T E C H N I C A L
R E P O R T

0483

TESTING WEAK EXOGENEITY IN THE EXPONENTIAL FAMILY: AN APPLICATION TO FINANCIAL POINT PROCESSES

DOLADO, J., RODRIGUEZ-POO, L. and D. VEREDAS

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

CORE DISCUSSION PAPER
2004/49

TESTING WEAK EXOGENEITY IN THE EXPONENTIAL FAMILY: AN APPLICATION TO FINANCIAL POINT PROCESSES

Juan J. Dolado[1], Juan Rodriguez-Poo[2] and David Veredas[3]

Abstract

In this paper, two tests for weak exogeneity in the econometric modelling of financial point processes are proposed. They are motivated by the common practice in many econometric studies of tick-by-tick data of making inference on the joint density of durations and marks through the conditional (marks given durations) density. However, this inference is only valid if the process of the marginal (durations) is weakly exogenous for the parameters of the conditional density, a hypothesis which is often left untested. Under standard pseudo-maximum likelihood conditions, we first derive a simple parametric score/LM test statistic when the potential dependence between the parameters of interest in the conditional model and the marginal process is assumed to be linear. Next, an alternative consistent test is proposed when the functional form of the dependence is left unspecified. To illustrate the use of these tests, we analyze two types of financial point processes, linked with market microstructure theory and stealth trading hypothesis, for five stocks traded at NYSE: (i) the relationship between trade size and trade durations and (ii) the relationship between volume and price durations. In general we reject the null hypothesis of weak exogeneity, therefore questioning some results in the literature which rely on separate estimation of each density.

Keywords: Weak exogeneity, pseudo-maximum likelihood, semiparametric models, point processes, high-frequency data. stealth trading, mixture of distribution hypothesis

JEL classification: C12, C41, C52, G10.

[1]Departamento de Economía, University Carlos III de Madrid; email: dolado@eco.uc3m.es

[2]Departamento de Análisis Económico, Universidad de Zaragoza; email: rodrigjm@unizar.es

[3]Econometrics and Finance Group, CentER, Tilburg Univeristy and CORE, Université catholique de Louvain; email: d.veredas@uvt.nl. Corresponding address: David Veredas, Tilburg University, Dept Econometrics & OR. P.O. Box 90153. 5000LE Tilburg, The Netherlands. Phone: +31 13 466 3151. Fax: +31 13 466 3280.
Work supported in part by the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, [MICFINMA] and by the DGCYT under research grant BEC2001-1121. With the usual disclaimer, we are grateful to Luc Bauwens and Roberto Pascual for useful comments.

1 INTRODUCTION

It is a common practice in econometrics to analyze a joint process, say $\{y, x\}$, through the analysis of the parameters of interest, ψ, via either the conditional density of y given x, $f(y|x)$, or the opposite, $f(x|y)$. It has the following justification: If we denote by $f(x, y)$ the joint density of the process $\{y, x\}$, then it can be always written as the product of the conditional times the marginal densities, i.e. $f(y, x) = f(y|x)f(x)$ or $f(y, x) = f(x|y)f(y)$.[1] In a parametric setup, researchers make distributional assumptions for $f(y|x)$ and $f(x)$, depending on parameter sets, θ^y for $f(y|x)$ and θ^x for $f(x)$. Then, if i) x is weakly exogenous for ψ and ii) the respective densities are correctly specified, the inference procedure described above is valid; cf. Engle et al. (1983). Otherwise, if either weak exogeneity or correctness of specification of the joint distribution fails, the resulting estimators would be inconsistent; c.f. Huber (1967), White (1981, 1982). However, the distributional sufficient condition has been somehow relaxed by assuming correct specification of the conditional mean instead of demanding it for the whole density function (see Gourieroux et al.1984 a,b, for details). This last result is based in a well known property of the exponential densities.

Regarding weak exogeneity, most tests have been derived under the assumption of correct specification of the joint density function of the process of interest. Furthermore, the form of the joint density has been commonly taken as Gaussian. Then, using standard maximum-likelihood (ML) approaches several tests for exogeneity have been proposed in the literature; cf. Engle et al. (1983), Engle and Hendry (1983) and Boswijk and Urbain (1997). Unfortunately, neither normality nor knowledge of the joint density of the process are appropriate assumptions when modelling financial point processes, which constitute our focus in this paper. This situation arises when no distribution may capture correctly the data features, because of skewness or lepto(platy)kurtosis, but the first moment is still correctly specified.

A financial point process may be, in general, decomposed in two sets of variables. The first, denoted by x, is the vector of durations - defined as the time interval between two financial events. The second, denoted by y, a matrix of variables associated to each arrival time - known, in statistics, as the marks of the process. Examples of financial durations are the time elapsing between trades of equity assets (see e.g., Engle, 2000) or between changes in the intervention interest rates set by central banks in the interbank reserve market (see e.g., Dolado and Maria-Dolores, 2002). The marks, in turn, may refer to changes in the price or volume of the traded assets or the change in the interest rate itself, whose probability of change is modelled conditional on a given duration of time where no variations occur. In these examples, the nature (i.e. support) of the mark differs among applications.

[1] For ease of exposition we focus on the factorization $f(y, x) = f(y|x)f(x)$. This is often the case in empirical applications of market microstructure. Specifically, when modelling financial point process we are interested in the distributions of durations, denoted by x, and marks, denoted by y. One advantange of these financial point processes is that only one of the two possible factorizations of $f(y, x)$ is valid. In effect, because of the time ordering of the process, it only makes sense to factorize the joint density as the conditional of the marks times the marginal of the duration, i.e. $f(y, x) = f(y|x)f(x)$. Alternatively, the factorization $f(y, x) = f(x|y)f(y)$ is meaningless as the conditional density in this case would be the probability that a duration dies conditional on the mark that is only observed when the duration dies.

In a similar tick-by-tick data[2] context to ours, Engle (2000) discusses the need of weak exogeneity if one wishes to perform a separate estimation of each density. If some of the parameters of the marginal process are present in the conditional density, his suggestion is to estimate them consistently from the marginal pdf and then replace these estimates in the conditional density to obtain estimates of the parameters of interest. This would provide consistent estimates of the parameters of interest (as long as the conditional mean is still correctly specified). Yet they will be inefficient, given the use of first-step estimates of the marginal mean. However, no explicit test for weak exogeneity is proposed, as we do here. Of course, this type of test would be a necessary requirement before implementing the two-step procedure.

For the previous reasons, we propose two tests for weak exogeneity using a pseudo-maximum likelihood (PML) framework. First, we propose a simple score-type test for weak exogeneity considering the class of Quasi Generalized Pseudo Maximum Likelihood (QGPML) estimators when the dependence between the parameters of the conditional and marginal means of the densities is assumed to be linear. In this case, it is easy to show that a pseudo-score test (equivalent to the standard score/LM test) suffices to achieve consistency. Secondly, since, under misspecification of the functional dependence, this score test may reject the weak exogeneity assumption when indeed the data generating process might present this feature, we also derive a consistent specification likelihood-ratio test that shows nontrivial power against nonparametric alternatives. A bootstrap procedure to approximate the asymptotic distribution of the latter test is developed.

In our empirical application, we consider two financial point processes. The financial durations correspond to the time intervals between i)trades and ii)accumulated price changes. The marks are i) trade size[3] and ii)accumulated volume per price changes respectively. Data are tick-by-tick of 5 stocks traded at NYSE in 1996. The choice of volume as the mark of the point process can be motivated by reference to the microstructure literature.

In effect, in many empirical market microstructure studies, durations have been modelled conditional on past values of different marks, such as the spread or trade size; cf. Engle (2000) or Bauwens and Giot (2000). On the other hand, modelling the marks conditional on the durations has important economic implications. For example, the first application, where we study the relation between trade size and trade durations, is useful to test the hypothesis of stealth-trading; cf. Barclay et al. (1993) and Chakravarty (2001). According to this hypothesis, informed traders trade medium size blocks to avoid the disclosure of their private information to the market. This is related with Easley and O'Hara (1992) model that predicts that tranquile periods are informative as no informed traders are in the market. In the second application we examine the relationship between accumulated volume per price change and price duration. This is linked with the Clark's mixture of distribution hypothesis (1973) that predicts a positive covariance between volume size and squared returns, a proxy for volatility. To the extent that price durations are as well a proxy of the instantaneous volatility of the market (cf. Engle and Russell, 1998), in this fashion we study the instantaneous relation

[2] Denomination of a point process among stock market practitioners.

[3] Trade size is also known as trade volume. However, we stick to trade size as it is the usual terminology among market microstructure empiricists.

between volume size and volatility.

In both examples, we test weak exogeneity, that is, if the durations are weakly exogenous for the parameters of the volume density. From our results, out of ten estimations, we reject the presence of weak exogeneity in seven (eight) cases using the parametric (nonparametric) test. Therefore, we conclude that it is often not valid to estimate separately the parameters of conditional and the marginal densities. And it seems useful to apply our weak exogeneity test before drawing strong conclusions from these exercises.

The rest of the paper is structured as follows. Section 2 introduces notation and offers a brief summary of the concept of weak exogeneity. Section 3 reviews the score/LM test in ML and PML frameworks, and discusses its application in several examples where the dependence between the parameters in the means of the conditional and marginal densities is assumed to be linear. Section 4 proposes a consistent specification test for weak exogeneity when the functional form of the dependence is left unspecified. Section 5 contains two applications and section 6 concludes.

2 WEAK EXOGENEITY

Following the seminal contribution of Engle et al. (1983), let $\{z_i\}_{i=1}^N = \{y_i, x_i\}_{i=1}^N$ be a bivariate stochastic process with joint density $f(z_i|I_i;\theta)$ where I_i is a sigma field consisting of past values of the process, $\underline{z_{i-1}}$ and current and past values of other valid conditioning variables $\underline{w_i}$, i.e. $I_i = (\underline{z_{i-1}}, \underline{w_i})$. The joint density can be factorized as

$$f_z(z_i|I_i;\theta) = f_{y|x}(y_i|x_i, I_i;\theta^y) f_x(x_i|I_i;\theta^x), \tag{1}$$

where $(\theta^y, \theta^x) \in \Theta$. The choice of this factorization, and not the reverse one i.e. $f_z = f_{x|y}f_y$, is given by economic information about the processes (x, y), as discussed above for marked-point processes.

Let $\psi = f(\theta) \in \Psi$ be the parameters of interest which are assumed to be only present in the conditional density. The key issue, addressed by Engle et al. (1983), is to know under which conditions it is possible to estimate ψ just as a function of θ^y and without loss of information. In other words, that all the information needed for the estimation of ψ is in $f_{y|x}$. This would be possible if x_i is *weakly exogenous* for ψ, namely if i) $\psi = f(\theta^y)$ and ii) θ^y and θ^x are variation free, i.e. $(\theta^y, \theta^x) \in \Theta^y \times \Theta^x$.[4]

The next step is to derive a simple test for weak exogeneity. For this, we need to specify some functional forms for the densities, an issue that shall be discussed at length later on. For

[4]Note that, albeit not considered in this article, further definitions of exogeneity arise if one is additionally interested in prediction or policy analysis, besides inference. For example, in the case of prediction conditional on forecasts of the exogenous variables, besides weak exogeneity, one also needs Granger noncausality of y with respect to x, leading to the concept of *strong* exogeneity. How to test for Granger noncausality in a financial point process has been analyzed by Renault and Werker (2004). Finally, when considering policy analysis, *strict* exogeneity requires weak exogeneity and that θ^y is invariant to interventions affecting θ^x; cf. Engle et al. (1983).

the time being, let us just assume that the parameters of interest are in the conditional mean of y given x and the information set I, with the following structure

$$E[y_i|x_i, I_i; \theta^y] = \mu_i^{y|x,I_i} = g\left[\alpha(\mu_i^{x|I})x_i + \lambda' U_i\right], \quad (2)$$

while the conditional mean of x_i has the form

$$E[x_i|I_i; \theta^x] = \mu_i^{x|I} = h\left[\beta' U_i\right], \quad (3)$$

where $U_i = (1, w_i, w_{i-1}, \ldots, z_{i-1}, \ldots)$; $g(\cdot)$, $\alpha(\cdot)$ and $h(\cdot)$ are assumed to be known functions. Their specific functional forms will depend on the the nature of the problem.[5] For notational simplicity, we suppress hereafter the information set I, always keeping in mind the dependence of the previous expressions on it.

