Ludwig Fahrmeir
Brian Francis
Robert Gilchrist
Gerhard Tutz (Eds.)

Advances in GLIM and Statistical Modelling

Proceedings of the GLIM92 Conference and the 7th International Workshop on Statistical Modelling, Munich, 13-17 July 1992

Springer-Verlag
New York Berlin Heidelberg London Paris
Tokyo Hong Kong Barcelona Budapest
1992

Table 2 contains the results of the Poisson estimation. We display the parameter estimates and the asymptotic absolute t-values using Poisson-, *White*– and MN– standard errors. The Poisson t-values are extremely high throughout, and inference based on them leads to a rejection of all but one hypothesis test against zero at the 5% confidence level. Considering the robust t-values it becomes evident that this would be a falacious decision in most of the cases. They are roughly one tenth of the Poisson t-values indicating a high degree of overdispersion in the model. The t_{MN}-values are mostly smaller than the t_{White}-values. Based on the *White* or MN t-values, only MARKET DIVERSIFICATION and the CONSUMPTION GOODS dummy are significant for both years. For the firm size / market concentration debate, the robust t-values indicate no substantial influence of market power as measured by the HERFINDAHL INDEX on inventive activity. FIRM SIZE has a significant effect only in 1982. Since the estimated coefficient for FIRM SIZE squared is negative, we observe a maximum in the relevant range with first positive but diminishing marginal returns and then negative marginal returns. However, the estimated robust t-values in contrast to the standard Poisson case are too low to rely on this finding.

5 Conclusion

Using results on the asymptotic distribution of a QMLE, we were able to derive that overdispersion will yield Poisson standard errors which underestimate the true standard errors of the estimator and that underdispersion will do the reverse. A Monte Carlo study for overdispersed data demonstrated how serious the problem is already in the presence of a modest violation of equidispersion. As an alternative, we proposed to base inference on robust standard errors. We studied two alternative approaches: The *White*-standard errors require only the assumption of a correctly specified mean function, whereas the MN-standard errors also need an assumption with respect to the variance function. The Monte Carlo evidence suggests that they both behave well for overdispersed data already in medium sized samples.

An important share of the existing literature has used more general parametric models, like for instance the negative binomial, to account for a violation of the Poisson assumption. This might, however, lead to inconsistent parameter estimates if the specific assumptions made about the departure from equidispersion are not fulfilled. Moreover, the computations might be cumbersome. A viable alternative model strategy is to use the Poisson quasi-likelihood and base inference on robust standard errors. We used this strategy to analyse a data set on patent activities, demonstrating that the approach might help to find more sound conclusions for issues of substantial policy relevance.

References

Acs, Z. and D. Audretsch 1991 *Innovation and technological change. An international comparison.* New York: Harvester-Wheatsheaf.

Gourieroux, C., A. Montfort and A. Trognon 1984 *Pseudo maximum likelihood methods: Theory.* Econometrica 52, 681-700.

Hausman, J., B.H. Hall and Z. Griliches 1984 *Econometric models for count data with an application to the patents-R&D relationship.* Econometrica 52, 909-938.

McCullagh, P. and J.A. Nelder 1989 *Generalized linear models.* 2nd ed., London: Chapman and Hall.

Schwalbach, J. and K.F. Zimmermann 1991 *A Poisson model of patenting and firm structure in Germany.* in: Acs, Z. and D. Audretsch (eds.), *Innovation and technological change. An international comparison.* New York: Harvester-Wheatsheaf.

White, H. 1982 *Maximum likelihood estimation of misspecified models.* Econometrica 50, 1-25.

PARAMETRIC AND SEMINONPARAMETRIC ANALYSIS OF NONLINEAR TIME SERIES

Stefan Mittnik
Department of Economics
State University of New York
Stony Brook, NY 11794-4384

Bruce Mizrach
Department of Finance
The Wharton School
Philadelphia, PA 19104-6337

1. Introduction

In many applications the functional form of a nonlinear process is neither likely to be known nor fit conveniently into commonly used parametric frameworks, such as bilinear models (Granger and Andersen, 1978), threshold autoregressions (Tong, 1983), exponential autoregressions (Ozaki, 1981), random-coefficient autoregressions (Tsay, 1987), or ARCH models (Engle, 1982). In this case it seems to be more appropriate to work with suitable approximations of the underlying process.

