THE RELATIVE POWER OF TESTS FOR CORRELATION IN THE SECOND MOMENT OF POSITIVE DATA[*]

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

and KEITH FREELAND[§]

June 17, 2009

Abstract

This paper investigates the relative power of a squares-based test for short-memory correlation in the second moment of strictly positive, skewed data. The comparison test is a quasi-locally most powerful test derived under the assumption of conditionally gamma data. Analytical asymptotic relative efficiency calculations show that the squares-based test has negligible relative power in empirically relevant scenarios. Finite sample simulation results confirm the poor performance of the squares-based test for fixed alternatives. The conclusion to be drawn from the results is that substantial power gains can be produced by incorporating appropriate distributional information in the derivation of the test.

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

JEL Codes: C12, C16, C22.

1

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, Gourieroux and Jasiak, 2004), with the autocorrelation function (ACF) of the squared data being used as a preliminary 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 statistic.

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 assumed to be a 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 (LMP) 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 ARE results show that the test based on the squares of the data 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 LMP test in some instances. Robustness of the finite sample power results to misspecification of the conditional distribution is also demonstrated.

The outline of the paper is as follows. In Section 2, the LMP test statistic is derived. Section 3 gives details of the asymptotic theory of both the LMP 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 LMP test highlighted. These local power results are supplemented with fixed alternative power comparisons, via Monte Carlo simulations, in Section 5. Some conclusions are provided in Section 6.

---

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

2 Derivation of the Score Test

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

$$f(\mathbf{y}) = \int \ldots \int f(\mathbf{y}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda}) d\boldsymbol{\lambda} = E_{\boldsymbol{\lambda}} \left[f(\mathbf{y}|\boldsymbol{\lambda}) \right], \tag{1}$$

where $\boldsymbol{\lambda}$ denotes a $(T \times 1)$ vector of latent variables $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_T]'$. Each λ_t, $t = 1, 2, \ldots, T$, 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,

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

with the $(T \times 1)$ vector $\mathbf{x}$ defined as $\mathbf{x} = [x_1, x_2, \ldots, x_T]'$, and $\boldsymbol{\Sigma_x} = Var_{\mathbf{x}}[\mathbf{x}]$. We also assume that

$$f(\mathbf{y}|\boldsymbol{\lambda}) = f(y_1|\lambda_1) f(y_2|\lambda_2) \ldots f(y_T|\lambda_T), \tag{3}$$

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 $\rho > 0$, so that λ_t is a correlated sequence with short memory.

Following Cox (1983) and McCabe and Leybourne (2000)[2], we define $f^*(\mathbf{y})$ as the second-order Taylor series expansion of $f(\mathbf{y}|\boldsymbol{\lambda})$ about $\boldsymbol{\mu_\lambda} = [\mu_\lambda, \mu_\lambda, \ldots, \mu_\lambda]' = E_{\boldsymbol{\lambda}}[\boldsymbol{\lambda}]$. Defining $L(\boldsymbol{\lambda}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\lambda})$ and $\ell = \log L$, we may write

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

where $\boldsymbol{\Sigma_\lambda} = Var_{\boldsymbol{\lambda}}[\boldsymbol{\lambda}]$ and $\mathbf{M} = \left(\frac{\partial \ell}{\partial \boldsymbol{\lambda}} \frac{\partial \ell}{\partial \boldsymbol{\lambda}'} + \frac{\partial^2 \ell}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \right)\Big|_{\boldsymbol{\lambda}=\boldsymbol{\mu_\lambda}}$.[3] Note that the second term in the expression for $\mathbf{M}$, $\frac{\partial^2 \ell}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'}\Big|_{\boldsymbol{\lambda}=\boldsymbol{\mu_\lambda}}$, is a diagonal matrix with elements,

$$\mathbf{r} = \left\{ \partial^2 \ell / \partial \lambda_t^2 \big|_{\lambda_t = \mu_\lambda} ; t = 1, \ldots, T \right\}, \tag{4}$$

