

Lucchetti, Riccardo; Rossi, Eduardo

Working Paper

Artificial regression testing in the GARCH-in-mean model

Quaderni di Dipartimento - EPMQ, No. 163

Provided in Cooperation with:

University of Pavia, Department of Economics and Quantitative Methods (EPMQ)

Suggested Citation: Lucchetti, Riccardo; Rossi, Eduardo (2004) : Artificial regression testing in the GARCH-in-mean model, Quaderni di Dipartimento - EPMQ, No. 163, Università degli Studi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi (EPMQ), Pavia

This Version is available at:

<https://hdl.handle.net/10419/87148>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Artificial Regression Testing in the GARCH-in-mean model

Riccardo Lucchetti
(Università di Ancona)

Eduardo Rossi
(Università di Pavia)

163 (03-04)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Marzo 2004

Artificial Regression Testing in the GARCH-in-mean model*

Riccardo Lucchetti
University of Ancona

Eduardo Rossi
University of Pavia

23rd February 2004

Abstract

The issue of finite-sample inference in GARCH-like models has seldom been explored in the theoretical literature, although its potential relevance for practitioners is self-evident. In some cases, asymptotic theory may provide a very poor approximation to the actual distribution of the estimators in finite samples.

The aim of this paper is to propose the application of the so-called *double length regressions* (DLR) to GARCH-in-mean models for inferential purposes. As an example, we focus on the issue of Lagrange Multiplier tests on the risk premium parameter. Simulation evidence suggests that DLR-based LM test statistics provides a much better testing framework than OPG-based LM test statistics, which is commonly used, in terms of actual test size, especially when the GARCH process exhibits high persistence in volatility. This result is consistent with previous studies on the subject.

1 Introduction

A standard practice in applied econometrics is to check the adequacy of the estimated models. The hypothesis of no conditional heteroskedasticity, no error autocorrelation, linearity, and parameter constancy are tested using various methods and procedures. In the case of GARCH regression models, one of the key issues in model selection is to test for the absence of feedback effects from the conditional variance to the conditional mean. When the latter includes a conditional variance term, we have a GARCH-in-mean model.

*We are indebted to Gianni Amisano, Tim Bollerslev, Nunzio Cappuccio, Carlo Gianini and James G. MacKinnon for comments on a previous version of this article. Needless to say, the usual disclaimer applies.

An attractive feature of a GARCH-in-mean model is that it can be interpreted as a reduced form model, because theoretical models (see (17)) make the market return a function of volatility, i.e. the risk premium should be larger when the asset return is more volatile. Moreover, for an agent making decisions at time t , the appropriate concept of volatility is the conditional variance of the asset return over the holding period for the asset. If the conditional variance is parametrised as in a GARCH model, this leads to a GARCH-in-mean specification. A relevant problem in this context is to test for a GARCH model versus a GARCH-in-mean model, for a given conditional variance specification. In general, analysis of GARCH-in-mean models is much more complex than that of pure GARCH models (see (19)). In the latter, the residuals obtained from the conditional mean regression might be used for diagnostic checking and specification. However, in the GARCH-in-mean model it is impossible to estimate the disturbance term without first specifying a valid model for the conditional variance, so that pre-estimation investigation is very difficult. The usual practice relies on post-estimation investigation. A possible strategy to tackle with this problem is the LM test.

There are several asymptotically equivalent ways to compute LM-type tests; some of them require the explicit calculation of the Hessian matrix, while some others do not. Since the latter are often computationally more convenient, we will focus on two methods for computing LM-type tests which do not require the computation of the Hessian, known as the OPG (Outer Product of Gradients Regression) and the DLR (Double-Length Regression) methods (see (3; 4; 6)).

OPG regressions were first used as a computational shortcut in the BHHH algorithm for maximising likelihoods, but were later also employed in the computation of LM-type statistics ((11)). The Double-Length Regression is less known. The DLR is computational device that can be used to evaluate quantities related to maximum likelihood estimation and testing, such as the information matrix.

As is well known, asymptotic theory may provide a very poor approximation to the actual distribution of test statistics in finite samples. Since there is some evidence that the DLR performs much better than the OPG in many different settings, it is interesting to compare them on the ground of model selection in a GARCH setting. In this paper, therefore, we will compare, by means of a Monte Carlo experiment, the OPG- and DLR-based LM tests in testing hypotheses on the ‘risk-premium’ parameter of a GARCH-in-mean model. Such hypotheses are particularly relevant in empirical finance applications where this parameter has a structural interpretation ((9; 1; 20) refer to these models as models with risk terms). Since the two tests are asymptotically equivalent, the analysis will focus on their behaviour on sample sizes typical of ap-

plied finance work.

The plan of the paper is as follows: section 2 introduces briefly the concept of *artificial regressions* and their connection to Lagrange Multiplier tests; in section 3 the model that we will analyse is described and the formulae for the gradient are provided. In section 4 simulation results are shown and discussed. The conclusions, with some indications for extending the basic model, are given in section 6.

2 Artificial regressions

The concept of an *artificial regression* was introduced by Davidson and MacKinnon (1984). Given a parametric model characterized by an unknown vector of parameters θ which belongs to a parameter space ($\Theta \subseteq \mathbb{R}^k$) and which can be estimated by minimizing a criterion function $Q(\theta)$ using T observations. A generic artificial regression can be written as:

$$r(\theta) = R(\theta)b + \text{residuals}, \quad (1)$$

where b is a k -vector of coefficients. Residuals are used here as a neutral term to avoid any implication that (1) is a statistical model. For (1) to constitute an artificial regression, the vector $r(\theta)$ and the matrix $R(\theta)$ must satisfy certain defining properties. The two principal conditions are the following:

$$\begin{aligned} R(\theta)'r(\theta) &= \mathbf{g}(\theta) \\ \frac{1}{T}R(\bar{\theta})'R(\bar{\theta}) &\xrightarrow{P} \mathcal{I}(\theta) \end{aligned}$$

where $\bar{\theta}$ is any consistent estimator of θ , $\mathbf{g}(\cdot)$ is the score vector and $\mathcal{I}(\theta)$ is the information matrix; see (6).