In this paper, we have two objectives. The first is to examine to what degree nonlinear generalized autoregressions (Mittnik, 1991a,b) can improve upon linear autoregressions in modeling financial data. The second is the application of seminonparametric models, as suggested by Gallant and Nychka (1987) and Gallant and Tauchen (1989), to modeling non-normality and heterogeneity in time series. In particular, we investigate whether the departures from Gaussianity financial data can be interpreted as model misspecifications. The empirical analysis is based on a time series of high-frequency, real-time Standard and Poor's 500 cash-index prices from the Chicago Mercantile Exchange. Experimental evidence in Mayfield and Mizrach (1991) seems to indicate that this data series might have a low-dimensional nonlinear structure.

The paper is organized as follows. The next section provides some background for representations of nonlinear processes. Section 3 briefly describes the generalized autoregression approach. Seminonparametric approximations are discussed in Section 4. The empirical application is presented in Section 5. Concluding remarks are given in the final section.

2. The Volterra Expansion

Let $\{x_t\}_{t=0}^{\infty}$ be a zero-mean, covariance stationary process. Taking the linear projection of x_t on past realizations of x_t, we obtain a series of white noise residuals, $\epsilon_t \equiv x_t - P(x_t|x_{t-1}, x_{t-2}, \ldots)$, where $P(\cdot|\cdot)$ is the projection operator. The Wold representation of $\{x_t\}_{t=0}^{\infty}$ is given by $x_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j} + \eta_t$ with $\sum_{j=0}^{\infty} b_j^2 < \infty$, where $b_0 = 1$ and η_t is a linearly deterministic process. The existence of the Wold representation relies crucially on the linearity of the data generating mechanism for x_t. If the x's were the realizations of a nonlinear transform, the ϵ's, though uncorrelated, would not be independent.

An analogous representation for nonlinear time series is the Volterra expansion. It closely resembles a Taylor series expansion, and will exist whenever a convergent Taylor series approximation of the data generating mechanism exists. Assuming that $\{\epsilon_t\}_{t=0}^{\infty}$ is an independent and identically distributed stochastic process defined on the finite interval $[\underline{\epsilon}, \bar{\epsilon}]$, consider the transform

$$x_t = f(x_{t-1}, \ldots, x_{t-p}) + \epsilon_t, \qquad (1)$$

where $f: \mathbf{R}^p \to \mathbf{R}$ is a stochastic difference equation.[1] The Volterra-series expansion of f, given by

$$x_t = f(0) + \sum_{i_1=1}^{p} f_{i_1} x_{t-i_1} + \sum_{i_1=1}^{p} \sum_{i_2=i_1}^{p} f_{i_{12}} x_{t-i_1} x_{t-i_2} + \sum_{i_1=1}^{p} \sum_{i_2=i_1}^{p} \sum_{i_3=i_2}^{p} f_{i_{123}} x_{t-i_1} x_{t-i_2} x_{t-i_3} + \ldots + \epsilon_t, \quad (2)$$

where $f_{i_1} = \frac{\partial f}{\partial x_{t-i_1}}\Big|_0$, $f_{i_{12}} = \frac{\partial^2 f}{\partial x_{t-i_1} \partial x_{t-i_2}}\Big|_0$ etc. and $i_{12\ldots n}$ stands for the multi-index $i_1 i_2 \ldots i_n$, provides an arbitrarily precise local approximation through a polynomial in lags of x_t for finite x_t's.

Expansion (2) is not particularly attractive for applied work. Even if p is of finite order, one may need an infinite number of kernels for the series expansion. However, this problem arises also in the linear case. The Wold representation is also infinite dimensional. We will follow the linear ARMA literature in assuming that x_t can be well approximated by a finite number of parameters, truncating our series expansion at some finite degree.