---

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

[3] Clearly, $f^*(\mathbf{y})$ is an approximation to $f(\mathbf{y})$ for which the error is $O\left\{ E_{\boldsymbol{\lambda}} \left[\|\boldsymbol{\lambda} - \boldsymbol{\mu_\lambda}\|^3 \right] \right\}$. 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 $a(.)$, $b(.)$ and $c(.)$, $\ell(\lambda_t|y_t)$ may be written as

$$\ell(\lambda_t|y_t) = a(y_t) + b(\lambda_t) + c(y_t, \lambda_t), \tag{5}$$

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

$$u(y_t) = c'(y_t, \lambda_t)|_{\lambda_t = \mu_\lambda}, \tag{6}$$

we can write that

$$\ell'(\lambda_t|y_t)|_{\lambda_t = \mu_\lambda} = \{u(y_t) - \mu_u\} = u_c(y_t), \tag{7}$$

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 test (see, for example, Cox and Hinkley, 1979, Sect. 4.8) of

$$H_0 : \rho = 0 \text{ against } H_1 : \rho > 0, \tag{8}$$

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

$$S = tr\left(\mathbf{M} \frac{\partial \boldsymbol{\Sigma}_\lambda}{\partial \rho}\bigg|_{\rho=0}\right) \bigg/ \left(2 + \sigma_\lambda^2 tr\mathbf{M}\right), \tag{9}$$

where σ_λ^2 is the variance of λ_t under the null. It is well-known that $\frac{\partial \boldsymbol{\Sigma}_\mathbf{x}}{\partial \rho}\big|_{\rho=0} \propto \mathbf{A}$, 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 $\frac{\partial \boldsymbol{\Sigma}_\lambda}{\partial \rho}\big|_{\rho=0} \propto \mathbf{A}$. Hence, apart from constants, (9) can be re-expressed as

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

where $\mathbf{u}_c$ is defined as a $(T \times 1)$ vector with tth element $u_c(y_t)$ (as given in (7)), $\mathbf{i}$ is a $(T \times 1)$ vector of 1's and $\mathbf{r}$ is defined in (4). Standardizing in the usual way we obtain

$$S_T = T^{-1/2}\mathbf{u}_c'\mathbf{A}\mathbf{u}_c / \left[2/T + \sigma_\lambda^2 \left[\mathbf{u}_c'\mathbf{u}_c + \mathbf{r}'\mathbf{i} \right] /T \right] \tag{10}$$

and, using a suitable weak law of large numbers (WLLN), the denominator in (10) 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

$$S_u = \mathbf{u}_c'\mathbf{A}\mathbf{u}_c, \tag{11}$$

the statistic,

$$S_{u,T} = T^{-1/2}S_u = T^{-1/2} \sum_{t=2}^{T} u_c(y_t)u_c(y_{t-1}), \tag{12}$$

may be used to test $H_0 : \rho = 0$ and this is asymptotically equivalent to the statistic S_T in (10).

It is a simple matter to identify the function $u(\cdot)$ in (6) for any particular conditional distribution. 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,

$$f(y_t|\theta, \lambda_t) = \frac{1}{y_t} \times \frac{1}{\Gamma(\frac{1}{\lambda_t})} \left(\frac{1}{\lambda_t}\right)^{\frac{1}{\lambda_t}} \times (-\theta y_t)^{\frac{1}{\lambda_t}} \exp\left(\frac{\theta y_t}{\lambda_t}\right). \tag{13}$$

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, $G(\alpha, \beta)$, we have $\alpha = \lambda_t^{-1}$ and $\beta = -\lambda_t/\theta$. We adopt the parameterization in (13) 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 second order dependence in y_t. Noting that $\mu_y = E_y[y_t] = E_{y|\lambda}[y_t] = -\frac{1}{\theta}$, we see (with reference to (5)) that

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

with

$$g(y_t) = \frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right). \tag{14}$$

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. Setting $u(y_t)$ equal to

$$d(y_t) = \left(y_t - \mu_y\right)^2 \tag{15}$$

yields the squares-based test.