Under the assumption that the parameters of interest depend solely on the parameters of the conditional mean $\mu_i^{y|x}$ i.e. $\psi = (\alpha, \lambda)$, it suffices to test that $\alpha(\mu_i^x) = \alpha$ (a constant) in order to test weak exogeneity. If such a case, the parameters of interest are not subject to cross-equation restrictions and ψ is not subject to variations in θ^x. More specifically, if $\alpha(\mu_i^x)$ is assumed to be linear, i.e. $\alpha(\mu_i^x) = \alpha_0 + \alpha_1 \mu_i^x$, then testing weak exogeneity reduces to test the null hypothesis $H_0 : \alpha_1 = 0$. In the sequel, we will refer to this case as LD (linear dependence). Note, however, that the approach easily extends to more general functional forms for dependence. For instance, $\alpha(\mu_i^x)$ may be assumed to be a nonlinear (e.g. quadratic) function of μ_i^x.

By contrast, if $\alpha(\cdot)$ is left unspecified, (2) becomes the so called semiparametric generalized partially linear model with time varying coefficients (see Severini and Staniswallis, 1994 and Cai, Fan and Li, 2000). In this framework, testing for weak exogeneity again reduces to test the null hypothesis of $H_0 : \alpha(\mu_i^x) = \alpha$. That is, under the null hypothesis of weak exogeneity no variation in $\alpha(\cdot)$ is allowed. This is a nonparametric test. The next two sections explain the parametric and the nonparametric tests, respectively.

3 THE SCORE PARAMETRIC TEST

In this Section we first introduce some further notation and briefly review the derivation of the score parametric test, first under ML and then under PML, using general forms of g and h in (2) and (3). Next, we discuss the choice of specific functional forms for g and h depending on the supports of y and x. We highlight that, by appropriate choice of the density and the functions g and h, the first order conditions are the same for all choices, up to a weighting factor. Finally, we derive the test for two simple examples, stressing its similarity with the very well-known case of linear conditional means and Gaussian distributions. In the first example, we assume that $f(y|x)$ follows a Poisson distribution and that $f(x)$ is Gaussian whilst in the second one we consider the case of two exponential distributions. Lastly, we briefly disgress about how to implement the test when the parameters of interest are not in the conditional mean, but in the conditional variance.

[5] Note that their arguments are linear. However, this assumption can be easily relaxed by substituting $g\left[\alpha(\mu_i^{x|I})x_i + \lambda' U_i\right]$ by $g\left[\alpha(\mu_i^{x|I})x_i, U_i(\lambda)\right]$ and $h\left[\beta' U_i\right]$ by $h\left[U_i(\beta)\right]$.

3.1 The LM test in ML and PML frameworks

As is well known, assuming that the densities are correctly specified and that LD holds, the general theory of ML leads to a simple score test for $\alpha_1 = 0$. This is the popular Lagrange Multiplier (LM) test based on the constrained model under $H_0 : \alpha_1 = 0$. Given (1), the log-likelihood is

$$L(\theta^y, \theta^x) = \ell_{y|x}(y|x; \mu^x(\psi), \theta^{y*}) + \ell_x(x; \theta^x)$$

where $\ell_{y|x}$ and ℓ_x are the conditional and marginal log-likelihood functions and $\theta^{y*} = \theta^y \setminus \psi$ i.e. the parameters not present in the conditional mean and hence of no interest. These may be parameters of higher moments, such as the variance.

Defining $\hat{i}(\theta) = \frac{\partial L_N}{\partial \theta}$ and $\hat{I}(\theta) = \frac{\partial^2 L_N}{\partial \theta \partial \theta'}$ to be the empirical score and Hessian, they can be partitioned as

$$\hat{i}(\theta) = \begin{pmatrix} \hat{i}_{y|x}(\theta) \\ \hat{i}_x(\theta) \end{pmatrix} \text{ and } \hat{I}(\theta) = \begin{pmatrix} \hat{I}_{y|x}(\theta) & \hat{I}_{y|x,x}(\theta) \\ \hat{I}_{x,y|x}(\theta) & \hat{I}_x(\theta) \end{pmatrix}.$$

Denote by $\widehat{\theta}_c$ the restricted ML estimator (i.e. under $\alpha_1 = 0$) and equivalently for $\widehat{\theta}_c^y$ and $\widehat{\theta}_c^x$. Then, under standard regularity conditions, they are asymptotically normally distributed, consistent and efficient, in the sense that they reach the Cramer-Rao bound. Under the null hypothesis $\hat{I}_{y|x,x}(\theta) = \hat{I}_{x,y|x}(\theta) = 0$ and

$$\hat{i}\left(\widehat{\theta}_c\right) = \begin{pmatrix} \hat{i}_{y|x}\left(\widehat{\theta}_c\right) \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{i}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c\right) \\ \hat{i}^{\theta^{y\dagger}}_{y|x}\left(\widehat{\theta}_c\right) \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{i}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c\right) \\ 0 \\ 0 \end{pmatrix},$$

where $\theta^{y\dagger} = \theta^y \setminus \alpha_1$. The FOCs imply that all the components of $\hat{i}_{y|x}\left(\widehat{\theta}_c^y\right)$ are zero, except the one pertaining to α_1. Equivalently, the submatrix $\hat{I}_{y|x}\left(\widehat{\theta}_c\right)$ can be partitioned

$$\hat{I}_{y|x}\left(\widehat{\theta}_c^y\right) = \begin{pmatrix} \hat{I}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c^y\right) & \hat{I}^{\alpha_1,\theta^{y\dagger}}_{y|x}\left(\widehat{\theta}_c^y\right) \\ \hat{I}^{\theta^{y\dagger},\alpha_1}_{y|x}\left(\widehat{\theta}_c^y\right) & \hat{I}^{\theta^{y\dagger}}_{y|x}\left(\widehat{\theta}_c\right) \end{pmatrix}.$$

Hence, the LM test has the familiar form (see e.g. Harvey, 1981)

$$S_{ML} = -\hat{i}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c\right)' \hat{I}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c^y\right)^{-1} \hat{i}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c\right) \tag{4}$$

where $\hat{I}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c^y\right)^{-1}$ is the inverse of the component corresponding to α_1 of the partitioned matrix $\hat{I}_{y|x}\left(\widehat{\theta}_c^y\right)$ i.e. $\hat{I}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c^y\right)^{-1} = \left[\hat{I}^{\alpha_1}_{y|x}\left(\widehat{\theta}_c^y\right) - \hat{I}^{\alpha_1,\theta^{y\dagger}}_{y|x}\left(\widehat{\theta}_c^y\right) \hat{I}^{\theta^{y\dagger}}_{y|x}\left(\widehat{\theta}_c^y\right)^{-1} \hat{I}^{\theta^{y\dagger},\alpha_1}_{y|x}\left(\widehat{\theta}_c^y\right)\right]^{-1}$. Under mild regularity conditions regarding the densities, it is well known that the limiting distribution of the test is $\chi^2(1)$ under H_0; cf. Silvey (1975).

In the derivations above, no specific assumptions have been made regarding the functional form of the joint density. It has just been specified as the product of the conditional and the marginal. In practice, however, we often encounter cases in which the nature of y may differ substantially from that of x. For example, y may be a binomial process and x a gamma process which makes it difficult to write down the true joint density. In such cases, if the parameters of interest are exclusively present in the mean of the conditional density and our focus lies on the estimation of a parameter present in the first conditional moment then, following the important results of Gourieroux et al. (1984a,b) we can write the pseudo-log-likelihood function as

$$Q^{LEF}(\psi, \eta) = \log q_{y|x}^{LEF}\left(y; \mu^{y|x}\right) + \log q_x^{LEF}(x; \mu^x), \tag{5}$$

where $q_{y|x}^{LEF}$ and q_x^{LEF} are the conditional and marginal pseudo log-likelihood functions that belong to the Linear Exponential Family (LEF hereafter), i.e.

$$\begin{aligned} q_{y|x}^{LEF} &= \exp\left\{A_{y|x}\left(\mu_i^{y|x}\right) + B_{y|x}(y_i) + C_{y|x}\left(\mu_i^{y|x}\right) y_i\right\} \text{ and} \\ q_x^{LEF} &= \exp\left\{A_x(\mu_i^x) + B_x(x_i) + C_x(\mu_i^x) x_i\right\} \end{aligned}$$

being A, B and C are scalars. The parameters are now only present in the conditional means. For $q_{y|x}^{LEF}$ they are the parameters of interest ψ. For q_x^{LEF} the parameters in the conditional mean are in the subset η. Therefore, $\theta^y = \psi \cup \theta^{y*}$ and $\theta^x = \eta \cup \theta^{x*}$.

As is well known, the LEF contains many important laws, such as Gaussian, Poisson, binomial, multinomial, gamma (and hence exponential), negative binomial, multivariate Gaussian and multivariate Poisson distributions, which turn out to be particularly useful in modelling durations and marks of a financial point process. However, some of these laws have some additional parameters which are present in higher moments. These parameters would be in θ^{y*} and θ^{x*}. In our setup, they will be assumed to be known since no assumptions are made on higher moments. For example, if q_x^{LEF} is *i.i.d.* Gaussian, $\theta^x = \left(\mu^x, \sigma^2\right)$, $\eta = \mu^x$ and $\theta^{x*} = \sigma^2$.

The estimates that maximize Q^{LEF} are the PML estimates. These estimates are consistent, asymptotically normal distributed but not efficient. Indeed, its variance-covariance matrix is of the classical *sandwich* form $J^{-1}IJ^{-1}$ where, for ψ,

$$J_{y|x,LEF} = E\left[\frac{\partial^2 Q^{LEF}}{\partial\psi\partial\psi'}\right] = E\left[\frac{\partial\mu^{y|x}}{\partial\psi}\Sigma_{y|x,0}^{-1}\frac{\partial\mu^{y|x}}{\partial\psi'}\right],$$

where $\Sigma_{y|x,0}$ is the variance-covariance matrix of the LEF density of the conditional model evaluated at θ_0^y, the true parameters. By the properties of LEF it equals $\frac{\partial C_{y|x}}{\partial\mu^{y|x}}$, and

$$I_{y|x,LEF} = E\left[\frac{\partial Q^{LEF}}{\partial\psi}\frac{\partial Q^{LEF}}{\partial\psi'}\right] = E\left[\frac{\partial\mu^{y|x}}{\partial\psi}\Sigma_{y|x,0}^{-1}\Omega_{y|x,0}\Sigma_{y|x,0}^{-1}\frac{\partial\mu^{y|x}}{\partial\psi'}\right],$$

where $\Omega_{y|x,0}$ is the variance-covariance matrix of the true density, evaluated at θ_0^y, of the conditional model. Equivalent matrices can be constructed for η.