3. Generalized Autoregressions

In Mittnik (1991a,b), truncated versions of (2) are referred to as generalized autoregressions (GARs). A generalized autoregression of degree r and order p, in short, a GAR(r,p) process, is defined by

$$x_t = \sum_{i_1=0}^{r} \sum_{i_2=0}^{r} \cdots \sum_{i_p=0}^{r} f_{i_{12\ldots p}} \prod_{j=1}^{p} x_{t-j}^{i_j} + \epsilon_t. \quad (3)$$

It follows from (3) that a process is a GAR(r,p) process if all coefficients associated with monomials involving lags of order $p+1$ and higher or powers of degree $r+1$ and higher are zero and if at least one coefficient of monomials involving x_{t-p} and one involving x_{t-i}^r, for $i \in \{1, 2, \ldots, p\}$, are nonzero.

Often it is more convenient to work with a state-space or first-order Markovian representation than the higher-order difference representation (3). As is shown in Mittnik (1991b), (3) has the state space representation

$$z_{t+1} = Az_t + N(\mathbf{x}_t \otimes z_t) + B\mathbf{x}_t, \qquad x_t = Cz_t + \epsilon_t, \quad (4)$$

where $\mathbf{x}_t = (x_t, x_t^2, \ldots, x_t^r)'$ and z_t is the state vector at time t.

Because a GAR model is linear in the parameters, conditional least squares may be used for estimation. In view of potential overparameterization problems, subset regressions or other forms of restrictions may be required in practice.

4. Seminonparametric Approximation of Density Functions

Phillips (1983) introduced to the econometrics literature an approach to approximating probability densities which underlies seminonparametric methods. Elbadwi et al. (1983) coined the term seminonparametric to indicate an estimation approach that is partly parametric and partly nonparametric. The parametric component consists of specifying a functional form for the conditional mean; the nonparametric part is the approximation of the higher moments of the conditional density.

Phillips showed that any probability density function (pdf), denoted by $h(z)$ can be approximated

[1] Under fairly weak conditions, f can be locally approximated by an infinite series expansion. If f is an analytic function, then all partial derivatives of f exist and x_t has a uniformly convergent Taylor-series expansion.

arbitrarily well by the family of extended rational approximants,

$$h(z) = \Phi(z)\frac{P_m(z)}{Q_n(z)} = \Phi(z)\frac{a_0 + a_1 z + \cdots + a_m z^m}{b_0 + b_1 z + \cdots + b_n z^n}, \quad (5)$$

where $\Phi(z)$ denotes the normal density and $P_m(z)$ and $Q_n(z)$ are polynomials of degree m and n, respectively. Gallant and Nychka (1987) prove consistency of the maximum likelihood estimator by allowing the degree of the polynomials, m and n, to grow with the sample size.

Our implementation of this estimator follows closely the work of Gallant and Tauchen (1989). In the linear case, the conditional mean of y_t is modeled by linear autoregression $y_t = \mathbf{y}_{t-1}'\beta + z_t$, where $\mathbf{y}_{t-1} = (1, y_{t-1}, \ldots, y_{t-p})'$. The pdf of the residuals, denoted by $h(z_t)$, is approximated by the Hermite polynomial expansion[2]

$$h(z_t; K_n) = \frac{(\sum_{j=0}^{K_n} \alpha_j z_t^j)^2 \Phi(z_t)}{\int (\sum_{j=0}^{K_n} \alpha_j u^j)^2 \Phi(u) du}. \quad (6)$$

Polynomial $\sum_{j=0}^{K_n} \alpha_j z_t^j$ reflects departures from Gaussianity in the conditional density. If $K_n = 0$, $h(z_t)$ reduces to the normal distribution.

Heterogeneities in the conditional density are permitted by allowing the coefficients α_j to depend on lagged dependent variables. Analogous to (3) we express α_j in terms of a K_h-degree polynomial in y_{t-i}, for $i = 1, \ldots, p$, i.e.,

$$\alpha_j(\mathbf{y}_{t-1}, K_h) = \sum_{i_1=0}^{K_h} \sum_{i_2=0}^{K_h} \cdots \sum_{i_p=0}^{K_h} a_{i_{12\ldots p}} \prod_{j=1}^{p} y_{t-j}^{i_j}. \quad (7)$$