The two artificial regressions that we consider are the OPG (Outer Product of Gradients) and the DLR (Double-Length Regression). The OPG regression is based on the fact that the information matrix is defined as the covariance matrix of the score vector. Given a sample of T observations, define $\ell_t(\theta)$ as the t -th contribution to the sample log-likelihood, where θ is parameter vector with k elements and \mathbf{G} is the gradient contribution matrix, ie a $T \times k$ matrix such that

$$\mathbf{G}_{ti}(\theta) = \frac{\partial \ell_t(\theta)}{\partial \theta_i}.$$

The Lagrange Multiplier (LM) test can be written in the so-called ‘score form’ as

$$LM = T^{-1}\mathbf{g}(\theta)'\mathcal{I}(\theta)^{-1}\mathbf{g}(\theta), \quad (2)$$

where

$$\mathbf{g}(\theta) = \mathbf{G}(\theta)' \iota$$

is the score vector for the log-likelihood summed over the whole sample and $\mathcal{I}(\theta)$ is the information matrix (ι is a vector of ones).

The OPG regression can be written as

$$\iota = \mathbf{G}b_O + \text{residuals.}$$

Let $\tilde{\theta}$ denote the constrained ML estimates obtained by imposing r restrictions when maximising the log-likelihood. Then the explained sum of squares (ESS) from the OPG regression

$$\iota = \tilde{\mathbf{G}}b_O + \text{residuals,}$$

where $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\theta})$ is the OPG form for the LM statistic, which is equal to T times the uncentered R^2 . In other words, the OPG-based LM test simply replaces the information matrix with its sample equivalent.

The OPG regression applies to a wide variety of models and requires only first derivatives. In general, however, both estimated covariance matrices and test statistics regression are not very reliable in finite sample. A large number of papers has shown that, in finite samples, LM tests based on the OPG regression tend to overreject the null hypothesis. Therefore, the double-length regression (DLR), which has been shown to have better finite sample properties, has been proposed as an alternative device.

The class of models to which the DLR applies may be written as

$$f_t(w_t, \theta) = u_t \quad t = 1, \dots, T \quad (3)$$

where $f_t(\cdot)$ is a smooth function that depends on the random vector w_t and on the parameter vector θ ; w_t contains the dependent variable y_t , and some exogenous and/or predetermined variables and/or lagged dependent variables x_t . Given a sequence of information sets \mathfrak{S}_t (which typically include x_t), it is not essential that $y_t|\mathfrak{S}_t$ follows the normal distribution, although it is essential that the model can be transformed so that (3) holds, and $f_t|\mathfrak{S}_t \stackrel{\text{iid}}{\sim} N(0, 1)$ under the null hypothesis.

The t -th contribution to the log-likelihood can be written as

$$\ell_t(\theta) = \text{const} - \frac{1}{2}f_t^2(\theta) + k_t(\theta), \quad (4)$$

where $k_t(\theta)$ is a Jacobian term, which is $\log \left| \frac{\partial f_t}{\partial y_t} \right|$.

It is useful to consider the derivatives of f_t and k_t with respect to the parameter vector θ , i.e. the two Jacobian vectors such that

$$\begin{bmatrix} df_t \\ dk_t \end{bmatrix} = \begin{bmatrix} F_t(\theta) \\ K_t(\theta) \end{bmatrix} d\theta. \quad (5)$$

From the definition of F_t and K_t it is clear that the score vector for observation t can be written as

$$g_t(\theta) = -f_t(\theta)F_t(\theta) + K_t(\theta), \quad (6)$$

so that the matrices \mathbf{F} , \mathbf{K} and the vector \mathbf{f} can be trivially defined, and the gradient vector equals:

$$\mathbf{g} = -\mathbf{F}'\mathbf{f} + \mathbf{K}'\iota.$$

In terms of equation (1), for the DLR we have :

$$\begin{aligned} R(\theta) &= \begin{bmatrix} -\mathbf{F}(\theta) \\ \mathbf{K}(\theta) \end{bmatrix} \\ r(\theta) &= \begin{bmatrix} \mathbf{f}(\theta) \\ \iota \end{bmatrix}. \end{aligned}$$

The double-length regression is therefore

$$\begin{bmatrix} \mathbf{f} \\ \iota \end{bmatrix} = \begin{bmatrix} -\mathbf{F} \\ \mathbf{K} \end{bmatrix} b_D + \text{residuals}$$

The fundamental result that makes the DLR possible is that, for this class of models, the information matrix satisfies (5):

$$\frac{1}{T} (\mathbf{F}'\mathbf{F} + \mathbf{K}'\mathbf{K}) \xrightarrow{p} \mathcal{I}(\theta),$$

provided that the matrices \mathbf{F} and \mathbf{K} satisfy appropriate regularity conditions.

If we run the DLR with the quantities $\mathbf{F}(\theta)$, $\mathbf{f}(\theta)$ and $\mathbf{K}(\theta)$ evaluated at $\tilde{\theta}$, then DLR-based LM test can be written as

$$\frac{1}{T} \tilde{\mathbf{g}}' \left(\tilde{\mathcal{I}}^{-1} \right) \tilde{\mathbf{g}} = \left(\iota' \tilde{\mathbf{K}} - \tilde{\mathbf{f}}' \tilde{\mathbf{F}} \right) \left(\tilde{\mathbf{F}}' \tilde{\mathbf{F}} + \tilde{\mathbf{K}}' \tilde{\mathbf{K}} \right)^{-1} \left(\tilde{\mathbf{K}}' \iota - \tilde{\mathbf{F}}' \tilde{\mathbf{f}} \right)$$

where the right-hand side is the explained sum of squares from the DLR, and

$$\tilde{\mathcal{I}} = \frac{1}{T} \left(\tilde{\mathbf{F}}' \tilde{\mathbf{F}} + \tilde{\mathbf{K}}' \tilde{\mathbf{K}} \right)$$

is a consistent estimator of \mathcal{I} under the null hypothesis.