From the above PML functions a pseudo-score test can be easily constructed. Note that it is a pseudo-score test because it differs the the classical score test, in the sense that its variance-covariance matrix does not reach the Cramer-Rao bound. The matrices $J_{y|x,LEF}$ and $I_{y|x,LEF}$, as well as the scores, may be partitioned exactly as in ML. Under the null, the test is equivalent to (4) but with different weighting matrix

$$S_{LEF} = -\hat{i}^{\alpha_1}_{y|x}\left(\hat{\psi}\right)' \hat{W}^{\alpha_1}_{y|x}\left(\hat{\psi}\right)^{-1} \hat{i}^{\alpha_1}_{y|x}\left(\hat{\psi}\right), \tag{6}$$

where (omitting the arguments and the subscripts for ease of exposition)

$$\hat{W}^{\alpha_1} = \left[\hat{I}^{\alpha_1} - \hat{I}^{\alpha_1,\theta^{y\dagger}}\hat{J}^{\theta^{y\dagger}-1}\hat{J}^{\theta^{y\dagger},\alpha_1} - \hat{J}^{\alpha_1,\theta^{y\dagger}}\hat{J}^{\theta^{y\dagger}-1}\hat{I}^{\theta^{y\dagger},\alpha_1} + \hat{J}^{\alpha_1,\theta^{y\dagger}}\hat{J}^{\theta^{y\dagger}-1}\hat{I}^{\theta^{y\dagger}}\hat{J}^{\theta^{y\dagger}-1}\hat{J}^{\theta^{y\dagger},\alpha_1}\right]^{-1},$$

such that $\hat{I}$ and $\hat{J}$ are consistent estimators of I and J respectively. They are obtained by replacing the linear operator E by an empirical mean and ψ by $\hat{\psi}$. The test is distributed as $\chi^2(1)$. This expression is quite cumbersome since nothing has been assumed regarding the variance. The weighting matrix of the pseudo-score test differs substantially from that of the ML score test. Only in the case where the variance of LEF density equals the true variance, $\Omega_0 = \Sigma_0$ we will get a variance-covariance matrix that reaches the semiparametric efficiency bound, i.e. $J^{-1}IJ^{-1}$ reduces to

$$J_{y|x,LEF} = E\left[\frac{\partial \mu^{y|x}}{\partial \psi}\Omega^{-1}_{y|x,0}\frac{\partial \mu^{y|x}}{\partial \psi'}\right]. \tag{7}$$

Alternatively, if we have some information about how Ω looks like, we can estimate it consistently in a first step and then replace it in the LEF, giving rise to Generalized PML (GPML) discussed below.

As mentioned before, some of the distributions belonging to the LEF have parameters, θ^{y*} and θ^{x*}, that are not present in the conditional means but in higher moments, in particular in the true variances, denoted by $\nu^{y|x} = \nu^{y|x}\left(\theta^{y*}\right)$ and $\nu^{x} = \nu^{x}\left(\theta^{x*}\right)$. It could even be the case that some of the parameters of the conditional mean are also present in the true variances: $\nu^{y|x} = \nu^{y|x}\left(\theta^{y}\right)$ and $\nu^{x} = \nu^{x}\left(\theta^{x}\right)$. For example, in the case of the Poisson distribution, one does not need to estimate any variance parameter as the variance equals the mean. However, for the negative binomial one needs to estimate a second parameter that is present in the variance but not in the mean; cf. Gourieroux et al.(1984b) for more details on this specific example. In any case, the parameters involved in the variance need to be estimated consistently ad hoc. Denote by $\hat{\nu}^{y|x}$ the variance evaluated in $\hat{\theta^{y*}}$ or $\hat{\theta^{y}}$, which are strongly consistent estimates of θ^{y*} and θ^{y} respectively. Similarly for ν^{x}. The LEF may be extended to the case where the variance is considered to be a nuisance parameter. Such a family is the Generalized LEF (GLEF)

$$q^{GLEF}_{y|x} = \exp\left\{A_{y|x}\left(\mu^{y|x}_i, \hat{\nu}^{y|x}_i\right) + B_{y|x}\left(y_i, \hat{\nu}^{y|x}_i\right) + C_{y|x}\left(\mu^{y|x}_i, \hat{\nu}^{y|x}_i\right)y_i\right\},$$

and equivalently for the marginal density. The generalized-pseudo-log-likelihood (GPML) can be written as

$$Q^{GLEF}\left(\psi, \eta\right) = \log q^{GLEF}_{y|x}\left(y; \mu^{y|x}, \hat{\nu}^{y|x}\right) + \log q^{GLEF}_{x}\left(x; \mu^{x}, \hat{\nu}^{x}\right). \tag{8}$$

If the parameters involved in the variances can be consistently estimated, the estimated GPML parameters are also consistent, asymptotically normal and with variance-covariance matrix, for ψ,

$$I_{y|x,GLEF} = E\left[\frac{\partial \mu^{y|x}}{\partial \psi}\Omega_0^{-1}\frac{\partial \mu^{y|x}}{\partial \psi'}\right],$$

which is, indeed, the lowest bound we can get given that the distribution is not correctly specified. Notice that it is equivalent to the semiparametric efficiency bound which is attained since we know the form of the true variance and we estimate it consistently ad hoc.

As before, a score/LM test can be implemented by partitioning the matrix $I_{y|x,GLEF}$ and evaluating it under H_0, so that test becomes

$$S_{GLEF} = -\hat{i}^{\alpha_1}_{y|x}\left(\widehat{\theta}^y_c\right)' \hat{I}^{\alpha_1}_{y|x,GLEF}\left(\widehat{\theta}^y_c\right)^{-1}\hat{i}^{\alpha_1}_{y|x}\left(\widehat{\theta}^y_c\right) \tag{9}$$

where $\hat{I}_{y|x,GLEF}\left(\widehat{\theta}^y_c\right)$ is a consistent estimator of $I_{y|x,GLEF}$. As before, the test is asymptotically $\chi^2(1)$.

3.2 Specifying the conditional means and the first order conditions

The previous section has highlighted that one of the main advantages of working within the LEF or GLEF is that a correct specification of the mean, even if the distribution if misspecified, leads to estimates that are consistent and normally distributed, albeit not efficient. Nonetheless, this loss of efficiency may be minimized when the variance is correctly specified and estimated consistently. Therefore, the key issue in deriving a score/LM test is the correct specification of the conditional mean. In Section 3.1 we assumed some very general forms of both the conditional and marginal densities. In the sequel, however, we will focus only on the conditional density since it is the one of main interest for most empirical applications. Yet, the same sort of results extend to the marginal distribution.

Recall that the conditional mean has the form

$$\mu_i^{y|x} = g\left[\alpha(\mu_i^x)x_i + \lambda' U_i\right]$$

where, under LD, $\alpha(\mu_i^x) = \alpha_0 + \alpha_1\mu_i^x$. Therefore, the only remaining issue for deriving the weak-exogeneity test is how to specify g. Its functional form depends critically on the support of the data.

For example, if y_i has real support, the most appropriate functional form for g is the identity and the conditional mean is $\mu_i^{y|x} = (\alpha_0 + \alpha_1\mu_i^x)\,x_i + \lambda' U_i$. By constrast, if y_i has positive real support then its conditional expectation must be always positive. In this case, the exponential function is a natural choice for g, i.e. $\mu_i^{y|x} = \exp\left\{(\alpha_0 + \alpha_1\mu_i^x)\,x_i + \lambda_1' U_i\right\}$. Note that is adequate whether y_i is continuous or discrete. In effect, it is commonly used in duration and count regressions. Lastly, if one is interested in modelling the probability that a binary variable, y_i, takes values 0 and 1, then the conditional expectation must be bounded between 0 and 1, leading to the logistic function as a natural choice for g, namely

$$\mu_i^{y|x} = \frac{1}{1+\exp\left\{-\left(\alpha_0+\alpha_1\mu_i^x\right)x_i-\lambda' U_i\right\}}.$$

In this fashion, it is possible to choose convenient functional forms for the conditional mean for almost all the supports we may encounter in practice. For instance, if it happens to be positive discrete with a finite number of values, the multinomial-type conditional mean is the appropriate one, but this is just a straightforward extension of the binomial case discussed above. By assuming a particular density in accord with the nature of the support of the variable (namely Gaussian, exponential, binomial, multinomial or Poisson), the first-order conditions are the same, up to some weighting factor depending on nuisance parameters. For example, if y_i has real, positive discrete, positive continuous or binary support, we can assume a Gaussian, Poisson, exponential and binomial distributions, respectively. Thus, computation of the corresponding FOC's yields

$$\sum_{i=1}^{N}\frac{\partial\mu_i^{y|x}}{\partial\psi}\frac{1}{\sigma^2}\left[y_i-\mu_i^{y|x}\right]=0 \tag{10}$$

for a Gaussian distribution and

$$\sum_{i=1}^{N}\frac{\partial\mu_i^{y|x}}{\partial\psi}\left[y_i-\mu_i^{y|x}\right]=0 \tag{11}$$

for the other densities afore mentioned. In all cases the FOCs can be interpreted as the sum of residuals weighted by some function depending on the nuisance parameters and the derivative of the conditional mean with respect to ψ. They are indeed the same FOCs as in a weighted least squares problem. Hence, if we specify correctly the conditional mean and the choice of the density is constrained to the exponential family, the FOC's are equivalent, up to some weight, to the ones obtained in the Gaussian case.

This implies that, when constructing the score test, we can always use the same score function $i_{y|x}^{\alpha_1}\left(\widehat{\theta}_c^y\right)$. Unfortunately, the same result does not hold for the Hessian since, for a general LEF or GLEF, the second order condition is $\frac{\partial^2 A_{y|x}\left(\mu^{y|x}\right)}{\partial\psi\partial\psi'}+\frac{\partial^2 C_{y|x}\left(\mu^{y|x}\right)}{\partial\psi\partial\psi'}y$ which depends on the functional forms of $A_{y|x}\left(\mu^{y|x}\right)$, $C_{y|x}\left(\mu^{y|x}\right)$ and g.

In Section 3.3 we provide several specific examples of the effects of different choices of the previous functional forms and compute the score/LM test for weak exogeneity explicitly in each case. We also show that the LM test may be also interpreted as a t-test in an augmented regression model.

Some final clarifying remarks on consistency and efficiency are in order now. Throughout the paper we have stressed that as far as i)the conditional mean is correctly specified, and ii)the parameters of interest are in the first moment, the estimates are consistent in the PML framework. Otherwise, that is if the conditional mean is incorrectly specified, the estimates are inconsistent. The test that we propose is based on the correct specification of the conditional mean. For example, suppose that there is weak exogeneity, then $\mu_i^{y|x} = g\left[\alpha_0 x_i + \lambda' U_i\right]$ is a correctly specified conditional mean and the PML estimators of α_0 and λ are consistent. However, suppose that weak exogeneity does not hold but we estimate *as if* there were. Then $\mu_i^{y|x} = g\left[\alpha_0 x_i + \lambda' U_i\right]$ would be a misspecified conditional mean leading to inconsistent $\hat{\alpha}_0$ and $\hat{\lambda}$, as they converge to some pseudo-true values that depend on α_1. Hence, testing for weak exogeneity in a PML framework, where the parameters of interest are in the conditional mean, boils down to testing consistency. The fact that, in a parametric framework and under the alternative of lack of weak exogeneity, the estimates of ψ are not consistent while they are consistent under the null, opens the possibility of using a Hausman-type test. This is certainly an interesting avenue to explore. However, since our goal in this Section is to derive a parametric test under the null and, in contrast to the LM test, the Hausman test requires estimation under both the unconstrained and constrained parametric models, we do not pursue this line research here. Nonetheless, and in spite of that, in next Section we propose a semiparametric likelihood ratio test, which needs to estimate the model under both the null and the alternative.

