Testing for Dependence in Non-Gaussian Time Series Data

B.P.M. McCabe*, G.M. Martin† and R.K. Freeland‡

June 15, 2004

Abstract

This paper provides a general methodology for testing for dependence in time series data, with particular emphasis given to non-Gaussian data. A dynamic model is postulated for a continuous latent variable and the dynamic structure transferred to the non-Gaussian, possibly discrete, observations. Locally most powerful tests for various forms of dependence are derived, based on an approximate likelihood function. Invariance to the distribution adopted for the data, conditional on the latent process, is shown to hold in certain cases. The tests are applied to various financial data sets, and Monte Carlo experiments used to gauge their finite sample properties.

Key Words: Latent variable model; locally most powerful tests; approximate likelihood; correlation tests; stochastic volatility tests.

JEL Codes: C12, C16, C22.

*School of Management, University of Liverpool, UK.

†Department of Econometrics and Business Statistics, Monash University, Australia. Corresponding author: E-Mail address: Gael.Martin@buseco.monash.edu.au. Postal Address: Department of Econometrics and Business Statistics, Faculty of Business and Economics, Monash University, Clayton, Victoria 3800, Australia. Phone Number: +61 3 9905 1189. Fax Number: +61 3 9905 5474.

‡Department of Actuarial Science and Statistics, University of Waterloo, Canada.

1 Introduction

This paper provides a general methodology for testing for dependence in time series data. Particular emphasis is given to non-Gaussian data as the typically restricted sample space of the non-Gaussian variable leaves the use of existing test procedures open to question. As it is difficult to formulate structural models of dependence that apply generally to the wide range of data types that comprise the non-Gaussian class, a latent variable approach is adopted. That is, a dynamic model is postulated for a continuous latent variable and the dynamics transferred to the non-Gaussian, possibly discrete, observations, via a response function that defines stochastic parameters on which the non-Gaussian variable depends.

In an extension of the approach adopted in Cox (1983) and McCabe and Leybourne (2000), the methodology is based on an approximate likelihood function, whereby the expectation of the distribution of the data, conditional on the unobservable stochastic parameters, is evaluated in a region local to the mean of the parameter distribution. As the approximate likelihood is a function of only the first and second moments of the latent process, and not of the latent variables themselves, the full probabilistic structure of the unobservable process need not be specified. In addition, as the test statistics are derived from the approximate likelihood, there are no computational issues arising from the presence of a high dimensional vector of unobservable variables. The tests are derived as locally most powerful (LMP) tests and have maximum power in the region in which the approximation to the true likelihood is most accurate. Particular attention is paid to the case where the conditional distribution of the data is a member of the exponential family, as this allows for a unified treatment of random variables of many different types. The tests can be viewed as a preliminary step in the analysis of the data, with specific dynamic models being formulated and estimated should dependence need to be modelled.

The general framework includes cases in which the conditional mean of the observed variable is a function of the stochastic parameter, in which case the procedure produces a test for correlation in the levels of the variable. We demonstrate by constructing tests for both short and long memory correlation for any conditional distribution within the exponential family. The tests are shown to be invariant with respect to members of the exponential family. When using standardized variables to eliminate unknown means and variances, the statistics are in fact equivalent, under weak conditions, to the corresponding Locally Best Invariant (LBI) tests of the covariance matrix of a Gaussian unconditional distribution. Since such LBI statistics form the basis of well known tests for correlation, in effect this result demonstrates a form of optimality for standard correlation tests in a broader distributional setting. For cases in which the stochastic parameter is related to higher-order conditional moments, the methodology provides a mechanism for producing tests for higher-order dependence. In order to highlight this fact we produce tests for short and long memory stochastic volatility. In contrast with the tests for correlation, optimal tests for dependence in the second moment do depend on the particular member of the exponential family chosen to model the conditional distribution of the data.

The outline of the paper is as follows. In Section 2, the general latent variable model is defined and the nature of the transfer of dependence from the latent process to the data demonstrated. As is highlighted therein, the nature of the response function, as well as the values of various moments, are crucial in determining the extent of the dependence transfer from the latent to the observed process. Section 3 outlines the approximation to the likelihood function, based on a Taylor series expansion of the distribution of the data, conditional on the stochastic parameters. The general test procedure is then outlined in Section 4, including details of some simplifications that can occur. When the data is standardized, the approximate LMP statistic is shown to be equivalent to the exact Gaussian LBI statistic derived under weak conditions. In Section 5, we then derive specific tests for short memory and long memory correlation that are valid for any conditional distribution with the exponential family. The LMP short memory statistic, based on an $AR(1)$ process for the latent variable, is shown to be the first-order autocorrelation coefficient. The long-memory statistic, using a fractionally integrated process for the latent variable, is the statistic derived by Robinson (1994) and Tanaka (1999) under a Gaussian distributional assumption. Statistics for testing for short and long memory correlation in the variance of a process are also derived, again for any conditional distribution within the exponential family. When a conditional Gaussian distribution is adopted, the LMP statistics have the same structural form as the statistics for testing for correlation in the levels of the data, but now applied to the squares. In contrast, the adoption of a conditional gamma distribution, appropriate for data on the positive domain, leads to stochastic volatility test statistics based on a different transformation of the data. Section 6 demonstrates the application of the tests to several non-Gaussian financial time series, whilst Section 7 reports the results of Monte Carlo experiments used to assess the finite sample size and power properties of the tests. Some conclusions are provided in Section 8.

2 Induced Dependence

Let $\{y_t\}$ denote a sequence of arbitrary random variables. A continuous latent sequence $\{x_t\}$ is used to induce dependence in $\{y_t\}$, via a random parameter λ_t, and the distribution of y_t conditional on λ_t. In this section, we investigate the extent to which dependence in the latent continuous $\{x_t\}$ is manifested in the observed, possibly discrete, variable $\{y_t\}$.

Suppose there are T observations, $y_1, ..., y_T$ which are stacked into the $(T \times 1)$ vector $\mathbf{y} = [y_1, ..., y_T]'$. The vector $\mathbf{y}$ is assumed to depend on T random parameters $\lambda_1, ..., \lambda_T$, combined as the $(T \times 1)$ vector $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_T]'$. Each λ_t, $t = 1, 2, \ldots, T$, is, in turn, linked to an underlying scalar latent process x_t via the relations

$$\lambda_t = h(x_t), \quad t = 1, ..., T, \tag{1}$$

for a response function $h(.)$. The $(T \times 1)$ vector $\mathbf{x}$ is defined as $\mathbf{x} = [x_1, ..., x_T]'$.

Dependence in $\mathbf{y}$ is to be modelled indirectly via an assumed dynamic model for $\mathbf{x}$, with the dynamics in $\mathbf{x}$ transmitted to $\mathbf{y}$ through the response function. The response function is designed to ensure that $\mathbf{x}$ is mapped into the appropriate space for the random parameters $\boldsymbol{\lambda}$. We write the joint density/mass function of $(\mathbf{y}, \boldsymbol{\lambda})$ as

$$f(\mathbf{y}, \boldsymbol{\lambda}) = f(\mathbf{y}|\boldsymbol{\lambda})f(\boldsymbol{\lambda}),$$

with $f(\mathbf{y}|\boldsymbol{\lambda})$ being the conditional distribution of $\mathbf{y}$ given $\boldsymbol{\lambda}$ and $f(\boldsymbol{\lambda})$ the marginal distribution of $\boldsymbol{\lambda}$. The marginal distribution of $\mathbf{y}$ is therefore

$$\begin{aligned} f(\mathbf{y}) &= \int ... \int f(\mathbf{y}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \\ &= E\left[f(\mathbf{y}|\boldsymbol{\lambda})\right]. \end{aligned} \tag{2}$$

For notational convenience, the fact that $f(\mathbf{y})$ is a function of the fixed unknown parameters that characterize the marginal distribution of $\boldsymbol{\lambda}$, and, in some cases, the distribution of $\mathbf{y}$ conditional on $\boldsymbol{\lambda}$, is not made explicit. However, the marginal distribution in (2) clearly defines the likelihood function for that set of fixed parameters.

To investigate how the correlation in $\mathbf{x}$ is transmitted to $\mathbf{y}$ we begin by expressing the first two moments of $\mathbf{y}$ as a function of the moments of $\boldsymbol{\lambda}$. Let the first and second moments of the conditional distribution of $\mathbf{y}$ be denoted by $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}$ and $\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}$ respectively and those of the marginal of $\boldsymbol{\lambda}$ by $\boldsymbol{\mu}_{\lambda}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}$ respectively. The moments of the marginal distribution of $\mathbf{y}$ are

$$E[\mathbf{y}] = E\left[E_{\lambda}\left[\mathbf{y}|\boldsymbol{\lambda}\right]\right] = E\left[\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}\right], \tag{3}$$

$$\begin{aligned} E[\mathbf{y}\mathbf{y}'] &= E\left[E_{\lambda}\left[\mathbf{y}\mathbf{y}'|\boldsymbol{\lambda}\right]\right] = E\left[\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}} + \boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}\boldsymbol{\mu}'_{\mathbf{y}|\boldsymbol{\lambda}}\right] \\ &= E\left[\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}\right] + V\left[\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}\right] + E\left[\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}\right] E\left[\boldsymbol{\mu}'_{\mathbf{y}|\boldsymbol{\lambda}}\right], \end{aligned} \tag{4}$$

$$V[\mathbf{y}] = E\left[\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}\right] + V\left[\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}\right]. \tag{5}$$

Thus, we can see that the variation in $\mathbf{y}$ can be decomposed into individual components associated with the mean and variance of the conditional distribution of $\mathbf{y}|\boldsymbol{\lambda}$. Assume, for the sake of illustration, we parameterize such that $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}} = \boldsymbol{\lambda}$. In this case (3) and (5) become respectively

$$E[\mathbf{y}] = \boldsymbol{\mu}_{\lambda} \tag{6}$$

$$V[\mathbf{y}] = E\left[\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}\right] + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}. \tag{7}$$

That is, the unconditional mean of $\mathbf{y}$ is the same as that of $\boldsymbol{\lambda}$ while the unconditional variance-covariance matrix is the sum of the expectation of the variance-covariance matrix of the conditional distribution of $\mathbf{y}$ and the variance-covariance matrix of $\boldsymbol{\lambda}$. For example, when $\boldsymbol{\lambda}$ is stationary and when the elements of $\mathbf{y}$ are conditionally independent, as will be assumed at a later stage in the paper, the autocorrelation function (ACF) of y_t is

$$\begin{aligned} cor[y_t, y_{t-k}] &= \frac{cov[y_t, y_{t-k}]}{V[y_t]} \\ &= \frac{cov[\lambda_t, \lambda_{t-k}]}{E\left[V[y_t|\lambda_t]\right] + V[\lambda_t]} \\ &= cor[\lambda_t, \lambda_{t-k}].\frac{1}{E\left[V[y_t|\lambda_t]\right]/V[\lambda_t]] + 1}, \end{aligned} \tag{8}$$

for all lags $k > 0$. Hence, unambiguously, the ACF of y_t is less than that of the random parameter, λ_t. In addition, if the ratio $E\left[V[y_t|\lambda_t]\right]/V[\lambda_t]$ is large, the correlation in the observed process will tend to be small regardless of the dependence in the stochastic parameter.

Using (1), we may assess the effect of $\mathbf{x}$ on the moments of $\mathbf{y}$ by approximating, via a Taylor series expansion, the moments of $\boldsymbol{\lambda}$ as

$$E[\lambda_t] \approx h(x_t)|_{x_t=E[x_t]} + \frac{1}{2}\left.\frac{\partial^2 h}{\partial x_t^2}\right|_{x_t=E[x_t]} V[x_t] \tag{9}$$

$$cov[\lambda_t, \lambda_s] \approx \left.\frac{\partial h}{\partial x_t}\right|_{x_t=E[x_t]} \left.\frac{\partial h}{\partial x_s}\right|_{x_s=E[x_s]} cov[x_t, x_s], \tag{10}$$

for $t, s = 1, 2, \ldots, T$. Denoting the variance-covariance matrix of $\mathbf{x}$ by $\boldsymbol{\Sigma}_{\mathbf{x}}$, $\mathbf{h}_0$ as the $(T\times 1)$ vector with tth element $h(x_t)|_{x_t=E[x_t]}$, $\mathbf{h}_2$ as the $(T\times 1)$ vector with tth element $\left.\frac{\partial^2 h}{\partial x_t^2}\right|_{x_t=E[x_t]}$ and $\mathbf{H}_1$ as the $(T\times T)$ matrix with tsth element $\left.\frac{\partial h}{\partial x_t}\right|_{x_t=E[x_t]} \left.\frac{\partial h}{\partial x_s}\right|_{x_s=E[x_s]}$, (9) and (10) can be expressed more compactly as

$$\boldsymbol{\mu}_\lambda \approx \mathbf{h}_0 + \frac{1}{2}\mathbf{h}_2 \odot diag(\boldsymbol{\Sigma}_{\mathbf{x}}) \tag{11}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} \approx \mathbf{H}_1 \odot \boldsymbol{\Sigma}_{\mathbf{x}}, \tag{12}$$

with $\odot$ denoting the direct product and $diag(\boldsymbol{\Sigma}_{\mathbf{x}})$ a vector consisting of the diagonal elements of $\boldsymbol{\Sigma}_{\mathbf{x}}$. For example, if x_t is stationary, then $\boldsymbol{\Sigma}_{\mathbf{x}}$ is time invariant and all of the elements of $\mathbf{H}_1$ are the same (equal to h_1).[1] In this case, $\boldsymbol{\lambda}$ is also stationary and

[1] This same notational convention is used when $\mathbf{h}_0$ and $\mathbf{h}_2$ contain elements that are all equal to a particular constant.