Incorporating this into (6) describes the seminonparametric model with lag length p and polynomial degrees K_n and K_h or, in short, the SNP(p, K_n, K_h) model with conditional density

$$h(z_t; K_n, K_h) = \frac{[\sum_{j=0}^{K_n} \alpha_j(\mathbf{y}_{t-1}, K_h) z_t^j]^2 \Phi(z_t)}{\int [\sum_{j=0}^{K_n} \alpha_j(\mathbf{y}_{t-1}, K_h) u^j]^2 \Phi(u) du}. \quad (8)$$

5. Empirical Analysis

We apply the above techniques to a sample of 5,000 real-time realizations of the Standard and Poor's 500 cash index which spans approximately the first trading week of January 1987. Expressing returns in terms of log differences of the index level, the raw data, as can be seen in Table 1, are positively skewed and highly leptokurtic.

The empirical application consists of three parts. First, we examine the in-sample fit of AR and GAR models. Second, we investigate their out-of-sample forecasting performances and compare them to the one of a random walk (RW). Finally, we examine to what extent departures from Gaussianity or homogeneity can be due to misspecification of the functional form for the conditional mean.

[2] In the univariate case, it is easy to show that the expansion is orthogonal. Let H_n be the nth order Hermite polynomial and let H_m be the mth order. Then $E(H_n H_m) = 0$, for $m \neq n$. See Kendall and Stuart (1963) for explicit calculation of the first ten or so Hermite polynomials and proof of the assertion of orthogonality.

We take the square in the numerator to ensure that the density is everywhere positive and divide by the integral in the denominator to ensure that $\int h(z) dz = 1$.

Table 1: Statistical Analysis of S&P500 Raw Data and Model Residuals[a]

Statistic	Raw Data Full	In	Out	AR Residuals In	Out	GAR Residuals In	Out
Mean	1.07E-5	1.18E-5	8.51E-7	-1.68E-8	-3.61E-6	-1.93E-8	-3.06E-6
Std. Dev.	9.85E-5	9.73E-5	1.09E-4	8.78E-5	9.76E-5	8.62E-5	9.73E-5
Skew	1.3117	1.6947	-1.1071	1.5993	-0.2966	1.6607	-0.3603
Kurtosis	27.0907	29.7335	10.3106	36.7696	9.5923	38.6430	10.2308

[a] The full sample consists of 5000 observations. Models are fit over the in-sample period consisting of the first 4,500 observations. The initial 12 in-sample observations are used for estimation and models selection purposes and are dropped from the analysis. The forecasts are one-step-ahead predictions for the out-of-sample period consisting of the subsequent 500 observation.

Table 2: Forecasting Comparison[a]

	Linear AR	GAR	Random Walk
MSPE	9.756E-5	9.721E-5	3.131E-4
	AR vs. RW	GAR vs. RW	GAR vs. AR
Stat. (9)	3.272	3.272	0.812
p-value	0.001	0.001	0.417

[a] The comparison is based on 500 one-step-ahead out-of-sample-forecast without reestimation. Statistic (9) has an asymptotic standard normal distribution. The p-value is for a two-sided test.

5.1. In-Sample Estimation

With the goal of parsimony in mind, we use the Bayesian Information Criterion (BIC) for model selection, i.e., BIC $= \log \hat{\sigma}_\epsilon^2 + kT^{-1} \log T$, where $\hat{\sigma}_\epsilon{}^2$ is the standard error of the residuals, k is the number of parameters, and T is the sample size.

In the linear case, the minimum BIC value (-20.967) is obtained with an AR(8) model. Starting with this AR(8) model, we then used stepwise-regression methods to select additional nonlinear monomials from a GAR(2,3) process. The minimum BIC value (-20.995) was obtained when including the four monomials $y_{t-1}^2 y_{t-2}, y_{t-1} y_{t-3}^2, y_{t-1} y_{t-2} y_{t-3}^2$, and $y_{t-1}^2 y_{t-2} y_{t-3}^2$. The improvement of the fit over the linear case is fairly modest. The standard deviation of the in-sample residuals, reported in Table 1, is only 0.3% smaller for the GAR model.

5.2. Out-of-Sample Forecasting

We use the remaining 500 observations for an out-of-sample forecasting exercise comparing the AR, GAR and, as a naive benchmark, RW models. The latter is also of interest to financial economists, since it has some implications for market efficiency. One-step-ahead forecasts, with updating of the dependent variables, but no reestimation of the coefficients were computed.