3   Asymptotic Distribution Theory

Because of the u-transformation of the y_t embodied in the statistic in (12), 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 (14) and (15)) 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 (3) and (13), 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. When y_t is mixing,

$$T^{-1/2} \sum_{t=2}^{T} \left[u_c(y_t) u_c(y_{t-1}) - E\left[u_c(y_t) u_c(y_{t-1}) \right] \right] \to^d N(0, \omega^2), \tag{16}$$

where $\omega^2 > 0$ is the usual long run variance of the sum in (16).

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 (3) under the null hypothesis that $\rho = 0$ in (2). 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 $E\left[u_c(y_t) u_c(y_{t-1}) \right] \neq 0$. Consider

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

As the first term in $S_{u,T}$ converges in distribution by (16) and the second term diverges, $S_{u,T}$ also diverges, and a two-sided test based on $S_{u,T}$ is therefore consistent whenever $E\left[u_c(y_t) u_c(y_{t-1}) \right] \neq 0$. *A fortiori* there is consistency against the model (1) to (3) under the alternative when $\rho \neq 0$. This follows because $\{x_t\}$ in (2) 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 (3) under the alternative hypothesis $\rho \neq 0$ in (2).

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 (11)), in the case of a positive and positively skewed DGP, are quantified in the following section via ARE calculations.

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

$$ARE_{d,g} = \lim_{T} \left[\frac{\left. \frac{\partial \mu_{S_d}(\rho)}{\partial \rho} \right|_{\rho=0} / \left. \sigma_{S_d}(\rho) \right|_{\rho=0}}{\left. \frac{\partial \mu_{S_g}(\rho)}{\partial \rho} \right|_{\rho=0} / \left. \sigma_{S_g}(\rho) \right|_{\rho=0}} \right]^2, \tag{17}$$

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

$$\frac{S_u - \mu_{S_u}(\rho)}{\sigma_{S_u}(\rho)} \to^d N(0, 1) \tag{18}$$

in some local region $\{0 \leq \rho < \delta\}$, which includes the alternative hypothesis. The condition in (18) 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 (16).

To evaluate the expression in (17), we use the specifications of the model (1) to (3) with conditional density as given in (13). 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.
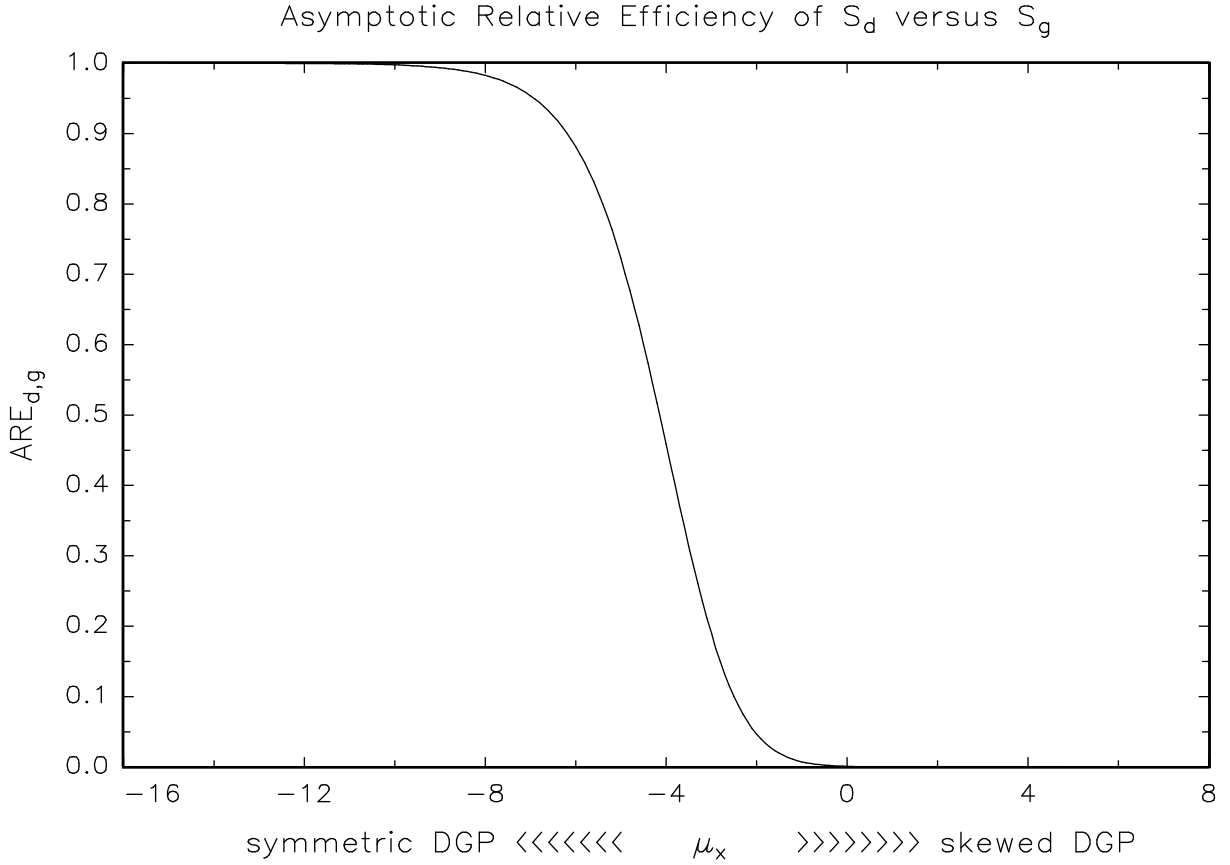
Figure 1: $ARE_{d,g}$ as a function of μ_x (for $\sigma_\eta = 1$). The degree of skewness in the representative conditional gamma distribution underlying the calculations is an increasing function of μ_x.