(12) implies that

$$cor[\lambda_t, \lambda_{t-k}] \approx cor[x_t, x_{t-k}] \tag{13}$$

for all k. To this order of approximation then, the effect of the response function, $h(.)$, disappears from the correlation in the random parameters. However, in terms of the observations, substitution of (13) and (10) into (8) produces

$$cor[y_t, y_{t-k}] \approx cor[x_t, x_{t-k}].\frac{1}{\frac{E[V[y_t|\lambda_t]]}{h_1 V[x_t]} + 1}, \tag{14}$$

for all k. Thus the qualitative nature of the ACF is transmitted from x_t to y_t, but with $V[x_t]$, h_1 and $E\left[V[y_t|\lambda_t]\right]$ playing an important role in the magnitude of the correlation transmission. Thus, the presence of correlation in the latent variable alone is not sufficient to induce correlation in the observed data of equivalent magnitude. The variation in the latent variable, the response function and the variation in the data conditional on λ_t also affect the degree of correlation in the observed data. The qualitative nature of these results remain valid when $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}} \neq \boldsymbol{\lambda}$. In this case, the marginal variance-covariance matrix of $\mathbf{y}$ in (5) can still be expressed as an approximate function of the moments of $\mathbf{x}$ using expressions of the same form as (11) and (12). However, in this case, the quantities being approximated are respectively $E\left[\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}\right]$ and $V\left[\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}\right]$, in which case the derivatives in $\mathbf{h}_0, \mathbf{h}_2$ and $\mathbf{H}_1$ are defined in terms of the compound function $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}$, where $\boldsymbol{\lambda} = h(\mathbf{x})$, rather than with respect to just $\boldsymbol{\lambda} = h(\mathbf{x})$.

To investigate the transfer of higher order dependence to $\mathbf{y}$, consider $Cov[w(y_t),$ $g(y_s)]$ for arbitrary functions $w(.)$ and $g(.)$. If this covariance is zero for all functions $g(.)$ and $w(.)$, then y_t and y_s are independent. To illustrate, we assume conditional independence and that the distribution of $y_t|\lambda_t$ has mean θ (independent of λ_t) and variance $\sigma^2(\lambda_t)$. Note that in this setup the y_t variables themselves are uncorrelated. Express $Cov\left[w(y_t), g(y_s)\right]$ as

$$Cov\left[w(y_t), g(y_s)\right] = E\left[w(y_t) g(y_s)\right] - E\left[w(y_t)\right] E\left[g(y_s)\right],$$

where

$$E\left[w(y_t) g(y_s)\right] = E\left[E_\lambda\left(w(y_t)\right) E_\lambda\left(g(y_s)\right)\right]$$

and

$$E\left[w\left(y_t\right)\right] = E\left[E_\lambda\left[w\left(y_t\right)\right]\right]; \quad E\left[g\left(y_s\right)\right] = E\left[E_\lambda\left[g\left(y_s\right)\right]\right].$$

Now,

$$E_\lambda\left[w\left(y_t\right)\right] \approx w(y_t)|_{y_t=\theta} + \frac{1}{2}\left.\frac{\partial^2 w}{\partial y_t^2}\right|_{y_t=\theta} \sigma^2(\lambda_t), \tag{15}$$

with a similar expression holding for $E_\lambda\left[g\left(y_s\right)\right]$. Noting that the terms in $w(.)$ in (15) do not depend on λ_t it follows that

$$Cov\left[w\left(y_t\right), g(y_s)\right] \approx k.Cov\left[\sigma^2(\lambda_t), \sigma^2(\lambda_s)\right] \tag{16}$$

for some constant k depending only on the second derivative terms like those in (15). For example, by setting $w(y_t) = g(y_t) = y_t^2$, we see that correlation in the conditional covariance is transmitted to the squares of the observed variables y_t. In addition, to this order of approximation, if the conditional covariance sequence is uncorrelated the observed y_t and y_s are independent. A further Taylor series expansion allows the analysis to be conducted in terms of the x_t variables and an expression corresponding to (14) to be produced for the more general functions of y_t.

The result in (14) makes it clear that a short memory process in x_t maps into a short memory process in y_t, long memory into long memory, etc. As such, a test for a particular form of correlation in y_t needs to be based on the specification of the corresponding correlation structure for x_t. Similarly, (16) suggests that (stochastic) volatility in x_t is transmitted to functions, including the squares, of y_t. This principle guides the construction of all test statistics in Section 5. However, given that the latent process and attendant response function, $h(.)$, are unobserved, it is crucial that the proposed test statistics do not depend on any explicit specification for $h(.)$. None of the tests suggested in Section 5 depend on the response function. That said, in any controlled experiments that assess the power of the tests, the response function does have an impact in that it contributes, via (14) and (16), to the extent of the dependence transfer from x_t to y_t under the alternative hypothesis.

Finally, the above setup is in fact more general than it may appear, as can be seen by a judicious redefinition of the quantities involved. Specifically, let y_t denote a $(p \times 1)$ random variable on which there are T observations, $y_1, ..., y_T$. The observations on y_t are stacked into the $(Tp \times 1)$ vector $\mathbf{y} = \left[y_1', ..., y_T'\right]'$. The vector $\mathbf{y}$ is assumed to depend on N random parameters, each of dimension $(q\times1)$, $\lambda_1, ..., \lambda_N$, combined as the $(Nq \times 1)$ vector $\boldsymbol{\lambda} = [\lambda_1', ..., \lambda_N']'$. Each element of $\boldsymbol{\lambda}_i = (\lambda_{1i}, ..., \lambda_{qi})'$, $i = 1, 2, \ldots, N$, is, in turn, linked to an underlying scalar latent process $x_{\ell i}$ via the relations

$$\lambda_{\ell i} = h_\ell(x_{\ell i}), \qquad \ell = 1, ..., q;\ i = 1, ..., N,$$

for response functions $h_\ell(.)$. The $(Nq \times 1)$ vector $\mathbf{x}$ is defined as $\mathbf{x} = [x_1', ..., x_N']'$, where x_i, $i = 1, 2, \ldots, N$, is the $(q \times 1)$ vector with ℓth element $x_{\ell i}$. As before

$$f(\mathbf{y}) = E\left[f(\mathbf{y}|\boldsymbol{\lambda})\right]$$

and the expressions (3) to (5) remain valid. Also, when $\boldsymbol{\mu}_{\mathbf{y}|\mathbf{x}} = \boldsymbol{\lambda}$, (9) and (10) can be replaced by

$$\begin{aligned} E\left[\lambda_{\ell i}\right] &\approx h_\ell(x_{\ell i})|_{x_{\ell i}=E[x_{\ell i}]} + \frac{1}{2}\left.\frac{\partial^2 h_\ell}{\partial x_{\ell i}^2}\right|_{x_{\ell i}=E[x_{\ell i}]} V[x_{\ell i}] \\ cov[\lambda_{\ell i}, \lambda_{mj}] &\approx \left.\frac{\partial h_\ell}{\partial x_{\ell i}}\right|_{x_{\ell i}=E[x_{\ell i}]} \left.\frac{\partial h_m}{\partial x_{mj}}\right|_{x_{mj}=E[x_{mj}]} cov[x_{\ell i}, x_{mj}], \end{aligned}$$

or by the corresponding approximations when $\boldsymbol{\mu}_{\mathbf{y}|\mathbf{x}} \neq \boldsymbol{\lambda}$. Thus the results above regarding dependence transfer, as well as the analysis to follow, apply in the more general case.

3 An Approximate Likelihood

In this section we suggest an approximation to the expectation in (2) that defines the marginal distribution of the data or, alternatively, the likelihood function. Following Cox (1983) and McCabe and Leybourne (2000), we take a Taylor Series expansion of

$f(\mathbf{y}|\boldsymbol{\lambda})$ about $\boldsymbol{\mu}_\lambda$, thereby producing,

$$\begin{aligned}
f(\mathbf{y}) &= E\left[f(\mathbf{y}|\boldsymbol{\lambda})\right] \\
&= E\left[f(\mathbf{y}|\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} + (\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda)' \left.\frac{\partial f(\mathbf{y}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} \right. \\
&\quad \left. +\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda)' \left.\frac{\partial^2 f(\mathbf{y}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} (\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda) + O\left\|\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda\right\|^3 \right] \\
&= f(\mathbf{y}|\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} + \frac{1}{2} E\left[(\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda)' \left.\frac{\partial^2 f(\mathbf{y}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} (\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda) \right] \\
&\quad + O\left\{E\left[\left\|\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda\right\|^3\right]\right\} \\
&= f(\mathbf{y}|\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} + \frac{1}{2} tr\left[\left.\frac{\partial^2 f(\mathbf{y}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} \boldsymbol{\Sigma}_{\boldsymbol{\lambda}} \right] \\
&\quad + O\left\{E\left[\left\|\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda\right\|^3\right]\right\}. \qquad (17)
\end{aligned}$$

Considering only the first two terms on the right-hand-side of the last line in (17), an approximation to the likelihood function, denoted by $f^*(\mathbf{y})$, is given by

$$f^*(\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} + \frac{1}{2} tr\left[\left.\frac{\partial^2 f(\mathbf{y}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} \boldsymbol{\Sigma}_{\boldsymbol{\lambda}} \right]. \qquad (18)$$

Since the approximation depends only on the first two moments of $\boldsymbol{\lambda}$, no additional distributional assumptions regarding $\boldsymbol{\lambda}$ are required. Defining

$$\ell(\boldsymbol{\lambda}|\mathbf{y}) = \log L(\boldsymbol{\lambda}|\mathbf{y}) = \log\left(f(\mathbf{y}|\boldsymbol{\lambda})\right)$$

and using

$$\frac{\partial^2 L}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'} = \left(\frac{\partial \ell}{\partial \boldsymbol{\lambda}}\frac{\partial \ell}{\partial \boldsymbol{\lambda}'} + \frac{\partial^2 \ell}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right) L(\boldsymbol{\lambda}|\mathbf{y}),$$

(18) becomes

$$\begin{aligned}
f^*(\mathbf{y}) &= L(\boldsymbol{\lambda}|\mathbf{y})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} \left[1 + \frac{1}{2} tr\left\{ \left.\left(\frac{\partial \ell}{\partial \boldsymbol{\lambda}}\frac{\partial \ell}{\partial \boldsymbol{\lambda}'} + \frac{\partial^2 \ell}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right)\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda} \boldsymbol{\Sigma}_{\boldsymbol{\lambda}} \right\}\right] \\
&= L^*(\boldsymbol{\mu}_\lambda|\mathbf{y}) \left[1 + \frac{1}{2} tr\left(\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\right)\right], \qquad (19)
\end{aligned}$$

where $L^*(\boldsymbol{\mu}_\lambda|\mathbf{y}) = L(\boldsymbol{\lambda}|\mathbf{y})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda}$ and

$$\mathbf{M} = \left.\left(\frac{\partial \ell}{\partial \boldsymbol{\lambda}}\frac{\partial \ell}{\partial \boldsymbol{\lambda}'} + \frac{\partial^2 \ell}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right)\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda}. \qquad (20)$$

In this paper we use the approximation in (19) as a vehicle for deriving tests of dependence. However, it is also possible to consider (19) directly as a starting point for estimation by treating it as a pseudo-likelihood. For example, we could take derivatives with respect to the unknown parameters and equate them to zero to provide a set of restrictions that could be treated as estimating equations for the unknown fixed parameters.[2] We could also employ the approximation in a Bayesian latent variable treatment to reduce the computational burden of dealing with the exact likelihood. In situations in which (19) is used as a basis for estimation, issues to do with the quality of the approximation throughout the entire parameter space, including the extent to which standard likelihood conditions hold for the pseudo-likelihood, would need to be addressed (see Heyde, 1997).

