



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**A QUASI-LOCALLY MOST POWERFUL
TEST FOR CORRELATION IN THE
CONDITIONAL VARIANCE OF POSITIVE
DATA**

Brendan P.M. MCCabe, Gael M. Martin and Keith Freeland

February 2010

Working Paper 02/10

A QUASI-LOCALLY MOST POWERFUL TEST FOR CORRELATION IN THE CONDITIONAL VARIANCE OF POSITIVE DATA*

BRENDAN P. M. MCCABE[†] GAEL M. MARTIN[‡]

and KEITH FREELAND[§]

February 9, 2010

Abstract

A test is derived for short-memory correlation in the conditional variance of strictly positive, skewed data. The test is quasi-locally most powerful (QLMP) under the assumption of conditionally gamma data. Analytical asymptotic relative efficiency calculations show that an alternative test, based on the first-order autocorrelation coefficient of the squared data, has negligible relative power to detect correlation in the conditional variance. Finite sample simulation results confirm the poor performance of the squares-based test for fixed alternatives, as well as demonstrating the poor performance of the test based on the first-order autocorrelation coefficient of the raw (levels) data. Robustness of the QLMP test, both to misspecification of the conditional distribution and misspecification of the dynamics is also demonstrated using simulation. The test is illustrated using financial trade durations data.

Key Words: Locally most powerful test; quasi-likelihood; asymptotic relative efficiency; durations data; gamma distribution; Weibull distribution.

JEL Codes: C12, C16, C22.

*This research has been supported by Australian Research Council Discovery Grant No. DP0664121. The authors would like to thank two anonymous referees for very detailed and constructive comments on an earlier draft of the paper.

[†]School of Management, University of Liverpool. *Email:* Brendan.McCabe@liverpool.ac.uk.

[‡]Corresponding author. Department of Econometrics and Business Statistics, P.O. Box, 11E, Monash University, Victoria, 3800, Australia. *Email:* Gael.Martin@Buseco.monash.edu.au.

[§]Department of Actuarial Science and Statistics, University of Waterloo, Canada. *Email:* rkfreeland@uwaterloo.ca.

1 Introduction

This paper contributes to an emerging literature in which second-order dependence in positive, highly skewed data is the focus of analysis. In the context of trade durations, a prime example of such data, the second moment represents a particular measure of liquidity risk. Only recently have the dynamics of this risk been modelled separately from the dynamics in the mean (e.g. Ghysels, Gouriéroux and Jasiak, 2004), with the autocorrelation function (ACF) of the squared data being suggested as a possible diagnostic tool.¹ While such a practice may have some merit, this paper quantifies the substantial power loss that can occur as a result of failing to incorporate information on the skewed nature of the data in the construction of a test for correlation in the conditional variance.

In order to quantify the potential power loss associated with the squares-based statistic, we derive a test for second-order dependence in a leading case. We consider a parameter-driven model (Cox, 1981) for dependent positive data, where the conditional distribution is gamma and the (positive) parameter of that distribution is a dependent log normal sequence, driven by a stationary autoregressive process of order one (AR(1)). The conditional variance is, in turn, parameterized as a function of the dependent log normal sequence. These assumptions allow us to produce exact analytic expressions for the asymptotic relative efficiency (ARE) of the new test, in comparison with the squares-based statistic. The new test is locally most powerful with respect to a quasi-likelihood function, which is used in order to avoid the well known computational difficulties associated with a latent variable structure.

The quasi-locally most powerful (QLMP) test statistic is shown to be the sample first-order autocorrelation coefficient constructed from a simple transformation of the data that is *not* equal to the square transformation. Both the new test and the test based on the (first-order autocorrelation of the) squares are shown to be consistent against appropriate α -mixing alternatives. This consistency result lends some legitimacy to the sort of preliminary testing cited in the opening paragraph above, in which the square transform is used as the default in positive data settings. However, the ARE results show that the squares-based test has *negligible* asymptotic efficiency relative to the new test, in empirically relevant settings. Finite sample results further highlight the inferiority of the squares-based test, with the latter shown to have empirical power that is several-fold less than that of the QLMP test in some instances. Robustness of the finite sample power results to misspecification of the conditional

¹See Bauwens *et al.* (2004) for related work.

distribution is also demonstrated.

For completeness, the finite sample performance of the QLMP test is also compared with that of the test based on the first-order autocorrelation coefficient of the *raw* (levels) data. Like the squares-based test, the latter test is shown to have negligible power to detect correlation in the conditional variance. When the data generating process (DGP) is parameterized to have correlation in *both* the conditional mean and variance, the empirical power of the QLMP test is only marginally less than that of the levels-based test, despite the fact that the former has been derived under the assumption of a fixed mean. The squares-based test is inferior, overall, to both alternative tests in this case. Even when the DGP has dependence *only* in the conditional mean, with the variance fixed, the QLMP test remains competitive with the levels-based test, and is still superior to the squares-based procedure.

Although gaining some motivation from the trade durations literature (see also Engle and Russell, 1998; Bauwens and Veradas, 2004; Strickland *et al.*, 2006, 2008; Bauwens and Hautsch, 2009), in which the second moment is of inherent interest as a risk measure, the results highlighted in the paper are relevant to any setting in which positive, skewed data is the focus. Some contributions to this general literature include Lawrance and Lewis (1980, 1985), Lewis, McKenzie and Hugus (1989) and Ristic (2005).

The outline of the paper is as follows. In Section 2, the QLMP test statistic is derived. Section 3 gives details of the asymptotic theory of both the QLMP and squares-based tests under α -mixing conditions. The ARE of the two tests is then investigated in Section 4, with the distinct power superiority of the QLMP test highlighted. These local power results are supplemented with fixed alternative power comparisons, via Monte Carlo simulations, in Section 5. The new test is illustrated in Section 6 using empirical trade durations data for IBM. Some conclusions are provided in Section 7.

2 Derivation of the Score Test

We begin with the class of models for the T -dimensional random vector $\mathbf{y} = [y_1, y_2, \dots, y_T]'$ with distribution given by

$$f(\mathbf{y}) = \int \dots \int f(\mathbf{y}|\boldsymbol{\lambda})f(\boldsymbol{\lambda})d\boldsymbol{\lambda} = E_{\boldsymbol{\lambda}} [f(\mathbf{y}|\boldsymbol{\lambda})], \quad (1)$$

where $\boldsymbol{\lambda}$ denotes a $(T \times 1)$ vector of latent variables $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_T]'$. The following results are well-known,

$$\begin{aligned} E_{\mathbf{y}}[\mathbf{y}] &= E_{\boldsymbol{\lambda}} [\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}], \\ \text{Var}_{\mathbf{y}}[\mathbf{y}] &= E_{\boldsymbol{\lambda}} [\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}] + \text{Var}_{\boldsymbol{\lambda}} [\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}], \end{aligned} \quad (2)$$

where $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}$ and $\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}$ are respectively the mean vector and variance-covariance matrix of the conditional distribution of $\mathbf{y}|\boldsymbol{\lambda}$. Each λ_t , $t = 1, 2, \dots, T$, in the present context assumed positive, is linked to an underlying scalar latent process x_t where we assume that $\lambda_t = e^{x_t}$. The latent process x_t is, in turn, assumed to follow a stationary Gaussian AR(1) process,

$$(x_t - \mu_x) = \rho(x_{t-1} - \mu_x) + \eta_t; \quad \eta_t \sim^{iid} N(0, \sigma_\eta^2); \quad t = 1, 2, \dots, T; \quad |\rho| < 1, \quad (3)$$

with the $(T \times 1)$ vector \mathbf{x} defined as $\mathbf{x} = [x_1, x_2, \dots, x_T]'$, and $\boldsymbol{\Sigma}_{\mathbf{x}} = \text{Var}_{\mathbf{x}}[\mathbf{x}]$. We also assume that

$$f(\mathbf{y}|\boldsymbol{\lambda}) = f(y_1|\lambda_1)f(y_2|\lambda_2) \dots f(y_T|\lambda_T), \quad (4)$$

so that dependence in \mathbf{y} is generated solely through $\boldsymbol{\lambda}$ (from the latent process \mathbf{x}). The null hypothesis is that $\rho = 0$, so that the elements of \mathbf{y} are independent (λ_t is an *i.i.d.* process under the null) and the alternative is that $\rho \neq 0$, so that λ_t is a correlated sequence with short memory.²

Following Cox (1983) and McCabe and Leybourne (2000)³, we define $f^*(\mathbf{y})$ as the second-order Taylor series expansion of $f(\mathbf{y}|\boldsymbol{\lambda})$ about $\boldsymbol{\mu}_{\boldsymbol{\lambda}} = [\mu_\lambda, \mu_\lambda, \dots, \mu_\lambda]'$ = $E_{\boldsymbol{\lambda}}[\boldsymbol{\lambda}]$. Defining $L(\boldsymbol{\lambda}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\lambda})$ and $\ell = \log L$, we may write

$$f^*(\mathbf{y}) = L(\boldsymbol{\lambda}|\mathbf{y})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}} \left[1 + \frac{1}{2} \text{tr}(\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}) \right],$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} = \text{Var}_{\boldsymbol{\lambda}}[\boldsymbol{\lambda}]$ and $\mathbf{M} = \left(\frac{\partial \ell}{\partial \boldsymbol{\lambda}} \frac{\partial \ell}{\partial \boldsymbol{\lambda}'} + \frac{\partial^2 \ell}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \right) \Big|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}}$.⁴ Note that the second term in the expression for \mathbf{M} , $\frac{\partial^2 \ell}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \Big|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}}$, is a diagonal matrix with elements,

$$\mathbf{r} = \left\{ \partial^2 \ell / \partial \lambda_t^2 \Big|_{\lambda_t=\mu_\lambda}; \quad t = 1, \dots, T \right\}, \quad (5)$$

²A “small σ ” Taylor series expansion (denoted by \approx) of $\lambda_t = e^{x_t}$ can be used to show that $\text{corr}(\lambda_t, \lambda_s) \approx \text{corr}(x_t, x_s)$ for all $t \neq s$. As such, the qualitative nature of the autocorrelation in x_t (short memory in this case) is transferred to λ_t .

