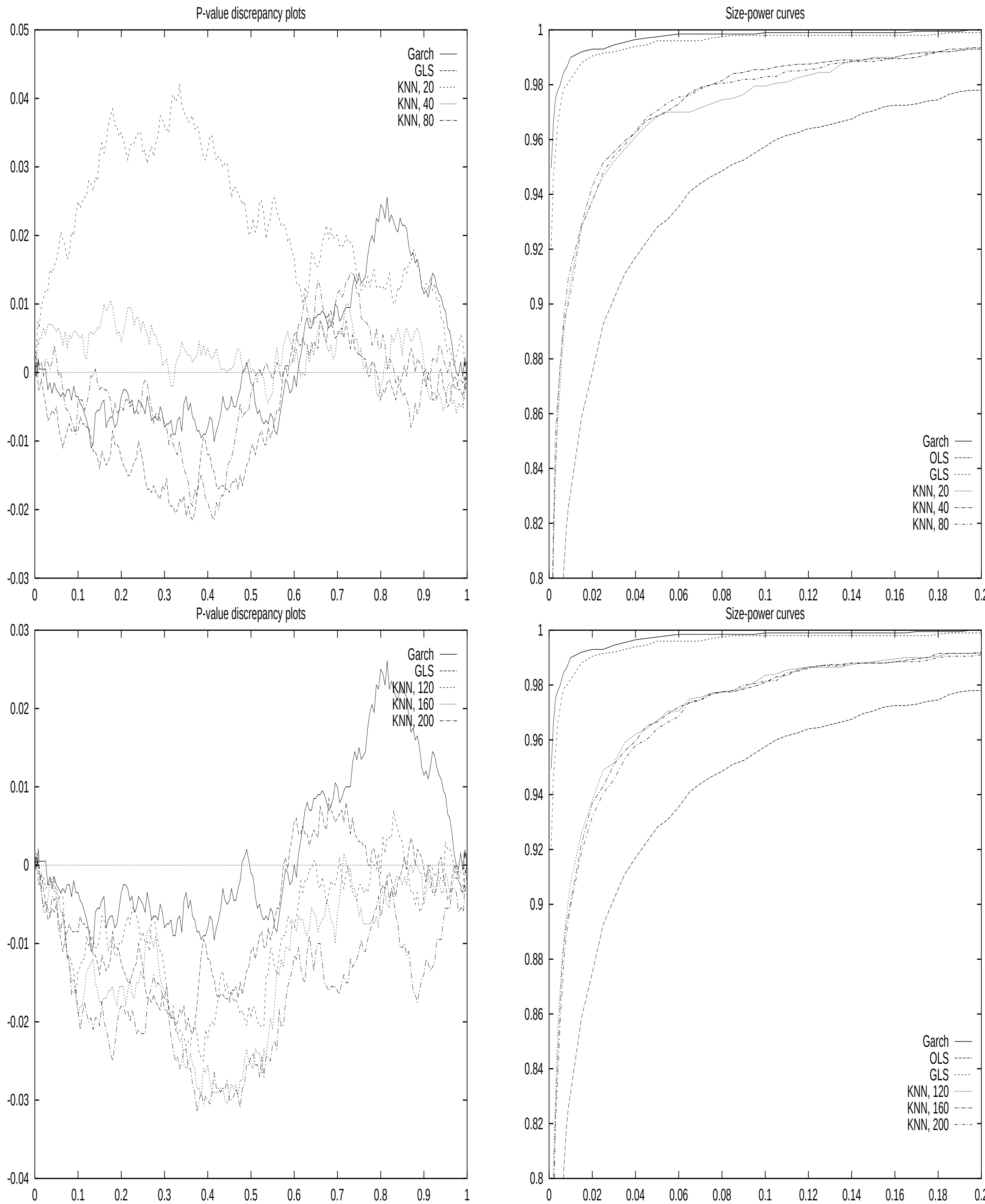

Figure 8: Regression on first lag, GARCH data, $n = 500$, $N = 2000$, $H1 : \beta = 0.25$.