Here, the conditional distribution of $\mathbf{y}|\boldsymbol{\lambda}$ is deemed to play a significant role in modelling the marginal distribution of $\mathbf{y}$ so that, for example, when data are counts we specify that $f(\mathbf{y}|\boldsymbol{\lambda})$ is Poisson. The role of the variable $\boldsymbol{\lambda}$ is to introduce dependence and possibly some limited overdispersion in $\mathbf{y}$. As such, $\boldsymbol{\lambda}$ is simply an artifact and may, for all practical purposes, be considered Gaussian with a small variance, in which case its higher order moments may safely be ignored. The approximation in (19), based on only the first two moments of $\boldsymbol{\lambda}$ and evaluated at the mean of $\boldsymbol{\lambda}$, is thus expected to be reasonable under such a scenario. However, by (14) and (16) the transfer of dependence from $\mathbf{x}$, via $\boldsymbol{\lambda}$, to $\mathbf{y}$ may possibly be small when there is little variation in $\boldsymbol{\lambda}$. Accordingly, we seek to construct tests that have maximum power in the region close to the null hypothesis of independence. This is accomplished in the next section by using a parameter π to control the degree of dependence and by constructing tests that are local in π.

[2] In related independent work by Huber, Ronchetti and Victoria-Fese (2003), the Laplace approximation of Tierney and Kadane (1986) is applied to produce an approximate likelihood function on which estimation of the parameters of the latent variable model is based. See also Davis and Rodriguez-Yam (2003).

4 Tests for Dependence

4.1 The General Form of the LMP Test

In this section a test procedure is developed using the density $f^*(\mathbf{y})$ in (19) to approximate that of $f(\mathbf{y})$ in (2). The procedure produces local tests of a scalar parameter π, where we assume that

$$\boldsymbol{\mu}_\lambda = \boldsymbol{\mu}_\lambda(\pi) \text{ and } \boldsymbol{\Sigma}_{\boldsymbol{\lambda}} = \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}(\pi).$$

That is, both the mean and variance-covariance matrix of the random parameter vector $\boldsymbol{\lambda}$ may depend on a single hyper-parameter π, with power being maximized local to the null hypothesis of

$$H_0 : \pi = 0. \tag{21}$$

The statistic for the LMP test against the one-sided alternative hypothesis,

$$H_1 : \pi > 0, \tag{22}$$

is given by

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

(See, for example, Casella and Berger, 1990, Section 8.3.4). Following King and Wu (1997) we can entertain multiple parameters π_i, $i = 1, 2, \ldots, s$, and construct the one-sided locally mean most powerful test by simply adding up the test statistics for the individual parameters. Thus for notational simplicity we continue with a single π. Taking the logarithm of (19) it follows that

$$S = \left.\frac{\partial\, \ell(\boldsymbol{\lambda}|\mathbf{y})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}}}{\partial \pi}\right|_{\pi=0} + \frac{tr\left[\left.\frac{\partial(\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}})}{\partial \pi}\right|_{\pi=0}\right]}{\left[2 + tr\left((\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}})|_{\pi=0}\right)\right]} \tag{23}$$

where

$$\frac{\partial\,(\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}})}{\partial \pi} = \mathbf{M}\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}}{\partial \pi} + \frac{\partial \mathbf{M}}{\partial \pi}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}.$$

Using the approximation for the moments of $\boldsymbol{\lambda}$ in terms of the latent vector $\mathbf{x}$ in (11) and (12), we can then derive a form of the test statistic that is a function of $\mathbf{x}$, which is the process with respect to which π is explicitly defined.

4.2 Maintained Assumptions

In this section we list the assumptions that we maintain throughout the rest of the paper, despite the fact that they are not necessary for the general theory of Section 4.1 above.

(a) We assume that only the marginal density of the latent variable, $f(\boldsymbol{\lambda})$, is a function of the parameter under test, π. Thus, the only way that π enters $f(\mathbf{y})$ is through $f(\boldsymbol{\lambda})$. The implication of this assumption is that $f(\mathbf{y}|\boldsymbol{\lambda})$ only depends on π through the influence of $\boldsymbol{\lambda}$, so that quantities like those in $\mathbf{M}$ depend on π only through $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$. This allows for several possible simplifications in (23). First, when $\frac{\partial \boldsymbol{\mu}_\lambda}{\partial \pi}|_{\pi=0} = \mathbf{0}$ it follows that

$$\left.\frac{\partial\, \ell(\boldsymbol{\lambda}|\mathbf{y})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}}}{\partial \pi}\right|_{\pi=0} = \left.\left(\frac{\partial\, \ell(\boldsymbol{\lambda}|\mathbf{y})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}}}{\partial \boldsymbol{\mu}'_\lambda}\frac{\partial \boldsymbol{\mu}_\lambda}{\partial \pi}\right)\right|_{\pi=0} = 0,$$

so only the second term of (23) remains. In addition, $\left.\frac{\partial m_{i,j}}{\partial \pi}\right|_{\pi=0} = \left.\frac{\partial m^*_{i,j}}{\partial \boldsymbol{\mu}'_\lambda}\frac{\partial \boldsymbol{\mu}_\lambda}{\partial \pi}\right|_{\pi=0} = 0$, for each element $m_{i,j}$ of the matrix $\mathbf{M}$ defined in (20), and thus (23) simplifies to

$$S = \frac{tr\left[\left(\mathbf{M}\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}}{\partial \pi}\right)\big|_{\pi=0}\right]}{\left[2 + tr\left((\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}})|_{\pi=0}\right)\right]}. \tag{24}$$

Secondly, when $\boldsymbol{\lambda}$ is nonstochastic under the null and $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}|_{\pi=0} = \mathbf{0}$ as a consequence, then (23) simplifies to

$$S = \left.\left(\frac{\partial\, \ell(\boldsymbol{\lambda}|\mathbf{y})|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}}}{\partial \boldsymbol{\mu}'_\lambda}\frac{\partial \boldsymbol{\mu}_\lambda}{\partial \pi}\right)\right|_{\pi=0} + tr\left[\left.\left(\mathbf{M}\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}}{\partial \pi}\right)\right|_{\pi=0}\right]. \tag{25}$$

Obviously, if both simplifications occur only the second term of (25) remains.

(b) We maintain that the null hypothesis, $H_0 : \pi = 0$, induces independence in $\boldsymbol{\lambda}$, whilst the alternative hypothesis, $H_1 : \pi > 0$, is associated with correlation of some sort in $\mathbf{x}$ and thus $\boldsymbol{\lambda}$. Hence under H_0, $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}|_{\pi=0} = \sigma^2_{\boldsymbol{\lambda}}\mathbf{I}$ for some $\sigma^2_{\boldsymbol{\lambda}} \geq 0$. We use the following additional notation for the trace terms that enter all versions of the test statistic, namely

$$tr\left[\left.\left(\mathbf{M}\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}}{\partial \pi}\right)\right|_{\pi=0}\right] = \mathbf{q}'\mathbf{D}_0\mathbf{q} + \mathrm{tr}\left(\mathbf{R}\mathbf{D}_0\right) \tag{26}$$

and

$$\begin{aligned} tr\left((\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}})|_{\pi=0}\right) &= \sigma^2_{\boldsymbol{\lambda}} tr(\mathbf{M}|_{\pi=0}) \\ &= \sigma^2_{\boldsymbol{\lambda}}\left[\mathbf{q}'\mathbf{q} + \mathrm{tr}\left(\mathbf{R}\right)\right], \end{aligned} \tag{27}$$

where

$$\mathbf{q} = \left.\frac{\partial \ell}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda, \pi=0}, \tag{28}$$

$$\mathbf{R} = \left.\frac{\partial^2 \ell}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'}\right|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_\lambda, \pi=0} \tag{29}$$

and

$$\mathbf{D}_0 = \left.\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}}{\partial \pi}\right|_{\pi=0}. \tag{30}$$

(c) We assume that $\mathbf{y}|\boldsymbol{\lambda}$ is conditionally independently and identically distributed. That is, for all $\pi \geq 0$, we assume that

$$f(\mathbf{y}|\boldsymbol{\lambda}) = f(y_1|\lambda_1)f(y_2|\lambda_2)\dots f(y_T|\lambda_T), \tag{31}$$

so that dependence in $\mathbf{y}$ is generated solely through $\boldsymbol{\lambda}$ (from the latent process $\mathbf{x}$), with the transfer of that dependence taking place via (14) or via (16). When the elements of $\boldsymbol{\lambda}$ are independent under H_0, (31) implies that the elements of $\mathbf{y}$ are also independent, that is,

$$\begin{aligned} f(\mathbf{y}) &= \int \dots \int f(\mathbf{y}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \\ &= \int \dots \int f(y_1|\lambda_1)f(y_2|\lambda_2)\dots f(y_T|\lambda_T)f(\lambda_1)f(\lambda_2)\dots f(\lambda_T)d\boldsymbol{\lambda} \\ &= \int f(y_1|\lambda_1)f(\lambda_1)d\lambda_1 \int f(y_2|\lambda_2)f(\lambda_2)d\lambda_2 \dots \int f(y_T|\lambda_T)f(\lambda_T)d\lambda_T \\ &= f(y_1)f(y_2)\dots f(y_T). \end{aligned} \tag{32}$$

Note without the independence of $\boldsymbol{\lambda}$, $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}|_{\pi=0} = \sigma^2_{\boldsymbol{\lambda}}\mathbf{I}$, $\sigma^2_{\boldsymbol{\lambda}} > 0$, is insufficient to ensure that $\mathbf{y}$ is independent, under the null, except when $\boldsymbol{\lambda}$ is Gaussian. However $\sigma^2_{\boldsymbol{\lambda}} = 0$, associated with $\boldsymbol{\lambda}$ being fixed, does imply independence in $\mathbf{y}$. Although not required for the general derivation of LMP tests based on the approximate likelihood function, the assumption in (31) has the advantage of clarifying the testing problem. Invoking this assumption, the null hypothesis is associated with independent data, and the alternative hypothesis with data that is dependent only through the influence of the correlated latent variable x_t, with correlation in x_t characterized by a positive value of π. As such, testing $H_0 : \pi = 0$ against $H_1 : \pi > 0$ is equivalent to testing for independent data against the alternative of dependent data. Moreover, the lack of any dependence structure in $f(\mathbf{y}|\boldsymbol{\lambda})$ is consistent with one of the motivations for the latent variable approach to dynamic modelling adopted in the paper, namely that when the data under consideration has a restricted sample space, the direct modelling of dependence via structural dynamic models can be difficult.

4.3 Comparison with an Exact Test

It is of interest to compare the test statistic based on the approximation with a corresponding statistic based on the exact likelihood function. Naturally, this may only be accomplished in special cases, since a closed-form solution for $f(\mathbf{y})$ in (2) is likely to be available for very few combinations of distributions for $\mathbf{y}|\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}$. The leading case in which an exact result is attainable is when it is assumed that $\mathbf{y}|\boldsymbol{\lambda} \backsim MN\left(\boldsymbol{\lambda}, \boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}\right)$, with $\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}$ not a function of $\boldsymbol{\lambda}$, and $\boldsymbol{\lambda} \backsim MN\left(\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\right)$, where MN denotes the multivariate normal distribution.[3] In this case standard algebra associated with the MN distribution produces

$$\mathbf{y} \backsim MN\left(\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\right).$$

Invoking Assumption (c), and assuming that $\mathbf{y}$ is conditionally homoscedastic, it follows that $\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}} = \sigma^2\mathbf{I}$, for some $\sigma^2 > 0$. Further assuming that $\boldsymbol{\lambda}$ is homoscedastic, it follows that $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}|_{\pi=0} = \sigma^2_{\boldsymbol{\lambda}}\mathbf{I}$, for some $\sigma^2_{\boldsymbol{\lambda}} > 0$. Hence, under the null, $\mathbf{y} \sim MN\left(\boldsymbol{\mu}_{\boldsymbol{\lambda}}|_{\pi=0}, \sigma_0^2\mathbf{I}\right)$, where $\sigma_0^2 = \sigma^2 + \sigma^2_{\boldsymbol{\lambda}}$. Defining

$$\mathbf{V} = \sigma^2\mathbf{I} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}},$$

it follows that

$$\begin{aligned}
S &= \left.\frac{\partial \log f(\mathbf{y})}{\partial \pi}\right|_{\pi=0} \\
&= \left.\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi'}\mathbf{V}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_{\lambda}\right)\right|_{\pi=0} - \left.\frac{1}{2}\left(\mathbf{y} - \boldsymbol{\mu}_{\lambda}\right)'\frac{\partial \mathbf{V}^{-1}}{\partial \pi}\left(\mathbf{y} - \boldsymbol{\mu}_{\lambda}\right)\right|_{\pi=0} \\
&= \left.\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi'}\mathbf{V}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_{\lambda}\right)\right|_{\pi=0} + \left.\frac{1}{2}\left(\mathbf{y} - \boldsymbol{\mu}_{\lambda}\right)'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \pi}\mathbf{V}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_{\lambda}\right)\right|_{\pi=0} \\
&= \mathbf{z}'\left.\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi}\right|_{\pi=0} + \frac{1}{2\sigma_0}\mathbf{z}'\mathbf{D}_0\mathbf{z}
\end{aligned}$$

[3] Note we can set $\boldsymbol{\lambda} = \mathbf{x}$ here without loss of generality.

where $\mathbf{z} = \left(\mathbf{y} - \left.\boldsymbol{\mu}_{\boldsymbol{\lambda}}\right|_{\boldsymbol{\pi}=\mathbf{0}}\right)/\sigma_0$ and $\mathbf{D}_0$ is as defined in (30).