Table 2 compares the three models on the basis of their mean-squared prediction error (MSPE). Of the three, the GAR produces the best point forecasts. The victory over the AR model is fairly modest, the MSPE is only 0.3% smaller. Both the linear and nonlinear models improve substantially over the RW though, with MSPE reductions of more than 300%.

While these are large improvements, the series is also highly leptokurtic, and much of the success may be attributed to just a few outliers. Mizrach (1991) proposes a framework for comparing MSPEs under weak populations assumptions. Consider two forecasts, $\tilde{y}_1$ and $\tilde{y}_2$, with respective forecast

Table 3: Seminonparametric Density Estimation[a]

p	K_n	K_h	AR Model k	AR Model BIC	GAR Model k	GAR Model BIC
8	0	0	10	-0.56674	14	-0.57885
8	2	0	12	-0.62230	16	-0.71839
8	2	1	36	-0.66595	40	-0.67731
8	2	2	143	-0.48164	147	-0.49354

[a] p: lag length in the conditional mean equation; K_n: degree of the polynomial allowing departures from normality; K_h: degree of the polynomial allowing heterogeneity; k: number of parameters.

errors, $e_1 \equiv y - \tilde{y}_1$ and $e_2 \equiv y - \tilde{y}_2$, drawn from some population (E_1, E_2). Letting MSPE$_i$ be the mean-squared prediction error of forecast i, i.e., MSPE$_i \equiv \frac{1}{n}\sum_{j=1}^{n} e_{ij}^2$, we test the null hypothesis H$_0$: MSPE$_1$ = MPSE$_2$. Defining $U = E_1 - E_2$ and $V = E_1 + E_2$, Mizrach (1991) shows that statistic

$$\frac{n^{-1}\sum_{j=1}^{n} u_j v_j}{\sqrt{n^{-1}\sum_{\ell=-\tau}^{\tau}(1 - \frac{|\ell|}{\tau+1})s_\ell}}, \tag{9}$$

with $s_\ell = \frac{1}{n}\sum_{j=1}^{n-|\ell|} u_j v_j u_{j+|\ell|} v_{j+|\ell|}$, is distributed asymptotically N(0,1) when the order of dependence is known to be ℓ.

Applying statistic (9), we find that the improvements over the RW for both the linear AR and GAR models are statistically significant at the 99.9% level. The improvement of the GAR over the linear AR model is not statistically significant.

5.3. Conditional Density Estimation

Using the AR(8) model for the conditional mean we considered SNP models with increasing degrees K_n and K_h. As Table 3 shows, there is strong evidence for non-normality and heterogeneity in the data. The linear model with the lowest BIC value (-0.66959) is an SNP(8,2,1) model. Using the subset GAR for the conditional mean, it appears that the GAR terms remove heterogeneity, but not non-normality. The lowest BIC value (-0.71839) was obtained for an SNP(8,2,0) model. As Table 1 indicates, the GAR model still leaves behind considerable excess kurtosis in the residuals.

6. Conclusions

Nonlinearities are omnipresent in time series modeling, especially, in economic time series. In many cases though, we can be quite ignorant about the functional form these nonlinearities take. One way to approach this problem, without completely abandoning parametric estimation, is to use a series estimator as in the GAR model.

We analyzed a high-frequency time series of financial asset returns that appears to be nonlinear. A GAR model captured these nonlinearities and improved upon a linear AR model regarding both in- and out-of-sample fit. The improvements were fairly modest though, and we cannot claim a statistically significant victory for the GAR forecasts.

A simple statistical analysis of the raw data reveals non-normality. Unconditionally, the return data are skewed and leptokurtic. The conditional densities are still leptokurtic. Higher-order polynomial transformations of the Gaussian density are required to model the data, regardless of whether a linear AR or a nonlinear GAR is used for the conditional mean.

The linear AR model leaves us with heterogenous residuals. This phenomenon is often modeled with ARCH or random-coefficient models. In our analysis we find evidence that the parameter variation may be due to misspecification of the conditional-mean equation. The inclusion of GAR terms removes the heterogeneity present in the data.

