Bull. Aust. Math. Soc. **95** (2017), 169–171
doi:10.1017/S0004972716000824

DETECTING AND MODELLING SERIAL DEPENDENCE IN NONGAUSSIAN AND NONLINEAR TIME SERIES

JIEYI HE

(Received 6 August 2016; first published online 2 November 2016)

2010 *Mathematics subject classification*: primary 62M10; secondary 62F12, 62G08, 62J12.

Keywords and phrases: observation-driven models, parameter-driven models, serial dependence, score test, supremum test, asymptotic theory.

Discrete time series data is seen in a wide variety of disciplines including biology, medicine, psychology, criminology and economics. However, traditional methods of detecting serial correlation in time series are not specifically designed for detecting serial dependence in discrete-valued time series. Thus new methods are needed to provide informative and implementable testing approaches.

This thesis is concerned with detection and estimation of serial dependence for a variety of observation-driven and parameter-driven models for regression analysis in binary and binomial time series. Generalised linear models (GLMs) are widely used for modelling discrete-valued data but do not allow for serial dependence and, as a result, inferences about regression effects may be invalid for time series application. Two classes of extended GLM have arisen to deal with this issue: observation-driven models and parameter-driven models, in which the serial dependence of the former relies on previous observations and residuals, and the serial dependence of the latter derives from an unobserved latent process. This thesis is structured in two parts corresponding to these two model classes. Chapter 1 provides a review of these models and existing methods for detecting and estimating serial dependence in them. Chapters 2 to 4 focus on observation-driven models and Chapters 5 to 7 focus on parameter-driven models. Chapter 8 distils the main results and conclusions from the thesis and suggests future research opportunities. The thesis proposes the use of score tests because they can be implemented using standard GLM fitting software or, for the parameter-driven models, software for fitting generalised linear mixed models, which is also readily available in most advanced statistical packages.

Thesis submitted to the University of New South Wales in August 2015; degree awarded on 24 March 2016; supervisor William Dunsmuir.

Chapter 2 reviews two subclasses of the observation-driven model group: generalised linear autoregressive moving average models (GLARMA) and binomial autoregressive and moving average models (BARMA). Score tests for detecting serial dependence are derived for a variety of specifications of the innovations used to define the state equation for these models. The score test can be implemented using the GLM estimates of regression parameters. A likelihood ratio test (LRT) and a Wald test are also defined. Two situations are considered: firstly, the regular case where there are no nuisance parameters arising in the specification of serial dependence under the null hypothesis of no serial dependence and, secondly, the irregular case, where there are nuisance parameters. In this latter case, we show how to implement supremum-type tests based on the score, likelihood ratio and Wald tests.

Chapter 3 is devoted to establishing the asymptotic properties of the various proposed test statistics as well as the standard Box–Pierce–Ljung (BLP) test based on residuals from the GLM fit, this latter being a standard method used by practitioners. To establish the asymptotic properties, we first derive the asymptotic distribution of the standard GLM estimators with a proof specifically tailored to the types of time series regressors that we consider throughout the thesis. This result is sufficient for us to establish the asymptotic distribution of the score and the BLP test for the regular case of no nuisance parameters. For the LRT and Wald tests, the distribution of the GLARMA and BARMA maximum likelihood estimators are required under the null hypothesis of no serial dependence. This is a new result. For the irregular case, where there are nuisance parameters under the null hypothesis, we show that the general large sample distribution proposed in the literature for the supremum-type tests provides a good approximation to this setting. We do not prove the asymptotic result in this case and rely, instead, on simulation evidence. Chapter 4 investigates the finite sample performance of these tests using some small-scale simulation studies, and application to several data sets is made to demonstrate their effectiveness in practice.

For parameter-driven models, the state equation consists of a regression component and a latent process component which is typically serially dependent. Fitting such models using maximum likelihood requires estimation of a large-dimensional integral or, less efficiently in a statistical sense, using composite likelihood methods. Such methods are complicated to apply and general purpose software is still not readily available to practitioners. This motivates our focus on simpler methods to firstly detect the existence of a latent process and, secondly, if one is detected, detect serial dependence in it. In Chapter 5, we have proposed easily implementable two-step score-type tests in which the first step detects the existence of a latent process and the second detects the serial dependence. The score test statistic of the first step has nuisance parameters that cannot be estimated under the null hypothesis. This thesis develops and applies supremum-type versions of the score tests to detect the presence of the latent process. Its asymptotic distribution, conditional on a fixed nuisance value and unconditional on the nuisance parameter space, is presented. The asymptotic distribution of the second step test requires the asymptotic normality of the marginal likelihood estimators, which maximise the likelihood under the assumption that the latent process is generated as an independent sequence. Using marginal likelihood estimation, practitioners can at least obtain the estimates and their standard errors of regression coefficients of parameter-driven models without the need to approximate high-dimensional integrals.

Chapter 6 concentrates on the large-sample properties of marginal likelihood estimation. For this method, conditions under which the regression parameters β and the latent process variance τ are identifiable are provided. Then the marginal likelihood estimators are shown to be consistent and asymptotically normal. Based on these results, we prove that the score test for serial dependence developed under parameter-driven models has an asymptotic chi-squared distribution under the null hypothesis and that the latent process is independent.

In practice, the marginal likelihood estimation method, when applied to binary data, gives zero estimates of the variance of the latent process with a substantial frequency. When this happens, the marginal likelihood model degenerates to the standard GLM. This thesis provides a useful finite sample approximation for the probability of obtaining $\hat{\tau} = 0$, where τ is the variance of the latent process. As a consequence of this, we show that, under finite samples, the marginal likelihood estimators of regression parameters β has a distribution that is better approximated by a mixed distribution.

If serial dependence is detected in the parameter driven-model, then the marginal likelihood estimation method cannot provide an estimate of the serial dependence parameters of the latent process. To obtain the full parameter set of regression parameter β, latent process variance τ and serial dependence coefficients ψ, the composite likelihood estimation is a possible approach. This generalisation of marginal likelihood estimation is based on the marginal densities of a joint vector of observations $f(y_t, \ldots, y_{t+L})$ in the construction of the likelihood. This thesis also implements the composite likelihood estimation method on some real data sets. Chapter 7 investigates how well the asymptotic distributions of the score-type tests and marginal likelihood estimators perform for finite samples via simulation. Application to the same real data sets as used in the section on observation-driven models is made to corresponding parameter-driven models to compare the tests based on observation-driven specification with those based on parameter-driven specifications.

Acknowledgement

I would like to thank my supervisor, Professor William Dunsmuir, for his patience, continuous support and guidance. Throughout the doctoral process, Professor William Dunsmuir has devoted his time selflessly to supervising me, encouraged me to pursue my ideas and provided technical and general advice for my research. His help and support make it possible for me to complete this thesis.

JIEYI HE, School of Mathematics and Statistics,
University of New South Wales, Sydney, New South Wales 2052, Australia
e-mail: jieyi.he@yahoo.com.au