Proposition 1 *Under the model defined by (1) to (3) and (13), 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

$$b = \frac{\left\{1 - E_N\left[\Psi'\left(e^{-x_t}\right)\left(e^{-x_t}\right)\right]\right\}^2}{V_N\left[\Psi(e^{-x_t}) + x_t\right] - e^{\mu_x + \frac{1}{2}\sigma_\eta^2} + E_N\left[\Psi'(e^{-x_t})\right]}.$$
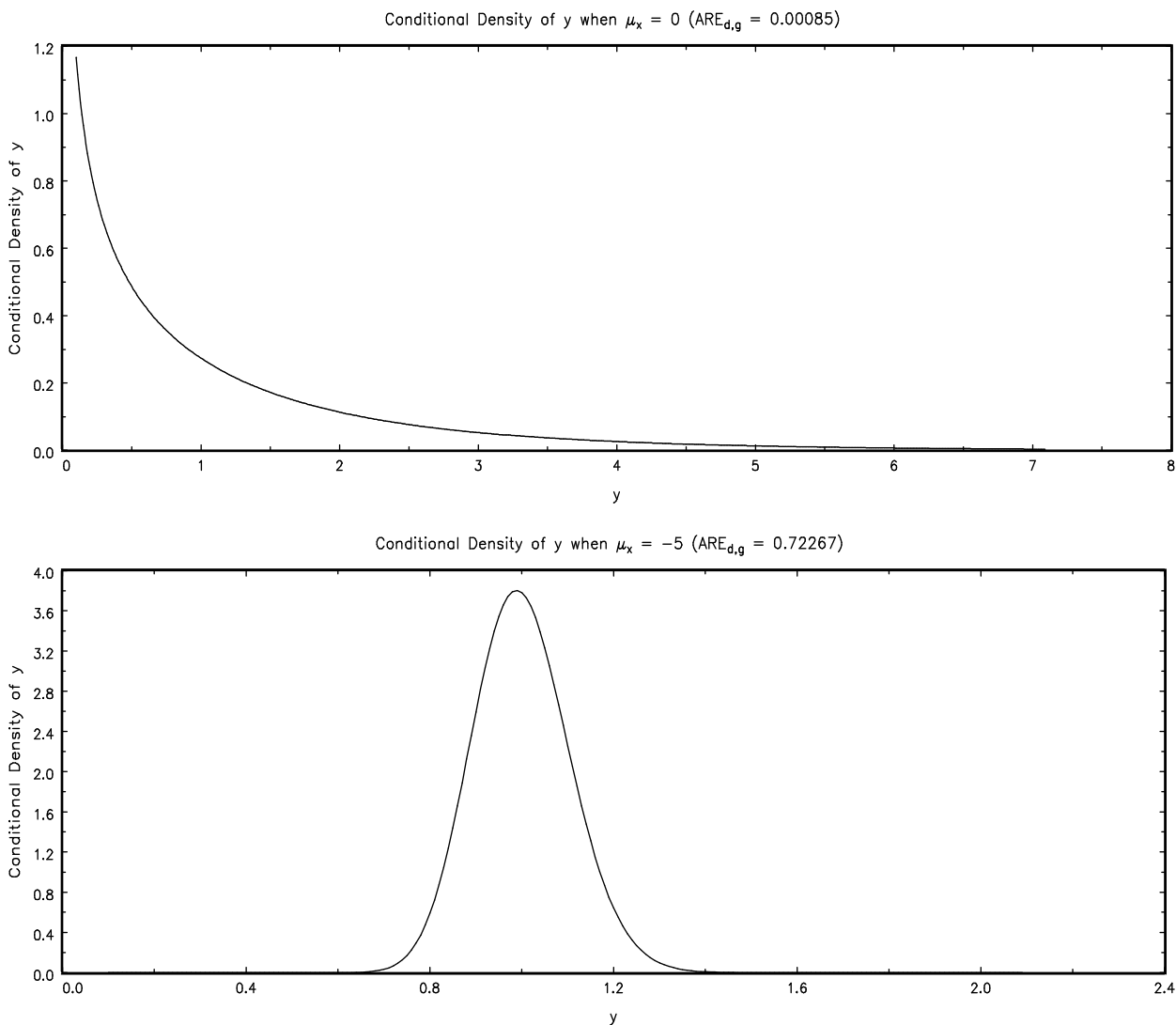
Conditional densities of y_t (and associated $ARE_{d,g}$ values) for selected values of μ_x.

