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

Russell Davidson
Queen's and GREQAM

Dirk Eddelbüttel
GREQAM

Canadian Econometric Study Group Meeting
McGill University, 23 September 1995

Introduction

Robinson (1987) considered the model

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + \epsilon_t \tag{1}$$
$$E(\epsilon_t^2) = f(\boldsymbol{Z}_t) \quad \text{and} \quad E(\epsilon_i\epsilon_j) = 0 \quad \text{for } i \neq j$$

with heteroskedasticity dependent on $\boldsymbol{Z}_t$, usually a subset of the regressors $\boldsymbol{X}_t$.

When 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 is known, its unknown parameters can be estimated consistently by regressing the squared residuals from (1) on the $\boldsymbol{Z}_t$ and feasible generalised least squares (FGLS) can be used to obtain an asymptotically efficient estimate of $\boldsymbol{\beta}$.

Robinson obtains asymptotic efficiency in estimating $\boldsymbol{\beta}$ using nonparametric estimates from a k nearest neighbours regression (kNN) of the squared OLS residuals on the $\boldsymbol{Z}_t$.

We apply this method to models with generalized autoregressive conditionally heteroskedastic (GARCH) error terms.

We consider the GARCH model of Bollerslev (1986)

$$\epsilon_t = y_t - \boldsymbol{X}_t \boldsymbol{\beta} \qquad \text{and} \qquad \epsilon_t | \Psi_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j} \equiv \boldsymbol{z}_t' \boldsymbol{\omega}$$

$$\boldsymbol{z}_t' = \left[1 \;\; \epsilon_{t-1}^2 \;\; \ldots \;\; \epsilon_{t-q}^2 \;\; h_{t-1} \;\; \ldots \;\; h_{t-p}\right]$$

$$\boldsymbol{\omega}' = [\alpha_0 \;\; \alpha_1 \;\; \ldots \;\; \alpha_q \;\; \gamma_1 \;\; \ldots \;\; \gamma_p]$$

$$\boldsymbol{\theta}' = \left[\boldsymbol{\beta}' \;\; \boldsymbol{\omega}'\right].$$

The contribution to the log-likelihood is

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

with derivatives

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

where

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

and

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

where

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

ML estimation of a correctly specified GARCH model is more efficient than any form of GLS, even asymptotically, as GLS ignores the dependence of h_t on $\boldsymbol{\beta}$.

GARCH models can be criticized for

- their messy asymptotics
- their parameter restrictions
- their *ad hoc* nature

We follow Robinson (1987) in modelling the variance by an auxiliary nonparametric regression. The square roots of the fitted values from this regression are used as weights in a feasible GLS estimation of the original model.

We propose two methods for the kNN regression.

1. The explanatory variables is regressed on the first lag of these same squared residuals. This is a nonparametric version of an ARCH(1) estimation procedure.

2. We regress the squared residuals on a grid of equidistant values. This is the "standard" nonparametric smoothing.

Monte Carlo Results

We use two different DGPs, a GARCH(1,1)

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

and an ARCH(1)

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

We use $n = 250$ observations and $N = 5000$ replications and concentrate on the

We use five different estimators

- maximum likelihood for either a GARCH(1,1) or an ARCH(1) using the analytical gradient and an expected Hessian

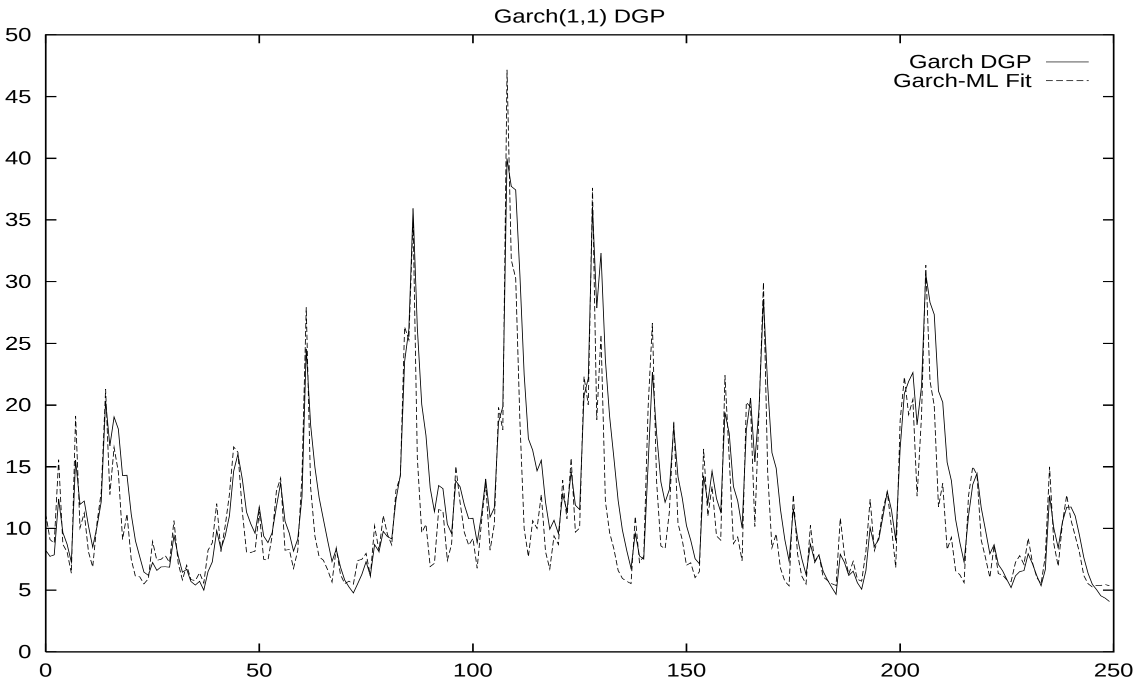
- infeasible GLS *à la* Aitken by premultiplying with the true conditional variance $h_t^{-1/2}$

- feasible GLS using a kNN regression of squared residuals on their first lag

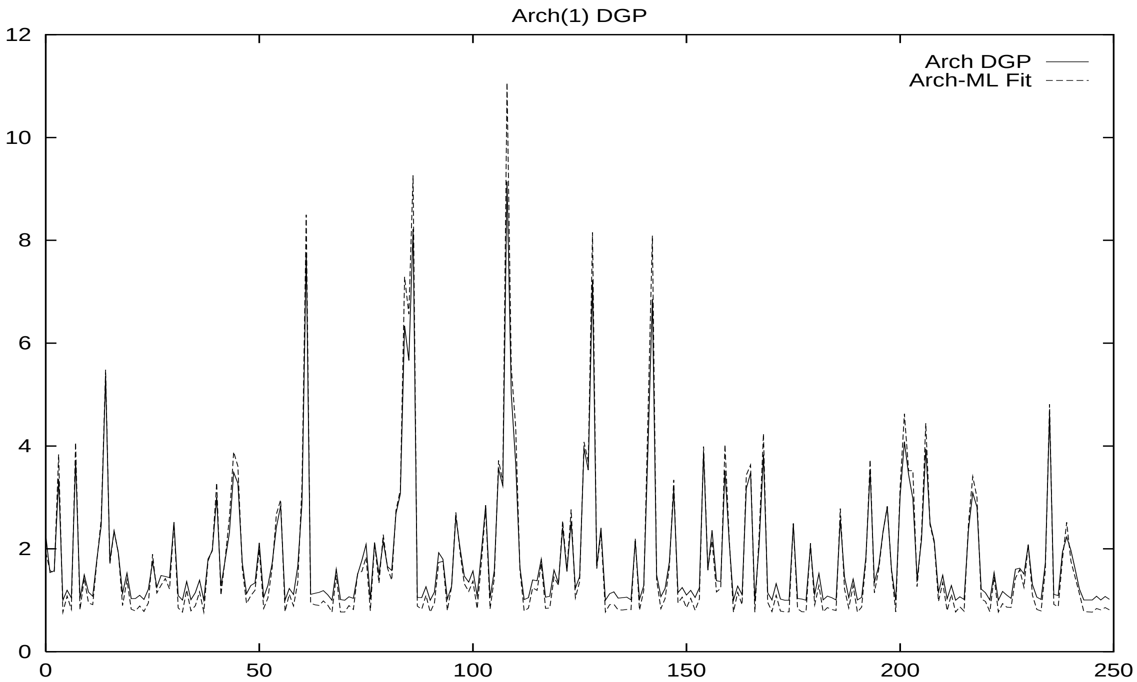
- feasible GLS using a kNN regression of squared residuals on an equidistant grid