In practice we need to estimate $\mathbf{z}$. Accordingly, define

$$\hat{\mathbf{z}} = (\mathbf{y} - \overline{y})/s, \tag{33}$$

with $\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ and $s = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \overline{y})^2}{T}}$. We then base the test on S evaluated using $\hat{\mathbf{z}}$. In Gaussian applications, we are typically interested in tests of covariance of the distribution and $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$ does not depend on π. In this case, or more generally when $\left.\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi}\right|_{\pi=0} = \mathbf{0}$, we obtain

$$S = \hat{\mathbf{z}}'\mathbf{D}_0\hat{\mathbf{z}}, \tag{34}$$

which is the LBI procedure, under location and scale invariance, for tests of the covariance matrix of the Gaussian distribution; see King and Hillier (1988).

Similar arguments allow us to deal with the multivariate lognormal-normal mixture, that is where $\log(\mathbf{y})|\boldsymbol{\lambda} \backsim MN\left(\boldsymbol{\lambda}, \boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}\right)$ and $\boldsymbol{\lambda} \backsim MN\left(\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\right)$; in which case it follows that $\log(\mathbf{y}) \backsim MN\left(\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \sigma^2\mathbf{I} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\right)$. The test statistic has the same form as that in (34), but based on $\log(\mathbf{y})$ rather than $\mathbf{y}$. It is possible to derive the marginal distribution of $\mathbf{y}$ in some other special cases, for example the Poisson-Gamma mixture, but additional independence assumptions are required.

We wish to compare the exact test in (34) with the approximate test based on the same assumptions as invoked above for the conditional distribution, namely $\mathbf{y}|\boldsymbol{\lambda} \backsim MN\left(\boldsymbol{\lambda}, \boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}\right)$, but with the distribution of $\boldsymbol{\lambda}$ left unspecified. Again focussing on the case where $\left.\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi}\right|_{\pi=0} = \mathbf{0}$, the version of the approximate statistic in (24) is appropriate. In this case, it follows that

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\lambda}} &= \sigma^{-2}(\mathbf{y} - \boldsymbol{\lambda}) \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'} &= -\sigma^{-2}\mathbf{I}, \end{aligned}$$

and that

$$S = \frac{\sigma^{-2}\left[\mathbf{z}'\mathbf{D}_0\mathbf{z} - \operatorname{tr}(\mathbf{D}_0)\right]}{2 + \sigma^2_{\boldsymbol{\lambda}}\left[\sigma_0\mathbf{z}'\mathbf{z} - \sigma^2 T\right]},$$

where $\mathbf{z} = \left(\mathbf{y} - \left.\boldsymbol{\mu}_{\boldsymbol{\lambda}}\right|_{\boldsymbol{\pi}=\mathbf{0}}\right)/\sigma_0$. Since $\hat{\mathbf{z}}'\hat{\mathbf{z}} = T$, replacing $\mathbf{z}$ by $\hat{\mathbf{z}}$ in (33) means that the test statistic reduces (by deleting multiplicative and additive constants) to

$$S = \hat{\mathbf{z}}'\mathbf{D}_0\hat{\mathbf{z}}, \tag{35}$$

which is equivalent to (34). Thus, in this particular case the test suggested by the approximation is equivalent to the LBI test based on the exact Gaussian likelihood.

5 Testing for Dependence; The Exponential Family

5.1 The Natural Exponential Family

Of course, while S in (23) may form the basis of a test procedure it cannot be used as a statistic until all nuisance parameters have been eliminated and critical values found. To this end we derive the form of S based on more specific assumptions about the conditional distribution of $\mathbf{y}$, namely that it falls within the exponential family of distributions. The exponential family of distributions allows for a unified treatment of random variables of different types such as the exponential (positive), the Poisson (discrete) and even the Gaussian random variable itself. We shall specify that the conditional distribution $f(\mathbf{y}|\boldsymbol{\lambda})$ is a member of the exponential family and, hence, that $f(\mathbf{y})$ has a distribution in the larger class that mixes the exponential family over the marginal distribution of $\boldsymbol{\lambda}$. This class contains the multivariate Gaussian distribution as shown in Section 4.3 when $f(\mathbf{y}|\boldsymbol{\lambda})$ and $f(\boldsymbol{\lambda})$ are specified as normal, with $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}} = \boldsymbol{\lambda}$ and $\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}$ functionally independent of $\boldsymbol{\lambda}$. However, in general $f(\boldsymbol{\lambda})$ will be non-Gaussian. Hence, the statistics derived in this section are applicable to a very broad range of data types.

For specified functions $a(.)$, $b(.)$ and $c(.)$, the density of the multivariate natural exponential family is given by (see Fahrmeir and Tutz, 1994)

$$f\left(\mathbf{y}|\boldsymbol{\theta},\gamma\right) = c\left(\mathbf{y},a(\gamma)\right)\exp\left\{\frac{\mathbf{y}'\boldsymbol{\theta} - b\left(\boldsymbol{\theta}\right)}{a(\gamma)}\right\}, \tag{36}$$

where $\boldsymbol{\theta}$ is a vector parameter and γ is a positive scalar dispersion parameter. Under the assumption of conditional independence, this exponential family includes, among others, the multivariate Poisson, Bernoulli, exponential, uniform and Gaussian distributions. In applications where we wish to test for correlation in the $\mathbf{y}$, we set $a(\gamma) = 1$ and $\boldsymbol{\theta}$ corresponds to $\boldsymbol{\lambda}$ of the previous sections. In the volatility applications γ plays the role of $\boldsymbol{\lambda}$.

In the next sub-section we derive tests of the null of independence against short memory and long memory correlation in $\mathbf{y}$ and in the final sub-section, tests for stochastic volatility, that is, tests that the variation in $\mathbf{y}$ is serially correlated.

5.2 Testing for Memory

In this section we use the simplified form of the exponential family distribution to test for correlation. That is $\boldsymbol{\theta} = \boldsymbol{\lambda}$ and $a(\gamma) = 1$, in which case (36) becomes

$$f(\mathbf{y}|\boldsymbol{\lambda}) = c(\mathbf{y}) \exp\{\mathbf{y}'\boldsymbol{\lambda} - b(\boldsymbol{\lambda})\}. \tag{37}$$

From (37) it follows that

$$\frac{\partial \ell}{\partial \boldsymbol{\lambda}} = \mathbf{y} - \frac{\partial b}{\partial \boldsymbol{\lambda}} = \mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}} \tag{38}$$

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} = -\frac{\partial^2 b}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} = -\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}. \tag{39}$$

Note that, apart from the moments involved, the expressions in (38) and (39) are both independent of any particular distributional form. It follows that $\mathbf{M}$ in (20) and the first term in (23) are also independent of distributional forms. Thus the statistic (23) is the same for all members of the exponential family in this setting.

The process $\mathbf{x}$ is used to generate correlation which is transferred to $\boldsymbol{\lambda}$ via a suitable response function $h(.)$. Assumptions *(a), (b)* and *(c)* of Section 4.2 are presumed to hold. The mean of the conditional exponential family is $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}$ and $\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}} = \sigma^2 \mathbf{I}$, with σ^2 independent of $\boldsymbol{\lambda}$. That is, we parameterize the conditional distribution in such a way that only the conditional mean is a function of the stochastic parameter vector $\boldsymbol{\lambda}$. Thus, the covariances of $\mathbf{y}$ are determined only by the second component of (5). Define

$$\mathbf{z} = \left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}} \Big|_{\boldsymbol{\lambda} = \boldsymbol{\mu}_{\boldsymbol{\lambda}}, \pi = 0} \right) / \sigma_0,$$

where $\sigma_0^2 \mathbf{I}$ is the variance-covariance matrix of $\mathbf{y}$ under the null. The statistic (23) has three components:

$$\begin{aligned} \left. \frac{\partial\, \ell(\boldsymbol{\lambda}|\mathbf{y})|_{\boldsymbol{\lambda} = \boldsymbol{\mu}_{\boldsymbol{\lambda}}}}{\partial \pi} \right|_{\pi=0} &= \mathbf{y}' \left. \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi} \right|_{\pi=0} - \left. \frac{\partial b}{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}'} \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi} \right|_{\pi=0} \\ &= \sigma_0 \, \mathbf{z}' \left. \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi} \right|_{\pi=0}, \end{aligned}$$

$$tr\left(\left(\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\right)\big|_{\pi=0}\right) = \sigma_{\lambda}^{2}\left[\sigma_0^2\mathbf{z}'\mathbf{z} - \sigma^2 T\right]$$

and

$$tr\left[\left.\frac{\partial\left(\mathbf{M}\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\right)}{\partial\pi}\right|_{\pi=0}\right] = tr\left[\mathbf{M}\left.\frac{\partial\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}}{\partial\pi}\right|_{\pi=0}\right] + \sigma_{\boldsymbol{\lambda}}^2 tr\left[\left.\frac{\partial\mathbf{M}}{\partial\pi}\right|_{\pi=0}\right].$$

In addition,

$$\sigma_{\boldsymbol{\lambda}}^2 tr\left[\left.\frac{\partial\mathbf{M}}{\partial\pi}\right|_{\pi=0}\right] = -2\sigma_{\boldsymbol{\lambda}}^2\sigma_0^2\,\mathbf{z}'\left.\frac{\partial\,\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}\big|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}}}{\partial\pi}\right|_{\pi=0}.$$

In many cases, for example when $\boldsymbol{\lambda}$ is weakly stationary, $\left.\frac{\partial\,\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}|_{\boldsymbol{\lambda}=\boldsymbol{\mu}_{\boldsymbol{\lambda}}}}{\partial\pi}\right|_{\pi=0}$ and $\left.\frac{\partial\boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial\pi}\right|_{\pi=0}$ are constant vectors, in which case the statistic (23) reduces to $tr\left[\mathbf{M}\frac{\partial\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}}{\partial\pi}\big|_{\pi=0}\right]$ (ignoring constants). Using (26), (38) and (39) and replacing $\mathbf{z}$ by $\hat{\mathbf{z}}$ in (33), the final form of the statistic is thus

$$S = \hat{\mathbf{z}}'\mathbf{D}_0\hat{\mathbf{z}}. \tag{40}$$

Hence, as suggested earlier we obtain the same statistic for all members of the exponential family, with (40) in turn equivalent to the LBI statistic in (34) derived under unconditional Gaussianity. Thus, the optimal procedures for testing for correlation are based on the usual sample ACF with no further cognizance needed to be taken of the underlying nature of the data. In particular, the procedures are just as applicable to data with a restricted sample space, such as discrete or positive data, as to data that is defined on the whole real line.

5.2.1 Testing for Short Memory Correlation

In this section we derive the form of the test statistic when x_t is modelled as a Markov process, that is as a stationary autoregressive model of order one ($AR(1)$),

$$(x_t - a) = \rho\left(x_{t-1} - a\right) + \eta_t, \tag{41}$$

with $|\rho| < 1$ and

$$\eta_t \sim iid(0,1), \tag{42}$$

for $t = 1, 2, \ldots, T$.[4] Given (14), the model in (41) is an appropriate starting point for the construction of a test for short memory correlation in y_t. With reference to the

[4] As the final version of the statistic is invariant to the variance of η_t in (42), we set the variance equal to 1.

parameter π defined in Section 4.1, we test the null hypothesis

$$H_0 : \pi = \rho = 0 \tag{43}$$

against the alternative hypothesis

$$H_1 : \pi = \rho > 0. \tag{44}$$

Note because of (42) and (31) the y_t are independent under the null. Under the alternative the y_t are correlated by (14). Given the $AR(1)$ model in (41), $E[x_t] = a$ for all t and the covariance matrix of $\mathbf{x}$ is

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & & \rho^{T-1} \\ \rho & 1 & \rho & & \rho^{T-2} \\ \rho^2 & \rho & 1 & & \rho^{T-3} \\ & & & & \\ & & & & \rho \\ \rho^{T-1} & & & \rho & 1 \end{bmatrix}. \tag{45}$$

Using (11), (12) and (45) we obtain

$$\boldsymbol{\mu}_{\boldsymbol{\lambda}} \approx \left[h_0 + 1/2h_2(1-\rho^2)^{-1}\right]\mathbf{i} \tag{46}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} \approx h_1 \boldsymbol{\Sigma}_{\mathbf{x}}, \tag{47}$$

where $\mathbf{i}$ denotes a column of 1's and h_0, h_1 and h_2 are suitable response functions constant in a. From (47) it follows that $\mathbf{D}_0 \approx h_1\mathbf{A}$, with $\mathbf{D}_0$ as defined in (30), where

$$\mathbf{A} = \left.\frac{\partial \boldsymbol{\Sigma}_{\mathbf{x}}}{\partial \rho}\right|_{\rho=0} = \begin{bmatrix} 0 & 1 & 0 & & & 0 \\ 1 & 0 & 1 & & & \\ 0 & 1 & 0 & 1 & & \\ & & 1 & & & \\ & & & & & 1 \\ 0 & & & & 1 & 0 \end{bmatrix}.$$

Thus, we obtain

$$S = \hat{\mathbf{z}}'\mathbf{A}\hat{\mathbf{z}}. \tag{48}$$

With scaling, the final form of the statistic becomes the normalized ordinary first-order correlation coefficient,

$$S = \frac{T^{1/2}\sum_{t=1}^{T-1}(y_t - \overline{y})(y_{t+1} - \overline{y})}{\sum_{t=1}^{T-1}(y_t - \overline{y})^2}, \tag{49}$$

which has previously been shown to be a locally optimal test for $AR(1)$ correlation in the Gaussian model; see, for example, Anderson (1971, Chp. 6). Under mild regularity (see, for example, Fuller, 1996, Corollary 6.3.6.2.) S is asymptotically distributed as $N(0,1)$ under the null of independence. Crucially, the specific form of the test statistic in (49) is independent of the response function $h(.)$.