³See Chesher (1984), Bera and Kim (2002), Huber *et al.* (2004) and Davis and Rodriguez-Yam (2005) for related work.

⁴Clearly, $f^*(\mathbf{y})$ is an approximation to $f(\mathbf{y})$ for which the error is $O \left\{ E_{\boldsymbol{\lambda}} \left[\|\boldsymbol{\lambda} - \boldsymbol{\mu}_{\boldsymbol{\lambda}}\|^3 \right] \right\}$. Indeed, $f^*(\mathbf{y}) > 0$ is a valid density in its own right as it integrates to unity.

because of conditional independence. Also, $\ell(\boldsymbol{\lambda}|\mathbf{y}) = \sum_t \ell(\lambda_t|y_t)$ and, without loss of generality, for some functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$, $\ell(\lambda_t|y_t)$ may be written as

$$\ell(\lambda_t|y_t) = a(y_t) + b(\lambda_t) + c(y_t, \lambda_t), \quad (6)$$

where we allow for the possibility that $a(y_t)$ and $b(\lambda_t)$ may be zero. The (conditional) score is then

$$\ell'(\lambda_t|y_t) = b'(\lambda_t) + c'(y_t, \lambda_t),$$

(where the prime denotes differentiation w.r.t. λ_t) and this has (conditional) expectation zero for all λ_t and, hence, unconditional expectation zero. Thus defining

$$u(y_t) = c'(y_t, \lambda_t)|_{\lambda_t=\mu_\lambda}, \quad (7)$$

we can write that

$$\ell'(\lambda_t|y_t)|_{\lambda_t=\mu_\lambda} = \{u(y_t) - \mu_u\} = u_c(y_t), \quad (8)$$

where $\mu_u = E_y[u(y_t)]$ and the notation $u_c(y_t)$ is used to denote the mean-corrected version of $u(y_t)$.

The locally most powerful one-sided test (see, for example, Cox and Hinkley, 1979, Sect. 4.8) of

$$H_0 : \rho = 0 \text{ against } H_1 : \rho > 0, \quad (9)$$

based on the quasi-score, is given by

$$S = \left. \frac{\partial \log f^*(\mathbf{y})}{\partial \rho} \right|_{\rho=0}.$$

From the properties of the log-normal distribution it follows that $\partial \boldsymbol{\mu}_\lambda / \partial \rho|_{\rho=0} = 0$ and so

$$S = \text{tr} \left(\mathbf{M} \left. \frac{\partial \boldsymbol{\Sigma}_\lambda}{\partial \rho} \right|_{\rho=0} \right) / (2 + \sigma_\lambda^2 \text{tr} \mathbf{M}), \quad (10)$$

where σ_λ^2 is the variance of λ_t under the null. It is well-known that $\left. \frac{\partial \boldsymbol{\Sigma}_\mathbf{x}}{\partial \rho} \right|_{\rho=0} \propto \mathbf{A}$, where \mathbf{A} is a tridiagonal matrix with zeros on the main diagonal and unity on the off-diagonals. Using the log-normal assumption for λ_t , it also follows that $\left. \frac{\partial \boldsymbol{\Sigma}_\lambda}{\partial \rho} \right|_{\rho=0} \propto \mathbf{A}$. Hence, apart from constants, (10) can be re-expressed as

$$S = \mathbf{u}'_c \mathbf{A} \mathbf{u}_c / (2 + \sigma_\lambda^2 [\mathbf{u}'_c \mathbf{u}_c + \mathbf{r}' \mathbf{i}]),$$

where \mathbf{u}_c is defined as a $(T \times 1)$ vector with t th element $u_c(y_t)$ (as given in (8)), \mathbf{i} is a $(T \times 1)$ vector of 1's and \mathbf{r} is defined in (5). Standardizing in the usual way we obtain

$$S_T = T^{-1/2} \mathbf{u}'_c \mathbf{A} \mathbf{u}_c / [2/T + \sigma_\lambda^2 [\mathbf{u}'_c \mathbf{u}_c + \mathbf{r}' \mathbf{i}] / T] \quad (11)$$

and, using a suitable weak law of large numbers (WLLN), the denominator in (11) converges in probability to a constant under the *i.i.d.* null. Convergence to the same constant also occurs under local alternatives by LeCam's 3rd Lemma (see van der Vart, 1998, Sect 6.2). Thus, defining

$$S_u = \mathbf{u}'_c \mathbf{A} \mathbf{u}_c, \quad (12)$$

and recognizing that $\mathbf{u}'_c \mathbf{A} \mathbf{u}_c = 2 \sum_{t=2}^T u_c(y_t) u_c(y_{t-1})$, the statistic,

$$S_{u,T} = T^{-1/2} \sum_{t=2}^T u_c(y_t) u_c(y_{t-1}), \quad (13)$$

may be used to test $H_0 : \rho = 0$ and the test based on this statistic is asymptotically equivalent to the test based on S_T in (11).⁵ As $S_{u,T}$ is shown in Section 3 to be asymptotically normal, a two-sided test of H_0 against $H_1 : \rho \neq 0$ is conducted by rejecting in either tail of the normal distribution.

It is a simple matter to identify the function $u(\cdot)$ in (7) for any particular conditional distribution, $f(y_t | \lambda_t)$. The choice of parameterization of the conditional distribution determines the dependence structure of the observed y_t . Given the focus of the paper, we parameterize the conditional distribution in such a way that the conditional mean does not depend on λ_t , while the conditional variance does. By (2), and by the assumption of conditional independence, it follows that y_t is uncorrelated, but with correlation in the conditional variance.⁶

For the positive, highly positively skewed data that is the focus of this paper, the gamma distribution is a suitable choice of conditional, with density,

$$f(y_t | \theta, \lambda_t) = \frac{1}{y_t} \times \frac{1}{\Gamma(\frac{1}{\lambda_t})} \left(\frac{1}{\lambda_t} \right)^{\frac{1}{\lambda_t}} \times (-\theta y_t)^{\frac{1}{\lambda_t}} \exp \left(\frac{\theta y_t}{\lambda_t} \right). \quad (14)$$

The conditional mean and variance are given respectively by $E_{y|\lambda}[y_t] = -\frac{1}{\theta}$ and $V_{y|\lambda}[y_t] = \frac{1}{\theta^2} \lambda_t$, with $\theta < 0$ a scalar constant. In the textbook notation for the gamma distribution, $G(\alpha, \beta)$,

⁵The scale factor of 2 that appears in the expression for $\mathbf{u}'_c \mathbf{A} \mathbf{u}_c$ is cancelled when the test statistic is used in its standardized form; hence its omission from the expression in (13).

⁶Using the same sort of small σ argument as sketched in Footnote 2 it can be shown that the short memory correlation in λ_t is transferred (approximately) to the conditional variance of λ_t even if the conditional variance is a non-linear function of λ_t .

with probability density function (pdf), $f(y_t|\alpha, \beta) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)} y_t^{(\alpha-1)} \exp(-y_t/\beta)$, we have $\alpha = \lambda_t^{-1}$ and $\beta = -\lambda_t/\theta$. We adopt the parameterization in (14) to ensure that the conditional mean of $y_t|\lambda_t$ is a not function of λ_t and that correlation in λ_t induces correlation in $V_{y|\lambda}[y_t]$ only. Noting that $\mu_y = E_y[y_t] = E_{y|\lambda}[y_t] = -\frac{1}{\theta}$, we see (with reference to (6)) that

$$c(y_t, \lambda_t) = -\frac{1}{\lambda_t} g(y_t),$$

with

$$g(y_t) = \frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right). \quad (15)$$

Thus, $u(y_t) = \mu_\lambda^{-2} g(y_t)$ and the constant μ_λ^{-2} may be ignored in the construction of the test, so we set $u(y_t) = g(y_t)$ in this case. Formally, setting $u(y_t)$ equal to

$$d(y_t) = (y_t - \mu_y)^2 \quad (16)$$

yields a test based on the correlation coefficient of the squared data, the primary comparator in the paper.

3 Asymptotic Distribution Theory

Because of the u -transformation of the y_t embodied in the statistic in (13), α -mixing is a natural environment in which to analyze the asymptotic behaviour of $S_{u,T}$. We therefore adopt the, now standard (see McCabe and Tremayne, 1993, Sec 10.8), α -mixing assumption for a central limit theorem (CLT) to hold for stationary y_t . Inspection of the $u \in \{g, d\}$ transformations (with $g(\cdot)$ and $d(\cdot)$ as given respectively in (15) and (16)) shows that there are no greater moment requirements than when $u = d$, as the transformation in this case depends on y_t^2 . Thus, the existence of moments of y_t slightly larger than 8 is a sufficient condition, that applies to both transformations, for the CLT to hold for the mixing product sequence $\{u_c(y_t) u_c(y_{t-1})\}$. From now on, references to mixing processes assume that sufficient moment conditions hold. Note under the model (1) to (4) and (14), the moment conditions are satisfied, since the conditional gamma and marginal log normal distributions have finite moments of all orders and, therefore, so too have the y_t .⁷ When y_t is mixing,

$$T^{-1/2} \sum_{t=2}^T [u_c(y_t) u_c(y_{t-1}) - E[u_c(y_t) u_c(y_{t-1})]] \rightarrow^d N(0, \omega^2), \quad (17)$$

⁷This point also applies to the conditional Weibull distribution that underlies some of the simulation results in Section 5.

where $\omega^2 > 0$ is the usual long run variance of the sum in (17).