Note that both b_O and b_D can be written, apart from a scale factor, as the product of a consistent estimate of \mathcal{I}^{-1} times \mathbf{g} . In practice, the OPG and DLR regressions lead to asymptotically equivalent formulations of a LM test. However, they differ because they use a different estimator of the information matrix. (16) provide Monte Carlo evidence which suggests that tests based on the DLR generally perform very much better than tests based on the OPG regression in finite samples, although no

analytical motivation for this is available. Since we cannot invoke any formal proof for the superiority in finite samples of DLR-LM statistics with respect to OPG-LM statistics, it is even more interesting to assess the relative performances of the two procedures in specific situations. In particular, we look at the behavior of these two statistics in a very popular setting in applied finance, like GARCH-in-mean models, where OPG-based LM test statistics are widely used.

3 The model

Let us consider a GARCH(1,1)-in-mean process, i.e. a process y_t such that

$$y_t = x_t\pi + h_t\phi + e_t = \mu_t + e_t \quad (7)$$

where

$$\begin{cases} e_t | \mathfrak{S}_{t-1} \sim N(0, h_t) \\ h_t = c + ae_{t-1}^2 + bh_{t-1} \end{cases} \quad (8)$$

where \mathfrak{S}_t is the σ -field generated by $\{w_t, w_{t-1}, \dots\}$. The process is assumed to be weakly stationary.

After gathering all the parameters in a vector $\theta = (\pi, \phi, c, a, b)'$, equation (8) means that a process f_t can be defined as

$$f_t = \frac{e_t}{\sqrt{h_t}},$$

and that it is a Gaussian white noise with unit variance. By doing so, equation (4) applies, where k_t equals $-\frac{1}{2} \log h_t$ in the present case.

The vectors F_t and K_t defined in equation (5) have to be evaluated recursively, due to the recursive terms in (8). Let a “state vector” z_t be defined as

$$z_t = \begin{bmatrix} e_t \\ h_t \end{bmatrix}.$$

For F_t we have

$$F_t = J_t^f J_t^z = \frac{\partial f_t(\theta)}{\partial z_t(\theta)} \frac{\partial z_t(\theta)}{\partial \theta}$$

and

$$K_t = J_t^k J_t^z = \frac{\partial k_t(\theta)}{\partial z_t(\theta)} \frac{\partial z_t(\theta)}{\partial \theta},$$

so that $df_t = J_t^f dz_t$ and $dk_t = J_t^k dz_t$. From the definition of f_t and k_t it can be immediately shown that

$$J_t^f = h_t^{-1/2} \begin{bmatrix} 1 & -\frac{e_t}{2h_t} \end{bmatrix} \quad (9)$$

$$J_t^k = \begin{bmatrix} 0 & -\frac{1}{2h_t} \end{bmatrix} \quad (10)$$

The recursive term enters J_t^z , since we have from (8):

$$dz_t = \begin{bmatrix} de_t \\ dh_t \end{bmatrix} = \begin{bmatrix} 1 & -\phi \\ 0 & 1 \end{bmatrix} \left\{ \begin{bmatrix} 0 & 0 \\ 2ae_{t-1} & b \end{bmatrix} \begin{bmatrix} de_{t-1} \\ dh_{t-1} \end{bmatrix} + \begin{bmatrix} -x_t & -h_t & 0 & 0 & 0 \\ 0 & 0 & 1 & e_{t-1}^2 & h_{t-1} \end{bmatrix} d\theta \right\}$$

or more compactly

$$dz_t = M_t dz_{t-1} + Q_t d\theta,$$

where M_t and Q_t are straightforwardly defined. By an induction argument, J_t^z must obey

$$J_t^z = M_t J_{t-1}^z + Q_t$$

and therefore, given a starting point¹ for J_0^z , all the relevant quantities can be calculated recursively, making it possible to evaluate F, K and G for any given θ .

4 Simulations

In order to assess the finite sample properties of OPG- and DLR-based LM tests on the parameter ϕ , we set up the following DGP:

$$y_t = 0.2y_{t-1} + e_t \quad (11)$$

$$h_t = 1 + ae_{t-1}^2 + bh_{t-1}, \quad (12)$$

where $e_t \sim N(0, h_t)$, namely an ordinary GARCH(1,1) model identical to that shown in equation (7), with $\phi = 0$.

The hypothesis we test by means of the OPG- and DLR-based LM test is the absence of risk premium $H_0 : \phi = 0$. We do not consider the power of tests, but only their size².

Each replication was carried out on sample sizes of the order of magnitude of those normally encountered in applied work. 250 observations are roughly a year of daily data. Moreover, samples with 100 observations only were also considered, in order to consider the behaviour of the statistics under extreme conditions.

The range of parameters for which simulations took place can be basically explained as follows:

¹It is convenient to assume as a starting point $h_0 = c/1 - a - b$, ie the unconditional variance of e_t (see (18)). Its derivative is therefore:

$$\frac{\partial h_0}{\partial \theta} = \begin{bmatrix} 0 & 0 & \frac{1}{1-a-b} & \frac{c}{(1-a-b)^2} & \frac{c}{(1-a-b)^2} \end{bmatrix}$$

²The same test could be obviously carried out in a number of ways, possibly less complicated, for the present model; the purpose of investigating this specific hypothesis is to use it as a test case to show that different ways of computing an LM test can have dramatic consequences on finite-sample inference.

1. In one set of simulations (reported in table 1), we focussed on the issue of volatility persistence, to assess whether the finite-sample size distortion for the two tests is significantly affected by the memory of the variance process. Given that the unconditional variance of e_t is given by

$$h_u = \frac{1}{1 - a - b},$$

we analyse six cases; in each case, the parameters a and b were adjusted³ so that $b = 2a$, with $(a+b)$ ranging from 0.66 to 0.99. The magnitude h_u can be seen as the ratio between unconditional and minimum conditional variance that, and is therefore interpreted as a measure of the degree of conditional heteroskedasticity.

2. The other set of simulations (reported in table 2) keeps the unconditional variance constant, in order to explore what effects different values of the parameters a and b have on the test size for a given degree of conditional heteroskedasticity.

In practice, the LM test presupposes that a full-rank estimate of \mathcal{I} is available. This, however, may not happen. In fact, it is possible that the maximum likelihood estimates lie on the boundary of the parameter space; when $a = 0$, it is possible to show that the information matrix \mathcal{I} is singular, and the model is not identified. For $a = 0$, the G matrix is not full rank, and therefore $G'G$ is singular. The same happens to $F'F + K'K$. In these cases, however, there would be little point in using a classical test, because such tests are not applicable to points on the boundary.