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 $\alpha \to \infty$.[4] Figure 1 plots $ARE_{d,g}$ over

---

[4]The 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 $2\alpha^{-1/2}$.

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 applications[5]. 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 $\mu_x \to -\infty$ (for fixed σ_η^2) and $\alpha \to \infty$ as a consequence, both tests are equally efficient according to this measure.

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 $u \in \{d, g\}$,

$$\hat{\rho}_u = T^{-1/2} s_u^{-2} \sum_{t=2}^{T} \hat{u}_c(y_t)\hat{u}_c(y_{t-1}) \tag{19}$$

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, $s_u^2 \to^p \sigma_u^2$ by the WLLN. Hence, by the continuous mapping theorem and the CLT we have that $\hat{\rho}_u \to^d N(0, 1)$ 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 (19) under both conditional gamma and conditional Weibull distributions. The generating process for x_t is the AR(1) process in (2), with $\lambda_t = e^{x_t}$. With reference to the conditional gamma density in (13) and the AR(1) process for x_t in (2), we impose parameter values that ensure that the generated data is qualitatively similar to typical positive and very positively skewed data. Specifically, we produce data with a mean approximately equal to 1 and

---

[5]See, for example, the shape of the empirical distributions of trade durations in Engle and Russell (1998) and Bauwens *et al.* (2004).

Table 1:

Finite Sample Sizes and Powers of Tests of $H_0 : \rho = 0$ against $H_1 : \rho > 0$ under Strictly Positive Conditional Distributions

Empirical Size and Power

| | Gamma $f(y_t \| \lambda_t)$ | | | | | | Weibull $f(y_t \| \lambda_t)$ | | | | | |
| | $N = 200$ | | $N = 500$ | | $N = 2000$ | | $N = 200$ | | $N = 500$ | | $N = 2000$ | |
	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_g$	$\hat{\rho}_d$	$\hat{\rho}_g$	$\hat{\rho}_d$
$\rho =$												
0.0	0.047	0.034	0.053	0.034	0.056	0.032*	0.046	0.047	0.048	0.058	0.048	0.065*
0.1	0.088	0.058	0.102	0.060	0.171	0.067	0.089	0.054	0.119	0.055	0.255	0.057
0.3	0.213	0.070	0.327	0.081	0.693	0.099	0.233	0.060	0.424	0.064	0.907	0.075
0.5	0.410	0.086	0.642	0.104	0.966	0.139	0.448	0.067	0.780	0.072	1.000	0.093
0.7	0.644	0.119	0.869	0.141	0.993	0.191	0.632	0.070	0.927	0.077	1.000	0.099