Suppose that $\{y_t\}$ is an *i.i.d.* sequence, it follows that $\{u(y_t)\}$ is also *i.i.d.* Hence, the CLT implies that $S_{u,T}$ is asymptotically $N(0, \sigma_u^4)$ where σ_u^2 is the (short run) variance of $u(y_t)$. Thus, for example, $S_{u,T}$, for $u \in \{g, d\}$, is asymptotically normal for all $\{y_t\}$ that are independent. *A fortiori* this includes the case where $\{y_t\}$ is generated by the model (1) to (4) under the null hypothesis that $\rho = 0$ in (3). So, for example, it follows that $S_{g,T}$ is asymptotically normal regardless of whether $f(y_t|\lambda_t)$ is specifically gamma or not. The corresponding comment applies to $S_{d,T}$.

Now suppose that $\{y_t\}$ is mixing with $E[u_c(y_t)u_c(y_{t-1})] \neq 0$. Consider

$$S_{u,T} = T^{-1/2} \sum_{t=2}^T \{u_c(y_t)u_c(y_{t-1}) - E[u_c(y_t)u_c(y_{t-1})]\} + T^{-1/2}(T-1)E[u_c(y_t)u_c(y_{t-1})].$$

As the first term in $S_{u,T}$ converges in distribution by (17) and the second term diverges, $S_{u,T}$ also diverges, and a two-sided test based on $S_{u,T}$ is therefore consistent whenever $E[u_c(y_t)u_c(y_{t-1})] \neq 0$. *A fortiori* there is consistency against the model (1) to (4) under the alternative when $\rho \neq 0$. This follows because $\{x_t\}$ in (3) is a mixing process and this implies that $\{\lambda_t\}$ is also mixing and hence, so too is $\{y_t\}$ by conditional independence. It is also straightforward to show that $E[u_c(y_t)u_c(y_{t-1})] \neq 0$ for $u \in \{g, d\}$. Thus, two-sided tests based on $S_{u,T}$, $u \in \{g, d\}$ are consistent against the model (1) to (4) under the alternative hypothesis $\rho \neq 0$ in (3).

Thus far then, both tests are equally good. However, the point of the transformations is, of course, to obtain greater power when some knowledge of an appropriate DGP is available. To illustrate this, the power gains associated with use of S_g rather than S_d (with S_u , $u \in \{g, d\}$ as defined in (12)), in the case of a positive and positively skewed DGP, are quantified in the following section via ARE calculations. These theoretical power results are supplemented with numerical power comparisons in Section 5, based on Monte Carlo experiments.

4 Asymptotic Relative Efficiency of S_u , $u \in \{d, g\}$

The ARE of a test based on the squares-based statistic S_d , relative to a test based on statistic S_g , under a sequence of local alternatives, is a measure of the (asymptotic) relative local power of the two tests. We are interested in linking the loss of efficiency of the squares-based test with the degree of skewness in the underlying DGP. To this end, the location and scale parameters of the underlying conditional gamma DGP are used to control the degree of skewness. Under regularity conditions (see for example, Stuart *et al.*, 1998, Chp. 26), the

ARE can be represented as

$$ARE_{d,g} = \lim_T \left[\frac{\left. \frac{\partial \mu_{S_d}(\rho)}{\partial \rho} \right|_{\rho=0} / \left. \sigma_{S_d}(\rho) \right|_{\rho=0}}{\left. \frac{\partial \mu_{S_g}(\rho)}{\partial \rho} \right|_{\rho=0} / \left. \sigma_{S_g}(\rho) \right|_{\rho=0}} \right]^2, \quad (18)$$

where $\mu_{S_u}(\rho)$ and $\sigma_{S_u}(\rho)$, $u \in \{d, g\}$, are means and standard deviations such that

$$\frac{S_u - \mu_{S_u}(\rho)}{\sigma_{S_u}(\rho)} \rightarrow^d N(0, 1) \quad (19)$$

in some local region $\{|\rho| < \delta\}$, which includes the alternative hypothesis. The condition in (19) is valid in our case, as the correlation coefficients that underlie our tests have an asymptotic normal distribution under α -mixing conditions for y_t , as demonstrated in (17).

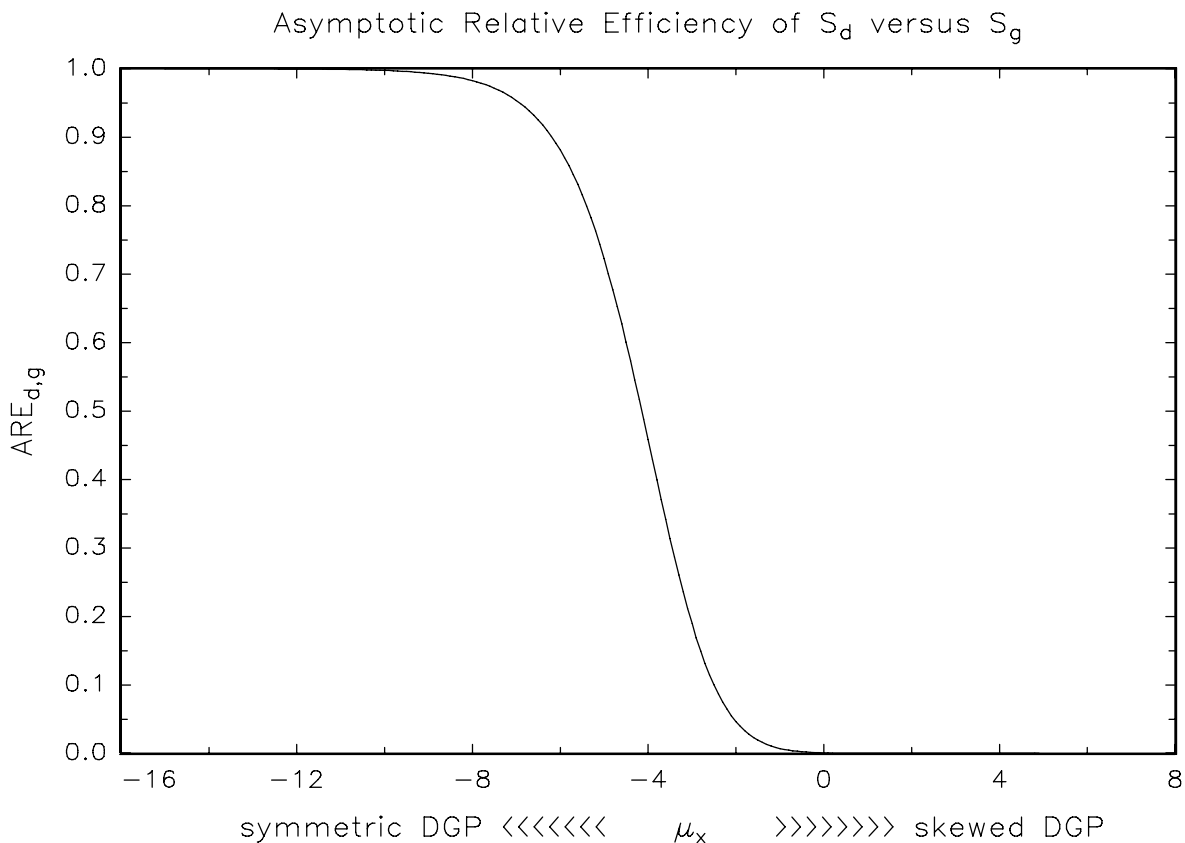


Figure 1: $ARE_{d,g}$ as a function of μ_x (for $\sigma_\eta = 1$). The degree of skewness in the representative conditional gamma distribution underlying the calculations is an increasing function of μ_x .

To evaluate the expression in (18), we use the specifications of the model (1) to (4) with conditional density as given in (14). That is, the relative performance of the S_g test,

which is derived via the quasi-likelihood, is assessed with respect to the true model. In the following proposition, expectations with respect to the $N(\mu_x, \sigma_\eta^2)$ distribution are denoted with a subscript N , e.g. E_N and V_N , while Ψ is the derivative of the log-gamma function, $\Psi(z) = \frac{\partial}{\partial z} \log \Gamma(z)$, with Ψ' being the derivative of Ψ , the digamma function. The proof is given in the Appendix.

Proposition 1 *Under the model defined by (1) to (4) and (14), the ARE of S_d to S_g is given by*

$$ARE_{d,g} = \left[\frac{a}{b} \right]^2,$$

where

$$a = \frac{\left[e^{2\mu_x + \sigma_\eta^2} \right]}{\left[6e^{3\mu_x + \frac{9}{2}\sigma_\eta^2} + 3e^{2\mu_x + 2\sigma_\eta^2} - e^{2\mu_x + \sigma_\eta^2} \right]}$$

and

$$b = \frac{\{1 - E_N[\Psi'(e^{-x_t})(e^{-x_t})]\}^2}{V_N[\Psi(e^{-x_t}) + x_t] - e^{\mu_x + \frac{1}{2}\sigma_\eta^2} + E_N[\Psi'(e^{-x_t})]}.$$

From Proposition 1 it can be seen that $ARE_{d,g}$ does not depend on the conditional mean $\mu_{y|\lambda} = -1/\theta$ ($= \mu_y$) but, rather, depends only on μ_x and σ_η^2 . The values of these parameters can be used to characterize the nature of the conditional gamma DGP underlying the relative power calculations. Specifically, setting $\alpha = [E_\lambda[\lambda_t]]^{-1}$ and $\beta = -E_\lambda[\lambda_t]/\theta$ in $G(\alpha, \beta)$, where $E_\lambda[\lambda_t] = e^{\mu_x + \frac{1}{2}\sigma_\eta^2}$, the representative conditional gamma DGP approaches a symmetric normal DGP with a mean of one as $\alpha \rightarrow \infty$.⁸ Figure 1 plots $ARE_{d,g}$ over μ_x for $\sigma_\eta^2 = 1$, whilst Figure 2 plots the corresponding (representative) DGP's for $\mu_x = 0$ and $\mu_x = -5$ respectively. Clearly, the dominance of the optimal test over the squares-based test is very pronounced for distributions at the skewed end of the spectrum ($\mu_x = 0$ for example), with the relative efficiency of the squares-based test being virtually *zero* for distributions that describe the positive, highly skewed data that is typical of that observed in relevant empirical applications⁹. As the underlying DGP becomes less skewed, the ratio increases, with the relative efficiency of the squares-based test reaching approximately 70% for data that is close to symmetric ($\mu_x = -5$ for example). As $\mu_x \rightarrow -\infty$ (for fixed σ_η^2) and $\alpha \rightarrow \infty$ as a consequence, both tests are equally efficient according to this measure.