Thus, notwithstanding the possible non-Gaussian nature of the data, use of the common first-order correlation coefficient is seen to be an optimal procedure for testing for $AR(1)$ correlation in the model (2) in the context of conditional distributions within the exponential family defined by (36). Further, as is clear from a comparison of (48) and (35), and invoking the arguments presented in Section 4.3, the approximate LPM test derived here corresponds to the exact LBI test based on a Gaussian likelihood.

5.2.2 Testing for Long Memory Correlation

We adopt the fractional white noise model as a representation of long memory, defined to be the case where the covariances of a stationary process are not absolutely summable. Thus,

$$(1-L)^d(x_t - a) = \varepsilon_t, \tag{50}$$

where

$$\varepsilon_t \backsim iid(0,1). \tag{51}$$

The difference operator in (50) is defined as

$$(1-L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + ..., \tag{52}$$

where L denotes the lag operator and the expansion in (52) is valid for $d > -1$.

We wish to test the null hypothesis

$$H_0 : \pi = d = 0 \tag{53}$$

against the alternative hypothesis

$$H_1 : \pi = d > 0 \tag{54}$$

Since the test is local in d, terms that are $O(d^2)$ may be ignored without loss of generality. Thus, expressing (50) as

$$(x_t - a) = (1 - L)^{-d}\varepsilon_t$$

and collecting terms in d we obtain

$$(x_t - a) = \left\{1 + dL + \frac{d}{2}L^2 + \frac{d}{3}L^3 + ... + O(d^2)\right\}\varepsilon_t. \tag{55}$$

Hence $E[x_t] = a$ and, given (55),

$$E\left[(x_t - a)(x_{t-j} - a)\right] = \left.\begin{array}{cc} 1 & j = 0 \\ \frac{d}{j} & j \geq 1 \end{array}\right\} + O(d^2). \tag{56}$$

The y_t are independent under the null by (51) and (31), and correlated under the alternative by (14). Clearly,

$$\left.\frac{\partial E\left[(x_t - a)(x_{t-j} - a)\right]}{\partial d}\right|_{d=0} = \left.\begin{array}{cc} 0 & j = 0 \\ \frac{1}{j} & j \geq 1 \end{array}\right\} \tag{57}$$

and $\mathbf{D}_0 \approx h_1\mathbf{B}$ where the $(T \times T)$ matrix $\mathbf{B}$ contains the derivative functions in (57). Hence,

$$S = \hat{\mathbf{z}}'\mathbf{B}\hat{\mathbf{z}} \tag{58}$$

which, with appropriate scaling, may be written as

$$S = T^{1/2}\sum_{j=1}^{T-1}\frac{1}{j}\hat{\rho}_j, \tag{59}$$

where $\hat{\rho}_j = \hat{\gamma}_j/\hat{\gamma}_0$ with $\hat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T}(y_t - \overline{y})(y_{t-j} - \overline{y})$. Under the mild regularity conditions specified in Tanaka (1999), Theorem 3.1, S is asymptotically distributed as $N\left(0, \frac{\pi^2}{6}\right)$ when the null hypothesis of independence holds. In particular, this asymptotic result does not depend on the assumption of normality for y_t invoked by Tanaka in the derivation of the statistic and remains valid for the non-Gaussian data types that are the focus here. Hence, the statistic in (59) is shown to be the approximate LMP procedure for testing for long memory in (2) in all cases in which a conditional distribution in the exponential class is adopted. As in the case of the short memory statistic derived in the previous section, the long memory statistic is independent of the response function $h(.)$ used to transmit correlation into (2) via $\boldsymbol{\lambda}$. Also as in the short memory case, the statistic in (58) corresponds to the exact LBI statistic for long memory derived under an unconditional Gaussian assumption.

5.3 Testing for Stochastic Volatility

We now treat the dispersion parameter, γ in (36) as stochastic while keeping $\boldsymbol{\theta}$ constant. Under conditional independence,

$$f(y_t|\theta,\lambda_t) = c(y_t, a(\lambda_t)) \exp\left\{\frac{y_t\theta - b(\theta)}{a(\lambda_t)}\right\}, \tag{60}$$

where θ is now assumed to be a fixed scalar parameter for all t and $\gamma = \lambda_t$ is a random parameter linked to the underlying latent variable x_t via (1). The parameter λ_t will produce randomness in the conditional variance of y_t, whilst the fixed parameter θ will ensure that the conditional mean is nonstochastic. In this case $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}}$ is independent of $\boldsymbol{\lambda}$ so that (5) has only a single component due to the diagonal matrix $\boldsymbol{\Sigma}_{\mathbf{y}|\boldsymbol{\lambda}}$.

The null hypothesis is that $\mathbf{y}$ has an overdispersed exponential family distribution (since λ_t is an $i.i.d$ process under the null) against the alternative that there is volatility clustering, i.e. that λ_t is a correlated sequence with either short or long memory. Again, assumptions *(a), (b)* and *(c)* of Section 4.2 are deemed to hold. The null and alternative hypotheses are characterized by (21) and (22) respectively, with the latent process x_t assumed to follow either the $AR(1)$ process in (41) or the long memory process in (50). Expectations are evaluated, under the null, at $E[\lambda_t] = \mu_\lambda$. Given (60), the elements of the vector $\mathbf{q}$ are given by

$$q_t = \left.\frac{\partial \log(c)}{\partial \lambda_t}\right|_{\lambda_t=\mu_\lambda} + [y_t\theta - b(\theta)]\left[-\frac{1}{a(\lambda_t)^2}\left.\frac{\partial a}{\partial \lambda_t}\right|_{\lambda_t=\mu_\lambda}\right]$$

and the diagonal elements of the matrix $\mathbf{R}$ given by

$$r_{tt} = \left.\frac{\partial^2 \log(c)}{\partial \lambda_t^2}\right|_{\lambda_t=\mu_\lambda} + [y_t\theta - b(\theta)]\left[\frac{2}{a(\lambda_t)^3}\left.\frac{\partial a}{\partial \lambda_t}\right|_{\lambda_t=\mu_\lambda} - \frac{1}{a(\lambda_t)^2}\left.\frac{\partial^2 a}{\partial \lambda_t^2}\right|_{\lambda_t=\mu_\lambda}\right],$$

with zero off-diagonal elements. Notice the differences between q_t and r_{tt} derived in this section and the corresponding quantities computed from (38) and (39). In particular, in the present case q_t is typically a non-linear function of the observed variable y_t and r_{tt} is not free of the observations. This means that, in contrast to tests for correlation in the y_t, the optimal statistic for stochastic volatility will depend on the particular member of the exponential family chosen for the analysis. This is borne out below where the optimal statistic in the Gaussian case is based on the (standardized) squares of the observations while that of the Gamma distribution is not.

As $\left.\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\lambda}}}{\partial \pi}\right|_{\pi=0} = \mathbf{0}$ for the particular memory models being entertained, the form of the test statistic in (24) applies. Also, given that the diagonal elements of $\mathbf{D}_0$ are equal to zero, for both the short and long memory models described in the previous sections, the statistic then reduces to

$$S = \frac{\mathbf{q}'\mathbf{D}_0\mathbf{q}}{2 + \sigma_\lambda^2 \left[\mathbf{q}'\mathbf{q} + \sum_{t=1}^{T} r_{tt}\right]}. \tag{61}$$

5.3.1 Conditional Gaussian Distribution

If we consider the case where $y_t|\theta, \lambda_t \sim N(\theta, \lambda_t)$, we have that $E[y_t] = \theta = \mu_y$ and $V[y_t] = E[\lambda_t] = \mu_\lambda = \sigma_y^2$, giving $\log\left(c\left(y_t, \lambda_t\right)\right) = -\frac{1}{2}\left[\frac{y_t^2}{\lambda_t} + \log(2\pi\lambda_t)\right]$, $b(\theta) = \theta^2/2$ and $a(\lambda_t) = \lambda_t$. Thus,

$$\begin{aligned}
q_t &= \left.\frac{1}{2\lambda_t}\left(\frac{y_t^2}{\lambda_t} - 1\right)\right|_{\lambda_t=\mu_\lambda} - \left.\left[y_t\theta - \theta^2/2\right]\frac{1}{\lambda_t^2}\right|_{\lambda_t=\mu_\lambda} \\
&= \left.\frac{1}{2\lambda_t^2}\left[(y_t-\theta)^2 - \lambda_t\right]\right|_{\lambda_t=\mu_\lambda} \\
&= \frac{1}{2\sigma_y^4}\left[\left(y_t - \mu_y\right)^2 - \sigma_y^2\right] \\
&= k.\frac{d_t - E[d_t]}{V[d_t]^{1/2}},
\end{aligned} \tag{62}$$

where $d_t = \left(y_t - \mu_y\right)^2$ and k is some constant. Similarly,

$$\begin{aligned}
r_{tt} &= -\left.\frac{1}{\lambda_t^3}\left[\left(y_t - \mu_y\right)^2 - \frac{\lambda_t}{2}\right]\right|_{\lambda_t=\mu_\lambda} \\
&= -\frac{1}{\sigma_y^6}\left[\left(y_t - \mu_y\right)^2 - \frac{\sigma_y^2}{2}\right] \\
&= k_1\left[\frac{d_t - E[d_t]}{V[d_t]^{1/2}} + k_2\right],
\end{aligned} \tag{63}$$

for some constants k_1 and k_2. Next define an empirical version of d_t, $d_t = (y_t - \overline{y})^2$, with the vector $\mathbf{d}_s$ composed of the standardized elements $(d_t - \overline{d})/s_d$, where $\overline{d}$ and s_d are the mean and standard deviation of d_t. Empirical versions of q_t and r_{tt} are then produced by replacing the term $\left(d_t - E[d_t]\right)/V[d_t]^{1/2}$ by $(d_t - \overline{d})/s_d$ in (62) and

(63) respectively. Since $\mathbf{d}_s'\mathbf{d}_s = T$ and $\mathbf{d}_s'\mathbf{i} = 0$, it follows that $\mathbf{q}'\mathbf{q}+\sum_{t=1}^{T} r_{tt}$ in (61) is equal to a constant and we may use

$$S = \mathbf{d}_s'\mathbf{D}_0\mathbf{d}_s \tag{64}$$

as a test statistic. Note that we gain robustness to normality by using s_d as a consistent estimator of the fourth moment. It follows that the tests for short run and long run correlation in the volatility may be constructed by calculating (49) and (59) but using the standardized squares of the observations, $\left(\frac{d_t-\overline{d}}{s_d}\right)$, in place of the standardized observations themselves. Under the null hypotheses in (43) and (53) respectively and when the regularity conditions of Fuller (1996) and Tanaka (1999) are applied to d_t, $\overline{d}$ and s_d converge by the weak law of large numbers, and hence it is easy to show that S is asymptotically normal when scaled appropriately.

Thus, under the assumption of conditional normality, the approximate LMP test for $AR(1)$ correlation in the conditional variance is the first-order correlation coefficient constructed from the standardized squared data and, under the null of independence, asymptotically $N(0,1)$ critical values may be used. This test is asymptotically equivalent to the Lagrange Multiplier test for an Autoregressive Conditionally Heteroscedastic (ARCH) process of order one, as proposed in Engle (1982), also under the assumption of conditional normality; see also McLeod and Li (1983). The approximate LMP test for long memory (given conditional normality) is the Tanaka (1999) test but applied to the standardized squared data.