The probability of obtaining such an estimate can be shown to vanish asymptotically for θ_0 , the true vector of parameters, inside the parameter space; however, for a small sample size this probability is not negligible. In fact, we have obtained many of these cases. Since our objective is to evaluate the performance of tests that would be inapplicable anyway in those situations, we rejected those simulations where the estimate for a was less than 1.0E-08.

Each set of simulations consists of 10000 replications. For each value of h_u two statistics were computed⁴:

OPG_E ESS from the OPG regression, evaluated at $\tilde{\theta}$

DLR_E ESS from the DLR regression, evaluated at $\tilde{\theta}$

³Since it is common in applied financial work to find $b > a$, these constraints on parameters were chosen to ensure that our simulations reflected “real life” conditions.

⁴All computations were carried out using Ox for Linux 3.20 (see (7)); source listings are available on request.

where $\tilde{\theta}$ is the constrained maximum likelihood estimate (namely the unconstrained GARCH estimate, without the *in-mean* part). The simulation results are reported in tables 1–3. Each table lists the frequency of the event

$$S > \chi_1^2(\alpha)$$

where S is the test statistic and α is the significance level. In each table, entries in boldface indicate experiments in which the rejection frequency was outside the upper and lower bound of a 95% confidence interval; these are computed as

$$CB = \alpha \pm 1.96 \times \sqrt{\frac{\alpha(1-\alpha)}{10000}},$$

by using the normal approximation to the binomial.

It is apparent that the DLR versions of the LM statistic always outperform their OPG counterparts (see table 1 and figure 1). This is true also for the case of parameter values ($a + b = 1$), that empirically corresponds to the that of Integrated GARCH.

This is particularly true at smaller sample sizes, as was to be expected; moreover, the greater the conditional heteroskedasticity, the more apparent the size distortion of the OPG LM is. On the other hand, there seems to be little influence on the tests' size by the values of a and b for a given level of h_u , as can be seen from tables 2–3.

It is also worth pointing out that, for heavily heteroskedastic processes, the OPG-based LM statistic shows a significantly distorted actual size even for a sample size of 1000, whereas samples as big as 250 are sufficient to bring the DLR-based statistic close to its asymptotic size.

It may be conjectured that the good performance of the DLR regression could be hindered in the presence of misspecification of the error term distribution, since the very principle on which the DLR is based is specific to the Normal distribution. In order to ascertain the extent of this effect, some preliminary estimates (not reported here) were carried out where the data were generated with a t -distributed error term with 6 degrees of freedom; such a distribution has moments up to the fifth order, and a variance of 1.5; moreover, such a distribution is heavily leptokurtic, which is a characteristic commonly found in empirical data on returns. The results indicate that the accuracy of the test statistics is much worse. However, that the DLR-based LM statistic does not perform worse than the OPG-based one. In other words, the DLR test statistic appears to be more precise but not not less robust than the OPG test. This aspect, however, is still being investigated.

As far as more general models are concerned, the issue arises of differentiating the log-likelihoods analytically in a manner suitable to

Table 1: Variable h_u
Sample size = 100

Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.220	0.440	2.941	0.2505	0.2094	0.1395	0.1067	0.0776	0.0545	0.0201	0.0114
0.242	0.484	3.650	0.2531	0.2094	0.1395	0.1056	0.0780	0.0542	0.0200	0.0103
0.264	0.528	4.808	0.2543	0.2090	0.1402	0.1035	0.0780	0.0524	0.0204	0.0106
0.286	0.572	7.042	0.2560	0.2052	0.1409	0.1047	0.0801	0.0530	0.0206	0.0098
0.308	0.616	13.158	0.2541	0.2045	0.1425	0.1037	0.0794	0.0534	0.0219	0.0103
0.330	0.660	100.00	0.2792	0.2135	0.1645	0.1104	0.0998	0.0571	0.0292	0.0123

Sample size = 250

Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.220	0.440	2.941	0.2180	0.1997	0.1169	0.1029	0.0625	0.0521	0.0147	0.0112
0.242	0.484	3.650	0.2177	0.1994	0.1180	0.1030	0.0641	0.0521	0.0156	0.0113
0.264	0.528	4.808	0.2221	0.2000	0.1207	0.1016	0.0639	0.0510	0.0159	0.0105
0.286	0.572	7.042	0.2268	0.2022	0.1218	0.1013	0.0631	0.0511	0.0153	0.0109
0.308	0.616	13.158	0.2329	0.2055	0.1268	0.1029	0.0645	0.0515	0.0155	0.0111
0.330	0.660	100.00	0.2492	0.2098	0.1391	0.1038	0.0749	0.0517	0.0187	0.0114

Sample size = 500

Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.220	0.440	2.941	0.2058	0.1976	0.1077	0.1005	0.0522	0.0493	0.0112	0.0092
0.242	0.484	3.650	0.2079	0.1985	0.1069	0.1002	0.0549	0.0498	0.0120	0.0101
0.264	0.528	4.808	0.2140	0.1995	0.1065	0.0994	0.0552	0.0487	0.0123	0.0103
0.286	0.572	7.042	0.2148	0.2000	0.1088	0.0980	0.0554	0.0497	0.0131	0.0103
0.308	0.616	13.158	0.2163	0.1980	0.1106	0.0971	0.0557	0.0477	0.0128	0.0105
0.330	0.660	100.00	0.2292	0.2004	0.1211	0.0984	0.0629	0.0486	0.0145	0.0096

Sample size = 750

Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.220	0.440	2.941	0.2094	0.2017	0.1062	0.1013	0.0540	0.0500	0.0112	0.0099
0.242	0.484	3.650	0.2117	0.2031	0.1059	0.0999	0.0528	0.0500	0.0115	0.0105
0.264	0.528	4.808	0.2125	0.2028	0.1080	0.1005	0.0552	0.0494	0.0116	0.0106
0.286	0.572	7.042	0.2134	0.2019	0.1071	0.1018	0.0532	0.0485	0.0120	0.0104
0.308	0.616	13.158	0.2124	0.1998	0.1130	0.0996	0.0560	0.0492	0.0130	0.0103
0.330	0.660	100.00	0.2185	0.1994	0.1152	0.1000	0.0619	0.0473	0.0141	0.0094