3.3 Examples

In this Section we show two examples of some functional forms for the conditional and marginal densities. As a benchmark we consider the well-known case where both are Gaussian, giving rise to the classical t-test form of the LM test. The first example mixes two different distributions, Poisson and Gaussian for the conditional and marginal, respectively. Therefore, the resulting test would be the same if the marginal density would any other than Gaussian, though the marginal density must belong to the exponential family and its conditional mean needs to be correctly specified. The second example, closely related to the empirical applications below, considers the case of two exponential densities.

3.3.1 Benchmark: Normal densities with linear model

As a benchmark, we start by reviewing the simplest case where it is assumed that the conditional and marginal densities are Gaussian and that the conditional means are linear, i.e. g and h are the identity. This case has been worked out by Boswijk and Urbain (1997), where they examine how to test for weak exogeneity of variables for some set of parameters in cointegrated systems with Gaussian error terms, and also by Engle and Hendry (1993), where strict exogeneity is tested by augmenting the conditional mean with dummies capturing structural breaks (policy interventions) in the marginal process. The functions A, B and C in $q_{y|x}^{LEF}\left(y, \mu^{y|x}\right)$ are $A_{y|x}\left(\mu_i^{y|x}\right) = -(\log \sigma\sqrt{2\pi} + (\mu_i^{y|x}/2\sigma^2))$, $B_{y|x}\left(y_i\right) = -y_i^2/2\sigma^2$ and $C_{y|x}\left(\mu_i^{y|x}\right) = 2\mu_i^{y|x}/\sigma^2$, and equivalently for $q_x^{LEF}\left(x, \mu^x\right)$. Then the quasi log-likelihood function is

$$Q^{GLEF} \propto -\frac{1}{2\widehat{\sigma}_y^2}\sum_{i=1}^{N}\varepsilon_i^2 - \frac{1}{2\widehat{\sigma}_x^2}\sum_{i=1}^{N}\vartheta_i^2$$

where $\varepsilon_i = \left(y_i - \mu_i^{y|x}\right)$ and $\vartheta_i = (x_i - \mu_i^x)$, $\mu_i^{y|x} = \left(\alpha_0 + \alpha_1 \mu_i^{y|x}\right) x_i + \lambda' U_i$ and $\mu_i^x = \beta' U_i$, and $\widehat{\sigma}_y^2$ and $\widehat{\sigma}_x^2$ are strongly consistent estimates of the variances. In this simple case, those authors have shown that the score corresponds to that of an augmented regression when we augment the conditional mean $\mu_i^{y|x} = \alpha_0 x_i + \lambda' U_i$ to $\mu_i^{y|x} = (\alpha_0 + \alpha_1 \mu_i^x) x_i + \lambda' U_i$, so that the LM test becomes

$$S_{GLEF} = \hat{\alpha}_1' Var(\hat{\alpha}_1)^{-1} \hat{\alpha}_1,$$

which of course is equivalent to a t-test for $H_0 : \alpha_1 = 0$ in the augmented regression estimated by OLS (see Boswijk and Urbain, 1997, p 32).[6]

3.3.2 Poisson and Gaussian densities with exponential and linear models

In this example we assume that the conditional and marginal densities are Poisson and Gaussian respectively. Therefore, the appropriate conditional means are exponential and linear respectively, i.e. g is the exponential function and h is the identity. The functions A, B and C in $q_{y|x}^{LEF}(y, \mu^{y|x})$ are now $A_{y|x}\left(\mu_i^{y|x}\right) = -\mu_i^{y|x}$, $B_{y|x}(y_i) = -\log y_i$ and $C_{y|x}\left(\mu_i^{y|x}\right) = \log \mu_i^{y|x}$. And for $q_x^{LEF}(x, \mu^x)$ the expressions are like in the the previous example. Hence, the quasi log-likelihood function is

$$Q^{GLEF} \propto -\sum_{i=1}^{N} \mu_i^{y|x} + y_i \log \mu_i^{y|x} - \frac{1}{2\widehat{\sigma}_x^2} \sum_{i=1}^{N} \vartheta_i^2,$$

where $\mu_i^{y|x} = \exp\left\{\left(\alpha_0 + \alpha_1 \mu_i^{y|x}\right) x_i + \lambda' U_i\right\}$ and $\mu_i^x = \beta' U_i$ and $\widehat{\sigma}_x^2$ is a strongly consistent estimate of the variance of x. The score and the hessian, under the null hypothesis, are

$$\hat{\imath}\left(\hat{\theta}_c\right) = \begin{pmatrix} \hat{\imath}_{y|x}\left(\hat{\theta}_c\right) \\ \hat{\imath}_x\left(\hat{\theta}_c\right) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \sum_{i=1}^{N} \varepsilon_i x_i \mu_i^x \\ 0 \end{pmatrix}$$

and

$$\hat{I}\left(\hat{\theta}_c\right) = -\begin{pmatrix} \sum_{i=1}^{N} \mu_i^{y|x} U_i^2 & \sum_{i=1}^{N} \mu_i^{y|x} x_i U_i & \sum_{i=1}^{N} \mu_i^{y|x} x_i U_i \mu_i^x & 0 \\ \sum_{i=1}^{N} \mu_i^{y|x} x_i U_i & \sum_{i=1}^{N} \mu_i^{y|x} x_i^2 & \sum_{i=1}^{N} \mu_i^{y|x} x_i^2 \mu_i^x & 0 \\ \sum_{i=1}^{N} \mu_i^{y|x} x_i \mu_i^x U_i & \sum_{i=1}^{N} \mu_i^{y|x} x_i^2 \mu_i^x & \sum_{i=1}^{N} \mu_i^{y|x} (x_i \mu_i^x)^2 & -\sum_{i=1}^{N} \mu_i^x x_i U_i \varepsilon_i \\ 0 & 0 & -\sum_{i=1}^{N} \mu_i^x x_i U_i \varepsilon_i & \sum_{i=1}^{N} \mu_i^x U_i^2 \end{pmatrix},$$

[6] Note that this result remains unchanged if we replace the marginal Gaussian density by any other distribution since, under H_0, the score test is only based on the conditional density rather than on the joint density.

where the rows in the score are the FOC's with respect to λ, α_0, α_1 and β respectively. Asuming that $-\sum_{i=1}^{N}\mu_i^x x_i U_i \varepsilon_i \to 0$,[7] the score test is equivalent to the sum of two score tests, one for the conditional density and another for the marginal density.

To show that it yields the t-test, we simplify the problem considering only the score and the hessian for α_0 and α_1, since this simplification reduces greatly the complexity of the matrices. The hessian for the conditional density is now

$$\hat{I}_{y|x}\left(\hat{\theta}_c\right) = -\begin{pmatrix} \sum_{i=1}^{N}\mu_i^{y|x}x_i^2 & \sum_{i=1}^{N}\mu_i^{y|x}x_i^2\mu_i^x \\ \sum_{i=1}^{N}\mu_i^{y|x}x_i^2\mu_i^x & \sum_{i=1}^{N}\mu_i^{y|x}\left(x_i\mu_i^x\right)^2 \end{pmatrix},$$

where the inverted bottom-right element is

$$\hat{I}_{y|x}^{\alpha_1}\left(\hat{\theta}_c\right)^{-1} = -\frac{\sum_{i=1}^{N}x_i^2\mu_i^{y|x}}{-\sum_{i=1}^{N}\left(\mu_i^{y|x}x_i^2\mu_i^x\right)^2 + \sum_{i=1}^{N}\mu_i^{y|x}x_i^2\sum_{i=1}^{N}\mu_i^{y|x}x_i^2\mu_i^{x^2}},$$

which corresponds to the variance-covariance matrix of generalized least squares estimators, with weighting matrix equal to $\mu_i^{y|x}$ in the augmented regression. It is equivalent to the Gaussian case with heteroskedastic variance equal to $\mu_i^{y|x}$, as the variance equals the mean in a Poisson distribution. This is in line with with the above-mentioned result that the FOCs are the same, up to some factor depending on the higher moments, for all the densities in the exponential family. The score test in this case is

$$S_{LEF} = \left(\sum_{i=1}^{N}\varepsilon_i x_i \mu_i^x\right)^2 \hat{I}_{y|x}^{\alpha_1}\left(\hat{\theta}_c\right)^{-1}.$$

3.3.3 Exponential densities with exponential models

Lastly, we provide an example that we will use in the empirical applications. The densities are both exponential and the conditional means are exponential. Thus, the functions A, B and C in $q_{y|x}^{LEF}\left(y,\mu^{y|x}\right)$ are $A_{y|x}\left(\mu_i^{y|x}\right) = -\log\mu_i^{y|x}$, $B_{y|x}\left(y_i\right) = 0$ and $C_{y|x}\left(\mu_i^{y|x}\right) = -1/\mu_i^{y|x}$. And likewise for $q_x^{LEF}\left(x,\mu^x\right)$. The quasi log-likelihood function is

$$Q^{GLEF} = -\sum_{i=1}^{N}\log\mu_i^{y|x} + \frac{y_i}{\mu_i^{y|x}} + \log\mu_i^x + \frac{x_i}{\mu_i^x},$$

where $\mu_i^{y|x} = \exp\left\{\left(\alpha_0 + \alpha_1\mu_i^{y|x}\right)x_i + \lambda' U_i\right\}$ and $\mu_i^x = \exp\left\{\beta' U_i\right\}$. Following the same steps as in previous example, the score test is

$$S_{LEF} = \left(\sum_{i=1}^{N}\left(\frac{y_i}{\mu_i^{y|x}} - 1\right)\mu_i^x x_i\right)^2 \hat{I}_{y|x}^{\alpha_1}\left(\hat{\theta}_c\right)^{-1} \qquad (12)$$

[7] Boswijk and Urbain, 1997, justify this condition in the Gaussian case.

where $\hat{I}_{v|d}^{\alpha_1}\left(\hat{\theta}_c\right)^{-1}$ is the bottom right side element of the inverse of

$$\hat{I}_{y|x}\left(\hat{\theta}_c\right) = -\begin{pmatrix} \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} U_i^2 & \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} x_i U_i & \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} x_i U_i \mu_i^x \\ \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} x_i U_i & \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} x_i^2 & \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} x_i^2 \mu_i^x \\ \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} x_i \mu_i^x U_i & \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} x_i^2 \mu_i^x & \sum_{i=1}^N \frac{y_i}{\mu_i^{y|x}} \left(x_i \mu_i^x\right)^2 \end{pmatrix}.$$

3.4 The case when ψ is present in the variance

So far it has been assumed that either the variance of the conditional density is fully known, as in PML, or that it has a known functional form, which depends upon parameters that can be consistently estimated ad hoc, as in GPML. Thus, in either case, the parameters of interest, ψ, are only present in the conditional mean. However, we may find some cases, like in volatility models, where the parameters of interest are in the conditional variance. In this case, the weak-exogeneity test is no longer based on the first moment but on the second moment.[8] Further, a more general case can be envisaged when the parameters of interest are in both the mean and variance. More precisely, let

$$\mu_i^{y|x} = g_1\left[\alpha_1\left(\mu_i^x, \nu_i^x\right) x_i + \lambda' U_i\right] \text{ and } \nu_i^{y|x} = g_2\left[\alpha_2\left(\mu_i^x, \nu_i^x\right) h\left(x_i\right) + \eta' U_i\right],$$

where g_1 and g_2 are known functions.[9] Thus, for the conditional density we can use the quadratic exponential family, QEF, defined by

$$\begin{aligned} q_{y|x}^{QEF} &= \exp\left\{A_{y|x}\left(\mu_i^{y|x}, \nu_i^{y|x}\right) + B_{y|x}\left(y_i, \nu_i^{y|x}\right)\right. \\ &\left. + C_{y|x}\left(\mu_i^{y|x}, \nu_i^{y|x}\right) y_i + y_i' D_{y|x}\left(\mu_i^{y|x}, \nu_i^{y|x}\right) y_i\right\}, \end{aligned}$$

and, equivalently, for the marginal density

$$q_x^{QEF} = \exp\left\{A_x\left(\mu_i^x, \nu_i^x\right) + B_x\left(x_i, \nu_i^x\right) + C_x\left(\mu_i^x, \nu_i^x\right) x + x_i' D_x\left(\mu_i^x, \nu_i^x\right) x_i\right\}.$$

The quadratic-pseudo-log-likelihood (QPML) is now

$$Q^{QEF}\left(\theta^y, \theta^x\right) = \log q_{y|x}^{QEF}\left(y; \mu^{y|x}, \nu^{y|x}\right) + \log q_x^{QEF}\left(x; \mu^x, \nu^x\right),$$

where α_1 and α_2 are now functions of μ^x and ν^x. Following Engle and Hendry (1993, p. 125) we assume for the conditional mean

$$\alpha_1\left(\mu_i^x, \nu_i^x\right) = \alpha_{1,0} + \alpha_{1,1}\mu_i^x + \alpha_{1,2}\left(\mu_i^x\right)^2 + \alpha_{1,3}\nu_i^x + \alpha_{1,4}\nu_i^x \mu_i^x,$$

[8] For ease of exposition we still assume that the estimation of the marginal model focusses on the mean and not on the variance. However, this assumption may be relaxed by allowing for dynamic conditional variance.