⁸The invariance of $ARE_{d,g}$ to θ means that we can assign θ any arbitrary value. This value will, of course, affect the mean of $G(\alpha, \beta)$. This DGP is only representative of the conditional distribution underlying the ratio in that it is based on the substitution of $E(\lambda_t)$ into α and β , rather than the substitution of a particular value of λ_t . The standardized skewness coefficient for the $G(\alpha, \beta)$ distribution is $2\alpha^{-1/2}$.

⁹See, for example, the shape of the empirical distributions of trade durations in Engle and Russell (1998) and Bauwens *et al.* (2004), plus that of the data analysed in Section 6.

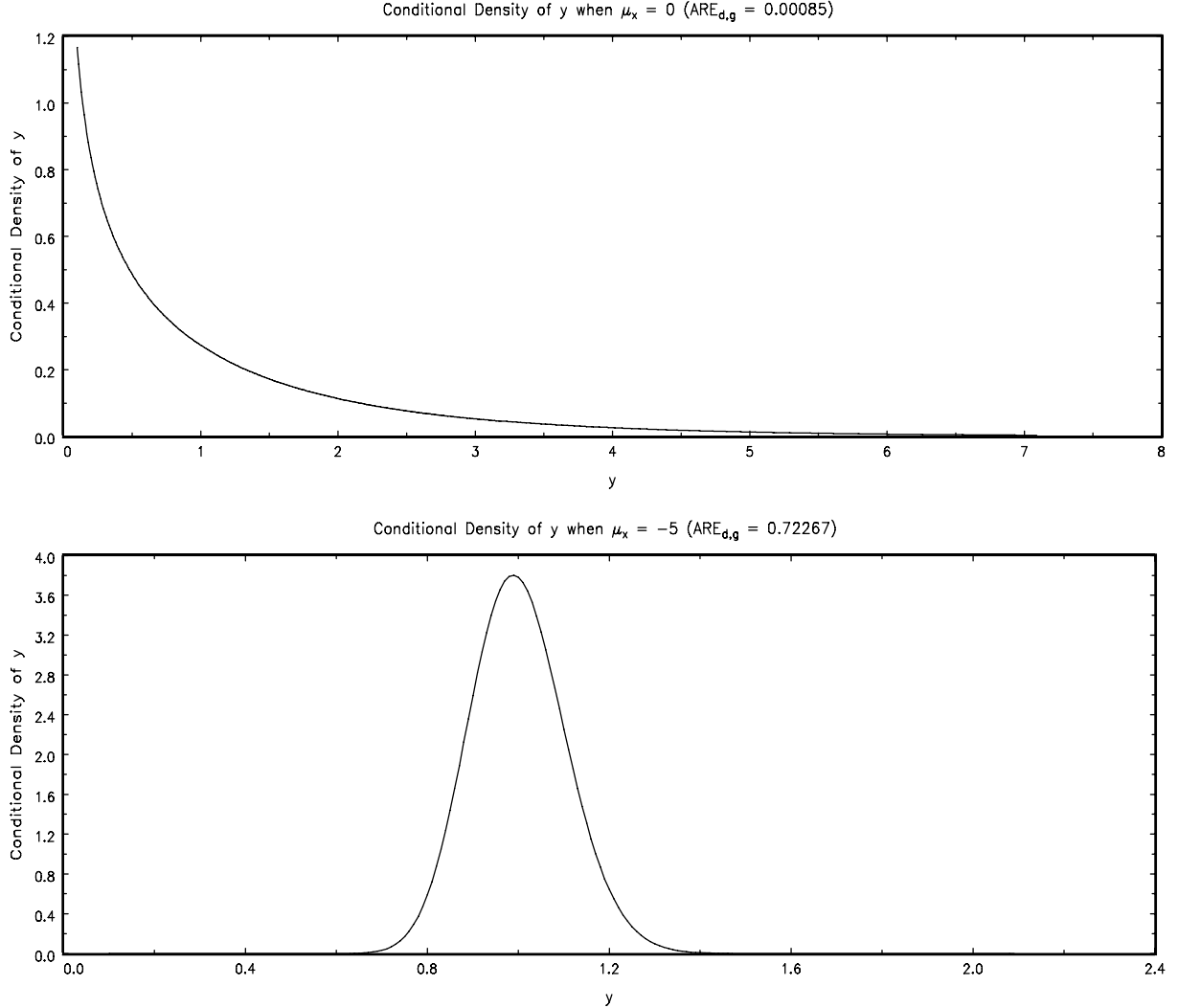


Figure 2: Conditional densities of y_t (and associated $ARE_{d,g}$ values) for selected values of μ_x .

5 Finite sample performance of empirical tests

In practice we may estimate $u(y_t)$ for $u \in \{d, g\}$ by substituting the sample mean \bar{y} for μ_y to obtain $\hat{u}(y_t)$. We may also estimate $\mu_u = E_y[u(y_t)]$ by the sample mean of $\hat{u}(y_t)$, denoted by $\hat{\mu}_u$. Finally, we may estimate the mean-corrected $u_c(y_t)$ by $\hat{u}_c(y_t) = \hat{u}(y_t) - \hat{\mu}_u$. When studentised by the variance, s_u^2 of $\hat{u}(y_t)$, the statistic, for $u \in \{d, g\}$,

$$\hat{\rho}_u = T^{-1/2} s_u^{-2} \sum_{t=2}^T \hat{u}_c(y_t) \hat{u}_c(y_{t-1}) \quad (20)$$

may be used to implement the tests in practice. It is easy to see that

$$T^{-1/2} \sum_{t=2}^T \hat{u}_c(y_t) \hat{u}_c(y_{t-1}) = T^{-1/2} \sum_{t=2}^T u_c(y_t) u_c(y_{t-1}) + o_p(1),$$

so that estimating $u_c(y_t)$ has no asymptotic effect. In addition, under the null of independence, $s_u^2 \rightarrow^p \sigma_u^2$ by the WLLN. Hence, by the continuous mapping theorem and the CLT we have that $\hat{\rho}_u \rightarrow^d N(0, 1)$ under the null and so normal critical values may be used to perform the test.

In Table 1, we report the empirical size and power of the tests based on $\hat{\rho}_d$ and $\hat{\rho}_g$ in (20) under both conditional gamma and conditional Weibull distributions. The generating process for x_t is the AR(1) process in (3), with $\lambda_t = e^{x_t}$ and $\rho \geq 0$.¹⁰ With reference to the conditional gamma density in (14) and the AR(1) process for x_t in (3), we impose parameter values that ensure that the generated data is qualitatively similar to typical positive and very positively skewed data. Specifically, we produce samples with an overdispersion ratio (= ratio of sample variance to sample mean) which averages 4.5 (approximately) across simulations, as matches the overdispersion ratio of the empirical high frequency trade durations data (measured in seconds) to which the test is applied in Section 6. This is achieved using the expressions for the unconditional moments: $\mu_y = -1/\theta$, $V_y[y_t] = E_\lambda[\lambda_t]/\theta^2$ and $E_\lambda[\lambda_t] = e^{\mu_x + \frac{1}{2}\sigma_\eta^2}$. The mean parameter μ_x in (3) is set at a value that ensures that for each value of ρ in (3) (and for $\sigma_\eta = 1$), the mean of the generated λ_t values approximates $E_\lambda[\lambda_t]$ in each case. $E_\lambda[\lambda_t]$ is, in turn, linked to the unconditional variance of the data as per the expression for $V_y[y_t]$, with values assigned to $E_\lambda[\lambda_t]$ and θ to ensure that the average value (over simulated samples) of the sample overdispersion ratio is the required value. The conditional Weibull distribution, parameterized to ensure that the mean is fixed and only the conditional variance is a function of λ_t , is calibrated in such a way that the artificial data is qualitatively similar to that generated under the conditional gamma distribution, for each value of ρ .¹¹

¹⁰Given that financial trade durations are a key data type to which the test may be applied, we choose to document finite sample power in the direction of positive values of ρ . This choice is motivated by the stylized feature of positive correlation seen in the second moment of financial returns data - associated with the ‘clustering’ behaviour of the variance.