Sample size = 1000

Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.220	0.440	2.941	0.2057	0.2010	0.1052	0.0998	0.0537	0.0506	0.0112	0.0110
0.242	0.484	3.650	0.2070	0.2001	0.1048	0.1003	0.0535	0.0511	0.0117	0.0107
0.264	0.528	4.808	0.2076	0.2001	0.1085	0.1016	0.0530	0.0500	0.0114	0.0106
0.286	0.572	7.042	0.2099	0.2003	0.1060	0.0981	0.0540	0.0489	0.0111	0.0100
0.308	0.616	13.158	0.2111	0.1981	0.1075	0.1001	0.0565	0.0512	0.0124	0.0098
0.330	0.660	100.00	0.2247	0.2039	0.1171	0.1019	0.0619	0.0518	0.0137	0.0103

Table 2: Fixed h_u
Sample size = 100

Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.10	0.60	3.33	0.2504	0.2111	0.1371	0.1082	0.0752	0.0558	0.0213	0.0119
0.15	0.55	3.33	0.2494	0.2078	0.1385	0.1084	0.0770	0.0553	0.0214	0.0113
0.20	0.50	3.33	0.2505	0.2093	0.1393	0.1078	0.0784	0.0549	0.0206	0.0114
0.25	0.45	3.33	0.2515	0.2067	0.1387	0.1049	0.0769	0.0537	0.0200	0.0106
0.30	0.40	3.33	0.2600	0.2082	0.1430	0.1050	0.0814	0.0524	0.0204	0.0100
0.35	0.35	3.33	0.2591	0.2039	0.1428	0.1039	0.0832	0.0523	0.0202	0.0099
0.40	0.30	3.33	0.2591	0.2023	0.1435	0.1026	0.0828	0.0513	0.0197	0.0094
0.45	0.25	3.33	0.2652	0.2015	0.1468	0.0990	0.0820	0.0507	0.0211	0.0099
0.50	0.20	3.33	0.2686	0.1995	0.1486	0.1016	0.0847	0.0526	0.0222	0.0104
0.55	0.15	3.33	0.2696	0.1968	0.1501	0.1006	0.0856	0.0519	0.0224	0.0106
0.60	0.10	3.33	0.2712	0.1935	0.1517	0.0982	0.0838	0.0476	0.0207	0.0089

Sample size = 250

Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.10	0.60	3.33	0.2173	0.2013	0.1170	0.1057	0.0611	0.0535	0.0148	0.0115
0.15	0.55	3.33	0.2150	0.1982	0.1180	0.1030	0.0602	0.0517	0.0145	0.0108
0.20	0.50	3.33	0.2162	0.2000	0.1164	0.1034	0.0620	0.0512	0.0146	0.0109
0.25	0.45	3.33	0.2182	0.1999	0.1181	0.1042	0.0633	0.0518	0.0155	0.0114
0.30	0.40	3.33	0.2245	0.2022	0.1179	0.1024	0.0649	0.0504	0.0145	0.0103
0.35	0.35	3.33	0.2282	0.2033	0.1214	0.1027	0.0657	0.0518	0.0146	0.0106
0.40	0.30	3.33	0.2327	0.2042	0.1238	0.1041	0.0681	0.0529	0.0148	0.0099
0.45	0.25	3.33	0.2357	0.2032	0.1254	0.1025	0.0705	0.0549	0.0149	0.0094
0.50	0.20	3.33	0.2383	0.2029	0.1301	0.1015	0.0707	0.0521	0.0144	0.0097
0.55	0.15	3.33	0.2419	0.2022	0.1337	0.1012	0.0730	0.0495	0.0150	0.0097
0.60	0.10	3.33	0.2454	0.2008	0.1355	0.1007	0.0745	0.0488	0.0151	0.0097