[9] Notice the difference between x_i and $h(x_i)$. The latter function appears in the conditional variance. For example, we may use $h(x_i) = x_i^2$ instead x_i.

and for the conditional variance we propose

$$\alpha_2\left(\mu_i^x,\nu_i^x\right)=\alpha_{2,0}+\alpha_{2,1}\left(\mu_i^x\right)^2+\alpha_{2,2}\nu_i^x+\alpha_{2,3}\nu_i^x\left(\mu_i^x\right)^2.$$

The estimates that maximize Q^{QEF} are consistent and asymptotically normal with variance-covariance matrix that can be found in Gourieroux et al. (1984a, p. 697).[10]

In this setup, testing for weak exogeneity reduces to test the set of restrictions $\alpha_{1,1}=\alpha_{1,2}=\alpha_{1,3}=\alpha_{1,4}=\alpha_{2,1}=\alpha_{2,2}=\alpha_{2,3}=0$. The form of the correponding test is equivalent to LEF with the appropriate variance-covariance expressions, leading to a test-statistic with a limiting χ^2 distribution with the corresponding degrees of freedom (equal 7, in the previous case).

4 A CONSISTENT TEST FOR WEAK EXOGENEITY

The problem of choosing a functional form for $\alpha(\cdot)$ appears in many examples. In Section 2, to implement the score test we propose LD, i.e, $\alpha(\mu_i^x)=\alpha_0+\alpha_1\mu_i^x$. In principle, there is no reason to choose this functional form. One might choose other alternatives such as $\alpha(\mu_i^x)=\alpha_0+\alpha_1\mu_i^x+\alpha_1(\mu_i^x)^2$ or any other known relationship. In Section 3.4 we have the same choice to make, where we have quadratic expressions for $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ in terms of μ_i^x and ν_i^x. But, again, other possible alternatives are suitable. It is important to realize that if the functional form for $\alpha(\cdot)$ is not correctly specified, the particular score test developed in the previous section may reject the weak exogeneity assumption when indeed the data generating process might present this feature.

To prevent against this failure in our testing procedure we propose to leave unspecified the functional form for $\alpha(\cdot)$. In the null hypothesis we will test that $\alpha(\cdot)$ is a constant function, that is, it does not depend on μ_i^x (weak exogeneity assumption). The alternative will be the complementary set, i.e. any type of dependence, $\alpha\left(\mu_i^x\right)$ (rejection of weak exogeneity assumption). Although other possible tests based in moment conditions are available, we propose a very simple test based in the Pseudo-likelihood ratio principle. Thus, under the null hypothesis we will compute the pseudo-log-likelihood function by imposing that $\alpha\left(\mu_i^x\right)=\alpha$ for any i. In fact, α and λ can be estimated under PML, as in previous section. Unfortunately, estimation under the alternative is much more cumbersome. Under the alternative hypothesis the conditional mean takes the following form

$$\mu_i^{y|x}=g\left[\alpha\left(\mu_i^x\right)x_i+\lambda'U_i\right],\quad i=1,\cdots,N,$$

and the pseudo-likelihood function for the i-th observation is $q_{y|x}^{LEF}\left(y_i,g\left[\alpha\left(\mu_i^x\right)+\lambda'U_i\right]\right)$. Note that standard pseudo-maximum likelihood estimation of λ may yield inconsistency and slow rates of convergence (for examples of inconsistency, see Kiefer and Wolfowitz, 1956, and Grenander, 1981, whereas for cases of slow rate of convergence, see Shen and Wong, 1994 and Birgé and Massart, 1994) in the presence of infinite "incidental" parameters, $\alpha\left(\mu_1^x\right),\cdots,\alpha\left(\mu_N^x\right)$.

In order to solve these proble ms many alternative solutions have been proposed. If the function $\alpha(\cdot)$ is estimated through splines, series, neural networks or wavelets then the method

[10]Since we do not use this family, and the variance-covariance expressions are quite cumbersome, we refer to the reader to the appropriate reference.

of sieve extremum estimation can be used. Chen and Shen (1998) give sufficient conditions for splines and Fourier series under regularly spaced dependent data. If, instead, other methods such as kernels or local polynomials are used, then the sieve method is no longer valid and other non-sieve ML estimation methods are recommended. Among others, we adopt here the so called generalized profile likelihood approach (see Severini and Wong, 1992 and Severini and Staniswalis, 1994). Taking the approach introduced in a different context by Cai, Fan and Li (2000) we proceed first to estimate $\alpha(\mu_i^x)$ using a local linear modelling scheme. The local linear fittings have several nice properties, such as high statistical efficiency, design adaptation and good boundary behavior (cf. Fan and Gijbels, 1996). Under the assumption that $\alpha(\mu_i^x)$ has a continuous second derivative, for each given point μ_0^x, it can be locally approximated by a linear function $\alpha(\mu_i^x) \approx a + b\,(\mu_i^x - \mu_0^x)$ for μ_i^x in a neighborhood of μ_0^x. Then, the following smooth pseudo-likelihood function is defined

$$g_{y|x}^{LEF}(a,b) = \sum_{i=1}^{N} \log\, q_{y|x}^{LEF}\left(y_i; g\left[\left\{a + b'\left(\mu_i^x - \mu_0^x\right)\right\} x_i + \lambda' U_i\right]\right) \frac{1}{h} K\left(\frac{\mu_i^x - \mu_0^x}{h}\right),$$

where $K(\cdot)$ is a kernel function, h is the bandwidth, $a = \alpha(\mu_0^x)$ and $b = \frac{\partial \alpha}{\partial \mu^{x\prime}}\Big|_{\mu^x = \mu_0^x}$. For given values of μ_i^x (estimated previously from the marginal model), λ and h the estimator of $\alpha(\mu_0^x)$ is defined as

$$\left(\hat{a}_{\lambda,h}, \hat{b}_{\lambda,h}\right) = \left(\hat{\alpha}_{\lambda,h}(\mu_0^x), \left.\frac{\partial \hat{\alpha}(\mu^x)}{\partial \mu^{x\prime}}\right|_{\mu^x = \mu_0^x, \lambda, h}\right) = \mathrm{argmax}_{a,b} \quad g_{y|x}^{LEF}(a,b),$$

and the local linear regression estimator is $\hat{\alpha}_{\lambda,h}(\mu_0^x) = \hat{a}_{\lambda,h}$. Repeating this procedure for $i = 1, \cdots, N$ we obtain N estimates of the unknown curve, $\hat{\alpha}_{\lambda,h}(\mu_1^x), \cdots, \hat{\alpha}_{\lambda,h}(\mu_N^x)$.

The estimator for the parametric part is obtained by maximizing the un-smooth likelihood function

$$\hat{\lambda} = \mathrm{argmax}_{\lambda} \quad \sum_{i=1}^{N} \log\, q_{y|x}^{LEF}\left(y_i; g\left[\hat{\alpha}_{\lambda,h}(\mu_i^x) x_i + \lambda' U_i\right]\right).$$

As in Section 3.2, assuming a convenient functional form for the conditional mean and the density in accord with the nature of the support of the variable, the FOCs for the local pseudo likelihood method are equivalent to (10) and (11) but now with the kernel function playing the role of an additional weighting factor:

$$\sum_{i=1}^{N} \frac{1}{h} K\left(\frac{\mu_i^x - \mu_0^x}{h}\right) \frac{\partial \mu_i^{y|x}}{\partial \psi} \frac{1}{\sigma^2} \left[y_i - \mu_i^{y|x}\right] = 0$$

for a Gaussian distribution and

$$\sum_{i=1}^{N} \frac{1}{h} K\left(\frac{\mu_i^x - \mu_0^x}{h}\right) \frac{\partial \mu_i^{y|x}}{\partial \psi} \left[y_i - \mu_i^{y|x}\right] = 0$$

for the other densities.

Under fairly standard conditions it is possible to show (see Cai, Fan and Li, 2000; Theorem 1 and Severini and Staniswalis, 1994; Proposition 1) that both $\left\{\hat{\alpha}_{\hat{\lambda},h}(\mu_i^x)\right\}_{i=1}^N$ and $\hat{\lambda}$ are consistent and asymptotically normal under the appropriate rates.

Therefore, the following nonparametric likelihood ratio test statistic is proposed

$$\begin{aligned} LR &= 2\left\{\sum_{i=1}^N \log q_{y|x}^{LEF}\left(y_i; g\left[\hat{\alpha}_{\hat{\lambda},h}(\mu_i^x)x_i + \hat{\lambda}'U_i\right]\right)\right. \\ &\quad \left. - \sum_{i=1}^N \log q_{y|x}^{LEF}\left(y_i; g\left[\hat{\alpha}x_i + \hat{\lambda}'U_i\right]\right)\right\}. \end{aligned} \tag{13}$$

For parametric models, the likelihood ratio statistic follows asymptotically a chi-square distribution with degrees of freedom $m - n$, where m and n are the number of parameters under the alternative and the null hypothesis respectively. Instead, when a nonparametric alternative is present the number of parameters tends to infinity. Therefore, the asymptotic distribution of LR will be gaussian and it will be independent of the parameters of interest.[11] This suggests the use of a conditional bootstrap to construct the null distribution of LR (see Cai, Fan and Li, 2000).

Finally, we briefly describe the bootstrap procedure. Since these models are dynamic with long memory, standard bootstrap is unfeasible. We rely on Hall and Yao (2003) that show how to bootstrap from a GARCH model with heavy tails.

1. Conditionally on the values of $\{x_i, U_i\}_{i=1}^N$ consider the pseudo-maximum likelihood estimators of α and λ under the hypothesis of weak exogeneity, $\hat{\lambda}$ and $\hat{\alpha}$.
2. Compute the residuals of the model, $\tilde{\varepsilon}_i$.
3. Draw ε_i^* by sampling randomly, with replacement, from $\tilde{\varepsilon}_i$.
4. Finally, compute the conditional means and the bootstrap sample y_i^*. We compute the test statistic LR^* in (13), substituting y_i^* for y_i.