5.3.2 Gamma Conditional Distribution

In the case where the data is restricted to the positive region, a Gamma distribution may be thought to be an appropriate choice of conditional distribution, where $y_t|\theta, \lambda_t$ in this case has density

$$f(y_t|\theta,\lambda_t) = \frac{1}{\Gamma(\frac{1}{\lambda_t})}\frac{1}{y_t}\left(\frac{-\theta y_t}{\lambda_t}\right)^{\frac{1}{\lambda_t}} \exp\left(\frac{\theta y_t}{\lambda_t}\right), \tag{65}$$

with $E[y_t|\theta,\lambda_t] = -\frac{1}{\theta}$, $V[y_t|\theta,\lambda_t] = \frac{1}{\theta^2}\lambda_t$, $E[y_t] = -\frac{1}{\theta} = \mu_y$ and $V[y_t] = \mu_y^2 E[\lambda_t] = \mu_y^2\mu_{\boldsymbol{\lambda}} = \sigma_y^2$. Also define $\mu_{\boldsymbol{\lambda}} = \sigma_y^2/\mu_y^2 = v_y$. In this case, it follows that $\log c\,(y_t,\lambda_t) =$

$\lambda_t^{-1}\log\left(\lambda_t^{-1}y_t\right) - \log(y_t) - \log\Gamma(\lambda_t^{-1})$, $b(\theta) = -\log(-\theta)$ and $a(\lambda_t) = \lambda_t$. Thus,

$$\begin{aligned} q_t &= \frac{\partial}{\partial\lambda_t}\left[\lambda_t^{-1}\log\left(\lambda_t^{-1}y_t\right)\right]\Big|_{\lambda_t=\mu_\lambda} - \frac{\partial}{\partial\lambda_t}\log\Gamma\left(\lambda_t^{-1}\right)\Big|_{\lambda_t=\mu_\lambda} \\ &\quad - \frac{[y_t\theta + \log(-\theta)]}{\lambda_t^2}\Big|_{\lambda_t=\mu_\lambda} \\ &= -\frac{1}{\lambda_t^2}\left\{\log\left(\lambda_t^{-1}y_t\right) + 1 + [y_t\theta + \log(-\theta)]\right\}\Big|_{\lambda_t=\mu_\lambda} - \frac{\partial}{\partial\lambda_t}\log\Gamma\left(\lambda_t^{-1}\right)\Big|_{\lambda_t=\mu_\lambda} \\ &\approx v_y^{-2}\left[\frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right) + \log(v_y) - 1\right] - v_y^{-2}\left[\frac{\nu_y}{2} + \log(v_y)\right] \\ &= v_y^{-2}\left[\frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right) - \left(1 + \frac{v_y}{2}\right)\right], \end{aligned}$$

using the approximation $\log\Gamma\left(\lambda_t^{-1}\right) \approx \left(\lambda_t^{-1} - 1/2\right)\log(\lambda_t^{-1}) - \lambda_t^{-1} + k$, for some constant k. Using (11), it follows that $E\left[\log(\frac{y_t}{\mu_y})\right] \approx -v_y/2$ and

$$E\left[\frac{y_t}{\mu_y} - \log\left(\frac{y_t}{\mu_y}\right)\right] \approx \left(1 + \frac{v_y}{2}\right). \tag{66}$$

Thus, defining

$$g_t = \frac{y_t}{\mu_y} - \log(\frac{y_t}{\mu_y}) \tag{67}$$

and using the approximation (66), we may write

$$q_t = k_1.\frac{(g_t - E[g_t])}{V[g_t]^{1/2}}, \tag{68}$$

for a constant k_1. Similarly,

$$\begin{aligned} r_{tt} &= \frac{2}{\lambda_t^3}\left[\log\left(\lambda_t^{-1}y_t\right) + \frac{1}{2} + [y_t\theta + \log(-\theta)]\right]\Big|_{\lambda_t=\mu_\lambda} \\ &\quad - \frac{\partial^2}{\partial\lambda_t^2}\log\Gamma\left(\lambda_t^{-1}\right)\Big|_{\lambda_t=\mu_\lambda} \\ &= k_2\left(\frac{g_t - E[g_t]}{V[g_t]^{1/2}} + k_3\right) \end{aligned}$$

for some constants k_2 and k_3. Define an empirical version of g_t in (67) as

$$g_t = \frac{y_t}{\bar{y}} - \log\left(\frac{y_t}{\bar{y}}\right) \tag{69}$$

and $\mathbf{g}_s$ as the vector of standardized elements, $(g_t - \overline{g})/s_g$, with $\overline{g}$ and s_g respectively the mean and standard deviation of g_t. We find, as in the Gaussian case above, that $\mathbf{q}'\mathbf{q} + \sum_{t=1}^{T} r_{tt}$ in (61) is equal to a constant and that

$$S = \mathbf{g}_s' \mathbf{D}_0 \mathbf{g}_s \tag{70}$$

can serve as a test statistic as a consequence. As in the conditional Gaussian case, tests for short run and long run correlation in the conditional variance may be constructed by calculating (49) and (59) but using the standardized variable $\mathbf{g}_s$ in place of the standardized observations themselves. Interestingly, and in contrast to the Gaussian case, the approximate optimal test is not based on the squares of the observations. When g_t satisfies the regularity of Fuller (1996), it is also straightforward to show that S (scaled) is still asymptotically $N(0,1)$ under the null of independence, for $\mathbf{D}_0$ equal to $\mathbf{A}$ as defined in Section 5.2.1. Similarly, the asymptotic $N(0, \frac{\pi^2}{6})$ distribution still holds under the null for $\mathbf{D}_0$ equal to $\mathbf{B}$ as defined in Section 5.2.2.

6 Illustrative Applications

6.1 Preliminaries

In this section we report the results of applying the four tests derived in Sections 5.2.1, 5.2.2, 5.3.1 and 5.3.2 respectively, to various non-Gaussian financial time series. Three of the data sets considered relate to trading on the Australian firm Broken Hill Proprietary (BHP) Limited, namely daily returns between 1998 and 2001, one-minute trade counts for 1 August, 2001, and trade durations for 1 August, 2001. The fourth data set comprises daily returns on the S&P500 index between 1994 and 1997. The purpose of the empirical exercise is two-fold. First, to confirm the existence of non-Gaussian data with the particular dynamic properties for which the procedures developed in the paper are designed to test. Secondly, to use the distributional features of the various data sets to motivate the design of the Monte Carlo experiments reported in Section 7 of the paper.

The empirical results associated with all four data sets are reported in Table 1, with the sample size on which each set of results is based reported in parentheses. Both the short memory (SM) correlation test and the long memory (LM) correlation

Table 1: Empirical Results

	SM TEST	LM TEST	SMSV TEST		LMSV TEST	
			NORM	GAM	NORM	GAM
BHP Returns ($T = 1011$)	2.736*[(a)]	0.688	6.192*	n.a.[(b)]	10.715*	n.a.
S&P500 Returns ($T = 949$)	1.748	0.318	1.497	n.a.	9.585*	n.a.
BHP Trade Durations ($T = 1432$)	3.106*	6.207*	n.a.	0.738	n.a.	2.147*
BHP Trade Counts ($T = 360$)	5.776*	17.159*	n.a.	n.a.	n.a.	n.a.

(a) * denotes significance at the 5% level.

(b) n.a. = not applicable.

test are applied to all four data sets. The stochastic volatility tests (both short memory (SMSV) and long memory (LMSV)), that assume conditional normality (NORM), are applied to the two returns data sets, whilst the stochastic volatility tests that assume a conditional gamma (GAM) distribution are applied to the durations data. Test statistics that are significant at the 5% level, using the appropriate asymptotic critical value, are indicated by an asterisk.

6.2 Empirical Results

In Figure 1, panels (a), (b) and (c) respectively, we present the times series plot, empirical distribution and sample ACF for BHP daily returns from 2 January, 1998 to 31 December, 2001. The sample ACF's for the squared and absolute returns are presented in panel (d). The asymptotic confidence bounds of $1.96 \times 1/\sqrt{T}$ are included

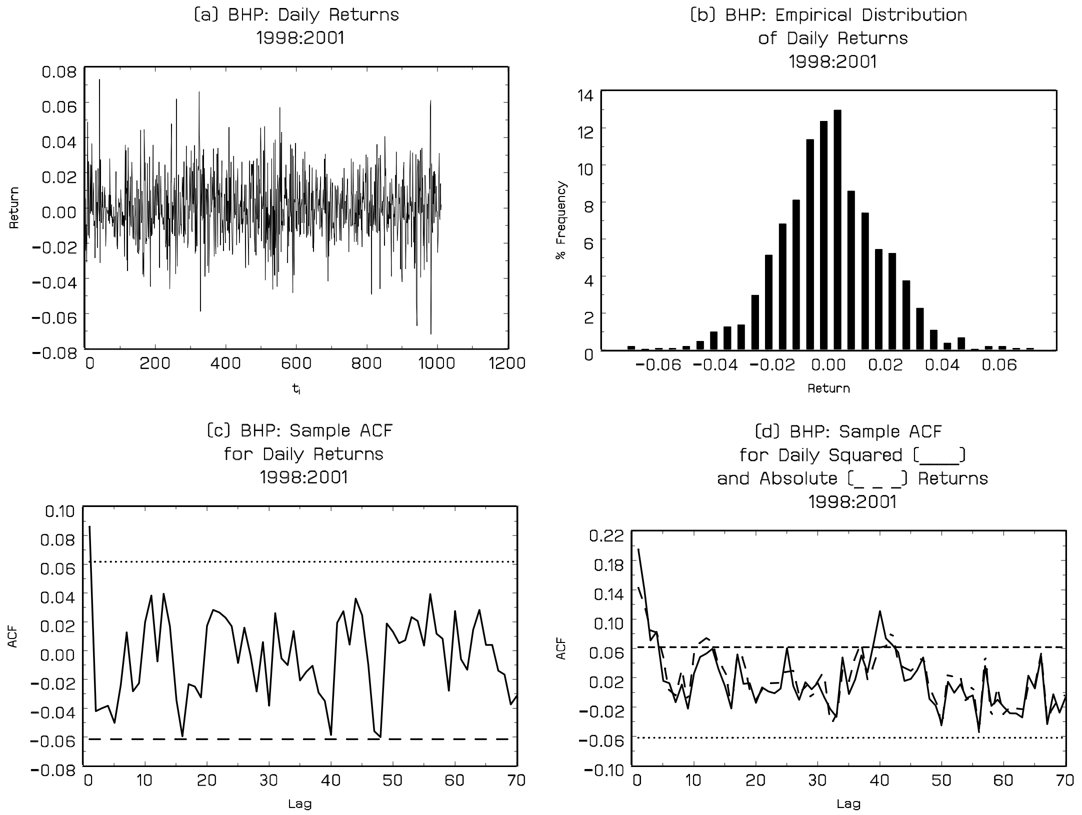

Figure 1:

on each ACF graph. As is very typical of such data, the sample ACF for the levels of the data indicates little correlation, in accordance with the efficient market theory. In contrast, more substantial correlation is evident in both squared and absolute returns, with there being a slower decline in the ACF for these quantities, indicating the possible presence of long memory in volatility; see, for example, Engle, Granger and Ding (1993) and Ray and Tsay (2000).

Linked to the time-varying volatility feature, the unconditional distribution of returns exhibits more kurtosis than is associated with a normal distribution, with the estimated kurtosis coefficient of 3.911 being significantly greater that the value of 3 associated with the normal distribution. That said, the degree of excess kurtosis is not extreme. In addition, there is little evidence of skewness, with the estimated skewness coefficient of 0.069 being insignificantly different from the value of zero associated with normality. As such, the assumption of conditional normality appears

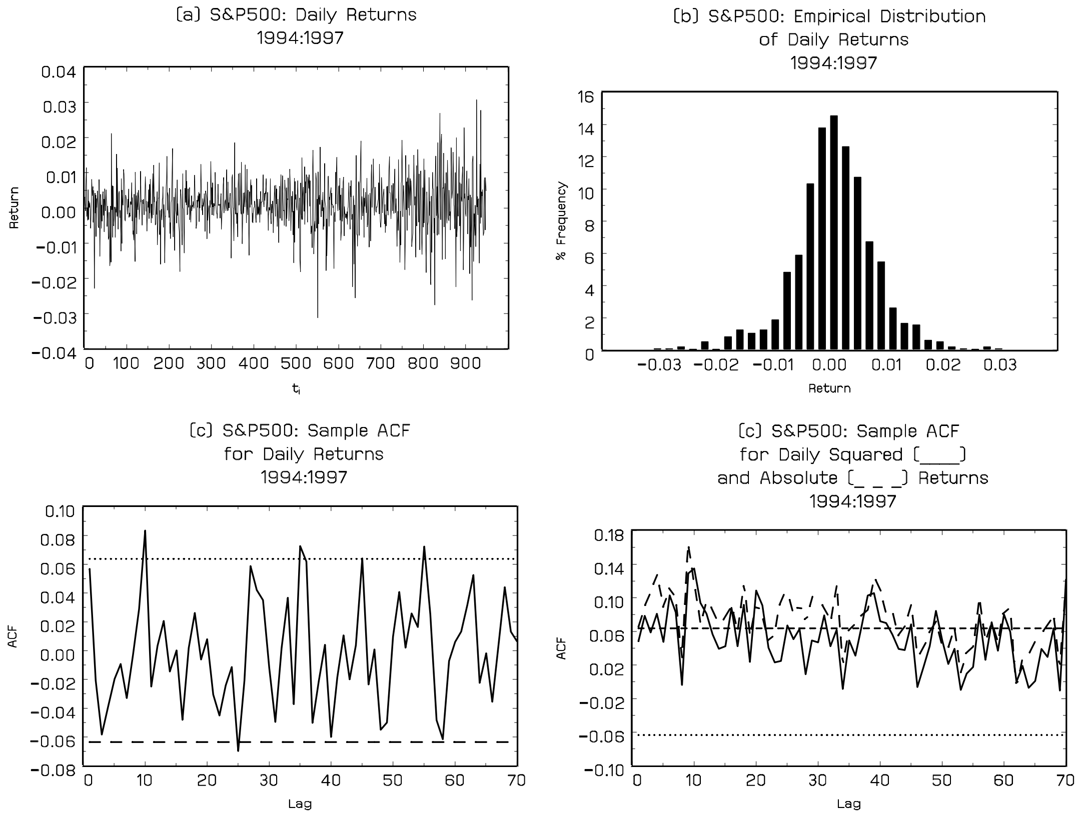

Figure 2:

justified, with the stochastic volatility test statistic in (64), as based on the squared data, being applicable as a consequence.