¹¹Specifically, we generate data from a distribution with density function

$$f(y_t|\theta, \lambda_t) = \lambda_t y_t^{\lambda_t-1} \left(\frac{\Gamma_t(1)}{\theta}\right)^{\lambda_t} \exp\left(-\left(\frac{y_t \Gamma_t(1)}{\theta}\right)^{\lambda_t}\right),$$

with conditional mean and variance given respectively by $E_{y|\lambda}[y_t] = \theta$ and $V_{y|\lambda}[y_t] = \theta^2 \left[\frac{\Gamma_t(2)}{\Gamma_t^2(1)} - 1\right]$, where $\Gamma_t(s) = \Gamma(1 + s\lambda_t^{-1})$. In the textbook notation for the Weibull distribution, $W(\alpha, \beta)$, with pdf $f(y_t|\alpha, \beta) =$

For comparison purposes, we also report the empirical size and power of the test based on the first-order sample autocorrelation coefficient of the levels data; i.e. the test based on the statistic in (20) with $\hat{u}(y_t) = y_t$ (denoted hereafter by $\hat{\rho}_y$). All calculations are based on 20,000 replications of the relevant process, using samples of size 200, 500 and 2000 and a nominal size of 5%. All powers are based on the empirical 5% critical values.

The results reported in Table 1 show that the empirical sizes of the QLMP test (as based on $\hat{\rho}_g$) are very close to the nominal value of 5%. This is in contrast with the size behaviour of the squares-based test, whereby the empirical sizes are substantially less than 5% in all cases. The results in Table 1 also demonstrate that, under a conditional gamma DGP, the $\hat{\rho}_g$ test is more powerful than the $\hat{\rho}_d$ test throughout the ρ -parameter space, with that dominance still evident for very large sample sizes. For example, when $\rho = 0.7$ and $N = 2000$, the $\hat{\rho}_g$ test has power that is more than eight times that of the squares-based $\hat{\rho}_d$ test, under the conditional gamma DGP. The power dominance of the $\hat{\rho}_g$ test over the $\hat{\rho}_d$ test continues to prevail even when the data is generated from a conditional Weibull distribution rather than the conditional gamma distribution under which $\hat{\rho}_g$ has been derived.

Very similar comments can be made with regard to the relationship between the $\hat{\rho}_g$ test and the test based on $\hat{\rho}_y$, with the power performance of the $\hat{\rho}_y$ test being very similar to that of the $\hat{\rho}_d$ test, throughout the ρ space.¹² The numerical results in Table 1 also illustrate the fixed-alternative consistency properties demonstrated theoretically in Section 3, with the power of all three tests increasing (in general) as T increases.¹³ That said, the power of both the squares- and levels-based tests under strictly positive data is still extremely low in large (but finite) samples.

Given that the $\hat{\rho}_g$ test has been derived as optimal under the assumption of a fixed mean, it is of interest to assess the robustness of the test to violation of this assumption. As such, we report results based on DGPs in which the conditional distribution (either gamma or Weibull) has dependence in both the conditional mean and variance (Table 2) and in the conditional mean only (Table 3) respectively.¹⁴ In brief, when correlation appears in the conditional mean,

$\alpha\beta^{-\alpha}y_t^{\alpha-1}\exp(-(y_t/\beta)^\alpha)$, we have $\alpha = \lambda_t$ and $\beta = \theta/\Gamma_t(1)$. The parameter μ_x in (3) is set at a value that ensures that for each value of ρ in (3) (and for $\sigma_\eta = 1$), the generated λ_t values yield values of y_t that, for given θ , have a sample overdispersion ratio that is similar to that of the simulated (conditional) gamma variates. Note that for the larger values of ρ , the average overdispersion ratio (over simulated samples) ranges between about 5 and 7.

¹²Interestingly however, the $\hat{\rho}_y$ test does not exhibit the distinct undersizing of the $\hat{\rho}_d$ test.

¹³The same reasoning as applied in Section 3 to the tests based on $\hat{\rho}_g$ and $\hat{\rho}_d$ can be used to demonstrate the consistency of the levels-based test.

¹⁴The parameterisations used in the simulation experiments in this case are as follows: 1) Correlation in

Table 1:

Finite Sample Sizes and Powers of Tests of $H_0 : \rho = 0$ against $H_1 : \rho > 0$ under Strictly Positive Conditional Distributions. Correlation in the Conditional Variance; Conditional Mean Fixed.

<u>Empirical Size and Power</u>									
<u>Gamma $f(y_t \lambda_t)$</u>									
	$N = 200$			$N = 500$			$N = 2000$		
	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$
$\rho =$									
0.0	0.050	0.032	0.047	0.051	0.032	0.051	0.054	0.031	0.053
0.1	0.085	0.051	0.051	0.113	0.058	0.054	0.187	0.060	0.051
0.3	0.216	0.062	0.056	0.363	0.068	0.060	0.764	0.077	0.066
0.5	0.434	0.073	0.066	0.703	0.083	0.075	0.980	0.098	0.082
0.7	0.683	0.089	0.086	0.908	0.103	0.091	1.000	0.122	0.103
<u>Weibull $f(y_t \lambda_t)$</u>									
	$N = 200$			$N = 500$			$N = 2000$		
	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$
$\rho =$									
0.0	0.045	0.027	0.047	0.046	0.027	0.050	0.054	0.038	0.058
0.1	0.097	0.055	0.054	0.128	0.058	0.055	0.242	0.061	0.057
0.3	0.276	0.065	0.063	0.472	0.073	0.066	0.920	0.085	0.069
0.5	0.536	0.073	0.072	0.837	0.087	0.074	1.000	0.112	0.078
0.7	0.734	0.079	0.081	0.969	0.105	0.084	1.000	0.139	0.083

Table 2:

Finite Sample Sizes and Powers of Tests of $H_0 : \rho = 0$ against $H_1 : \rho > 0$ under Strictly Positive Conditional Distributions. Correlation in the Conditional Variance and the Conditional Mean.

Empirical Size and Power									
Gamma $f(y_t \lambda_t)$									
	$N = 200$			$N = 500$			$N = 2000$		
	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$
$\rho =$									
0.0	0.049	0.042	0.052	0.054	0.044	0.055	0.056	0.047	0.055
0.1	0.090	0.066	0.087	0.114	0.072	0.118	0.215	0.091	0.237
0.3	0.255	0.122	0.243	0.417	0.158	0.438	0.867	0.277	0.910
0.5	0.534	0.225	0.535	0.816	0.343	0.852	1.000	0.652	1.000
0.7	0.820	0.419	0.835	0.988	0.642	0.991	1.000	0.916	1.000
Weibull $f(y_t \lambda_t)$									
	$N = 200$			$N = 500$			$N = 2000$		
	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$
$\rho =$									
0.0	0.047	0.034	0.051	0.049	0.036	0.054	0.055	0.040	0.055
0.1	0.061	0.083	0.147	0.068	0.094	0.236	0.087	0.122	0.600
0.3	0.143	0.195	0.563	0.225	0.260	0.891	0.493	0.485	1.000
0.5	0.401	0.379	0.921	0.680	0.557	0.999	0.991	0.920	1.000
0.7	0.788	0.630	0.995	0.980	0.864	1.000	1.000	1.000	1.000

Table 3:

Finite Sample Sizes and Powers of Tests of $H_0 : \rho = 0$ against $H_1 : \rho > 0$ under Strictly Positive Conditional Distributions. Correlation in the Conditional Mean; Conditional Variance Fixed.

<u>Empirical Size and Power</u>									
<u>Gamma $f(y_t \lambda_t)$</u>									
	<u>$N = 200$</u>			<u>$N = 500$</u>			<u>$N = 2000$</u>		
	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$
$\rho =$									
0.0	0.054	0.041	0.047	0.059	0.041	0.049	0.066	0.039	0.052
0.1	0.089	0.074	0.130	0.114	0.078	0.192	0.216	0.092	0.439
0.3	0.235	0.157	0.457	0.410	0.207	0.760	0.887	0.359	0.997
0.5	0.481	0.321	0.815	0.800	0.478	0.982	1.000	0.761	1.000
0.7	0.680	0.552	0.953	0.955	0.763	0.996	1.000	0.939	1.000
<u>Weibull $f(y_t \lambda_t)$</u>									
	<u>$N = 200$</u>			<u>$N = 500$</u>			<u>$N = 2000$</u>		
	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_y$
$\rho =$									
0.0	0.047	0.037	0.048	0.046	0.038	0.049	0.050	0.041	0.051
0.1	0.074	0.087	0.171	0.093	0.097	0.300	0.156	0.134	0.727
0.3	0.207	0.220	0.683	0.370	0.315	0.955	0.826	0.605	1.000
0.5	0.523	0.454	0.961	0.858	0.683	1.000	1.000	0.981	1.000
0.7	0.837	0.720	0.993	0.994	0.935	1.000	1.000	1.000	1.000

the $\hat{\rho}_y$ test now has better power behaviour than both the $\hat{\rho}_g$ and $\hat{\rho}_d$ tests. However, when correlation also appears in the conditional variance and the conditional distribution is gamma (results in the top panel in Table 2) the gain in power of the levels-based test over the $\hat{\rho}_g$ test is marginal only. Even under misspecification of the conditional distribution (results in the bottom panel in Table 2) the $\hat{\rho}_g$ test is still competitive with the levels-based test, in particular for large values of ρ and N . In the case where correlation enters *only* via the conditional mean, the $\hat{\rho}_g$ test still remains competitive with the levels-based test, including under the conditional Weibull DGP. The squares-based $\hat{\rho}_d$ is the worst performer overall in Tables 2 and 3, with some undersizing still in evidence. Most notably, the $\hat{\rho}_y$ and $\hat{\rho}_d$ tests do not exhibit the appallingly low power seen in Table 1, once correlation is allowed to enter the conditional mean (whether in tandem with correlation in the conditional variance, or not).