Note that although this test has been implemented based on PML estimation techniques, it can be easily extended to account for quadratic exponential families.

5 APPLICATIONS TO HIGH FREQUENCY DATA

In this section we apply the proposed testing approach for weak exogeneity to high- frequency data. More precisely, we model tick-by-tick data of five stocks, traded at the New York Stock Exchange (NYSE): Boeing, Coca Cola, Disney, Exxon and IBM. The data were extracted

[11] For the rigorous justification of this argument we refer to the articles by Shen and Wong (1994) and Shen, Shi and Wong (1999) that considered nonparametric likelihood ratio tests in a very general setting.

from the trades and quotes (TAQ) database pertaining to September, October, and November 1996. Trades and bid/ask quotes recorded before opening (9:30 am) and after closing (4:00 pm) were not used. This database consists of two parts: the first reports all trades, while the second lists the bid and ask prices posted by the specialist. Hence, it contains a great deal of information, as we may compute different types of durations (trade, price, volume) and the associated variables to each trade and quote (ask, bid, price, volume size, etc).[12]

This database has been widely used in the literature; cf. Engle (2000), Engle and Russell (1998) or Bauwens and Veredas (2004). In particular, Engle (2000) analyzes price volatility, using transaction data, *conditional* on current trade duration. The marginal density is modelled as an ACD type model (see below for a brief explanation of this model) while a GARCH model is estimated for the conditional density of price changes (measured by the midquote). The latter model includes the reciprocal of current duration and current expected duration, past long-run volatility and dummy variables for spread and trade size. From the viewpoint of our application, his most significant finding is that both current duration and expected duration are (significantly) negatively related with volatility. An increase in the trading activity, reflected by a shrunk of trade durations, implies an increase in volatility. By contrast, Engle and Russell (1998) analyze the conditional density of price durations (midquotes as well) *conditional* on several microstructure variables. They are interested in analyzing the argument made by Easley and O'Hara (1992) about how the number of transactions influences the price duration process. More concretely, they focus on the link between price durations and spread and volume per transaction. They find that the higher the volume is, the smaller the price duration is. A viceversa for the spread.

Here, despite using Engle's (2000) factorization, we adopt a different approach. Specifically, we study how market participant's reactions on volume, measured as trade size and volume per price changes, is affected by the timing of trades and price changes respectively. Timing is measured by the durations, denoted by d. In fact, in our first application we model trade size, defined as the volume traded in each transaction, *conditional* on trade durations, defined as the time intervals between consecutive trades. Volume size and trade durations are closely related. As a matter of fact, trade durations are a measure of the trading activity, i.e. the lower the duration, the higher the trading activity. The analysis of this relation may shed some light on the relation between trade size and trading intensity. In fact, two antagonist views may be defended. One may argue that informed traders are impatient as they are eager to trade quickly with large trade sizes as the private information they posses is highly perishable. By contrast, stealth trading theorists and practitioners, cf. Barclay et al. (1993) and Chakravarty (2001), argue that informed traders concentrate their trades of relatively small sizes (fragmentation of trade size), in order to not fully disclose their private information.

In our second application, we model volume, defined as accumulated share volume within a price durations, conditional on the latter duration. It is defined as the minimum duration that is required to observe a price change not less than a given amount. The price we focus on is the mid-price of the specialist's quote, i.e. the average of the bid and ask prices, and the threshold is equal to \$0.125 (in 1996 the minimum tick size was \$0.0625 and there was numerous tick changes in the mid price; they are not taken into account as they are mainly

[12]See Bauwens and Giot (2001) for further details on the TAQ database and on the functioning of the NYSE.

due to the short term component of the bid-ask quote updating process). The volume size is divided by the number of transactions within the duration. Otherwise, it will increase with the duration. Price durations are a measure of instantaneous volatility (see Engle and Russell, 1998, p 1153). Periods of high (low) volatility are related with periods of agitated (serene) activity and hence with periods of short (long) price durations. In effect, this second model tackles with the well-known issue of Clark's Mixture of Distribution Hypothesis (MDH) that establishes a joint dependence of returns and volume, as it hypothesize that both variables are driven by some unobserved information flow; cf. Clark (1973. Empirically a number of papers have found a positive relation between volume and volatility; cf. Lamoreux and Lastrapes (1990), and Giot (2000). Andersen (1996) develops a modified MDH where the covariance between square returns and volume size is positive, being a function of the common dependence of the information flow modeled by the intensity of information arrivals during the day. In a financial point process, price durations are a good proxy of the instantaneous volatility, see Bauwens and Giot (2001). We should therefore expect, following the above common dependence, that the parameters of the conditional and marginal densities derived from the joint process of volume size and price durations would not satisfy the property of free variation.

[TABLE 1 ABOUT HERE]

Descriptive statistics about the dataset are shown in Table 1. The number of observations is much greater for the trade process (ranging from 60,454 for IBM to 23,930 for Boeing) than for the price durations (ranging from 6,128 for IBM and 1,609 for Coca Cola). Durations are overdispersed, i.e. standard deviations are larger than the mean. These dispersion ratios cannot be captured by the exponential distribution. However, this is not a major concern for us since we will only focus on the specification of the means in both the conditional and marginal densities. The same reasoning applies to the two volume variables.

The sequences of trade size and volume per price change, as well as trade and price durations, are shown in the top panel of Figures 1 and 2 (because of space limitations we only report the plots for Boeing, but figures for other stocks are very similar). As is conventional with this kind of data, clustering of small and large durations can be observed, creating serial dependence. This can also be seen also through Spearman's correlation coefficients for serial dependence shown below the previous plots.

Volumes and durations also have a strong seasonal intra-daily pattern which should be considered when modelling both processes. In fact, both processes can be thought of as consisting of two parts: a stochastic component to be explained by some dynamic model, and a deterministic part, namely the seasonal intra-daily pattern. This effect arises from the systematic variation of the market activity during the day. We model time-of-day adjusted volumes $v_i = V_i/\phi_i^v\left(t_i^{'}\right)$ and durations $d_i = D_i/\phi_i^d\left(t_i^{'}\right)$, where $\phi_i^v\left(t_i^{'}\right)$ or $\phi_i^d\left(t_i^{'}\right)$ are the time-of-day effects at time $t_i^{'}$, the number of accumulated seconds since the opening, and V_i and D_i are the observed volume and durations. They are estimated by a nonparametric regression

$$\phi_i^v\left(t_0^{'}\right)=\frac{\frac{1}{Nh}\sum_{i=1}^{N}K\left(\frac{t_0^{'}-t_i^{'}}{h}\right)V_i}{\frac{1}{Nh}\sum_{i=1}^{N}\left(\frac{t_0^{'}-ti^{'}}{h}\right)} \tag{14}$$

where $K(\cdot)$ is a quartic kernel and $h = 2.78SN^{-1/5}$, S being the sample standard deviation. Equivalently for $\phi_i^d\left(t_i^{'}\right)$. Plots in the bottom of Figures 1 and 2 show the seasonal patterns for volume (left-hand side) and durations (right-hand side). The central line is the estimated seasonal pattern and the side lines show the 95% confidence intervals; cf., see Veredas et al. (2001) for the computations of these intervals. Durations exhibit an inverted U-shaped seasonality, meaning that trading activity during the opening and the closing is higher than for the rest of the day. The seasonal patterns of trade size and volume size per price duration present a decreasing pattern, meaning that the number of traded assets during the opening is higher than during the closing.

[FIGURES 1 AND 2 ABOUT HERE]

Once the most salient features of the data have been described, we now turn to the application of the proposed testing procedure. The process duration-volume is a point process, namely a stochastic process in which the position in the space of each realization is stochastic. It implies that not only the value of the variable we are interested in is stochastic, but also its position in the space. This space, in our case, is uni-dimensional and it is indeed the time line. Therefore, our observed process is a stochastic process in which each realization consists of two values: the moment of time in which the trade, or price change, occurred and the corresponding volume. In other words, the vector $\{z_i\}_{i=1}^N = \{y_i, x_i\}_{i=1}^N$ consists of a duration process, say x_i, and a volume process associated to each duration, y_i.

As discussed in the Introduction, for a financial point process only one factorization of the joint density is correct: the volume process must be conditional to the duration process. In other words, this means that in factorization (1) y_i is the mark and x_i the duration, that is, the same factorization as in Engle (2000). To homogenize notation hereafter $y_i = v_i$ and $x_i = d_i$.

The next step deals with how to specify the densities. Under the null of weak exogeneity, the conditional and the marginal models can be estimated independently. For ease of exposition we assume that the parameters of interest are only in the first moment (and hence we can use PML or GPML, but not QPML) and that $\dim(\theta^{v*}) = 0$. This implies that the variance is taken to be constant, or identical to the mean as is the case of the exponential or Poisson distributions. For the marginal density, we also assume that $\dim\left(\theta^{d*}\right) = 0$. However, as explained in Section 3.4, extensions to the case where parameters are present in higher moments are straightforward, making use of QLEF.

Since durations are strictly positive, an exponential conditional mean, i.e. $h(\cdot) = \exp(\cdot)$, is the most natural choice within the exponential family.[13] Further, we assume that the conditional mean of the durations follows an accelerated time model, $d_i = \mu_i^d \varepsilon_i^d = \exp\left(\beta U_i^d\right)\varepsilon_i^d$, where ε_i follows an exponential distribution with parameter equal to unity. This implies that durations also follow an exponential distribution with parameter $1/\exp\left(\beta U_i^d\right)$. Additionally, to account for the serial correlation, see Figure 1, we introduce lagged values of d in U_i^d. More specifically, under an exponential density, we adopt the Log-ACD(1,1) model proposed by Bauwens and Giot (2000), namely $U_i^d = \left(1, \ln d_{i-1}, \mu_{i-1}^d\right)'$. Therefore, the conditional mean to estimate is $\mu_i^d = \exp\left(\beta_1 + \beta_2 \ln d_{i-1} + \beta_3 \mu_{i-1}^d\right)$.

Trade size and volume per price duration also have positive and continuous support. Again an exponential distribution is a reasonable choice. Thus, we assume an exponential conditional mean. As with durations, trade size and volume per price duration are serially correlated (see Figures 1 and 2).[14] We choose a "volume" version of the Log-ACD model, the so-called Log-ACV model, whose plain version, ACV, was introduced by Manganelli (2002). Therefore, the Log-ACV model, including the component that is used to test weak exogeneity, is $v_i = \mu_i^{v|d}\varepsilon_i^v = \exp\left(\left(\alpha_0 + \alpha_1 \mu_i^d\right) d_i + \lambda U_i^v\right)\varepsilon_i^v$, where ε_i^v is exponentially distributed with parameter equal to 1. For U_i^v we chose a specification that contains lagged information on volume and a constant term, i.e. $U_i^v = \left(1, \ln v_{i-1}, \mu_{i-1}^{v|d}\right)'$. Therefore, the conditional mean to estimate is $\mu_i^{v|d} = \exp\left(\left(\alpha_0 + \alpha_1 \mu_i^d\right) d_i + \lambda U_i^v\right)$, where $\lambda = (\lambda_1, \lambda_2, \lambda_3)$. As both densities are exponential and the conditional expectations are also exponential, the parametric score test is (12). In turn, the nonparametric likelihood ratio test is (13), where the number of draws in the bootstrap is 500.

Estimates of the Log-ACV and Log-ACD models, under the null, are reported in Table 2. For the Log-ACD (bottom panel), they are in tune with the results found in the literature. All the parameters are significant and those that account for persistence, μ_{i-1}^d, are large, specially for trade durations. The sum $\beta_2 + \beta_3$ is very close to unity, meaning that the process is very persistent.