The results of the four tests as applied to the BHP returns data are reported in the top panel in Table 1. The SM test rejects the null hypothesis of independent returns at the 5% significance level, although the degree of first-order autocorrelation is not substantial, with the estimated coefficient having a value of 0.087. The LM test clearly fails to reject the null.[5] In contrast, significant short and long memory correlation is found in the conditional variance.

Figure 2 reproduces the graphical features of S&P500 returns from 2 January, 1994 to 31 December, 1997. In this case the evidence of excess kurtosis is slightly more marked, with significant kurtosis and skewness coefficients of 4.928 and -0.240 respectively. As with the BHP data however, the departure from normality is not

[5]The 5% asymptotic critical value for the LM test is $1.645 \times \sqrt{(\pi^2/6)} = 2.110$.

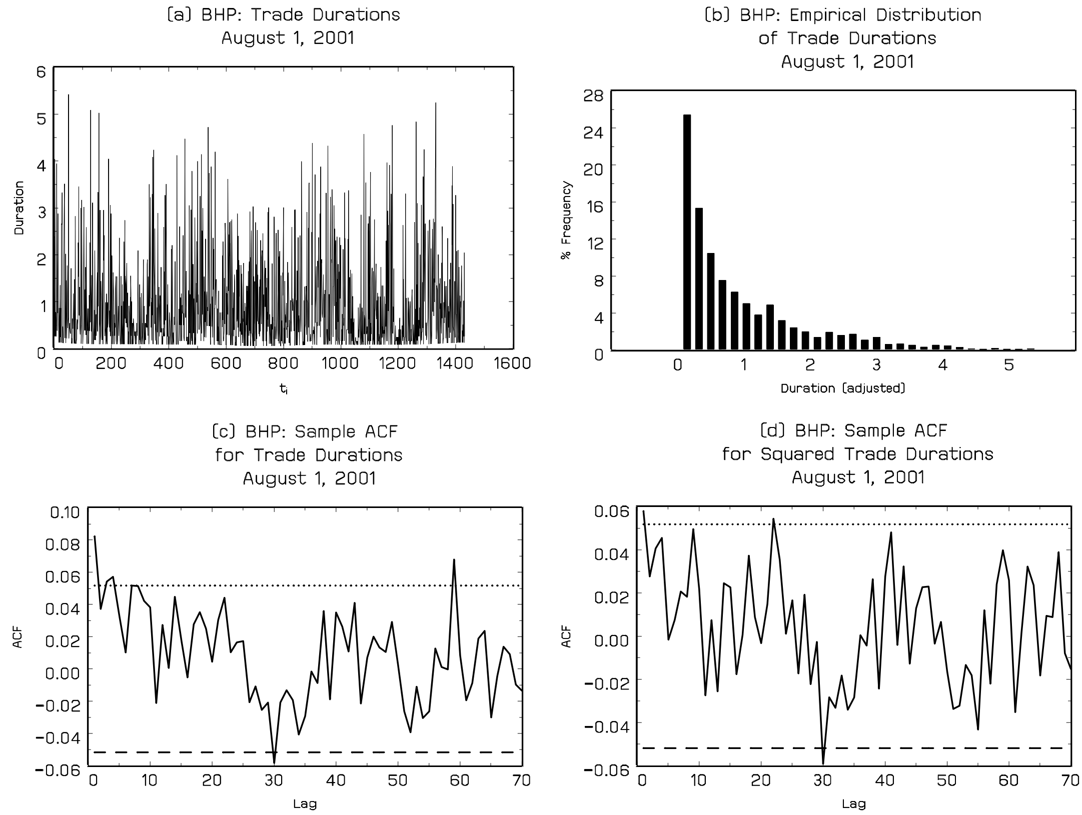

Figure 3:

extreme, indicating that a conditional Gaussian assumption is acceptable. The sample ACF's for the squared and absolute data provide much stronger evidence for a long memory structure in volatility, with significant autocorrelations still occurring after 60 lags. For this particular data set, the results (reported in the second panel in Table 1) tally exactly with the stylized facts associated with returns data. The level of returns display neither short nor long memory characteristics, according to the SM and LM correlation statistics, whilst the stochastic volatility tests clearly indicate that a long memory rather than a short memory structure exists.

In Figure 3 the empirical features of the BHP trade durations data is displayed. The data comprises the durations between trades on 1 August 2001, between 10.20am and 4.00pm, with zero trade durations omitted. The intraday pattern in the duration data is modelled using a cubic smoothing spline, with the roughness penalty chosen using generalized cross-validation; see also Engle and Russell (1998). The durations are then adjusted by dividing raw durations by the ordinate of the estimated spline function evaluated at the corresponding points. In modelling the correlation in such data using either an Autoregressive Conditional Duration model (Engle and Russell, 1998) or a Stochastic Conditional Duration Model (Bauwens and Veradas, 2004) the conditional distribution is typically specified as being either exponential or some variant thereof, such as the Weibull or gamma distributions. Certainly the empirical distribution in panel (b) indicates that any such distribution is a plausible choice. In particular, the adoption of a conditional gamma distribution means that the stochastic volatility statistic in (70) can be used to test for both short and long memory volatility in the data. From the statistics reported in Table 1 it is clear that as well as there being significant short and long memory correlation in (adjusted) durations over the day, there is evidence of a long memory structure in volatility. We include the sample ACF for the squared durations in panel (d) of Figure 3 in order to highlight the fact that the squared values of the data are not the appropriate quantity to consider in this case, with the graph giving no hint of the long memory discerned by the LMSV test.

Finally, in Figure 4 we present the graphical features of the one-minute trade count data for the six hours (360 minutes) between 10.00am and 4.00pm on 1 August 2001. In this case the data has not been adjusted for the intraday pattern. Perhaps as a consequence of this, more substantial memory is evident from the sample ACF. Both correlation statistics reported in Table 1 are also highly significant.

7 Finite Sample Properties

In this section we calculate finite sample sizes and powers for respectively : 1) the SM test, adopting an exponential conditional distribution; 2) the LM test, adopting a Poisson conditional distribution; 3) the SMSV test, adopting a gamma conditional distribution; and 4) the LMSV test under the assumption of conditional normality. All calculations in the Monte Carlo experiments are based on 10000 replications of the relevant process, with results reported for sample sizes of 400, 1000 and 1500. The latter are chosen to tally approximately with the sizes of the various empirical data sets analyzed in the previous section. Certain aspects of the experimental design

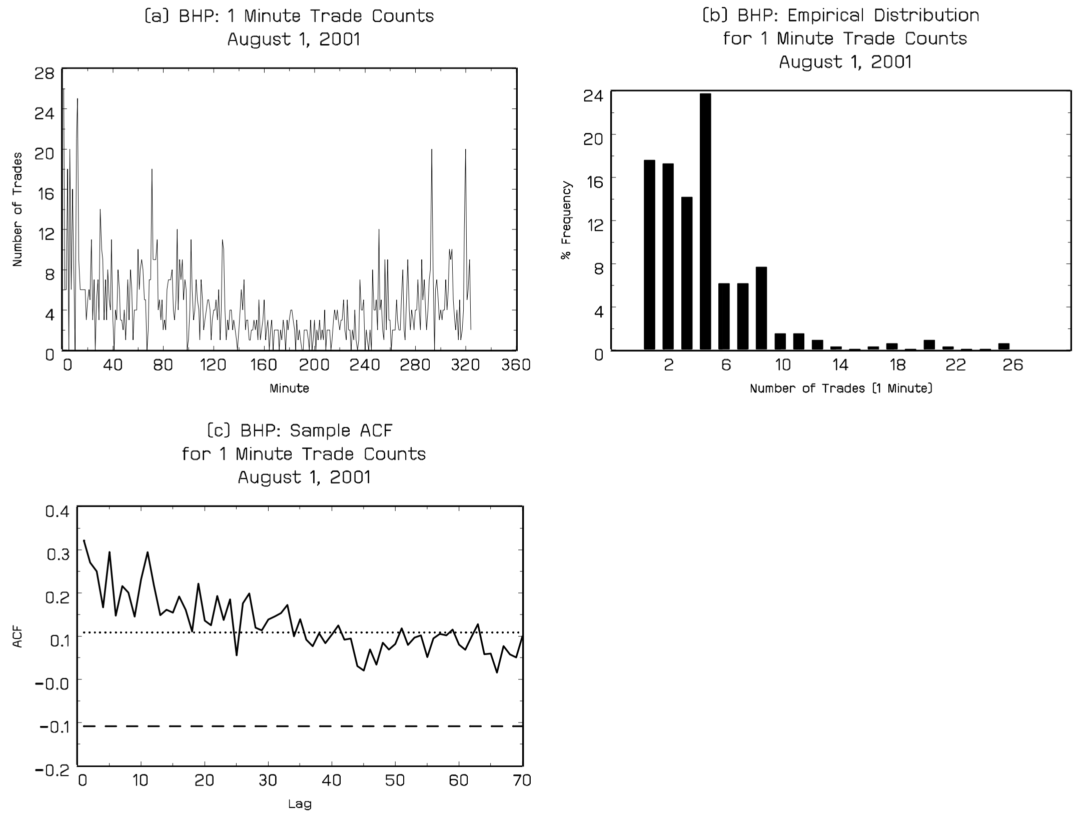

Figure 4:

for each set of experiments are determined by the features of the empirical data sets, with details of the calibration of the simulated data with the empirical data outlined in the Appendix. All results are presented in Table 2 below.

The hypotheses under test in the first set of experiments are given by (43) and (44) above. Under the null hypothesis, the statistic in (49) is asymptotically distributed as $N(0,1)$. Hence, the 5% empirical size of the test is estimated by the proportion of times that the calculated test statistic exceeds the 5% nominal critical value of 1.645. The conditional distribution of y_t is assumed to be exponential, hence

$$f(y_t|\lambda_t) = \lambda_t^{-1} e^{-y_t/\lambda_t} \tag{71}$$

for $t = 1, 2, \ldots, T$. In all replications, the values of x_t are generated from the $AR(1)$ process (41), with the response function in (1) assumed to be exponential, that is

$\lambda_t = h(x_t) = \exp(x_t)$.[6] The mean of x_t, a, is assigned a value that ensures that the mean value of the simulated data is approximately equal to 2, a value that is similar to the average value of the (adjusted) durations data analyzed in the previous section; see the Appendix for details.

The size and power results are recorded in the upper panel in Table 2. Corresponding to each value of ρ, the mean value of the first order correlation coefficient of y_t across the replicated samples of the particular size (ρ_y) is recorded.[7] For the sample size of $N = 400$, the power of the test for even low levels of correlation in the y_t values is very high. For the sample size closest to the size of the durations data set, $N = 1500$, the probability of the test correctly rejecting the null of independence in favour of the alternative of Markov dependence is approximately 90% even when the (estimated) degree of first-order correlation in the data is only 0.087, a value that is equivalent to the sample correlation coefficient for the empirical sample of durations. The empirical size of the test is reasonably close to the nominal level, for all sample sizes considered.

In the second set of experiments, the hypotheses under test are given by (53) and (54) above. Under the null, the statistic in (59) is asymptotically distributed as $N\left(0, \frac{\pi^2}{6}\right)$. Hence, the 5% empirical size of the test is estimated by the proportion of times that the calculated test statistic exceeds the 5% nominal critical value of $1.645 \times \sqrt{(\pi^2/6)} = 2.110$. The conditional distribution of y_t is assumed to be Poisson, hence

$$f(y_t|\lambda_t) = \frac{e^{-\lambda_t}\lambda_t^{y_t}}{y_t!} \tag{72}$$

for $t = 1, 2, \ldots, T$. In order to reduce the computational burden associated with the replication of (50) under H_1, the $AR(\infty)$ process invoked by the fractional operator, for $d \neq 0$, is approximated by an $AR(500)$ process,

$$x_t = a + d_1 x_{t-1} + d_2 x_{t-2} + \cdots + d_{500} x_{t-500} + \eta_t, \tag{73}$$

[6] In this and all other sets of experiments, the qualitative nature of the results was found to be robust to the choice of response function.