- OLS since both the ARCH and GARCH DGPs have *unconditionally* homoskedastic error terms

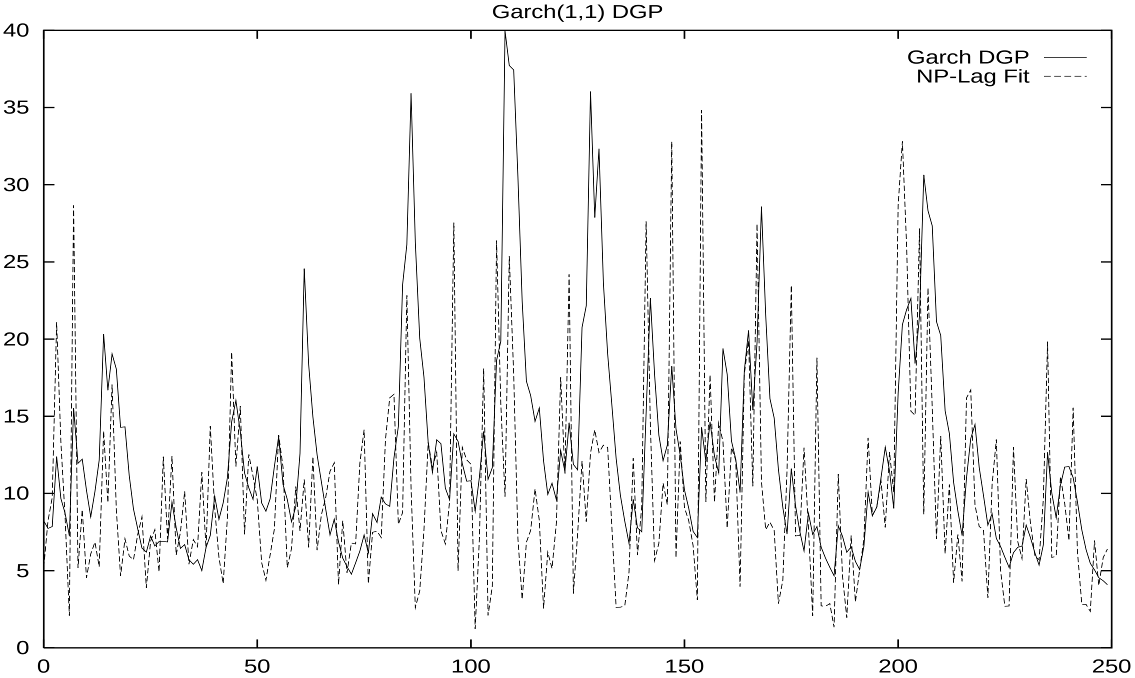
ML fitted values to GARCH(1,1)



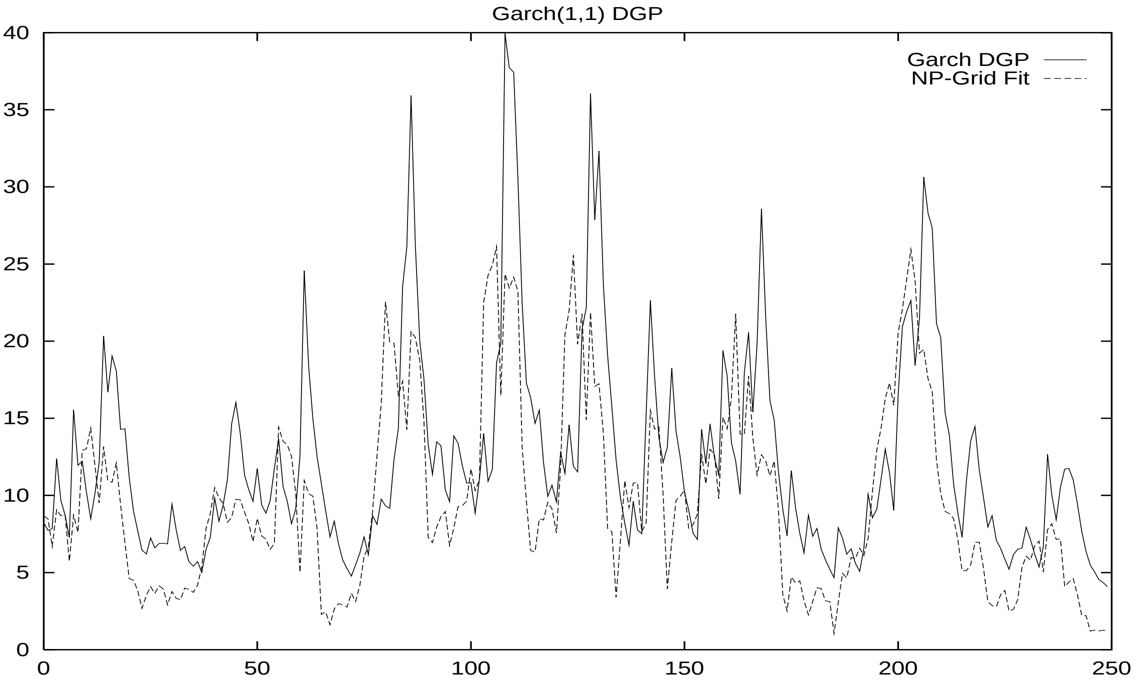
ML fitted values to ARCH(1)



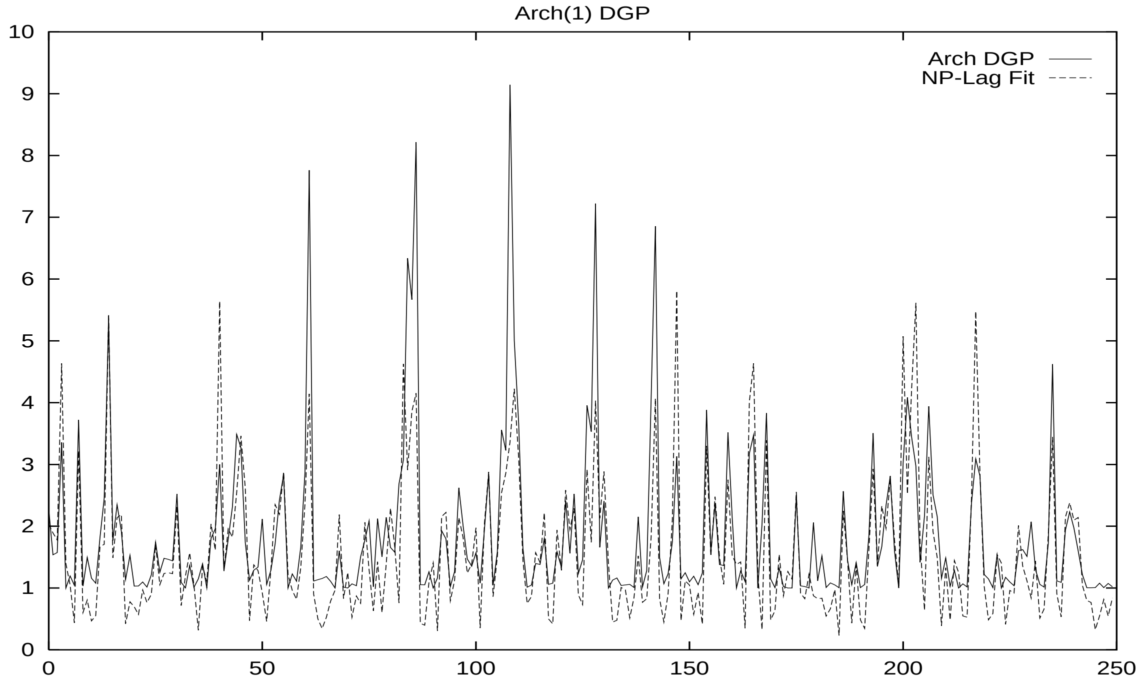
kNN-lag/GLS fitted values to GARCH(1,1)



