1. Introduction

A continuous time model reflects the natural evolution of a stochastic process indexed by an interval \( I \subseteq \mathbb{R}_+ \). However in reality, most frequently the process can only be observed at discrete times. Reconciling the discrete data with the proposed continuous model is the principal motivation for my current research.

Univariate continuous-time autoregressive moving average (CARMA) processes are the continuous time analogue of the widely employed discrete-time ARMA process. CARMA\((p, q)\) processes are the solutions of linear stochastic differential equations of the form

\[
D^p Y(t) + a_1 D^{p-1} Y(t) + \cdots + a_p Y(t) = b_0 D L(t) + b_1 D^2 L(t) + \cdots + b_q D^{q+1} L(t).
\]

They were introduced in [11] in a Gaussian setting and generalized in [5] to include Lévy driving processes.

The probabilistic properties of CARMA processes have received considerable attention. However, there has been little development in statistical inference for such models or in particular, goodness-of-fit. This thesis takes the first steps in developing rigorous statistical techniques for assessing goodness-of-fit of CARMA models, and complements recent work by Brockwell and Schlemm [8]. As pointed out in [8], if one decides to model a continuous time process using the CARMA framework, three main problems arise: a) the choice of the orders \( p \) and \( q \); b) estimation of the model coefficients \( a_i \) and \( b_j \); c) choosing an appropriate model for the Lévy driving process.

2. Summary of Previous Research

We have focused on analyzing b) and c) for the CARMA(1,0) (equivalently, CAR(1)) process. Because (under general assumptions) the CARMA\((p, q)\) process can be expressed as a sum of dependent CAR(1) processes (cf. [7]), we expect that in the future, our results will be useful in analyzing the more general model.

Formally, the Lévy-driven Ornstein-Uhlenbeck process (equivalently CAR(1) or CARMA(1,0) process, see (1)) is the stationary process, \( Y \), that satisfies the stochastic differential equation

\[
dY(t) = -aY(t)dt + \sigma dL(t), \quad a, \sigma > 0.
\]

This model was proposed by Barndorff-Nielsen and Shephard [3] as a continuous time stochastic volatility model (also known as the BNS model) by replacing the Brownian motion driving process (equivalently, noise) in the classical Ornstein-Uhlenbeck model (c.f. [20]) by a more general Lévy process \( L \). Lévy-driven CAR(1) processes not only allow one to consider non-Gaussian driving processes such as the gamma process, but can also incorporate discontinuities, or jumps, in the stochastic volatility models considered in many areas of application see ([13], [14], [15], [16],[17], for example). This model has become a very popular way to describe moderate and high frequency financial data, see ([10], [4], [3] and [2]).
There are several papers that discuss estimation of the coefficients for CAR(1) models; see e.g. [6] and [21]. In [6] and [8] the authors address the third issue, namely estimation of the parameters of a specified family of Lévy processes, assuming that the order and the coefficients of the model are known.

However, before one selects a parametric family of Lévy processes and/or estimates the model coefficients, one should verify whether it is reasonable to assume that the driving process is Lévy. From the point of view of exploratory data analysis, the first step would be to plot the sample covariances of the driving process at various lags. This procedure assumes a priori that the underlying driving process has finite second moment.

The driving process $L$ is unobservable and cannot be directly recovered if the CAR(1) process $Y$ is sampled at discrete times. Hence, the first topic addressed in my previous work (cf. [1]) was the development of statistical inference techniques for the sample covariances of the (approximately) recovered driving process $L$, assuming that the second moments of the driving process are finite. Initially, we assume that the model coefficients $a$ and $\sigma$ are known. We go beyond exploratory data analysis by providing a formal test of the hypothesis of uncorrelated increments. Our test statistic is shown to be asymptotically normal. En route, we proved several results of independent interest, including finding a precise bound on the approximation error of the unit increments of the recovered process, as well as providing an elementary proof of a central limit theorem for the integrated CAR(1) process $Y$.

Subsequently, we explored the performance of the test statistic when the model coefficient $a$ is unknown and must be estimated. Due to the complex relationship between the parameter $a$ and the (approximate) recovered increments of $L$, the choice of a suitable estimator is not straightforward. We have shown the consistency and asymptotic normality of our proposed estimator and then demonstrated its effect on the asymptotic behaviour of the test statistic. This work will be submitted very soon to Statistics and Probability Letters.

3. Summary of Current Research

If the hypothesis of independent increments of the driving process of a CAR(1) model has not been rejected, we then consider a test of goodness-of-fit for the unobserved Lévy driving process. Using the empirical process defined by the estimated unit increments of the driving process $L$, we provide a test of the composite hypothesis that the driving process belongs to a specific class of Lévy processes, such as Brownian motion or gamma processes. There are two main challenges that arise here: first, the hypothesis is composite, and so estimators of the parameters defining the distribution of $L$ must be substituted for the theoretical values in the usual empirical process. As is well known, even when exact values of $L$ are available, the resulting limiting distribution no longer leads to distribution-free tests. Second, we must use estimated values of $L$ both to define the empirical distribution and to calculate estimators. The first issue is resolved by using a simple yet powerful technique proposed by Burke and Gombay in [9]: if the parameters of the distribution are estimated using a single bootstrap sample drawn from the original observations, the limit of the empirical process is the same as when the correct theoretical values of the parameters are used. The second problem have been resolved with a high sampling frequency. This work will be submitted soon to Bernoulli.

In brief, if the CAR(1) process is observed at discrete times, we have constructed test statistics to test, initially, the Lévy assumption of uncorrelated increments of the driving process. Then, if the Lévy assumption is not rejected, we have created a more precise test statistic to examine which candidate of the Lévy family could be the driving process.
4. Future Work and Extensions

We proved the asymptotic normality of the test statistics for uncorrelatedeness of the increments of the driving process. It would be interesting to consider other tests for independence of the increments.

Another interesting topic for immediate extension is to further investigate the behavior of our goodness-of-fit test statistics if we replace the parameter $a$ by the estimator.

An extension to the Lévy-driven CARMA(2,1) process is an important problem. This model was employed by Todorov and Tauchen [19] and Todorov [18] to represent stochastic volatility in the Deutsche Mark/U.S Dollar daily exchange rate.

Subsequently, a natural question is whether our results are extendable to general CARMA($p,q$) or at least CARMA($p,1$) models. A close inspection of our proofs show that they rely on the inversion formula and second order properties of $Y$. Hence, our results should be extendable, but this will require detailed analysis.

Last but not least, in [12], Garcia et al. used an $\alpha$-stable Levy process as a driving process for the CARMA(2,1) process to model spot prices from the Singapore New Electricity Market. Accordingly, it would be desirable to extend some of our results to CARMA($p,q$) models driven by a stable Lévy process. Clearly our techniques will no longer be appropriate since they rely on an $L_2$ approximation of the noise $L$ by the process $Y$, and so a different approach will be needed.

While I would like to pursue this direction of research, I am also open to any other joint research projects that would be beneficial to my future department.

References