[7] This mean value is an estimate of the expected value of the first-order sample correlation coefficient for y_t, as based on a particular sample size. Since the sample correlation coefficient is a downwardly biased estimate of the population correlation coefficient in finite samples, (the estimate of) the expected value is likely to slightly understate the true degree of correlation present in the y_t process associated with the given latent process.

Table 2: Finite Sample Sizes and Powers

		$N = 400$		$N = 1000$		$N = 1500$	
SM TEST	ρ	ρ_y	Size/Power	ρ_y	Size/Power	ρ_y	Size/Power
	0.0	-0.003	0.053	-0.001	0.059	-0.001	0.057
	0.1	0.023	0.128	0.026	0.203	0.026	0.248
	0.3	0.084	0.447	0.087	0.775	0.087	0.901
	0.5	0.161	0.838	0.165	0.994	0.165	1.000
	0.7	0.267	0.987	0.276	1.000	0.277	1.000
LM TEST	d		Size/Power		Size/Power		Size/Power
	0.0		0.037		0.040		0.041
	0.1		0.365		0.696		0.852
	0.2		0.860		0.996		1.000
	0.3		0.990		1.000		1.000
	0.4		0.999		1.000		1.000
SMSV TEST	ρ		Size/Power		Size/Power		Size/Power
	0.0		0.047		0.052		0.054
	0.1		0.100		0.125		0.147
	0.3		0.271		0.447		0.570
	0.5		0.543		0.802		0.910
	0.7		0.795		0.958		0.983
LMSV TEST	d		Size/Power		Size/Power		Size/Power
	0.0		0.032		0.043		0.042
	0.1		0.089		0.152		0.192
	0.2		0.241		0.501		0.638
	0.3		0.520		0.878		0.956
	0.4		0.785		0.981		0.994

where $d_j = -\frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)}$. Simulation of x_t then occurs via the finite order AR model in (73), for a range of values of d, with $\lambda_t = \exp(x_t)$.[8] In this case the parameter a is used to produce simulated data with a mean value of approximately 5, which accords with the Poisson count data analyzed in the previous section.

The size and power results for the long memory test are recorded in the third panel of Table 2. As is evident from the results, the power of this test rises sharply near to the null hypothesis. For a sample size of 1000, the test has approximately 70% probability of correctly rejecting the null of independence when the underlying latent process is fractional with d equal to only 0.1. This probability increases to 85% for $N = 1500$. For the sample size closest to the size of the trade count data set, $T = 400$, there is close to 100% probability of correctly rejecting the null when the fractional parameter is equal to 0.3. The empirical size is somewhat less than the nominal size of 5% for the sample sizes considered here.

The third set of results relate to the application of the SMSV test to conditionally gamma data, with the simulated data being calibrated with the empirical durations data. The density of the conditional distribution is given by (65) and the underlying latent variable, x_t, assumed to follow the Markov process in (41), with $\lambda_t = \exp(x_t)$. All details of the specification of values for the parameters θ and a, are outlined in the Appendix.

Under the null hypothesis of (43), the statistic in (49), as applied to the standardized quantities in the vector $\mathbf{g}_s$, defined with reference to the quantity g_t in (69), is asymptotically distributed as $N(0, 1)$. The empirical size and power calculations are recorded in the fourth panel of Table 2. The powers are uniformly smaller than the corresponding powers for the SM correlation test applied to the conditionally exponential data, for each value of ρ. However, for the sample size that is closest to the size of the durations data set, $T = 1500$, the power is high for values of ρ far from the null.

The final set of experiments relates to the application of the LMSV test to condi-

[8] The alternative to this method of simulation is to generate the fractional process exactly using the known variance covariance matrix of the $(N \times 1)$ vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)'$. For large N, however, this exercise is computationally burdensome, given the need to calculate the Cholesky decomposition of an $(N \times N)$ matrix.

tionally normal data. The conditional density of y_t is thus given by

$$f(y_t|\theta,\lambda_t) = (2\pi\lambda_t)^{-1/2}\exp\left(-\frac{1}{2\lambda_t}(y_t-\theta)^2\right). \quad (74)$$

The conditional variance, $var(y_t|\theta,\lambda_t) = \lambda_t$, is linked to the underlying variable x_t via $\lambda_t = \exp(x_t)$, with x_t assumed to follow the fractional process in (50). Details of the way in which the simulated data is calibrated with the empirical S&P500 returns data via the specification of values for the fixed conditional mean, θ, and the parameter a of the model in (73), are provided in the Appendix.

Under the null hypothesis of (53), the statistic in (59), as applied to the standardized squares of the simulated data, is asymptotically distributed as $N\left(0,\frac{\pi^2}{6}\right)$. The empirical size and power calculations are recorded in the bottom panel of Table 2. The powers are uniformly smaller that the corresponding powers for the LM correlation test applied to the conditionally Poisson data, for each value of d, markedly so for the lower values of d. However, for the larger sample sizes, including that closest to the size of the returns data set, $T = 1000$, power is close to 100% for $d = 0.4$.

8 Conclusions

In this paper we have derived statistics for testing various forms of dependence in non-Gaussian data. The methodology is based on the modelling of dependence in the observed data indirectly via a dynamic structure for a latent process. In exploiting an approximation to the exact likelihood function, the computational issues associated with the unobservable variables are obviated. The tests are derived as locally most powerful tests, and, thus, exploit the accuracy of the approximation to the true likelihood function in the region of the null hypothesis of independence.

The short and long memory correlation statistics are invariant to the distribution adopted within the exponential family. Hence, the tests produced here have optimality properties in very broad distributional settings. The stochastic volatility statistics on the other hand have a form that is dependent on the particular distribution used in the exponential family. We conjecture that tests for correlation in higher-order moments would mimic this feature of the volatility tests, in the sense of being dependent in some way on the particular conditional distribution used to capture the basic features of the non-Gaussian data. The derivation of such higher-order dependence tests is left for future work. The application of the correlation and volatility tests to non-Gaussian financial data has been demonstrated, and their finite sample performance documented. The tests have been shown to possess high power, especially for the larger sample sizes typically associated with financial data sets, along with good size behaviour overall.

Acknowledgments

This research has been supported by an Australian Research Council Discovery Grant No. DP0450257. The authors would like to thank Chris Strickland for helping to produce the data used in the empirical section of the paper. We would also like to thank Rob Hyndman for some very constructive comments on an earlier draft of the paper. Thanks are due to participants at seminars at Monash University, the University of Sydney, the University of York, the University of Leeds and the University of Manchester Institute of Science and Technology.

References

[1] Anderson,T.W., (1971), *The Statistical Analysis of Time Series*, Wiley, NY.

[2] Bauwens, L. and Veradas, D., (2004), 'The Stochastic Conditional Duration Model: a Latent Variable Model for the Analysis of Financial Durations', *Journal of Econometrics,* 119, 381-412.

[3] Brockwell, P.J. and Davis, R.A., (1991), *Time Series: Theory and Methods,* Second Edition, Springer Verlag, New York.

[4] Casella, G. and Berger, R.L., (1990), *Statistical Inference*, Wadsworth Publishing, California.

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

[6] Davis, R.A. and Rodriguez-Yam, G., (2003), 'Estimation for State-Space Models; an Approximate Likelihood Approach', *Draft Paper,* Colorado State University.

[7] Ding, Z., Granger, C.W.J. and Engle, R.F., (1993), 'A Long-memory Property of Stock Market Returns and a New Model', *Journal of Empirical Finance,* 1, 83-106.

[8] Engle, R.F., (1982), 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica,* 50, 987-1008.

[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] Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modeling Based on Generalized Linear Models.* New York: Springer-Verlag.

[11] Fuller, W.A., (1996), *Introduction to Statistical Time Series*, Second Edition, J. Wiley, New York.

[12] Heyde, C. C., (1997), *Quasi-Likelihood and its Application,* Springer, New York.

[13] Huber, P., Ronchetti, E. and Victoria-Fese, M-P., (2003), 'Estimation of Generalized Linear Latent Variable Models', *Working Paper No. 2003.05*, University of Geneva.

[14] King, M.L. and Hillier, G. H., (1985), 'Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model', *Journal of the Royal Statistical Society, Series B*, 47, 98-102.

[15] King, M.L. and Wu, P.X., (1997), 'Locally Optimal One-sided Tests for Multiparameter Hypotheses', *Econometric Reviews,* 16, 131-156.

[16] Kwaitkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y., (1992), 'Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root', *Journal of Econometrics,* 54, 159-178.

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

[18] McLeod, A.L. and Li, W.K., (1983), 'Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations', *Journal of Time Series Analysis,* 4, 269-273.

[19] McCullagh, P. and Nelder, J.A., (1989), *Generalised Linear Models, Second Edition,* Chapman & Hall, New York.

[20] Ray, B.K. and Tsay, R.S., (2000), 'Long-range Dependence in Daily Stock Volatilities', *Journal of Business and Economic Statistics,* 18, 254-262.

[21] Robinson, P.M., (1994), 'Efficient Tests of Nonstationary Hypotheses', *Journal of the American Statistical Association,* 89, 1420-1437.

[22] Tanaka, K., (1999), The Nonstationary Fractional Unit Root, *Econometric Theory*, 15, 549-582.

[23] Tierney, L. and Kadane, J.B., (1986), 'Accurate Approximations for Posterior Moments and Marginal Densities', *Journal of the American Statistical Association,* 81, 82-86.

Appendix: Details of Monte Carlo Designs

<u>SM TEST</u>

The correspondence between the mean value of the simulated y_t data and the value of a in the $AR(1)$ process for x_t, (41), is approximated as follows. Given that for the exponential distribution in (71), $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}} = \boldsymbol{\lambda}$, and that by (6), $E(\mathbf{y}) = \boldsymbol{\mu}_{\boldsymbol{\lambda}}$ as a consequence, controlling the simulated values of y_t so as to maintain a particular value for the marginal mean of y_t is equivalent to simulating λ_t so as to maintain that same value for the marginal mean of λ_t. Given the exponential form of the response function and the moments of the stationary $AR(1)$ process for x_t in (41), it follows that

$$\begin{aligned} E[\lambda_t] &\approx e^{x_t}|_{x_t=E[x_t]} + \frac{1}{2}\frac{\partial^2 e^{x_t}}{\partial x_t^2}|_{x_t=E(x_t)} \times var[x_t] \\ &= e^a + \frac{1}{2}e^a(\frac{1}{1-\rho^2}). \end{aligned} \tag{75}$$

Hence, for any given value of ρ, the simulation of λ_t (and y_t) values with $E(\lambda_t) = 2$ is (approximately) achieved by setting the parameter a according to

$$a = \left[\ln(2) - \ln\left(1 + \frac{1}{2}\left(\frac{1}{1-\rho^2}\right)\right)\right]. \tag{76}$$

LM TEST

In the case of the LM test, with an exponential response function adopted, a in (73) is selected according to

$$a = \left[\ln\{E[\lambda_t]\} - \ln\left(1 + \frac{1}{2}var[x_t]\right)\right](1 - \sum_{j=1}^{500} d_j), \tag{77}$$

for any given value of d, where $var[x_t]$ is approximated as $var[x_t] \approx \sum_{j=1}^{500} \psi_j^2$, with $\psi_j = \frac{\Gamma(d+j)}{\Gamma(d)\Gamma(j+1)}$ being the jth coefficient in the finite order moving average approximation to (50),

$$x_t = (1-L)^{-d}a + \psi_1\eta_t + \psi_2\eta_{t-1} + \cdots + \psi_{500}\eta_{t-500}.$$

Since $\boldsymbol{\mu}_{\mathbf{y}|\boldsymbol{\lambda}} = \boldsymbol{\lambda}$ for the conditional Poisson distribution, it follows that $E(y_t) = E[\lambda_t]$, with $E[\lambda_t]$ in (77) set equal to the approximate sample mean of the Poisson trade count data, namely 5.

SMSV TEST

In the case of the conditional gamma distribution, with density as in (65), $E(y_t|\theta,\lambda_t) = \frac{-1}{\theta}$ and $var(y_t|\theta,\lambda_t) = \frac{1}{\theta^2}\lambda_t$. From (3) and (5) it then follows that $E(y_t) = -1/\theta$, $var(y_t) = E(\lambda_t)/\theta^2$ and

$$E(\lambda_t) = \frac{var(y_t)}{[E(y_t)]^2}. \tag{78}$$

Using the expression for $E(y_t)$, the parameter θ is equated with the negative of the reciprocal of the sample mean of the durations, whilst $E(\lambda_t)$ is set equal to the function of sample variance and sample mean corresponding to (78). The expression in (76) is used to produce a value of a for any given value of ρ.

LMSV TEST

In the case of the conditional Gaussian distribution, with density as in (74), $E(y_t|\theta, \lambda_t) = \theta$ and $var(y_t|\theta, \lambda_t) = \lambda_t$. From (3) and (5) respectively it then follows that $E(y_t) = \theta$ and $var(y_t) = E(\lambda_t)$. The parameter θ is thus equated with the sample mean of the S&P500 returns data. Setting $E(\lambda_t)$ equal to the sample variance of the S&P500 returns data, the value of a, for any given value of d, is determined using (77).