(a) * denotes significantly different from the nominal value of 5%, using an asymptotic critical value of 1.96 for the sample proportion.

variance between about 1.2 and 2, as matches high frequency trade durations data (adjusted for the intraday pattern); see, for example, Strickland *et al.* (2006). 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 (2) is set at a value that ensures that for each value of ρ in (2), 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 $V_y[y_t]$. The conditional Weibull distribution, reparameterized 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. 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 both tests are reasonably close to the nominal value of 5%, with the correct test (for the data type) tending to have better finite sample size behaviour than the alternative squares-based test, in all experiments. Only in two cases are the empirical sizes significantly different from the

nominal size, with both of those cases relating to the $\hat{\rho}_d$ test. 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 obtaining 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 five times that of the squares-based $\hat{\rho}_d$ test. 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. Indeed, the dominance is even more marked for this particular distributional specification, with the $\hat{\rho}_g$ test having excellent power, whilst the power of the $\hat{\rho}_d$ test is abysmal, even for $N = 2000$. The numerical experiments also confirm the fixed-alternative consistency properties demonstrated theoretically in Section 3, with the power of all tests increasing as T increases. That said, the power of the squares-based test under strictly positive continuous data is still very low in large (but finite) samples.

6 Conclusions

In this paper we have compared the sampling properties of a squares-based correlation statistic with those of a quasi-score statistic for testing for correlation in the second moment 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 the squares-based test being negligible for such data. The finite sample simulation results reported in Section 5 confirm the superior performance of the new test for fixed alternatives, even when the data is generated under a conditional Weibull distribution, rather than the conditional gamma distribution under which the test is derived. In summary, all results reported in the paper raise doubts about the wisdom of using the autocorrelation function of squared data to test for short-memory correlation in the variance of positive, highly skewed data, with a marked loss of power resulting from a failure to incorporate appropriate distributional information in the construction of the test statistic.

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] 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.

[4] Chesher, A. (1984) Testing for Neglected Heterogeneity, *Econometrica,* 52, 865-872.

[5] Cox, D.R. (1981) Statistical Analysis of Time Series: Some Recent Developments, *Scandanavian Journal of Statistics,* 8, 93-115.

[6] Cox, D.R. (1983) Some Remarks on Overdispersion, *Biometrika,* 69, 269-274.

[7] Cox, D.R. and Hinkley, D.V. (1979) *Theoretical Statistics,* Chapman and Hall, London.

[8] Davis, R.A. and Rodriguez-Yam, G. (2005) Estimation for State-Space Models Based on a Likelihood Approximation, *Statistica Sinica,* 15, 381-406.

[9] Engle, R.F. and Russell, J.R. (1998) Autoregressive Conditional Duration: a New Approach for Irregularly Spaced Transaction Data, *Econometrica,* 66, 987-1007.

[10] Ghysels, E., Gourieroux, C. and Jasiak, J. (2004) Stochastic Volatility Duration Models, *Journal of Econometrics,* 119, 413-433.

[11] 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

[12] 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 Festchrift for Heinz Neudecker,* eds. Heijmans, Pollock and Satorra, Kluwer, Amsterdam.

[13] McCabe, B.P.M. and Tremayne, A.R. (1993) *Elements of Modern Asymptotic Theory with Statistical Applications,* Manchester University Press, Manchester.

[14] 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.

[15] 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.

[16] 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 (13), the conditional moments of y_t are defined as

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

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,

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

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

and

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

Finally, the uncentered joint k_jth moments of the log-normal λ_t, $t = 1, 2, ...T$, are given by

$$
E_\lambda(\Pi_{j=1}^T \lambda_j^{k_j}) = e^{\left\{\mathbf{k}'\boldsymbol{\mu_x} + \frac{1}{2}\mathbf{k}'\boldsymbol{\Sigma_x}\mathbf{k}\right\}},
\tag{A.6}
$$

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

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

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 (3) and (13) 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*

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

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

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

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

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

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

and hence it follows that

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

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

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

where $\sigma_{u,0}^2$ is the variance of $u(y_t)$ under H_0, $m_{u,4,0}$ 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 tth 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*

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

and

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

respectively.