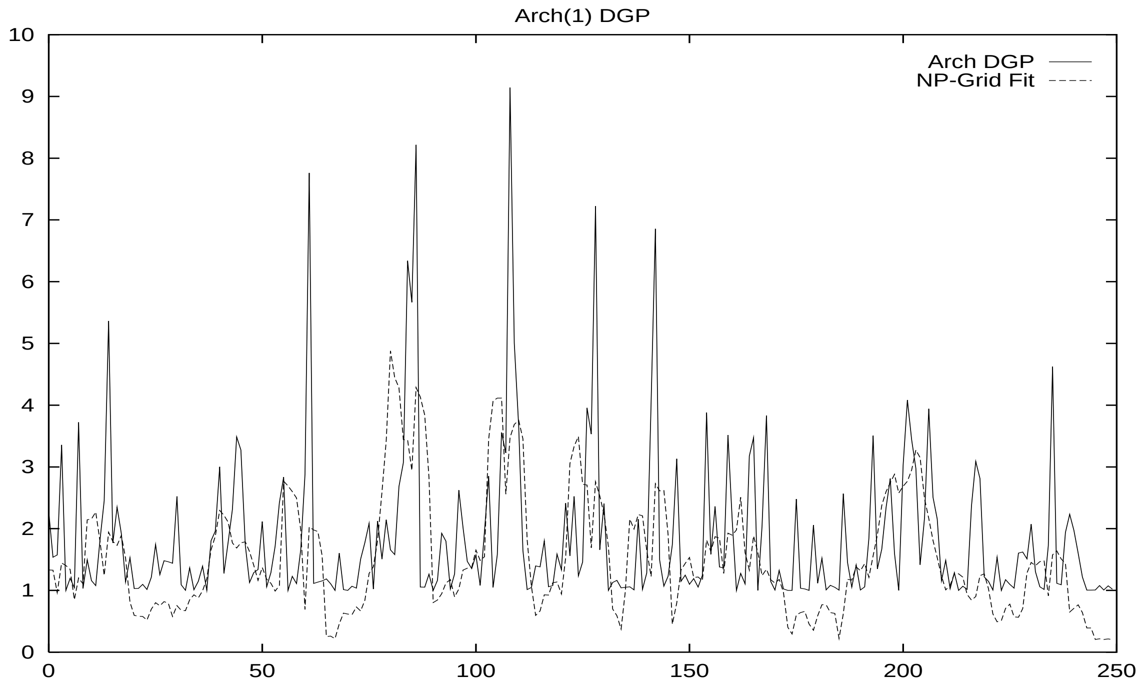
kNN-grid/GLS fitted values to GARCH(1,1)



kNN-lag/GLS fitted values to ARCH(1)



kNN-grid/GLS fitted values to ARCH(1)



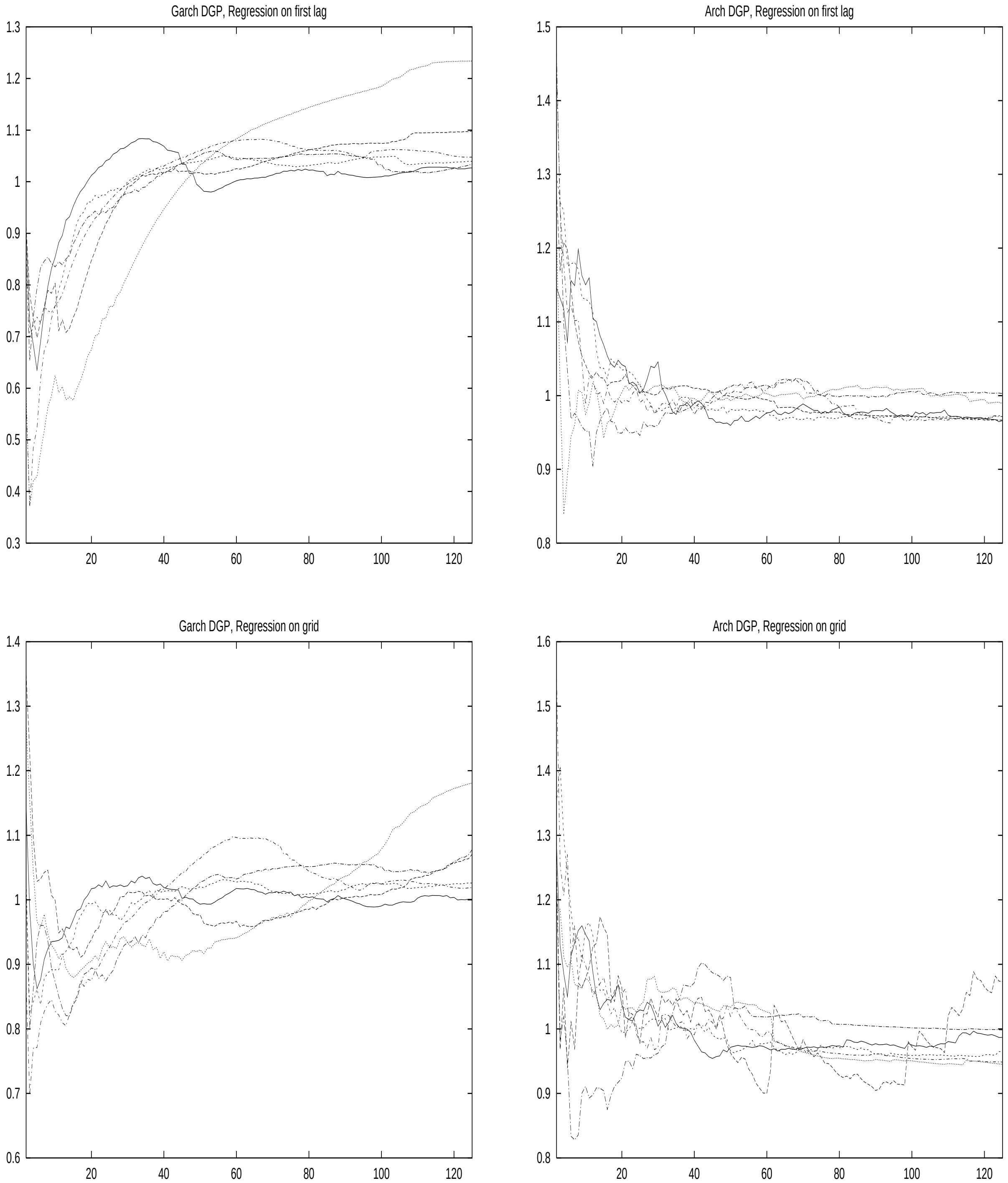
A fundamental problem with nonparametric estimation is the selection of the parameter that determines the “degree of localness” for the smoothing. Few theoretical results are available concerning the best choice of this parameter.

Cross-validation (CV) is commonly used as a data-driven mechanism for finding “good” parameter values.

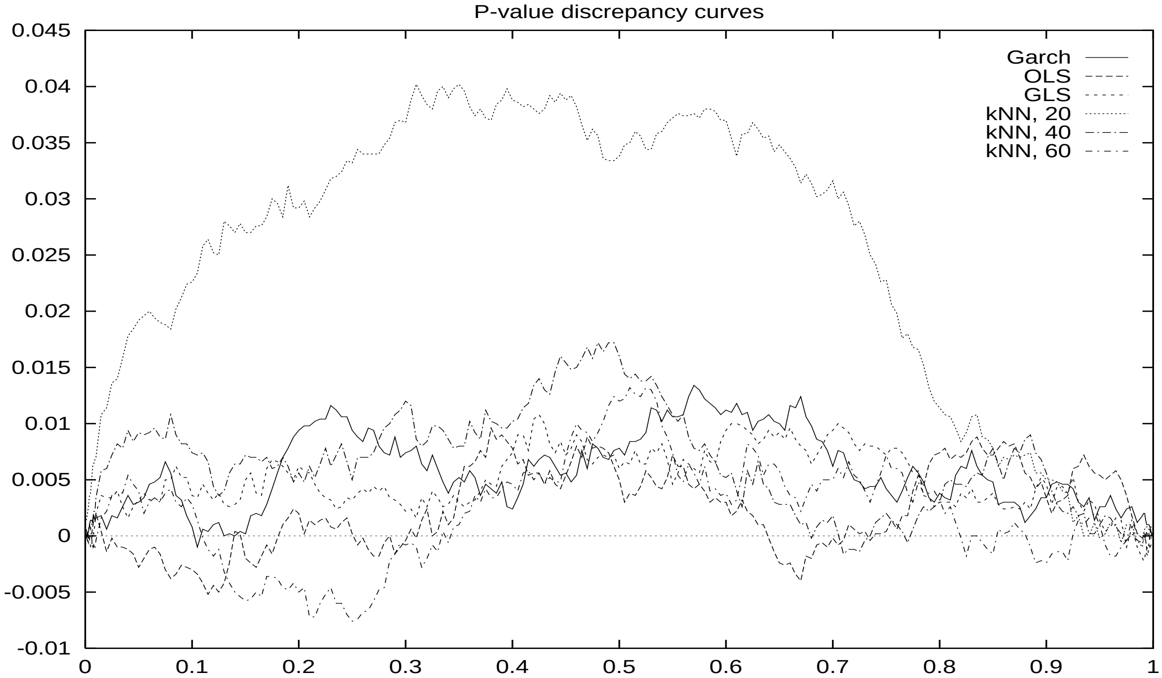
Do different realisations of the same DGP lead to cross-validation functions that are minimised by dramatically different values of k?