6 Empirical Application to Trade Durations Data

In this section we report the results of applying the new test to trade durations for IBM shares. In Figure 3 the empirical features of the IBM trade durations data are displayed. The data comprises the 6242 durations between trades on 15 January 2004, with zero durations omitted. In modelling the dependence in such durations data using either variants of the autoregressive conditional duration model (Engle and Russell, 1998) or versions of the stochastic conditional duration model (e.g. Bauwens and Veradas, 2004; Ghysels *et al.*, 2004; Strickland *et al.*, 2006), the conditional distribution is typically specified as being either exponential or some generalization thereof, such as the Weibull or gamma distribution. Certainly the empirical distribution in Panel (b) indicates that any such distribution is a plausible choice, and that the test based on the statistic in (20), with $u = g$, is an appropriate choice. The calculated value of $\hat{\rho}_g = 2.280$ is clearly significant at the 1% level, indicating the possibility of there being autocorrelation in the liquidity risk associated with trading in the given asset; see Ghysels *et al.* (2004). In this case, both the levels- and squares-based statistics are also significant ($\hat{\rho}_y = 12.68$; $\hat{\rho}_d = 7.15$) giving a strong indication that correlation in the mean is also a feature of the data.¹⁵

both the conditional mean and conditional variance: gamma: $G(\alpha = \lambda_t, \beta = \theta)$; Weibull: $W(\alpha = \theta, \beta = \lambda_t)$; 2) Correlation in the conditional mean only (conditional variance fixed): gamma: $G(\alpha = \lambda_t^2, \beta = \theta/\lambda_t)$; Weibull: $W(\alpha = \lambda_t, \beta = \theta/\sqrt{\Gamma_t(2) - \Gamma_t^2(1)})$. Once again, the parameter values are chosen so as to produce data with a given average overdispersion ratio (across simulations). Average dispersion ratios close to 4.5 are achieved in most cases; however, for the larger values of ρ , the average ratio ranges between about 5 and 10.

¹⁵This is not surprising given that the data has not been adjusted for any diurnal (intraday) pattern.

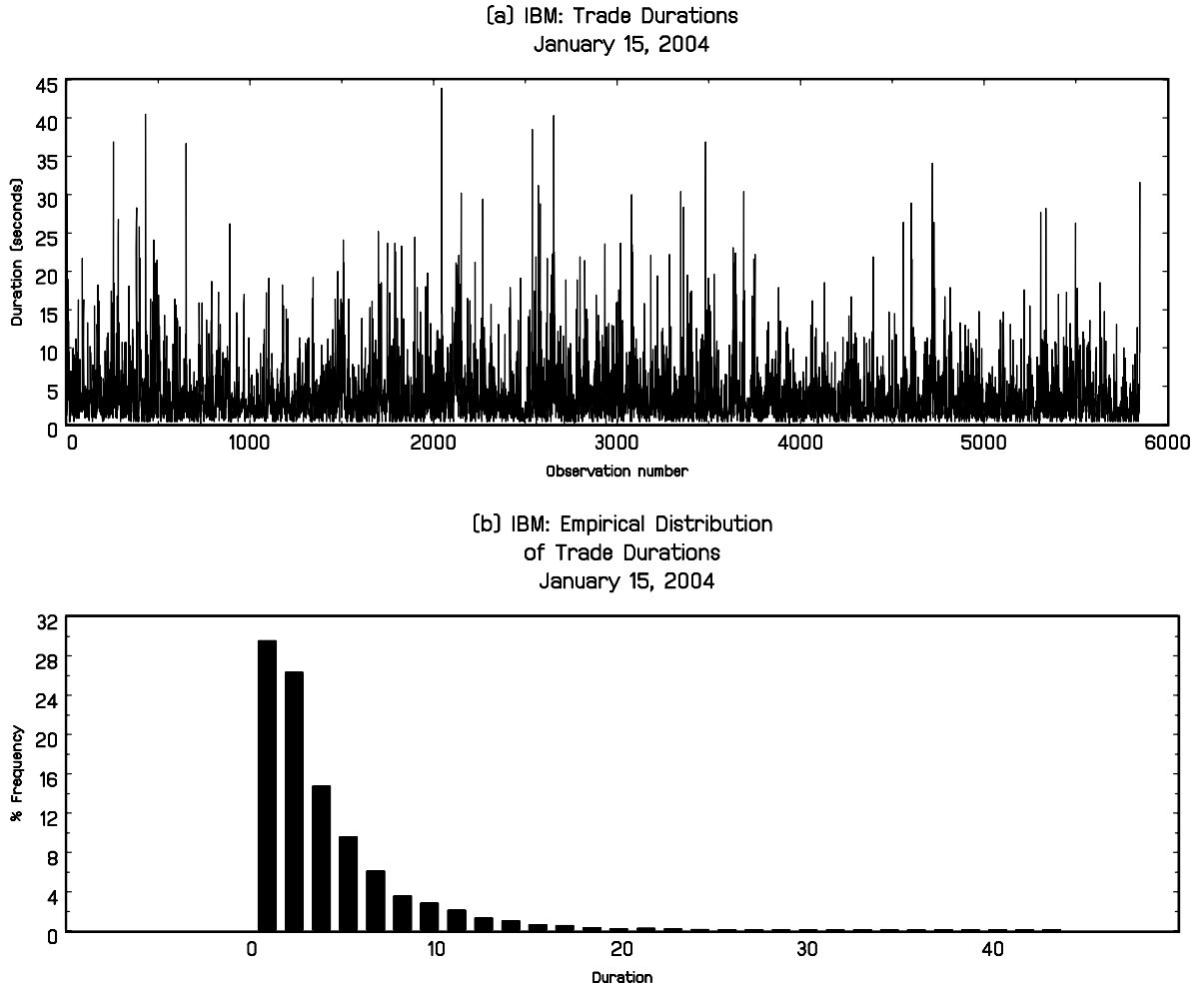


Figure 3: Duration (measured in seconds) between trades on IBM shares: January 15, 2004.

7 Conclusions

In this paper we have derived a quasi-locally most powerful (QLMP) test for testing for correlation in the conditional variance of data defined on the positive domain. For an analytically tractable leading-case model, the local power comparison conducted in Section 4 highlights the distinct benefit of applying a statistic that is adapted to positive, highly skewed data, with the relative power of an alternative squares-based test being shown to be negligible for such data. The finite sample simulation results reported in Section 5 confirm the superior performance of the QLMP test for fixed alternatives, relative to both the squares-based test and the test based on the first-order autocorrelation of the levels of the data. This superior performance obtains even when the data is generated under a conditional Weibull distribution,

rather than the conditional gamma distribution under which the QLMP test is derived. The new test is also shown to be robust to misspecification of the dynamics. In particular, its finite sample power is very close, in many cases, to that of the levels-based test, even in the case where the correlation affects the conditional mean only and the conditional variance - whose dynamics the QLMP test has been designed to detect - is fixed. This is in contrast with the finite sample power of the levels-based test, which is extremely low, even for large samples, when the conditional variance is autocorrelated and the conditional mean is not. In summary, all results reported in the paper highlight the gains to be had by incorporating appropriate information about the conditional distribution in the construction of the test statistic when testing for short-memory correlation in the conditional variance of positive, highly skewed data.

References

- [1] Anderson, T.W. (1971) *The Statistical Analysis of Time Series*, Wiley, NY.
- [2] Bauwens, L., Giot, P., Grammig, J. and Veredas, D. (2004) A Comparison of Financial Duration Models Via Density Forecasts, *International Journal of Forecasting*, 20, 589-609.
- [3] Bauwens, L. and Veredas, D. (2004), The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations, *Journal of Econometrics*, 199, 381-412.
- [4] Bauwens, L. and Hautsch, N. (2009), Modelling Financial High Frequency Data Using Point Processes, *Handbook of Financial Time Series*, Springer Berlin Heidelberg.
- [5] Bera, A.K. and Kim, S. (2002) Testing for Constancy of Correlation and Other Specifications of the BGARCH Model with an Application to International Equity Returns, *Journal of Empirical Finance*, 9, 171-195.
- [6] Chesher, A. (1984) Testing for Neglected Heterogeneity, *Econometrica*, 52, 865-872.
- [7] Cox, D.R. (1981) Statistical Analysis of Time Series: Some Recent Developments, *Scandinavian Journal of Statistics*, 8, 93-115.
- [8] Cox, D.R. (1983) Some Remarks on Overdispersion, *Biometrika*, 70, 269-274.

- [9] Cox, D.R. and Hinkley, D.V. (1979) *Theoretical Statistics*, Chapman and Hall, London.
- [10] Davis, R.A. and Rodriguez-Yam, G. (2005) Estimation for State-Space Models Based on a Likelihood Approximation, *Statistica Sinica*, 15, 381-406.
- [11] Engle, R.F. and Russell, J.R. (1998) Autoregressive Conditional Duration: a New Approach for Irregularly Spaced Transaction Data, *Econometrica*, 66, 987-1007.
- [12] Ghysels, E., Gouriou, C. and Jasiak, J. (2004) Stochastic Volatility Duration Models, *Journal of Econometrics*, 119, 413-433.
- [13] Huber, P., Ronchetti, E. and Victoria-Fese, M-P. (2004) Estimation of Generalized Linear Latent Variable Models, *Journal of the Royal Statistical Society, Series B*, 66, 893-908.
- [14] Lawrance, A.J. and Lewis, P.A.W. (1980) The Exponential Autoregressive-Moving Average (EARMA) Process, *Journal of the Royal Statistical Society, Series B*, 42, 150-161.
- [15] Lawrance, A.J. and Lewis, P.A.W. (1985) Modelling and Residual Analysis of Nonlinear Autoregressive Time Series in Exponential Variables, *Journal of the Royal Statistical Society, Series B*, 47, 165-202.
- [16] Lewis, P.A.W., McKenzie, E. and Hugus, D.K. (1989) Gamma Processes, *Communication in Statistics*, 5, 1-30.
- [17] McCabe, B.P.M. and Leybourne, S.J. (2000) A General Method of Testing for Random Parameter Variation in Statistical Models, in *Innovations in Multivariate Statistical Analysis: a Festschrift for Heinz Neudecker*, eds. Heijmans, Pollock and Satorra, Kluwer, Amsterdam.
- [18] McCabe, B.P.M. and Tremayne, A.R. (1993) *Elements of Modern Asymptotic Theory with Statistical Applications*, Manchester University Press, Manchester.
- [19] Ristic, M.M. (2005) A Beta-Gamma Autoregressive Process of the Second Order (BGAR(2)), *Statistics and Probability Letters*, 73, 403-410.
- [20] Strickland, C.M., C.S. Forbes and Martin, G.M. (2006) Bayesian Analysis of the Stochastic Conditional Duration Model, *Computational Statistics and Data Analysis (Special Issue on Statistical Signal Extraction and Filtering)*, 50, 2247-2267.