Sample size = 500

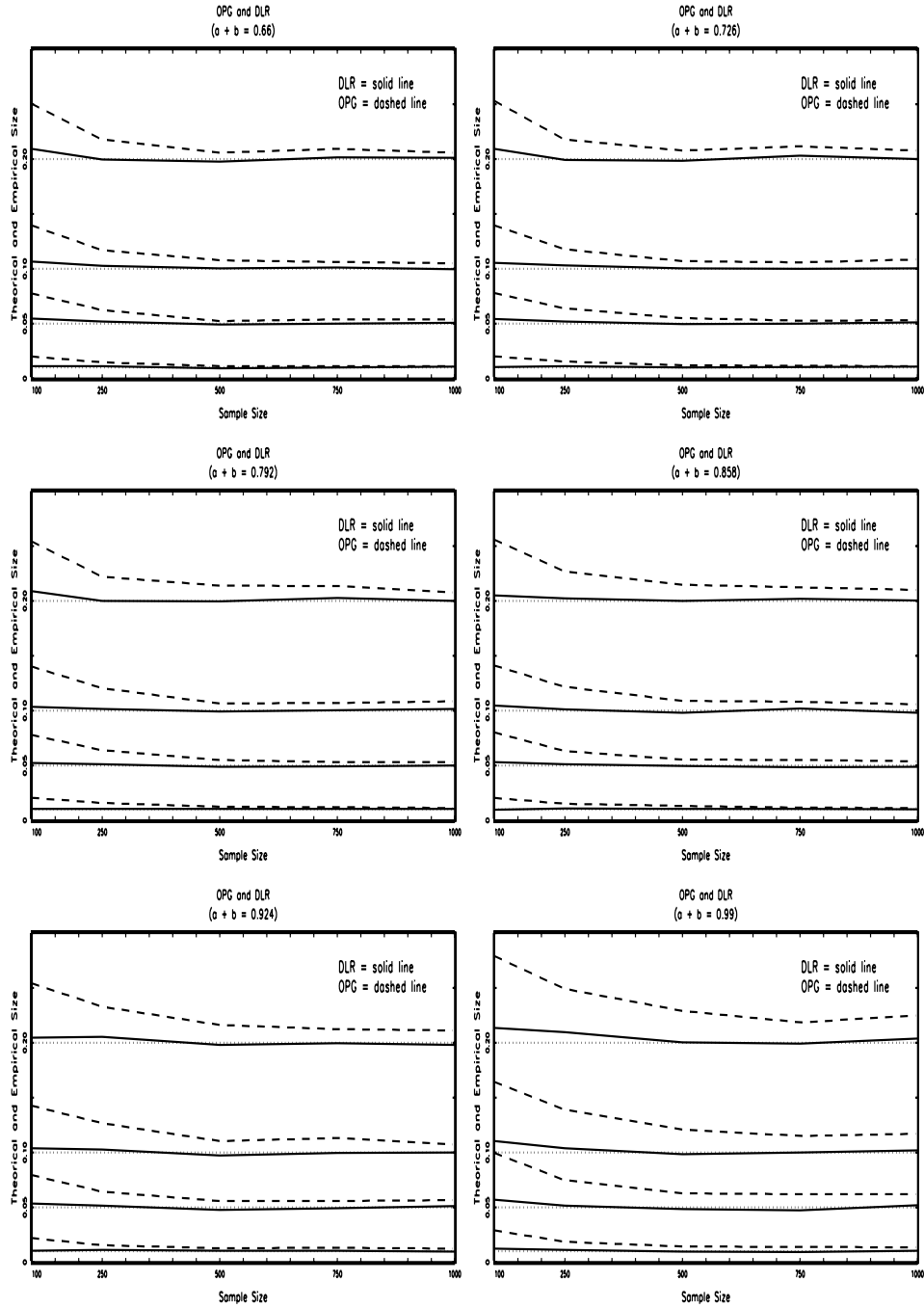
Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.10	0.60	3.33	0.2036	0.1937	0.1054	0.0999	0.0537	0.0498	0.0112	0.0090
0.15	0.55	3.33	0.2043	0.1952	0.1044	0.0980	0.0528	0.0500	0.0109	0.0090
0.20	0.50	3.33	0.2068	0.1978	0.1065	0.1000	0.0534	0.0501	0.0114	0.0095
0.25	0.45	3.33	0.2083	0.1985	0.1073	0.1005	0.0542	0.0497	0.0119	0.0100
0.30	0.40	3.33	0.2111	0.1991	0.1099	0.1006	0.0551	0.0493	0.0121	0.0097
0.35	0.35	3.33	0.2159	0.2007	0.1117	0.0999	0.0562	0.0492	0.0123	0.0097
0.40	0.30	3.33	0.2178	0.2024	0.1133	0.0997	0.0591	0.0495	0.0123	0.0093
0.45	0.25	3.33	0.2202	0.2002	0.1145	0.1002	0.0604	0.0506	0.0130	0.0090
0.50	0.20	3.33	0.2232	0.2008	0.1156	0.0997	0.0613	0.0507	0.0134	0.0093
0.55	0.15	3.33	0.2261	0.1986	0.1189	0.0976	0.0624	0.0489	0.0139	0.0091
0.60	0.10	3.33	0.2290	0.1955	0.1221	0.0938	0.0628	0.0468	0.0140	0.0091

Table 3: Fixed h_u (continued)

			Sample size = 750							
Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.10	0.60	3.33	0.2071	0.2020	0.1038	0.0998	0.0536	0.0502	0.0109	0.0098
0.15	0.55	3.33	0.2077	0.2025	0.1047	0.1004	0.0536	0.0503	0.0112	0.0099
0.20	0.50	3.33	0.2084	0.2017	0.1053	0.1002	0.0542	0.0509	0.0112	0.0099
0.25	0.45	3.33	0.2116	0.2041	0.1053	0.1005	0.0531	0.0494	0.0114	0.0103
0.30	0.40	3.33	0.2135	0.2035	0.1085	0.1014	0.0535	0.0499	0.0118	0.0113
0.35	0.35	3.33	0.2130	0.2023	0.1111	0.1043	0.0546	0.0496	0.0120	0.0112
0.40	0.30	3.33	0.2135	0.2020	0.1137	0.1045	0.0570	0.0491	0.0119	0.0105
0.45	0.25	3.33	0.2148	0.2007	0.1153	0.1036	0.0582	0.0486	0.0123	0.0103
0.50	0.20	3.33	0.2171	0.1990	0.1177	0.1021	0.0597	0.0505	0.0128	0.0100
0.55	0.15	3.33	0.2185	0.1984	0.1189	0.1007	0.0607	0.0503	0.0124	0.0097
0.60	0.10	3.33	0.2234	0.1957	0.1183	0.1010	0.0621	0.0497	0.0119	0.0089

			Sample size = 1000							
Model			Rejection frequency							
a	b	h_u	0.20		0.1		0.05		0.01	
			OPG	DLR	OPG	DLR	OPG	DLR	OPG	DLR
0.10	0.60	3.33	0.2046	0.2003	0.1047	0.1012	0.0538	0.0505	0.0113	0.0107
0.15	0.55	3.33	0.2042	0.1996	0.1047	0.1023	0.0526	0.0497	0.0119	0.0111
0.20	0.50	3.33	0.2058	0.2003	0.1051	0.1015	0.0534	0.0500	0.0117	0.0112
0.25	0.45	3.33	0.2072	0.2008	0.1042	0.0998	0.0537	0.0514	0.0114	0.0106
0.30	0.40	3.33	0.2065	0.2006	0.1049	0.0997	0.0541	0.0502	0.0118	0.0105
0.35	0.35	3.33	0.2095	0.2020	0.1053	0.1007	0.0547	0.0509	0.0108	0.0099
0.40	0.30	3.33	0.2115	0.2013	0.1088	0.1002	0.0553	0.0495	0.0113	0.0099
0.45	0.25	3.33	0.2134	0.1999	0.1100	0.1007	0.0561	0.0501	0.0122	0.0102
0.50	0.20	3.33	0.2131	0.1983	0.1096	0.0998	0.0577	0.0502	0.0127	0.0104
0.55	0.15	3.33	0.2121	0.1966	0.1096	0.0984	0.0587	0.0505	0.0133	0.0105

Figure 1: Empirical size of the OPG and DLR tests



implement the tests presented here. This can be difficult for some models. On the other hand, numerical differentiation can be used: in fact, there is little computational overhead in computing $\ell_t(\theta)$ as a function of $f_t(\theta)$ (most of which is typically a simple quadratic form) and $k_t(\theta)$. As a consequence, instead of differentiating $\ell_t(\theta)$ numerically, it is possible to differentiate numerically $f_t(\theta)$ and $k_t(\theta)$ to obtain $F_t(\theta)$ and $K_t(\theta)$ and then compute $g_t(\theta)$ via equation (6) whenever required.