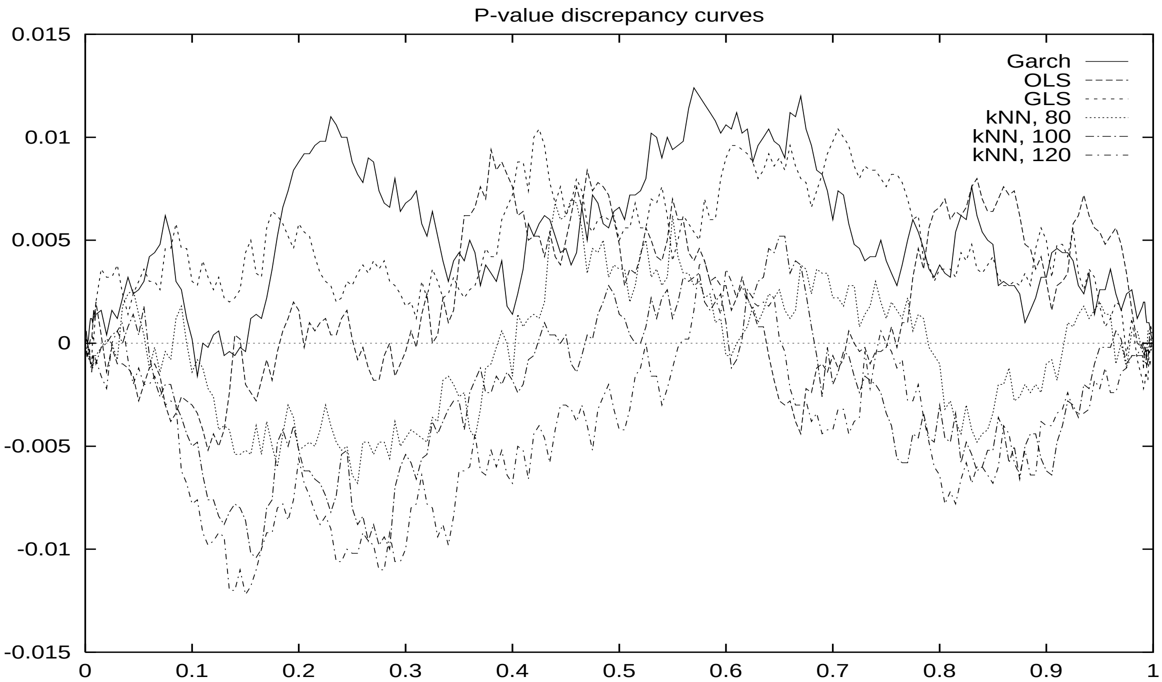
We calculated cross-validation curves for a small number of independent realisations of each DGP. The cross-validation functions are scaled appropriately so as to facilitate comparison.

Cross-validation functions

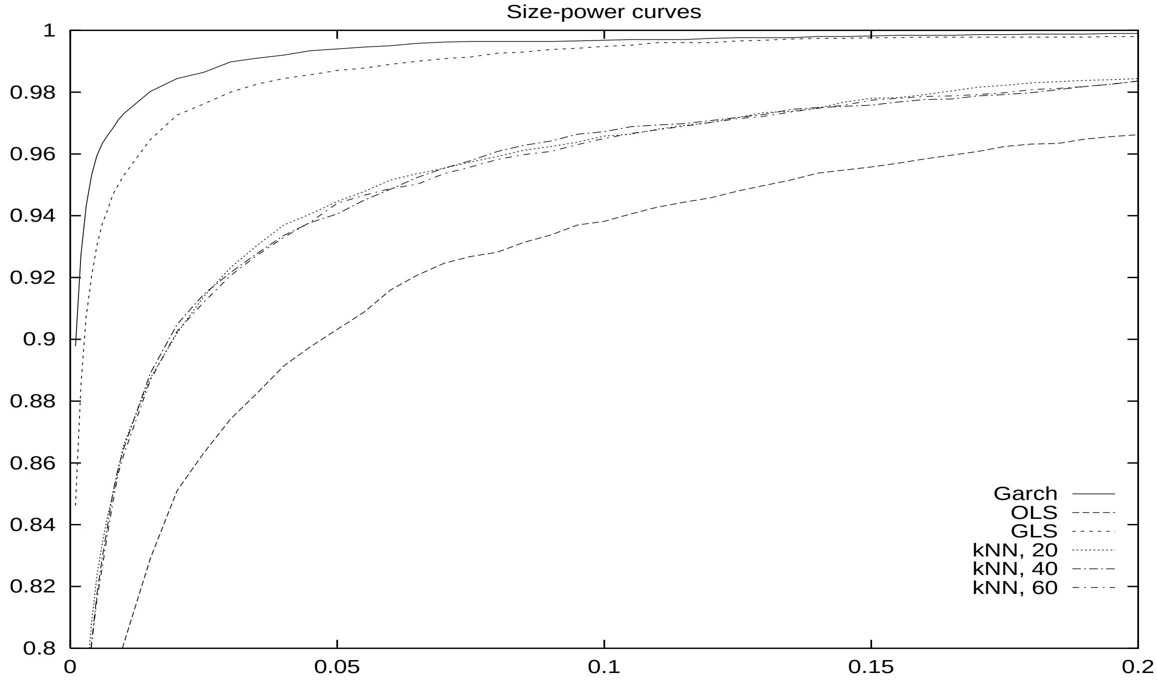


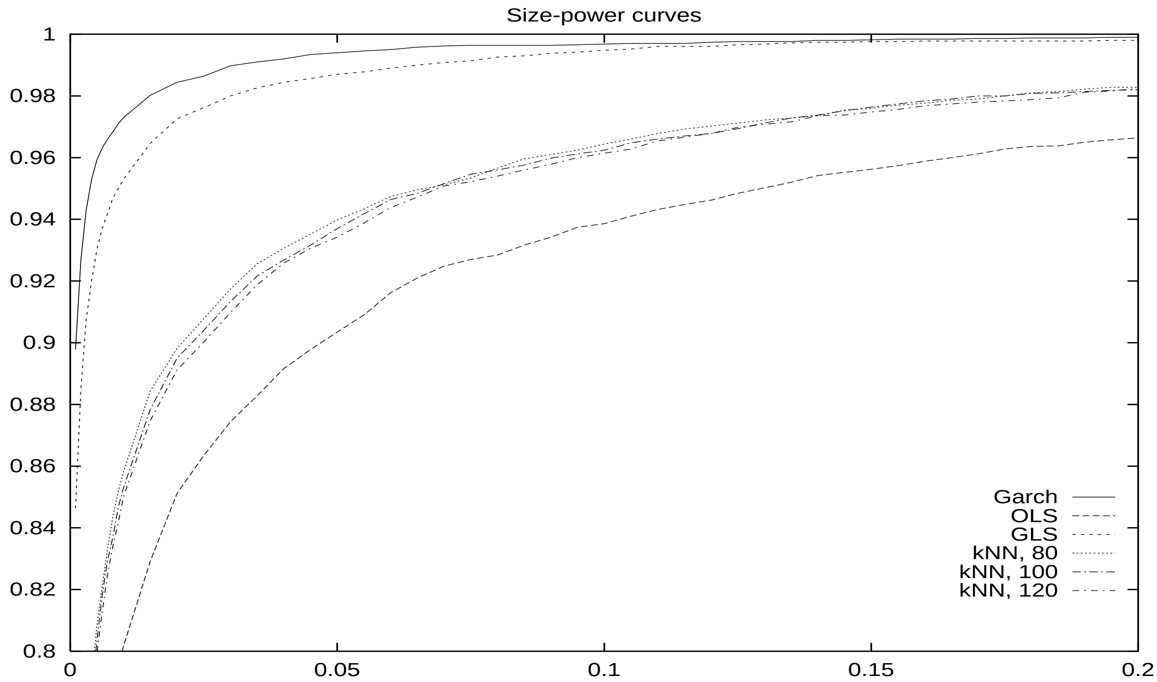
GARCH data; lagged squared errors.