- [21] Strickland, C.M., Martin, G.M. and Forbes, C.S., 2008, 'Parameterization and Efficient MCMC Estimation of Non-Gaussian State Space Models', *Computational Statistics and Data Analysis; Special Issue on Computational Methods in Finance*, 52, 2911-2930.
- [22] Stuart, A., Ord, K.J., and Arnold, S. (1998) *Kendall's Advanced Theory of Statistics, Volume 2A: Classical Inference and the Linear Model*, 6th Edition, Arnold.
- [23] Van der Vart, A. W. (1998) *Asymptotic Statistics*, Cambridge University Press, Cambridge.

Appendix: Proof of Proposition 1

Here we collect, for convenience, some background results used in the proof. With reference to the gamma distribution, with density in (14), the conditional moments of y_t are defined as

$$\begin{aligned}
 E_{y|\lambda} [y_t] &= \mu_{y|\lambda} = -\frac{1}{\theta} = \mu_y \\
 E_{y|\lambda} [y_t - \mu_{y|\lambda}]^2 &= \mu_y^2 \lambda_t \\
 E_{y|\lambda} [y_t - \mu_{y|\lambda}]^3 &= 2\mu_y^3 \lambda_t^2 \\
 E_{y|\lambda} [y_t - \mu_{y|\lambda}]^4 &= 3\mu_y^4 (2\lambda_t^3 + \lambda_t^2).
 \end{aligned} \tag{A.1}$$

Also under the gamma distribution, the conditional moment generating function of $\log(y_t)$ is given by

$$M_{\log(y_t)|\lambda}(s) = E[y_t^s | \theta, \lambda_t] = \frac{\Gamma\left(\frac{1}{\lambda_t} + s\right)}{\Gamma\left(\frac{1}{\lambda_t}\right)} \left(\frac{-\theta}{\lambda_t}\right)^{-s}. \tag{A.2}$$

Note that this expression also gives the raw (conditional) moments of y_t . The relevant conditional moments of $\log(y_t)$ are given by the following derivatives,

$$M'_{\log(y_t)|\lambda}(0) = E_{y|\lambda} [\log(y_t)] = -\log\left(\frac{-\theta}{\lambda_t}\right) + \Psi(\lambda_t^{-1}), \tag{A.3}$$

$$\begin{aligned}
 M'_{\log(y_t)|\lambda}(1) &= E_{y|\lambda} [y_t \log(y_t)] \\
 &= E_{y|\lambda} \left[\left. \frac{\partial}{\partial s} y_t^{s+1} \right|_{s=0} \right] \\
 &= \left(\frac{-\theta}{\lambda_t}\right)^{-1} \frac{1}{\Gamma(\lambda_t^{-1})} \Gamma(\lambda_t^{-1} + 1) \left[-\log\left(\frac{-\theta}{\lambda_t}\right) + \Psi(\lambda_t^{-1} + 1) \right]
 \end{aligned} \tag{A.4}$$

and

$$\begin{aligned}
 M''_{\log(y_t)|\lambda}(0) &= E_{y|\lambda} [\log^2(y_t)] \\
 &= E_{y|\lambda} \left[\left. \frac{\partial^2}{\partial s^2} y_t^s \right|_{s=0} \right] \\
 &= \log^2\left(\frac{-\theta}{\lambda_t}\right) - 2\log\left(\frac{-\theta}{\lambda_t}\right) \Psi(\lambda_t^{-1}) + \Psi'(\lambda_t^{-1}) + \Psi^2(\lambda_t^{-1}).
 \end{aligned} \tag{A.5}$$

Finally, the uncentered joint k_j th moments of the log-normal λ_t , $t = 1, 2, \dots, T$, are given by

$$E_{\lambda}(\prod_{j=1}^T \lambda_j^{k_j}) = e^{\{\mathbf{k}'\boldsymbol{\mu}_x + \frac{1}{2}\mathbf{k}'\boldsymbol{\Sigma}_x\mathbf{k}\}}, \tag{A.6}$$

where \mathbf{k} is the $(T \times 1)$ vector with j th element k_j . Note, this implies that

$$Cov(\lambda_i, \lambda_j) = \left\{ e^{\rho^{i-j}\sigma_x^2} - 1 \right\} e^{2\mu_x + \sigma_x^2},$$

where $\sigma_x^2 = \sigma_\eta^2 / (1 - \rho^2)$.

Proposition 1 is proved via a sequence of lemmata where it is assumed that the model (1) to (4) and (14) holds. Lemma 1 derives the form of the ARE quotient pertinent to the statistics at hand.

Lemma 1 *The ARE of S_d to S_g is*

$$ARE_{d,g} = \left[\frac{\left. \frac{\partial Cov_y[d(y_t), d(y_{t-1})]}{\partial \rho} \right|_{\rho=0} / \sigma_{d,0}^2}{\left. \frac{\partial Cov_y[g(y_t), g(y_{t-1})]}{\partial \rho} \right|_{\rho=0} / \sigma_{g,0}^2} \right]^2, \quad (\text{A.7})$$

where $\sigma_{u,0}^2$, $u \in \{d, g\}$, is the variance of $u(y_t)$ under the null $H_0 : \rho = 0$.

Proof of Lemma 1. The statistics S_u , $u \in \{d, g\}$, have the general form

$$S_u = (\mathbf{u} - E_y[\mathbf{u}])' \mathbf{A} (\mathbf{u} - E_y[\mathbf{u}]),$$

where the elements of \mathbf{u} are given by $u(y_t)$. It is well known that

$$\mu_{S_u}(\rho) = tr [\mathbf{A} E_y [(\mathbf{u} - E_y[\mathbf{u}]) (\mathbf{u} - E_y[\mathbf{u}])']]$$

and hence it follows that

$$\begin{aligned} \frac{\partial \mu_{S_u}(\rho)}{\partial \rho} &= tr \left[\mathbf{A} \frac{\partial E_y [(\mathbf{u} - E_y[\mathbf{u}]) (\mathbf{u} - E_y[\mathbf{u}])']}{\partial \rho} \right] \\ &= \sum_{t=2}^T \frac{\partial E_y [(u(y_t) - \mu_u) (u(y_{t-1}) - \mu_u)]}{\partial \rho} \\ &= (T-1) \frac{\partial Cov_y [u(y_2), u(y_1)]}{\partial \rho}. \end{aligned}$$

The variance of S_u under the null of independence is (see Anderson, 1971)

$$\sigma_{S_u}^2(\rho) \Big|_{\rho=0} = (m_{u,4,0} - 3\sigma_{u,0}^4) \sum_{t=1}^T a_{tt}^2 + 2\sigma_{u,0}^4 tr(\mathbf{A}^2) = 2\sigma_{u,0}^4 tr(\mathbf{A}^2),$$

where $m_{u,4,0}$ is the fourth centred moment of $u(y_t)$ under the null and we use the property that $a_{tt} = 0$ for all t , where a_{tt} is the t th diagonal element of \mathbf{A} . Inserting $u \in \{d, g\}$ in the definition of the ARE completes the proof. ■

Lemma 2 derives the variances in (A.7) under the null hypothesis, while Lemma 3 derives the mean shifts.

Lemma 2 Under the null hypothesis of independence ($H_0 : \rho = 0$), the variances of $d(y_t) = (y_t - \mu_y)^2$ and $g(y_t) = \frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right)$ are given by

$$\begin{aligned}\sigma_{d,0}^2 &= \mu_y^4 [6E_\lambda [\lambda_t^3] + 3E_\lambda [\lambda_t^2] - E_\lambda [\lambda_t]^2] \\ &= \mu_y^4 \left[6e^{3\mu_x + \frac{9}{2}\sigma_\eta^2} + 3e^{2\mu_x + 2\sigma_\eta^2} - e^{2\mu_x + \sigma_\eta^2} \right]\end{aligned}\quad (\text{A.8})$$

and

$$\begin{aligned}\sigma_{g,0}^2 &= V_\lambda [\Psi(\lambda_t^{-1}) + \log(\lambda_t)] - E_\lambda [\lambda_t] + E_\lambda [\Psi'(\lambda_t^{-1})] \\ &= V_N [\Psi(e^{-x_t}) + x_t] - e^{\mu_x + \frac{1}{2}\sigma_\eta^2} + E_N [\Psi'(e^{-x_t})]\end{aligned}\quad (\text{A.9})$$

respectively.