On multivariate models, several problems arise: it is not obvious how the concepts behind the DLR regressions can be generalised to multivariate models. Moreover, the task of analytically differentiating the log-likelihood of multivariate GARCH-in-mean models is a daunting one. However, analytical derivatives for the widely used BEKK specification (see (8)) have recently been presented by (12).

5 An illustrative example

One of the relevant results of our Monte Carlo simulations is that the DLR versions of the LM statistic always outperform their OPG counterparts, especially for small sample size. The practical implication is that when we test for the presence of a risk term, using a thousand or more observations, the difference between DLR and OPG can be irrelevant for the conclusions of the test. This is especially true when the sample frequency of returns is daily or intra-daily. The results could be slightly different when we use a sample size that does not exceed few hundreds data points. This is the case when we conduct the test using returns sampled at lower frequencies, e.g. weekly data. In this situation, the results of the two statistics can be markedly different, leading to opposite conclusions. We provide a simple example of a such case.

We analyze the DAX30 German stock index weekly excess returns, from 3/3/1999 to 10/12/2003, that is 250 observations. This is exactly the sample size for which the two test statistics show significant differences in performances. The series seems to be characterized by a leptokurtic distribution, as expected (see, tab.4). We propose two models that differ only for the conditional mean equation. The first is a simple GARCH(1,1)-in-mean model:

$$y_t = \pi_0 + \phi h_t + e_t \quad (13)$$

$$h_t = c + ae_{t-1}^2 + bh_{t-1} \quad (14)$$

while the second one is an AR(1)-GARCH(1,1)-in-mean model:

$$y_t = \pi_0 + \pi_1 y_{t-1} + \phi h_t + e_t \quad (15)$$

$$h_t = c + ae_{t-1}^2 + bh_{t-1} \quad (16)$$

Table 4: Descriptive statistics

T	250
mean	-0.236
st. dev.	4.173
skewness	-0.202
kurt	4.947

We estimate both models only under the null hypothesis. The estimation results are reported in table 5, where the DLR and OPG statistics for the null hypothesis $H_0 : \phi = 0$ are also reported. In both cases, we have contradictory results. In fact, looking at DLR we should reject the null hypothesis of absence of risk premium term, at a 5% significance level, while using the OPG results we should accept the null. Now, for this sample size, given the results of the Monte Carlo simulations shown above, the DLR-version of the LM test seems to be more reliable than the corresponding OPG version. Therefore, we conclude, contrary to the OPG results, that the conditional variance does not enter the conditional mean of this series.

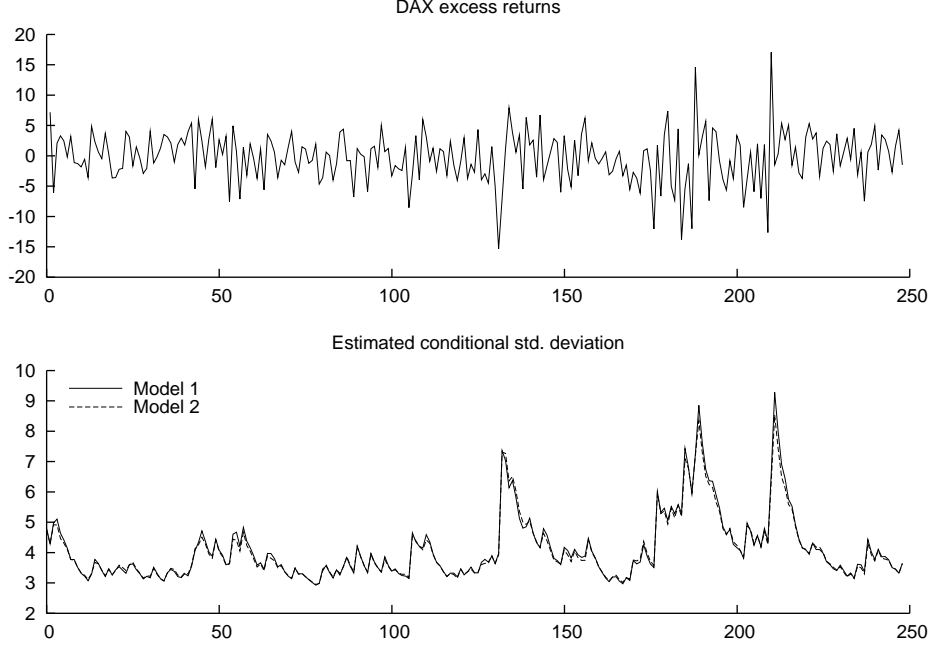
It is interesting to note that estimating equations (13)–(14) and (15)–(16) under the alternative hypothesis gives support to the conclusion that the DLR-LM test provides more accurate indications. As an experiment, we estimated the two models without the restriction $\phi = 0$ and then tested the restriction by means of a Wald test (using the robust variance matrix estimator by (2)); in these settings, the hypothesis is accepted for both models ($p = 0.1282$ for equation (13), $p = 0.1426$ for equation (15)).

Table 5: Estimation results

	Model 1				Model 2			
	Estimate	s.e.	z-stat	p-val	Estimate	s.e.	z-stat	p-val
Constant	-0.162	0.547	-0.295	0.768	-0.127	0.570	-0.223	0.823
y_{t-1}					-0.100	0.069	-1.440	0.150
c	2.000	1.228	1.629	0.103	1.925	1.215	1.584	0.113
a	0.179	0.078	2.287	0.022	0.164	0.074	2.209	0.027
b	0.715	0.117	6.089	0.000	0.729	0.116	6.259	0.000
OPG	4.149			0.042	4.440			0.035
DLR	3.780			0.052	3.649			0.056

On the basis of the Monte Carlo evidence (see table 1) and the indications of the Wald-type tests, we can safely conclude that the OPG-based LM tests leads to an incorrect rejection of the null, whereas the DLR-based test leads to its acceptance.