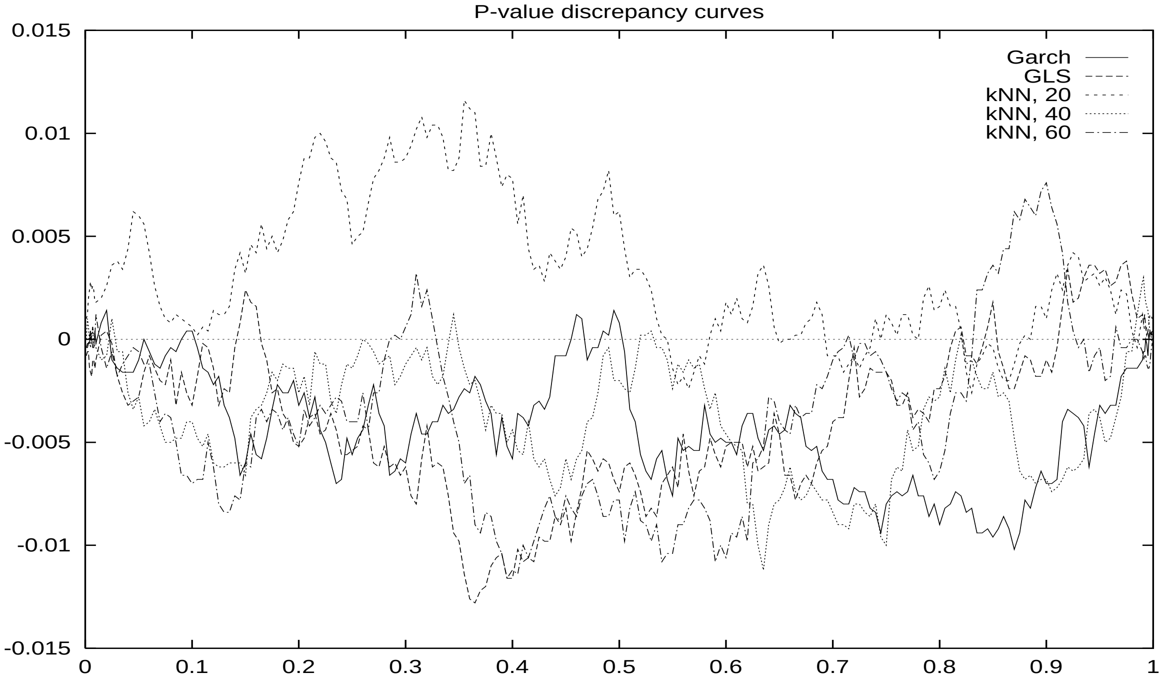


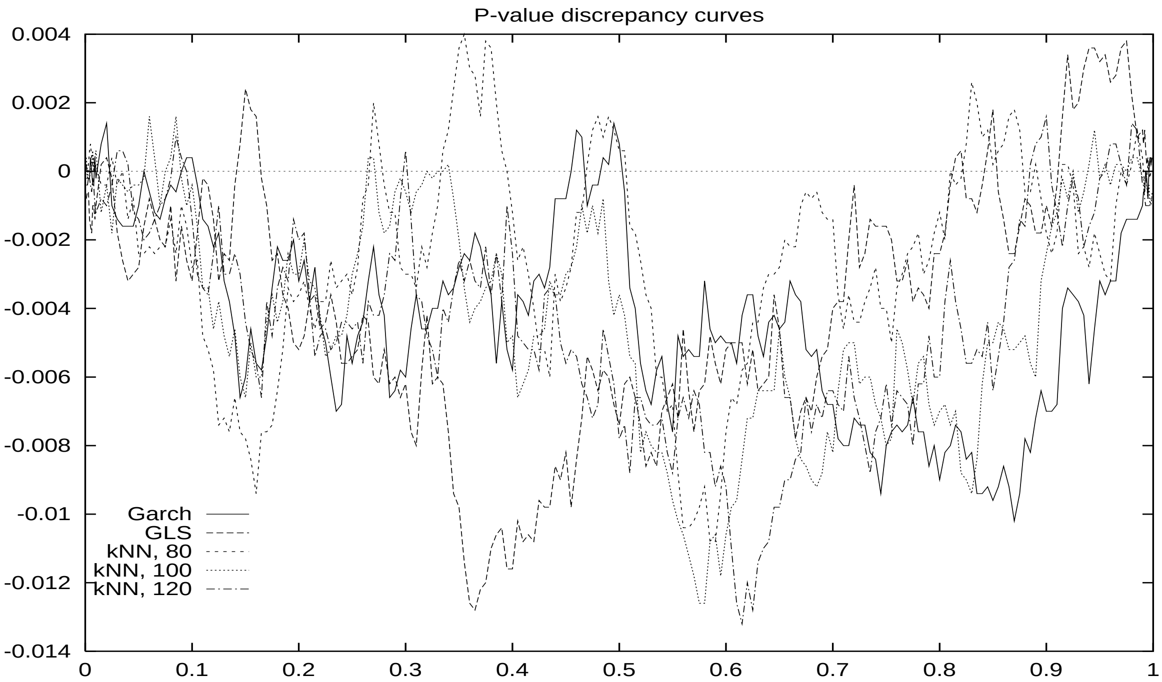
GARCH data; lagged squared errors.