Proof of Lemma 2. Using (A.1), it follows that

$$\begin{aligned}\sigma_{d,0}^2 &= E_y [d^2(y_t)] - \{E_y[d(y_t)]\}^2 \\ &= E_\lambda \left\{ E_{y|\lambda} \left[(y_t - \mu_y)^4 \right] \right\} - \left[E_\lambda \left\{ E_{y|\lambda} \left[(y_t - \mu_y)^2 \right] \right\} \right]^2 \\ &= E_\lambda [3\mu_y^4 (2\lambda_t^3 + \lambda_t^2)] - \mu_y^4 \{E_\lambda [\lambda_t]\}^2 \\ &= \mu_y^4 [6E_\lambda [\lambda_t^3] + 3E_\lambda [\lambda_t^2] - \{E_\lambda [\lambda_t]\}^2].\end{aligned}\quad (\text{A.10})$$

Now consider the corresponding function for $g(y_t)$,

$$\begin{aligned}\sigma_{g,0}^2 &= \text{Var}_y \left[\frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right) \right] \\ &= E_y \left\{ \left[\frac{y_t}{\mu_y} - \log(y_t) \right]^2 \right\} - \left\{ E_y \left[\frac{y_t}{\mu_y} - \log(y_t) \right] \right\}^2 \\ &= E_\lambda \left\{ E_{y|\lambda} \left[\frac{y_t^2}{\mu_y^2} - 2\frac{y_t}{\mu_y} \log(y_t) + \log^2(y_t) \right] \right\} - \left[E_\lambda \left\{ E_{y|\lambda} \left[\frac{y_t}{\mu_y} - \log(y_t) \right] \right\} \right]^2.\end{aligned}$$

Using (A.2) to (A.5), we obtain

$$\begin{aligned}\sigma_{g,0}^2 &= E_\lambda [\lambda_t] + 1 - 2\mu_y^{-1}\theta^{-1}E_\lambda [\log(-\theta\lambda_t^{-1}) - \Psi(\lambda_t^{-1} + 1)] \\ &\quad + E_\lambda [\Psi'(\lambda_t^{-1}) + \Psi^2(\lambda_t^{-1}) + \log^2(-\theta\lambda_t^{-1}) - 2\Psi(\lambda_t^{-1})\log(-\theta\lambda_t^{-1})] \\ &\quad - \{E_\lambda [1 - \Psi(\lambda_t^{-1}) + \log(-\theta\lambda_t^{-1})]\}^2.\end{aligned}$$

Simplifying, and using the fact that $\Psi(a+1) - \Psi(a) = a^{-1}$, we obtain

$$\sigma_{g,0}^2 = \text{Var} [\Psi(\lambda_t^{-1}) - \log(\lambda_t^{-1})] - E[\lambda_t] + E[\Psi'(\lambda_t^{-1})]. \quad (\text{A.11})$$

Finally, substitute $\lambda_t = e^{x_t}$ in (A.10) and (A.11), and use (A.6), to produce (A.8) and (A.9) respectively. ■

Lemma 3 *The derivatives of the covariances of d and g are given by*

$$\begin{aligned} \left. \frac{\partial}{\partial \rho} \text{Cov}_y [d(y_t)d(y_{t-1})] \right|_{\rho=0} &= \mu_y^4 \left. \frac{\partial}{\partial \rho} \text{Cov}_\lambda [\lambda_t \lambda_{t-1}] \right|_{\rho=0} \\ &= \mu_y^4 \sigma_\eta^2 e^{2\mu_x + \sigma_\eta^2} \end{aligned} \quad (\text{A.12})$$

and

$$\begin{aligned} \left. \frac{\partial}{\partial \rho} \text{Cov}_y [g(y_t)g(y_{t-1})] \right|_{\rho=0} &= \left. \frac{\partial}{\partial \rho} \text{Cov}_\lambda [\log(\lambda_t) + \Psi(\lambda_t^{-1}), \log(\lambda_{t-1}) + \Psi(\lambda_{t-1}^{-1})] \right|_{\rho=0} \\ &= \sigma_\eta^2 \{1 - E_N [\Psi'(e^{-x_t})(e^{-x_t})]\}^2, \end{aligned} \quad (\text{A.13})$$

where $d(y_t) = (y_t - \mu_y)^2$ and $g(y_t) = \frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right)$.

Proof of Lemma 3. Given conditional independence, and using the expressions in (A.1), it follows that

$$\begin{aligned} \text{Cov}_y [d(y_t)d(y_{t-1})] &= E_y [\{d(y_t) - E_y [d(y_t)]\} \{d(y_{t-1}) - E_y [d(y_{t-1})]\}] \\ &= E_\lambda [E_{y|\lambda} \{d(y_t) - E_y [d(y_t)]\} E_{y|\lambda} \{d(y_{t-1}) - E_y [d(y_{t-1})]\}] \\ &= E_\lambda [\{\mu_y^2 \lambda_t - \mu_y^2 E_\lambda [\lambda_t]\} \{\mu_y^2 \lambda_{t-1} - \mu_y^2 E_\lambda [\lambda_{t-1}]\}] \\ &= \mu_y^4 \text{Cov}_\lambda [\lambda_t \lambda_{t-1}]. \end{aligned} \quad (\text{A.14})$$

In the case of $g(y_t) = \frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right)$, using $E_y \left[\frac{y_t}{\mu_y}\right] = E_{y|\lambda} \left[\frac{y_t}{\mu_y}\right] = 1$, we obtain

$$E_{y|\lambda} \{g(y_t) - E_y [g(y_t)]\} = -E_{y|\lambda} [\log(y_t) - E_y \log(y_t)].$$

Hence, using (A.3) we obtain

$$\begin{aligned} \text{Cov}_y [g(y_t)g(y_{t-1})] &= E_\lambda [E_{y|\lambda} \{\log y_t - E_y(\log y_t)\} E_{y|\lambda} \{\log y_{t-1} - E_y(\log y_{t-1})\}] \\ &= \text{Cov}_\lambda [\log(\lambda_t) + \Psi(\lambda_t^{-1}), \log(\lambda_{t-1}) + \Psi(\lambda_{t-1}^{-1})]. \end{aligned} \quad (\text{A.15})$$

Under the log-normal assumption for λ_t , the expressions (A.14) and (A.15) respectively become

$$\text{Cov}_y [d(y_t)d(y_{t-1})] = \mu_y^4 (E[e^{x_{t-1}+x_t}] - \mu_\lambda^2) \quad (\text{A.16})$$

and

$$\begin{aligned} \text{Cov}_y [g(y_t)g(y_{t-1})] &= E[(x_t - \mu_x)(x_{t-1} - \mu_x)] + 2E[(x_t - \mu_x)(\Psi(e^{-x_{t-1}}) - \mu_\Psi)] \\ &\quad + E[(\Psi(e^{-x_t}) - \mu_\Psi)(\Psi(e^{-x_{t-1}}) - \mu_\Psi)], \end{aligned} \quad (\text{A.17})$$

where all expectations in (A.16) and (A.17) are with respect to the joint distribution of (x_{t-1}, x_t) and $\mu_\Psi = E_N[\Psi(e^{-x_t})]$ is the marginal expectation of $\Psi(e^{-x_t})$. The quantity we

are interested in is the derivative of each of these expressions with respect to ρ , evaluated at $\rho = 0$. Denoting the marginal and joint densities of x_t and (x_{t-1}, x_t) by $f(x_t)$ and $f(x_{t-1}, x_t)$ respectively, we note that

$$\left. \frac{\partial}{\partial \rho} f(x_t) \right|_{\rho=0} = 0$$

and

$$\left. \frac{\partial}{\partial \rho} f(x_{t-1}, x_t) \right|_{\rho=0} = \frac{(x_{t-1} - \mu_x)(x_t - \mu_x)}{\sigma_\eta^2} f_N(x_{t-1}) f_N(x_t),$$

where f_N denotes the Gaussian density with mean μ_x and variance σ_η^2 . Interchanging the order of differentiation and integration, and using Stein's Lemma for $N(\mu_x, \sigma_\eta^2)$ variables,

$$E_N[h(x_t)(x_t - \mu_x)] = \sigma_\eta^2 E_N[h'(x_t)],$$

we obtain

$$\begin{aligned} \left. \frac{\partial}{\partial \rho} \text{Cov}_y[d(y_t)d(y_{t-1})] \right|_{\rho=0} &= \mu_y^4 \sigma_\eta^{-2} \{E_N[(x_t - \mu_x) e^{x_t}]\}^2 \\ &= \mu_y^4 \sigma_\eta^{-2} \{\sigma_\eta^2 E_N[e^{x_t}]\}^2, \end{aligned}$$

where we have used the result that $\left. \frac{\partial}{\partial \rho} \mu_\lambda^2 \right|_{\rho=0} = 0$. Invoking (A.6), we obtain the expression in (A.12). Using similar analysis, we produce the expression in (A.13),

$$\begin{aligned} \left. \frac{\partial}{\partial \rho} \text{Cov}_y[g(y_t)g(y_{t-1})] \right|_{\rho=0} &= \sigma_\eta^2 + 2E_N[(x_t - \mu_x) \Psi(e^{-x_t})] + \sigma_\eta^{-2} \{E_N[(x_t - \mu_x) \Psi(e^{-x_t})]\}^2 \\ &= \sigma_\eta^2 + 2\sigma_\eta^2 E_N[\Psi'(e^{-x_t})(-e^{-x_t})] + \sigma_\eta^{-2} \{\sigma_\eta^2 E_N[\Psi'(e^{-x_t})](-e^{-x_t})\}^2 \\ &= \sigma_\eta^2 - 2\sigma_\eta^2 E_N[\Psi'(e^{-x_t})(e^{-x_t})] + \sigma_\eta^2 \{E_N[\Psi'(e^{-x_t})](e^{-x_t})\}^2 \\ &= \sigma_\eta^2 \{1 - E_N[\Psi'(e^{-x_t})(e^{-x_t})]\}^2. \end{aligned}$$

■

Proof of Proposition 1. The proof is a straightforward combination of Lemmata 1, 2 and 3. ■