Figure 2: DAX Index: values and estimated conditional SD



6 Conclusions

In this article we investigate the finite-sample properties of two alternative methods for computing LM-type tests: the OPG-based LM test statistics and the DLR-based LM test statistics.

Monte Carlo evidence for a set of univariate AR time series models with GARCH(1,1) errors indicates that the finite sample size of the DLR-LM test is much closer to its nominal value than the OPG-LM test: the OPG-LM test overrejects to an extent that can lead to false inferences on the risk premium parameter. This is especially true for data characterised by a high level of conditional heteroskedasticity: in these cases, the DLR-based LM statistics performs very well even for moderate sample sizes, whereas its OPG counterpart requires very large samples to match asymptotic critical values. We also provide a real-data example of a series of weekly returns in which the two test statistics give conflicting results and we show that the DLR-based test is to be preferred.

Although our experiment is somewhat limited, we expect our findings to be true in more general settings as well. Moreover, this experiment suggests that future research should closely look at the possible extensions of the DLR-based LM test to multivariate GARCH models. In general, there seems to be no reason to prefer the OPG-LM test over

the DLR-LM test, which has better properties under correct specification, and requires only a small additional computational effort.

References

- [1] BOLLERSLEV, T., R. F. ENGLE, AND J. WOOLDRIDGE (1988): "A capital asset pricing model with time varying covariances," *Journal of Political Economy*, 96, 116–131.
- [2] BOLLERSLEV, T., AND J. WOOLDRIDGE (1992): "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric Reviews*, 11(2), 143–172.
- [3] DAVIDSON, R., AND J. G. MACKINNON (1984): "Model specification tests based on artificial linear regressions," *International Economic Review*, 25, 485–502.
- [4] ——— (1988): "Double-length artificial regressions," *Oxford Bulletin of Economics and Statistics*, 50, 203–217.
- [5] ——— (1993): *Estimation and Inference in Econometrics*. Oxford University Press.
- [6] ——— (1998): "Artificial Regressions," Queen's University Institute for Economic Research Discussion Paper Number 978.
- [7] DOORNIK, J. A. (1999): *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, London, 3rd edn.
- [8] ENGLE, R. F., AND K. KRONER (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.
- [9] ENGLE, R. F., D. LILIEN, AND R. ROBINS (1987): "Estimating time varying risk premia in the term structure: the ARCH-M model," *Econometrica*, 55, 391–407.
- [10] ENGLE, R. F., AND V. NG (1993): "Measuring and testing the impact of news on volatility," *Journal of Finance*, 48, 1749–1777.
- [11] GODFREY, L. G., AND M. R. WICKENS (1981): "Testing Linear and Log-Linear Regressions for Functional Form," *Review of Economic Studies*, 48, 487–96.
- [12] LUCCHETTI, R. (2002): "Analytical Score for Multivariate GARCH models," *Computational Economics*, 19(2), 133–143.

- [13] LUMSDAINE, R. (1995): “Finite-Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models: A Monte Carlo Investigation,” *Journal of Business and Economic Statistics*, 13(1), 1–10.
- [14] LUMSDAINE, R., AND S. NG (1999): “Testing for ARCH in the Presence of a Possibly Misspecified Conditional Mean,” *Journal of Econometrics*, 93(2), 257–279.
- [15] LUNDBERGH, S., AND T. TERÄSVIRTA (2002): “Evaluating GARCH Models,” *Journal of Econometrics*, 110, 417–435.
- [16] MACKINNON, J. G., AND L. MAGEE (1990): “Transforming the Dependent Variable in Regression Models,” *International Economic Review*, pp. 315–39.
- [17] MERTON, R. (1973): “An intertemporal capital asset pricing model,” *Econometrica*, pp. 867–888.
- [18] NELSON, D. B., AND C. Q. CAO (1992): “Inequality Constraints in the Univariate GARCH Model,” *Journal of Business and Economic Statistics*, 10(2), 229–35.
- [19] PAGAN, A. R. (1996): “The Econometrics of Financial Markets,” *Journal of Empirical Finance*, 3, 15–102.
- [20] PAGAN, A. R., AND A. ULLAH (1988): “The econometric analysis of models with risk terms,” *Journal of Applied Econometrics*, 3, 87–105.

Dipartimento di economia politica e metodi quantitativi
Università degli studi di Pavia

List of the lately published Technical Reports
(available at the web site: "<http://economia.unipv.it/Eco-Pol/quaderni.htm>").

Quaderni di Dipartimento

#	Date	Author(s)	Title
147	01-03	A. Lijoi I. Pruenster	Practicable alternatives to the Dirichlet process
148	02-03	A. Lijoi Pruenster S.G.Walker	Extending Doob's consistency theorem to nonparametric densities
149	02-03	V.Leucari G.Consonni	Compatible Priors for Causal Bayesian Networks
150	02-03	L. Di Scala L. La Rocca G.Consonni	A Bayesian Hierarchical Model for the Evaluation of a Web Site
151	02-03	G.Ascari	Staggered Prices and Trend Inflation: Some Nuisances
152	02-03	G.Ascari P.Gagnepain	How inefficient are football clubs? An evaluation of the Spanish arms race
153	03-03	P.Dellaportas C.Tarantola	Categorical data squashing by combining factor levels
154	05-03	A. Lijoi Pruenster	A note on the problem of heaps
155	09-03	P.Berti P.Rigo	Finitely additive uniform limit theorems
156	09-03	P.Giudici C.Tarantola	Web Mining pattern discovery
157	09-03	M.A.Maggi U.Magnani M.Menegatti	On the Relationships between Absolute Prudence and Absolute Risk Aversion
158	10-03	P.Bertoletti	Uniform Pricing and Social Welfare
159	01-04	G.Ascari N.Rankin	Perpetual Youth and Endogenous Labour Supply: A Problem and a Possible Solution
160	02-04	C.Bianchi M.Menegatti	On the potential pitfalls in estimating convergence by means of pooled and panel data
161	02-04	P.Bertoletti	Logarithmic Quasi-Homothetic Preferences
162	02-04	L.Bonatti G.Felice	Endogenous growth and changing sectoral economies composition in advanced
163	03-04	R.Lucchetti E.Rossi	Artificial Regression Testing in the GARCH-in-mean model