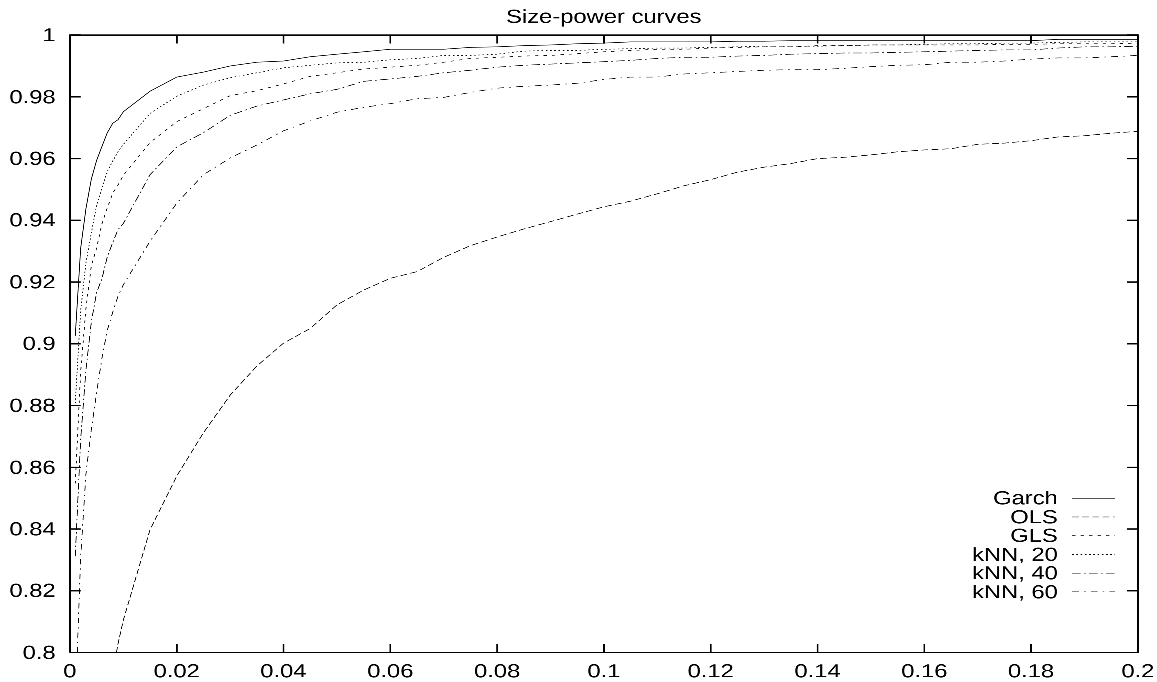


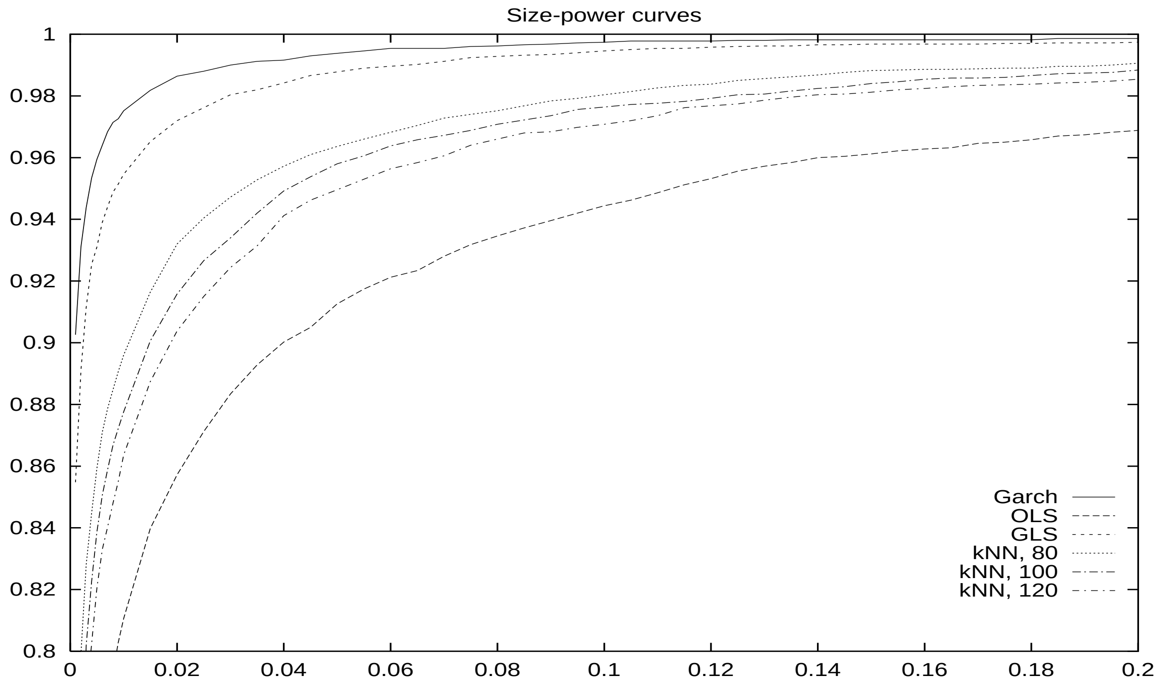
GARCH data; grid



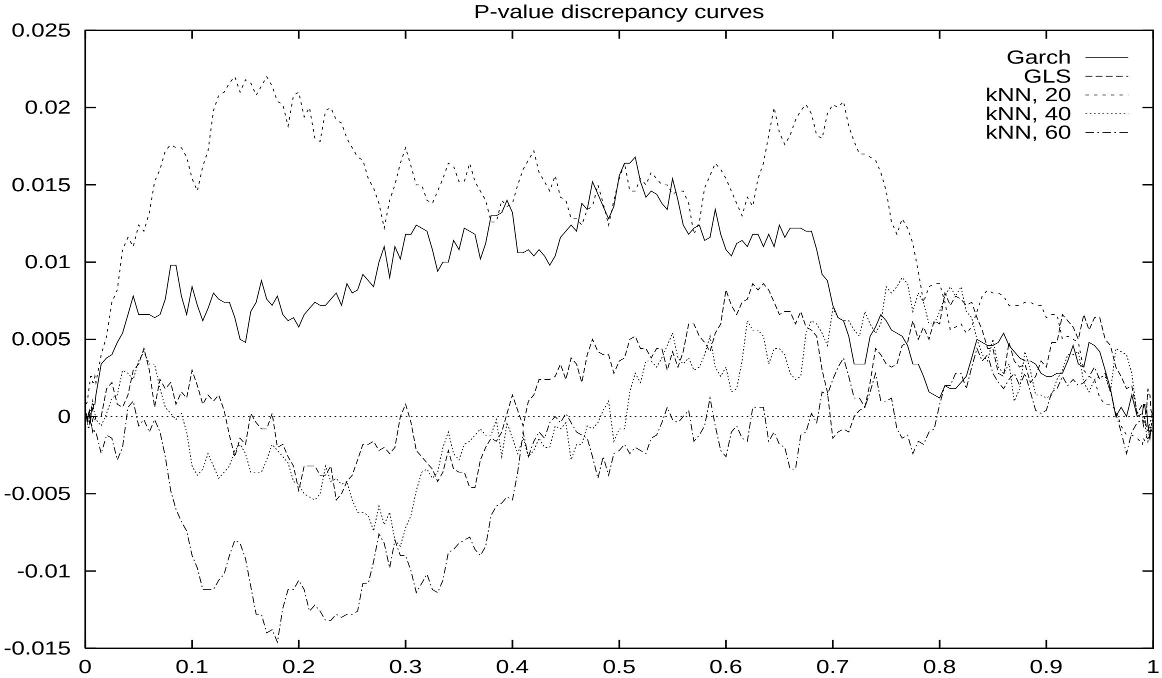


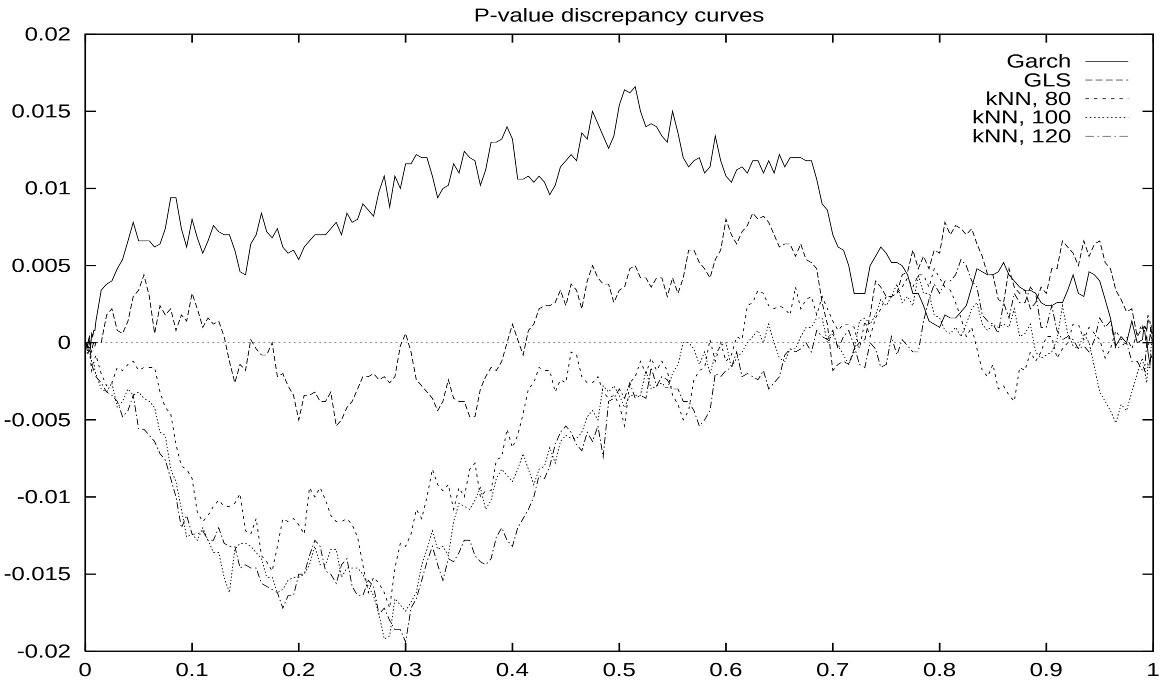
GARCH data; grid



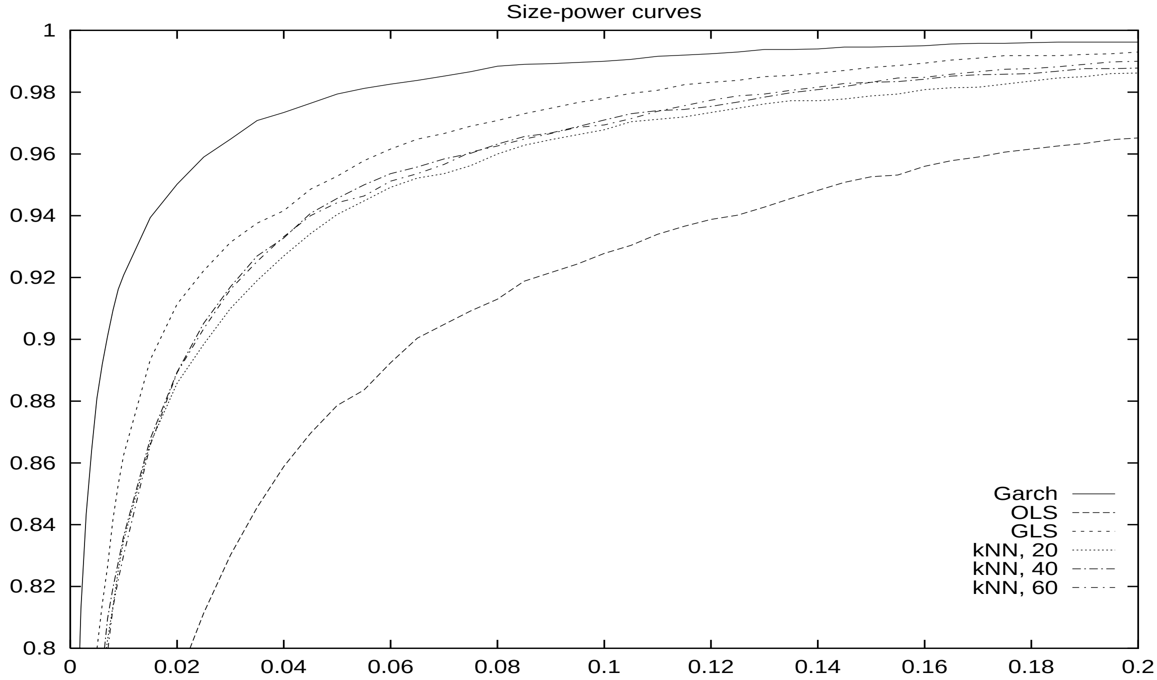


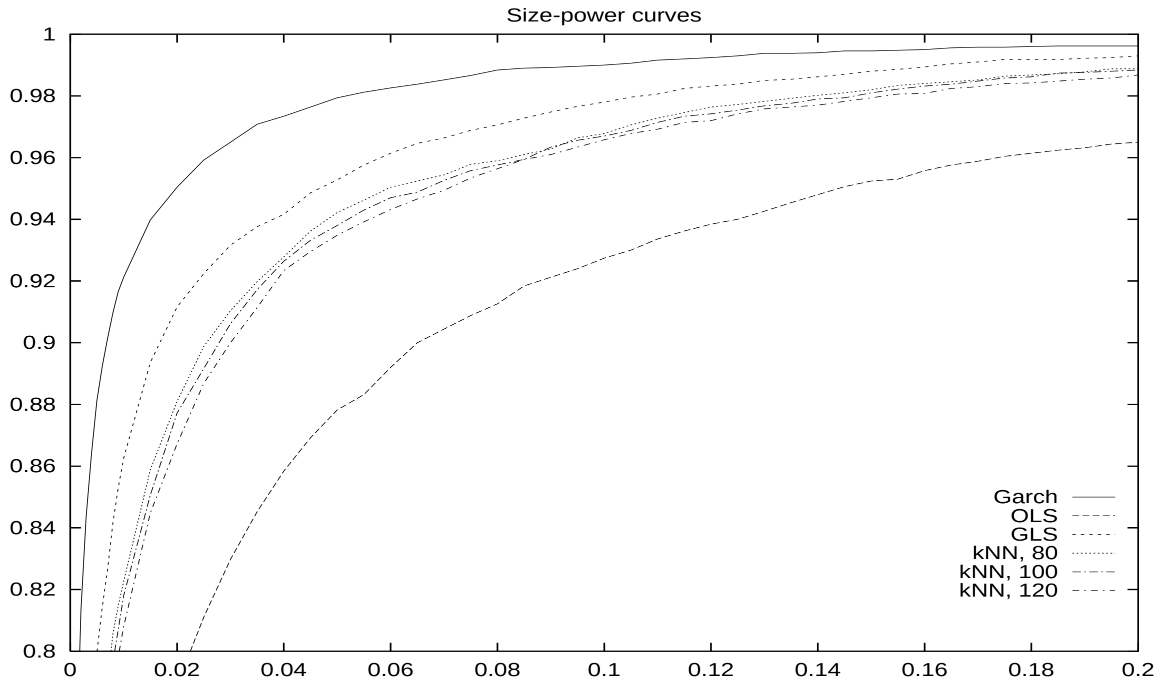
ARCH data; lagged squared errors



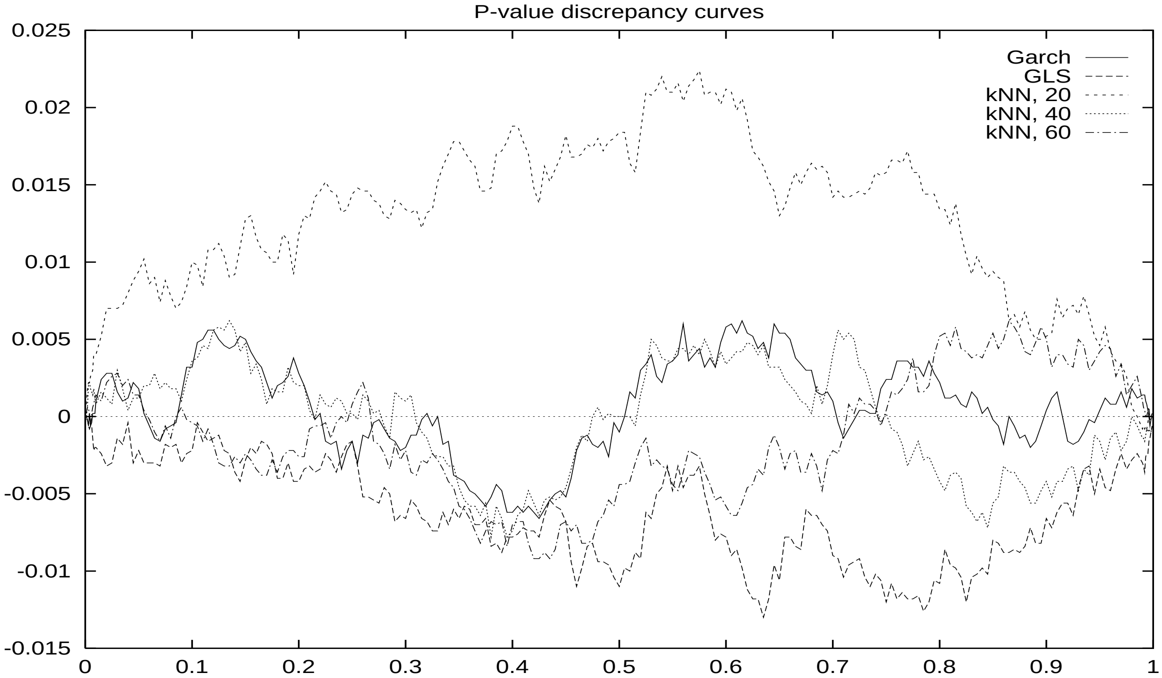


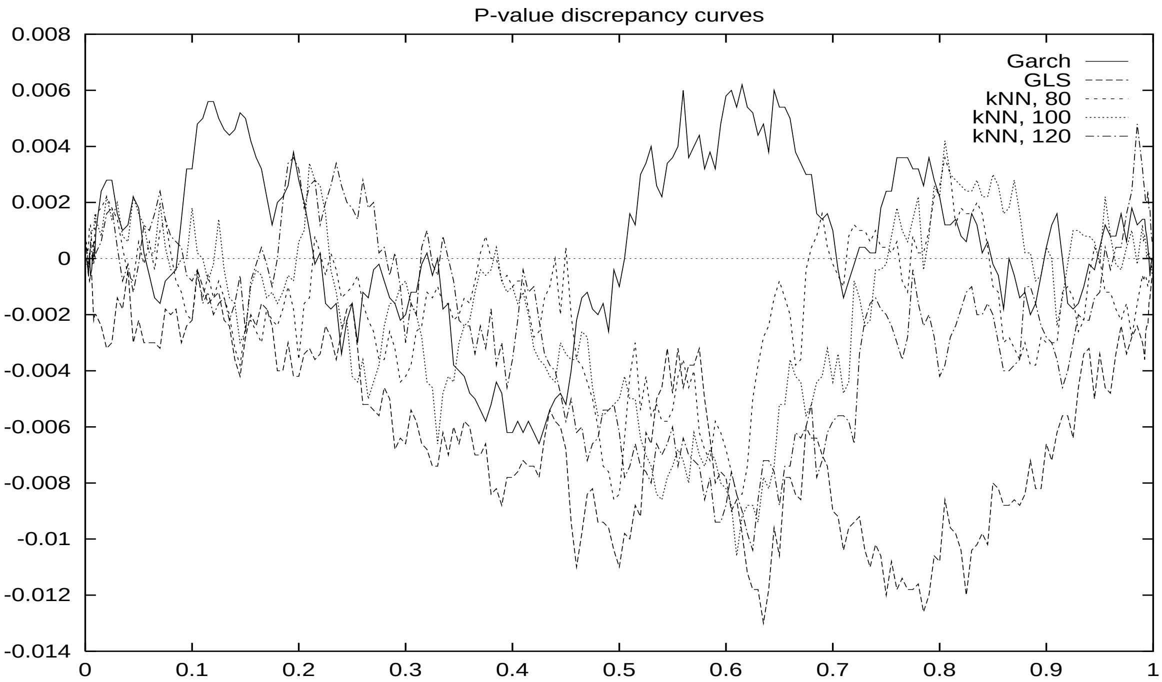
ARCH data; lagged squared errors



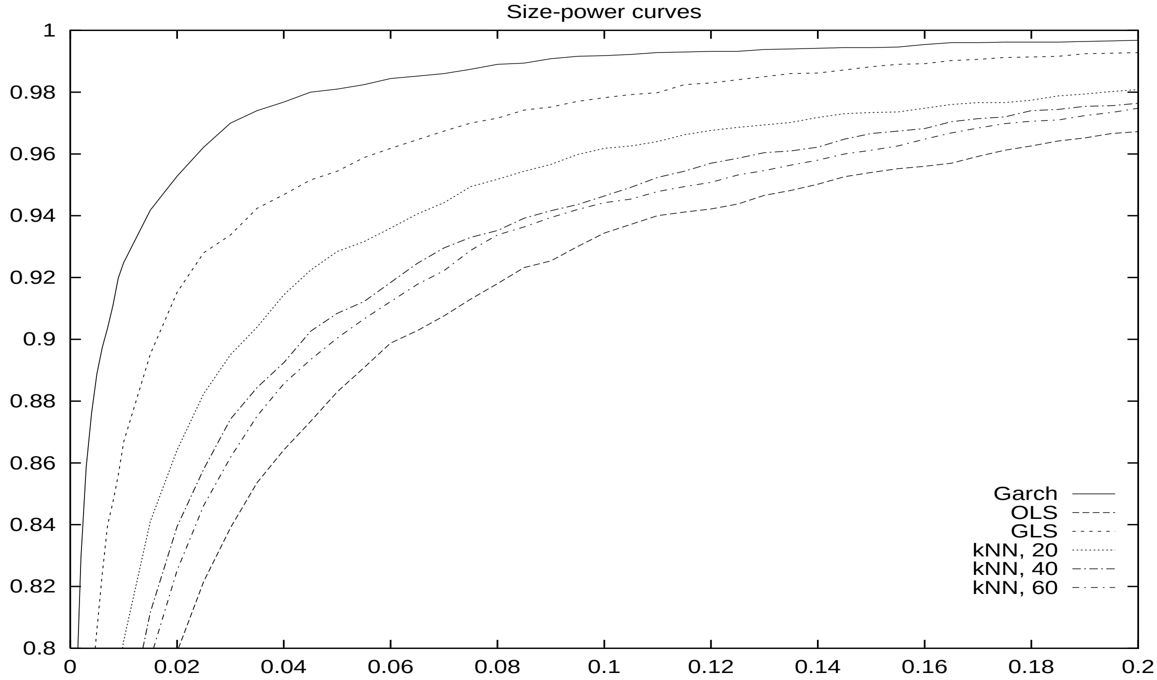


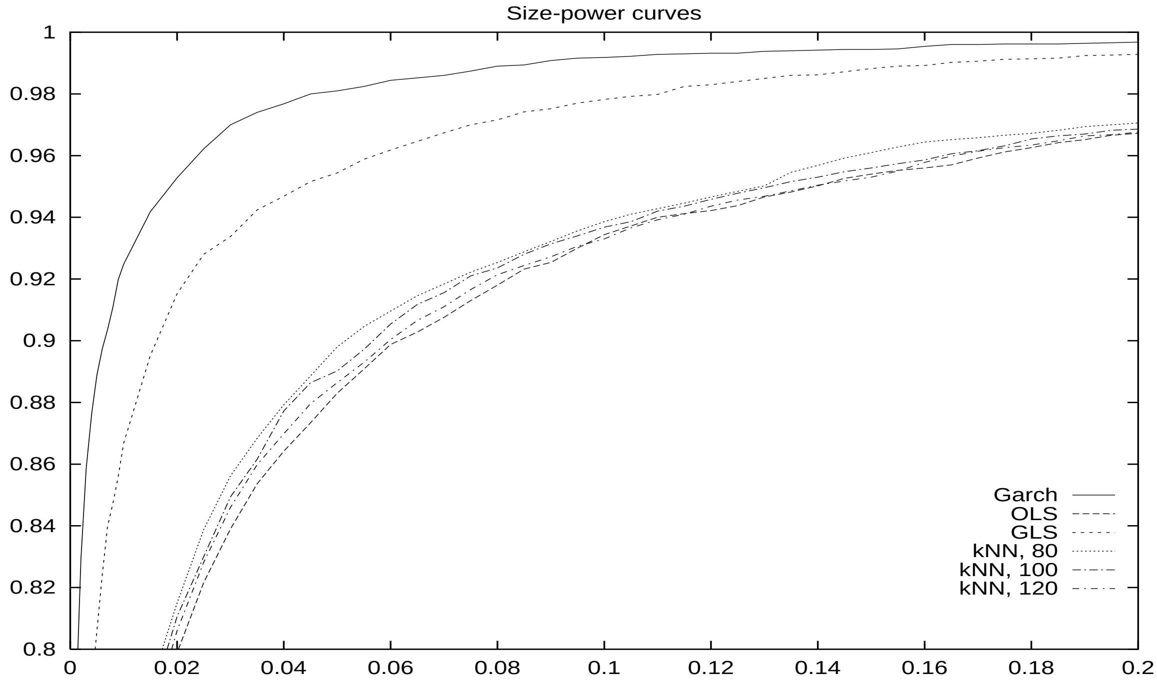
ARCH data, grid





ARCH data, grid





Conclusion

- The skedastic function is affected by the regression function parameters through the lagged squared errors. We cannot obtain full asymptotically efficiency.

- Our benchmark in the Monte Carlo studies is the less efficient infeasible GLS that was seen to be less efficient than ML

- Two different approaches are proposed. The conditional variance is regressed nonparametrically on lagged squared residuals. Alternatively, the conditional variance is modelled by a smooth of the squared residuals.

- With GARCH(1,1) data, the 'lagged squared errors' method is relatively inefficient. The smoothing procedure gives excellent results.

- With ARCH(1) data, the regression on lagged squared errors worked very well but the 'grid' regression gives very disappointing results.

- There is nothing to lose: both methods are always at least as good as ordinary least squares.

Future Research

- The advantage of nonparametric methods is that they can take account of nonlinear functional relations. We will see how the semiparametric estimators perform if the conditional skedastic function is nonlinear.

- We will use more lags of the squared residuals as the nonparametric regression seems to work much better when it is correctly specified.

- We will look at how much cross-validation can help in choosing the parameter k. A study in which the cross-validation function is minimised on each Monte Carlo replication could be extremely illuminating.