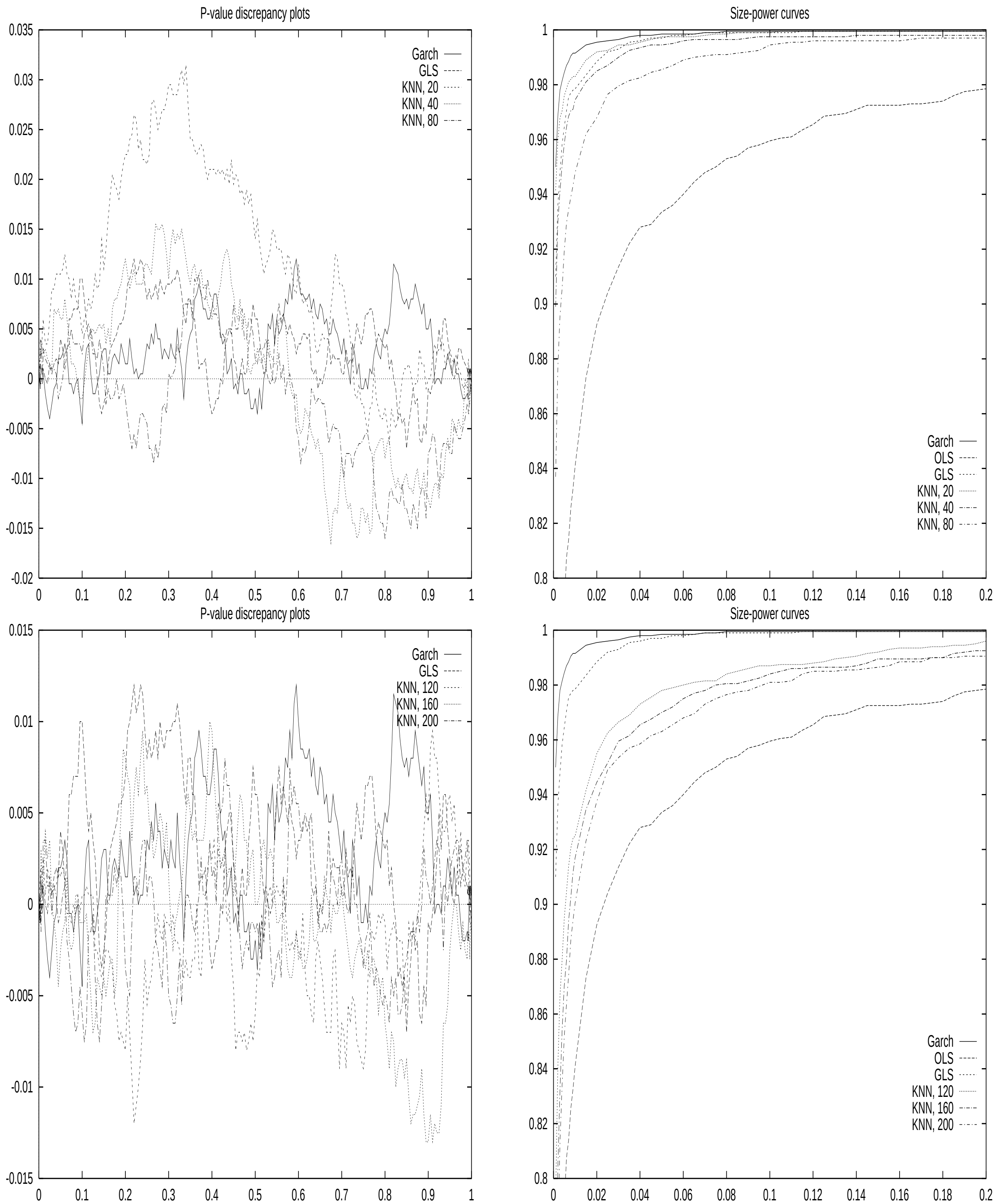

Figure 9: Regression on grid, GARCH data, $n = 500$, $N = 2000$, $H1 : \beta = 0.7$.