[TABLE 2 ABOUT HERE]

Regarding the Log-ACV model (top panel), some interesting conclusions can be drawn if weak exogeneity were to hold. As expected, the persistence parameter is relatively large for all trade sizes (Trade colummS in Table 2). Conversely, the results for standardized volume size per price duration (Price columns) are more mixed. While in three of the five models (Boeing, Disney and IBM) the persistence parameter is large and significant, for the other two models Coca Cola and Exxon) it is not. The coefficients on current duration are all negative

[13] The choice of exponential distributions for the conditional and marginal densities may be critized on the grounds that many papers on financial durations have shown that this distribution is too simple to capture the density aspects of the observed processes; cf. Bauwens and Veredas (2004) and Bauwens et al. (2004) among others. But even a three parameter distribution, like the generalized gamma, fails to fit correctly the trade durations density. However, since the exponential distribution belongs to the LEF and our parameters of interest are solely in the conditional mean, the choice of the exponential density suffices to test for weak exogeneity.

[14] Positive correlation of trade size has also been documented by Hasbrouck (1991).

and significant, except for Boeing. But interestingly the coefficients on current duration in Coca Cola and Exxon are much larger and negative. In all, it implies that an increase in the instantaneous volatility causes an increase of volume size. This evidence could be interpreted as supporting the MDH hypothesis.

Current trade duration has a positive effect on trade size for all cases except, again, Boeing. It supports the stealth-trading hypothesis. Shorter durations (higher trading activity) leading to smaller volume size implies that traders try to trade more quickly as it is more likely that another trader will buy a small volume size rather than a big block when the trading is intense. Conversely, that long durations lead to large volumes per transaction implies that in tranquil periods, large volume sizes are traded only by liquidity (uninformed) traders; cf. Easley and OHara (1992).

However, these results have been obtained, as in most of the literature, from separate estimation of the parameteres in the conditional and marginal means, i.e. under the null of weak exogeneity. It is therefore useful to test such a hypothesis using our proposed tests. On the one hand, the values of the score/LM test are reported in the S row in Table 2. They should be compared to 3.84, i.e., the 95% C.V. of a chi-squared with 1 degree of freedom. On the other, the values of the nonparametric LR test are displayed in the LR^* row and their bootstrapped CVs in the $95\%LR^*$ row just below. The results are fairly striking: the null hypothesis of weak exogeneity is rejected in seven out of ten cases with the S test and in eight cases with the LR^* test. For volume per price duration (Price), we only accept the null with both tests in just one case, Coca Cola. And for volume size duration (Trade), only Exxon and (almost) IBM pass the test. In the remaining cases, both parametric and nonparametric test values are very far from their CVs. On the whole, therefore, the most relevant finding is that weak exogeneity is rejected in seven of the ten models considered here. This means that if we want to analyze traded volume size of volume per price changes conditional to durations and the marginal distribution for durations, estimation should be done either done jointly (if possible) or using the two-step procedure suggested by Engle (2000). Otherwise, an inconsistency problem may arise.

6 CONCLUSIONS

In this article, we propose a test for weak exogeneity in misspecfied densities within the context of financial point processes. Following the important results in Gourieroux et al. (1984 a, b), the only conditions which are required are that the assumed distributions fall within the exponential family and that the conditional mean is correctly specified. Under these conditions, it is possible to derive a simple score/LM test-statistic to test weak exogeneity of the variables in the marginal distribution for the parameters of interest in the mean of the conditional distribution. This can be done for virtually any kind of point process, since the exponential family encompasses distributions that account for most possible supports of a random variable in the above-mentioned setup. Under a linearity assumption regarding functional dependence between the parameters of the means in the conditional and marginal densities, the test has a $\chi^2(1)$ limiting distribution and, in most cases, can be interpreted as the t-ratio on the coefficient of a linear combination of variables appearing in the mean of the marginal distribution (whose parameters are estimated separately from that distribution) in an augmented model of the mean of the conditional density. The second test is nonparametric, in the sense that it does not specify any specific functional form between the parameters of the means in both densities. The test is a nonparametric LR test whose distribution is calculated using bootstrap.

We have applied these tests to tick-by-tick data which are, statistically speaking, realizations of a point process, which is fully characterized by the duration and their marks. Two types of processes, linked with market microstructure theory, are considered. On the one hand, we model the relationship between trade size and trade durations. On the other, we examine the relationship between volume per price change and price durations. In general we reject the null hypothesis of weak exogeneity, therefore questioning results in the literature which rely on separate estimation of each density.

References

[1] Andersen, T. G. (1996). "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility", *Journal of Finance*, Vol. 51, No. 1, 169-204.

[2] Bauwens, L. and Giot, P. (2000). "The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks," *Annales d'Economie et de Statistique*, 60, 117-149

[3] Bauwens, L. Giot, P. Grammig, J. and Veredas, D. (2004). "A comparison of financial duration models via density forecast," CORE DP 2000/60. Université catholique de Louvain. *International Journal of Forecasting*, in press.

[4] Bauwens, L. and Veredas, D. (2004). "The Stochastic Conditional Duration Model: A latent factor model for the analysis of financial durations," *Journal of Econometrics*, 119, 381-412.

[5] Birgé, L. and P. Massart (1993). "Rate of Convergence for Minimum Contrast Estimators," *Probability Theory and Related Fields*, 97, 113-150

[6] Boswijk, H.P. and Urbain, J-P. (1997). "Lagrange-Multiplier Tests For Weak Exogeneity: A Synthesis", *Econometric Reviews*, 16(1), 21-38.

[7] Cai, Z., Fan, J. and Li, R. (2000). "Efficient estimation and inferences for varying-coefficient models" *J. Amer. Statist. Assoc.*, 95, 888-902.

[8] Chen, S.X. and Shen, S.X. (1998). "Sieve extremum estimates for weakly dependent data," *Econometrica*, 66, 289-314.

[9] Clark, P.K.(1973). " A subordinated stochastic process model with finite variance for speculative prices" *Econometrica Assoc.*, Vol. 41, No. 1, 135-155.

[10] Dolado, J and Maria-Dolores, R. (2002). "Evaluating changes in the Bank of Spains interst rate target: An alternative approach using marked poin processes", *Oxford Bulletin of Economics and Statistics*, 64 , 159-182.

[11] Easley, D. and O'Hara, M. (1992). "Time and the process of a security price adjustment", *Journal of Finance*, Vol. 19, No. 2, 69-90.

[12] Engle, R.F., Hendry, D.F. and Richard, J-F. (1983). "Exogeneity," *Econometrica*, Vol. 51, No. 2, 277-304.

[13] Engle, R.F., Hendry, D.F (1993). "Testing Superexogeneity and invariance in regression models," *Journal of Econometrics*, 56, 119-139.

[14] Engle, R.F. and Russell, J.R. (1998). "Autoregressive conditional duration: a new approach for irregularly spaced transaction data," *Econometrica*, Vol. 66, No, 5, 1127-1162.

[15] Engle, R.F. (2000). "The Econometrics of Ultra High Frequency Data," *Econometrica*, Vol. 68, No. 1, 1-22.

[16] Fan, J. and Gijbels, I. (1996). Local polynomial modelling and its applications. Monographs on Statistics and Applied Probability, 66. Chapman & Hall, London.

[17] Giot,P. (2000). "Time tranformations, intraday data and volatility models", *Journal of Computational Finance*, 4, 31-62.

[18] Gouriéroux, C., Monfort, A. and Trognon, A. (1984). "Pseudo Maximum Likelihood Methods: Theory," *Econometrica* Vol. 52, No. 3, 681-700.

[19] Gouriéroux, C., Monfort, A. and Trognon, A. (1984). "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica* Vol. 52, n. 3, 701-720.

[20] Grenander, U. (1981). *Abstract Inference*. New York: Wiley Series.

[21] Hall, P. and Yao, Q. (2003). "Inference in ARCH and GARCH models with heavy-tailed errors", *Econometrica*, Vol. 71, No. 1, 285-317.

[22] Harvey, A.C. (1981). Time Series Models, Philip Allan Publishers Ltd. (Oxford, UK)

[23] Hasbrouck, J. (1991). "Measuring the Information Content of Stock Trades", *Journal of Finance*, Vol. 46, No. 1, 179-207.

[24] Huber, P. J. (1967). "The behavior of maximum likelihood estimates under nonstandard conditions", in *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics*, 221–233, Univ. California Press, Berkeley, Calif.

[25] Kiefer, J. and Wolfowitz J. (1956) "Consistency of the maximum likelihood estimator in the presence of infinetely many incidental parameters," *Annals of Mathematical Statistics*, 28, 887-906.

[26] Lamoureux, C.G. and Lastrapes, W.D. (1990). "Heteroscedasticity in stock return data: volume versus GARCH effects", *Journal of Finance*, Vol. 45, No. 1, 220-229.

[27] Manganelli, S. (2002). "Duration, volume and volatility impact of trades", European Central Bank working paper No 125.

[28] Renault, E, and Werker, B. (2004). "Stochastic Volatility with Transaction Time", CentER WP.

[29] Severini, T.A. and Staniswalis, J.G. (1994). "Quasi-likelihood estimation in semiparametric models," *Journal of the American Statistical Association*, 89, 501-511.

[30] Severini, T.A. and Wong, W. H. (1992). "Profile likelihood and conditionally parametric models," *Annals of Statistics*, 20, 1768-1802.

[31] Shen, X., Shi, J. and Wong, W. H. (1999). "Random sieve likelihood and general regression models", *J. Amer. Statist. Assoc.*, 94 , 835–846.

[32] Shen, X. and Wong, W. H. (1994). "Convergence rates for sieve estimates," *Annals of Statistics*, 22, 580-615.

[33] Silvey,S.D. (1975). Statistical Inference (Chapman and Hall, Cambridge, UK).

[34] Veredas, D, Rodriguez-Poo, J., and A. Espasa (2001) "On the (intradaily) seasonality of a financial point process", CORE DP 2002/23, Université catholique de Louvain. Louvain-la-Neuve.

[35] White, H. (1981). "Consequences and detection of misspecified nonlinear regression models", *J. Amer. Statist. Assoc*, 376, 419–433.

[36] White, H. (1982). "Maximum likelihood estimation of misspecified models", *Econometrica*, 50, 1–25.