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

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

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

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

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

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

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

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

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*

$$\frac{\partial}{\partial \rho} Cov_y \left[d(y_t) d(y_{t-1}) \right] \Big|_{\rho=0} = \mu_y^4 \frac{\partial}{\partial \rho} Cov_\lambda \left[\lambda_t \lambda_{t-1} \right] \Big|_{\rho=0}$$

$$= \mu_y^4 \sigma_\eta^2 e^{2\mu_x + \sigma_\eta^2} \qquad (A.12)$$

and

$$\frac{\partial}{\partial \rho} Cov_y \left[g(y_t) g(y_{t-1}) \right] \Big|_{\rho=0} = \frac{\partial}{\partial \rho} Cov_\lambda \left[\log(\lambda_t) + \Psi(\lambda_t^{-1}), \log(\lambda_{t-1}) + \Psi(\lambda_{t-1}^{-1}) \right] \Big|_{\rho=0}$$

$$= \sigma_\eta^2 \left\{ 1 - E_N \left[\Psi' \left(e^{-x_t} \right) \left(e^{-x_t} \right) \right] \right\}^2, \qquad (A.13)$$

where $d(y_t) = \left(y_t - \mu_y \right)^2$ 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

$$Cov_y \left[d(y_t) d(y_{t-1}) \right] = E_y \left[\{ d(y_t) - E_y \left[d(y_t) \right] \} \{ d(y_{t-1}) - E_y \left[d(y_{t-1}) \right] \} \right]$$

$$= E_\lambda \left[E_{y|\lambda} \{ d(y_t) - E_y \left[d(y_t) \right] \} E_{y|\lambda} \{ d(y_{t-1}) - E_y \left[d(y_{t-1}) \right] \} \right]$$

$$= E_\lambda \left[\{ \mu_y^2 \lambda_t - \mu_y^2 E_\lambda \left[\lambda_t \right] \} \{ \mu_y^2 \lambda_{t-1} - \mu_y^2 E_\lambda \left[\lambda_t \right] \} \right]$$

$$= \mu_y^4 Cov_\lambda \left[\lambda_t \lambda_{t-1} \right]. \qquad (A.14)$$

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

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

Hence, using (A.3) we obtain

$$Cov_y \left[g(y_t) g(y_{t-1}) \right] = E_\lambda \left[E_{y|\lambda} \{ \log y_t - E_y (\log y_t) \} E_{y|\lambda} \{ \log y_{t-1} - E_y (\log y_{t-1}) \} \right]$$

$$= Cov_\lambda \left[\log \left(\lambda_t \right) + \Psi \left(\lambda_t^{-1} \right), \log \left(\lambda_{t-1} \right) + \Psi \left(\lambda_{t-1}^{-1} \right) \right]. \qquad (A.15)$$

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

$$Cov_y \left[d(y_t) d(y_{t-1}) \right] = \mu_y^4 \left(E \left[e^{x_{t-1} + x_t} \right] - \mu_\lambda^2 \right) \qquad (A.16)$$

and

$$Cov_y \left[g(y_t) g(y_{t-1}) \right] = E \left[\left(x_t - \mu_x \right) \left(x_{t-1} - \mu_x \right) \right] + 2E \left[\left(x_t - \mu_x \right) \left(\Psi \left(e^{-x_{t-1}} \right) - \mu_\Psi \right) \right]$$

$$+ E \left[\left(\Psi \left(e^{-x_t} \right) - \mu_\Psi \right) \left(\Psi \left(e^{-x_{t-1}} \right) - \mu_\Psi \right) \right], \qquad (A.17)$$

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

$$\frac{\partial}{\partial \rho} f(x_t) \bigg|_{\rho=0} = 0$$

and

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

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

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

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

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

∎

Proof of Proposition 1. The proof is a straightfoward combination of Lemmata 1, 2 and 3. ∎