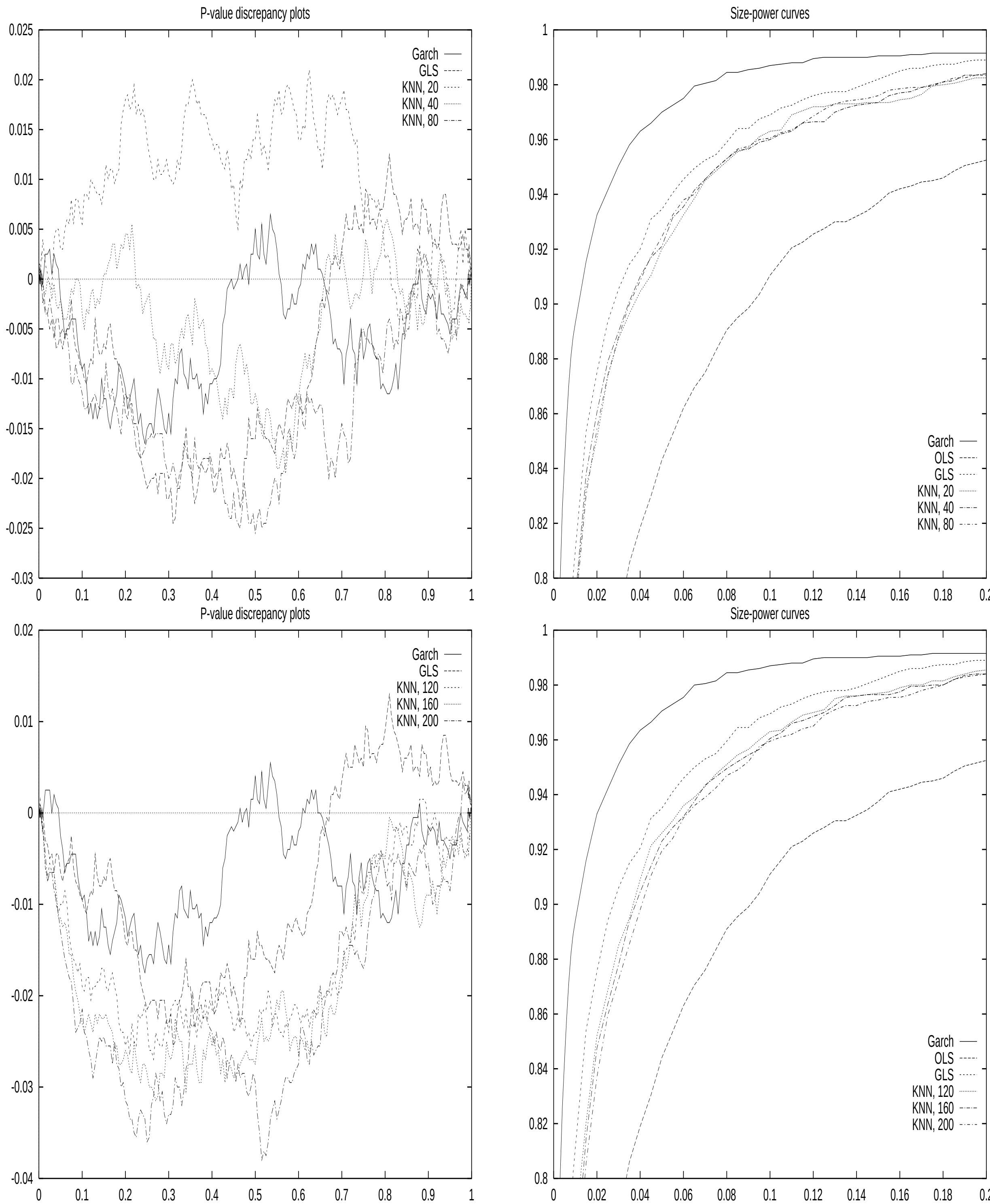

Figure 10: Regression on first lag, ARCH data, $n = 500$, $N = 2000$, $H1 : \beta = 0.25$.