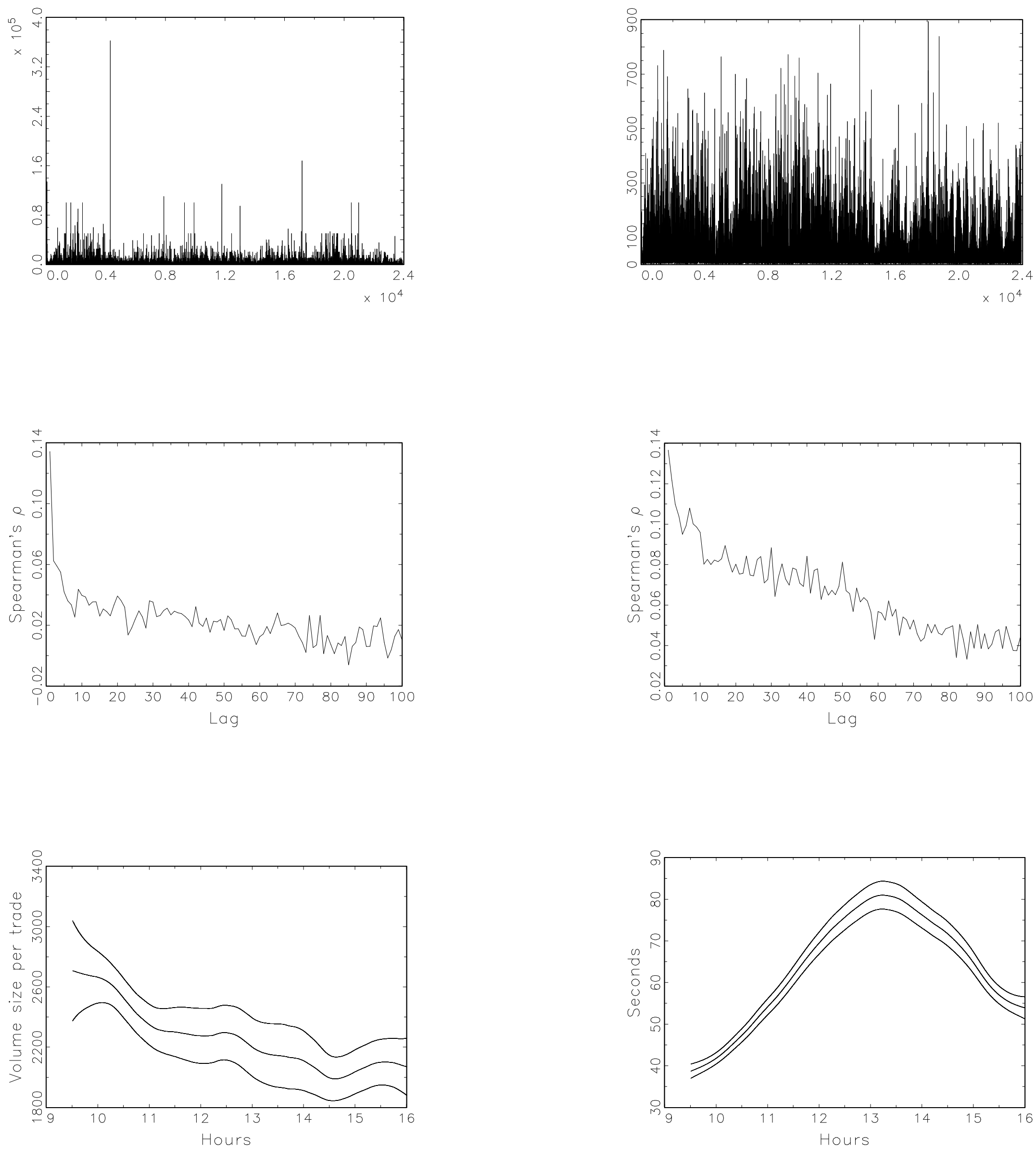
[37] White, H. (1984). Asymptotic Theory for Econometricians (Academic Press, NY).

[38] Wong, W. H. and Shen, X. (1995). "Probability inequalities for likelihood ratios and convergence rates of sieve MLEs", *Annals of Statististics*, 339–362.

Table 1: Data Information

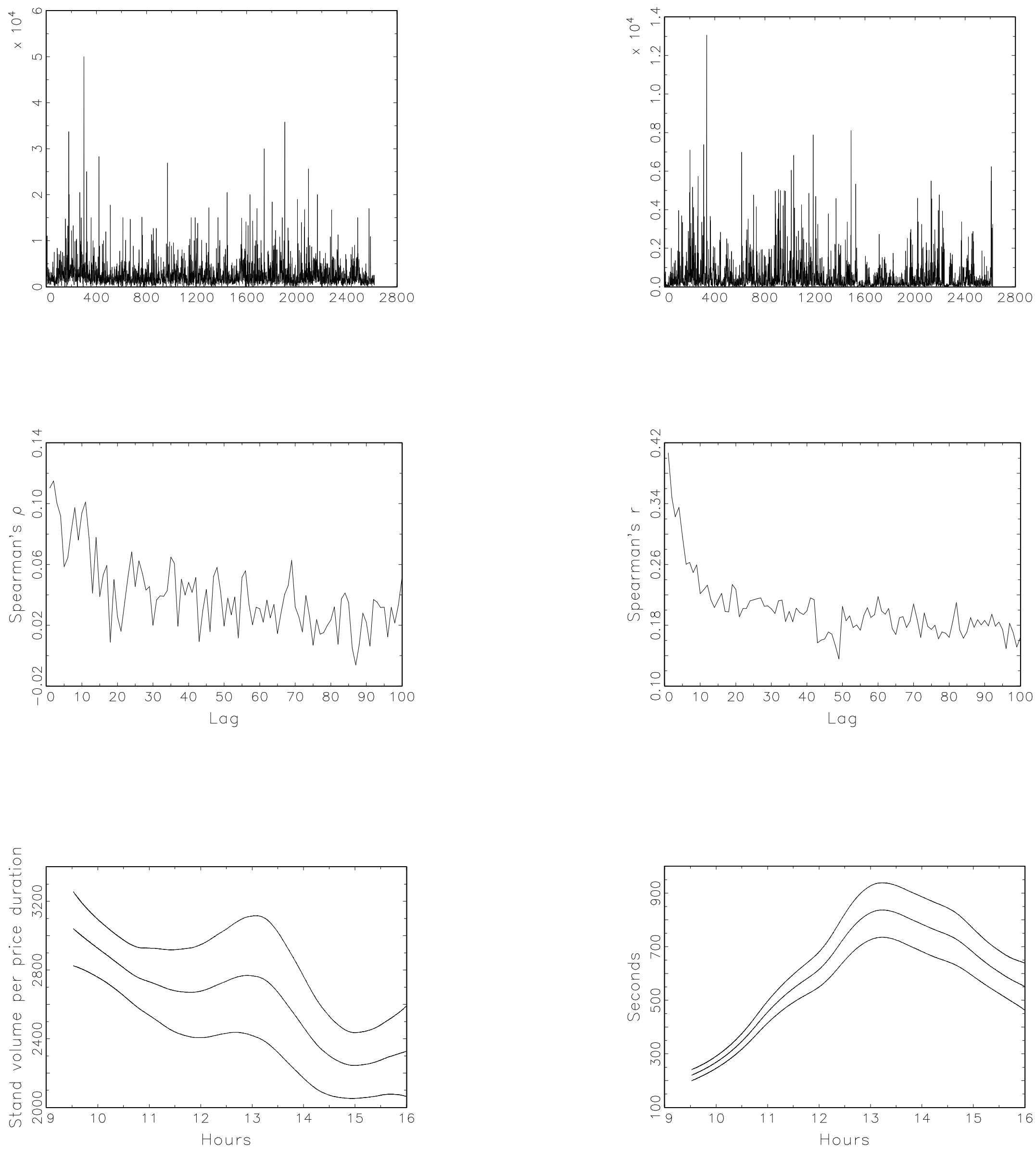
	Boeing		Coca Cola		Disney		Exxon		IBM	
	Trade	Price	Trade	Price	Trade	Price	Trade	Price	Trade	Price
N	23930	2620	39620	1609	32821	2160	28371	2717	60454	6728
mean	60.72	533.5	36.66	839.5	44.28	646.9	51.13	513.0	24.14	214.9
sd	76.83	867.8	45.01	1168	56.25	966.3	64.46	721.0	34.59	365.4
% $<$ mode	0.08	0.07	0.12	0.14	0.07	0.11	0.07	0.19	0.11	0.07
min	1.00	3.00	1.00	3.00	1.00	3.00	1.00	2.00	1	1
max	894	1359	772	11682	926	9739	892	10430	674	7170
N	23930	2620	39620	1609	32821	2160	28371	2717	60454	6728
mean	2293	2664	3062	3952	1790	2211	2362	2858	2436	2911
sd	2721	3066	7517	3546	4478	2750	4885	3351	5103	3329
% $<$ mode	0.18	0.25	0.19	0.39	0.23	0.24	0.15	0.29	0.42	0.34
min	101	1	101	1	101	1	101	1	101	1
max	362201	50001	278101	39494	209701	60765	195201	71446	567101	77501

Descriptive statistics. Top part of the table are for trade and price durations. Bottom part for trade size and volume per price change. N denotes the number of observations, sd the standard deviation, % <mode the proportion of observations smaller than the mode, min and max are the smallest and the largest values.



From top to bottom and from left to right, observed trade size, trade size Spearman's ρ coefficients, trade size seasonal pattern, observed durations, durations Spearman's ρ coefficients, durations seasonal pattern.

Figure 1: Descriptive data plots. Volume and trade durations for Boeing.



From top to bottom and from left to right, observed volume, volume Spearman's ρ coefficients, volume seasonal pattern, observed durations, durations Spearman's ρ coefficients, durations seasonal pattern.

Figure 2: Descriptive data plots. Volume and price durations for Boeing.

Table 2: Estimation results

	Boeing		Coca Cola		Disney		Exxon		IBM	
	Trade	Price	Trade	Price	Trade	Price	Trade	Price	Trade	Price
C	0.0771	0.0074	0.0957	0.1464	0.0312	0.0358	0.0146	0.0940	0.0565	0.0134
	[0.0022]	[0.0020]	[0.0019]	[0.0269]	[0.0005]	[0.0070]	[0.0004]	[0.0216]	[0.0008]	[0.0018]
$\ln v_{i-1}$	0.0857	0.0177	0.0928	0.1416	0.0299	0.0581	0.0184	0.0330	0.0716	0.0164
	[0.0016]	[0.0040]	[0.0012]	[0.0225]	[0.0004]	[0.0115]	[0.0005]	[0.0096]	[0.0008]	[0.0016]
$\mu_{i-1}^{y\|d}$	0.6589	0.9710	0.6155	0.2816	0.9566	0.8820	0.9676	−0.178	0.8140	0.9451
	[0.0074]	[0.0074]	[0.0062]	[0.1546]	[0.0007]	[0.0247]	[0.0011]	[0.1465]	[0.0030]	[0.0091]
d_i	0.0112	0.0004	0.0238	−0.097	0.0009	−0.011	0.0037	−0.059	0.0026	−0.004
	[0.1197]	[0.0005]	[0.0008]	[0.0231]	[0.0002]	[0.0038]	[0.0002]	[0.0075]	[0.0004]	[0.0009]
S	12.01	76.30	12.20	0.065	54.25	46.46	0.367	6.274	2.892	46.01
LR^*	6.122	18.08	7.954	3.504	10.03	9.923	2.951	19.95	5.009	12.38
$95\%LR^*$	5.079	3.127	4.263	7.330	3.853	2.879	5.134	7.136	4.665	3.028
C	0.0277	0.1293	0.0272	0.0938	0.0225	0.1055	0.0276	0.0756	0.0305	0.0871
	[0.0016]	[0.0158]	[0.0022]	[0.0174]	[0.0016]	[0.0202]	[0.0019]	[0.0148]	[0.0016]	[0.0066]
$\ln d_{i-1}$	0.0445	0.1773	0.0430	0.1360	0.0362	0.1520	0.0438	0.1149	0.0598	0.1566
	[0.0025]	[0.0211]	[0.0035]	[0.0210]	[0.0027]	[0.0276]	[0.0030]	[0.0214]	[0.0031]	[0.0114]
μ_{i-1}^{d}	0.9531	0.7944	0.9422	0.7331	0.9598	0.8208	0.9502	0.8535	0.9416	0.8269
	[0.0029]	[0.0290]	[0.0049]	[0.0509]	[0.0034]	[0.0378]	[0.0040]	[0.0360]	[0.0033]	[0.0140]

Entries are LEF estimates –using the exponential distribution. Estimates for Log-ACV model, under the null $\alpha_1 = 0$, in the top part of the table. Estimates for Log-ACD model in the bottom part of the table. Numbers in brackets are heteroskedastic-consistent standard errors. C denotes the constant of the model. S denotes the score test. LR^* denotes the nonparametric log-likelihood ratio test. $95\%LR^*$ is its 95% quantile.