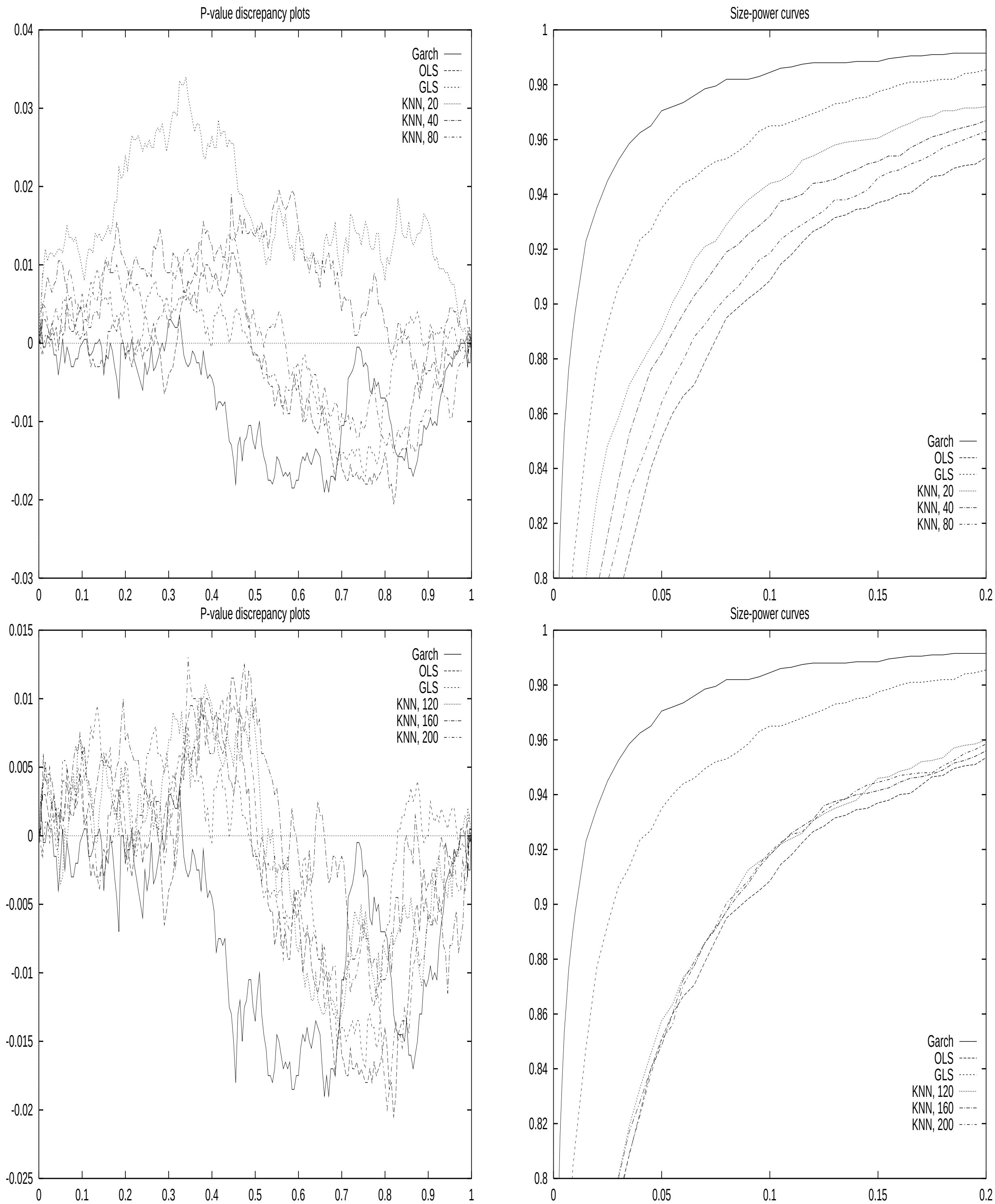

Figure 11: Regression on grid, ARCH data, $n = 500$, $N = 2000$, $H1 : \beta = 0.7$.

References

Bera, Anil K. and Matthew L. Higgins (1993). ARCH models: Properties, estimation and testing. *Journal of Economic Surveys 7(4)*, 305–366.

Berndt, Ernst K., Bronwyn H. Hall, Robert E. Hall, and Jerry H. Hausman (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement 4(3)*, 653–665.

Bollerslev, Tim (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 31*, 307–327.

Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner (1992). ARCH modelling in finance — A review of the theory and empirical evidence. *Journal of Econometrics 52(1-2)*, 5–59.

Bollerslev, Tim, Robert F. Engle, and Daniel B. Nelson (1994). ARCH models. In Robert F. Engle and Daniel McFadden (Eds.), *Handbook of Econometrics*, Volume IV. Amsterdam: North-Holland.

Davidson, Russell and James G. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Davidson, Russell and James G. MacKinnon (1994, September). Simulation methods for investigating the size and power of hypothesis tests. Discussion Paper 903, Queen's University Institute for Economic Research, Kingston, Ontario. Available as `ftp://qed.econ.queensu.ca/pub/faculty/mackinnon/graph.ps`.

Delgado, Miguel A. (1993). Computing nonparametric functional estimates in semiparametric problems. *Econometric Reviews 12(1)*, 125–128.

Engle, Robert F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50(4)*, 987–1007.

Friedman, Jerome H., Forest Baskett, and Leonard J. Shustek (1975). An algorithm for finding nearest neighbors. *IEEE Transactions on Computers C-24*, 1000–1006.

Guillaume, Dominique M., Michel M. Dacorogna, Rakhal R. Davé, Ulrich M. Müller, Richard B. Olsen, and Olivier V. Pictet (1994, December). From the bird's eye to the microscope: A survey of new stylized facts of the intra-day foreign exchange markets. Discussion Paper DMG.1994-04-06, Olsen & Associates, Zurich.

Guillaume, Dominique M., Olivier V. Pictet, and Michel M. Dacorogna (1995, March). On the intra-day performance of garch processes. In *First International Conference on High Frequency Data in Finance*, Volume 3, Zurich. Olsen & Associates.

Härdle, Wolfgang (1990). *Applied Nonparametric Regression*. Econometric Society Monographs No. 19. Cambridge: Cambridge University Press.

Härdle, Wolfgang and Oliver Linton (1994). Applied nonparametric methods. In Robert F. Engle and Daniel McFadden (Eds.), *Handbook of Econometrics*, Volume IV. Amsterdam: North-Holland.

Moshier, Stephen Lloyd Baluk (1989). *Methods and Programs for Mathematical Functions*. Ellis Horwood Series in Mathematics and its Applications. Chichester: Ellis Horwood. The routines from this book are availabe at `ftp://netlib.att.com/cephes/`.

Robinson, Peter M. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica 55(4)*, 875–891.