References

Elbadwi, I., A.R. Gallant, and G. Souza (1983) "An Elasticity Can Be Estimated Consistently without A Priori Knowledge of Functional Form," *Econometrica* 51, 1731-52.

Engle, R.F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation," *Econometrica* 50, 987-1007.

Gallant, A.R. and D.W. Nychka (1987) "Seminonparametric Maximum Likelihood Estimators," *Econometrica* 55, 363-90.

Gallant, A.R. and G. Tauchen (1989) "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica* 57, 1091-1120.

Granger, C.W.J. and A.P. Andersen (1978) *An Introduction to Bilinear Time Series Models*, Göttingen: Vandenhoeck and Ruprecht.

Kendall, M.G. and A. Stuart (1963) *The Advanced Theory of Statistics*, 2nd ed., London: Charles Griffin and Co.

Mayfield, S. and B. Mizrach (1991) "On Determining the Dimension of Real Time Stock Price Data," *Journal of Business Economics and Statistics*, forthcoming.

Mittnik, S. (1991a) "Analyzing Conditional Economic Dynamics with Nonlinear State Space Models," *Proceedings of the Business and Economic Statistics Section of the Annual Meeting of the American Statistical Association*, Alexandria: American Statistical Association, 61-66.

Mittnik, S. (1991b) "Nonlinear Time Series Analysis with Generalized Autoregressions: A State Space Approach," Working Paper, Department of Economics, SUNY-Stony Brook.

Mizrach, B. (1990) "Time Series Analysis of Ergodic Nonlinear Dynamical Systems," to appear in *Advances in Statistical Analysis and Statistical Computing*, R.S. Mariano ed.

Mizrach, B. (1991) "Forecast Comparison in L2," Working Paper, Department of Finance, The Wharton School, submitted to *Journal of Econometrics*.

Ozaki, M. (1981) "Nonlinear Threshold Autoregressive Models for Nonlinear Random Vibrations," *Journal of Applied Probability* 12, 443-51.

Phillips, P.C.B. (1983) "ERA's: A New Approach to Small Sample Theory," *Econometrica* 51, 1505-25.

Tong, H. (1983) *Threshold Models in Nonlinear Time Series Analysis*, New York: Springer-Verlag.

Tsay, R. (1987) "Conditionally Heteroscedastic Time Series Models," *Journal of the American Statistical Association* 82, 590-604.

Nonparametric Approaches to Generalized Linear Models

Wolfgang K. Härdle
Berwin A. Turlach
C.O.R.E. and Institut de Statistique
Université Catholique de Louvain
B-1348 Louvain-la-Neuve, Belgium

1. Introduction and Motivation

In this paper we consider classes of statistical models that are natural generalizations of *generalized linear models*. Generalized linear models cover a very broad class of classical statistical models including linear regression, ANOVA, logit, and probit models. An important element of generalized linear models is that they contain parametric components of which the influence has to be determined by the experimentator. Here we describe some lines of thought and research relaxing the parametric structure of these components.

In generalized linear models response variable and explanatory variables are related by predetermined functional forms, e.g., the logit model with a logistic link function and a linear form on the explanatory variables, see McCullagh and Nelder (1989). In this example the fixed parametric structures are the logistic distribution function and the (linear) form of the influence of the explanatory variables. Generalizing such a type of model means to abandon the form of either of these fixed components, i.e., the logistic (inverse) *link function* or the *linear predictor*. Generalizing the form of the link function means to allow for a flexible or parameter free form. Generalizing the form of the linear predictor means to allow for any unknown function of the explanatory variables.

Allowing for any functional form of influence for the predictor variables leads into well known dimensionality problems, commonly called the *curse of dimensionality* (Huber 1985). In order to avoid this curse of dimensionality Hastie and Tibshirani (1990) proposed to generalize the linear predictor by a sum of non-parametric univariate functions. This leads to so called *generalized additive models*. They contain generalized linear models as a special case when the link function is known and the univariate functions operating on the explanatory variables are linear.

Relaxing the form of the link function means to keep the linear predictor but to replace, in terms of our previous example, the logistic function by a non-parametric (preferably monotone) function. More generally several of these types of response models can be added, each using a different linear predictor and (non-parametric) link function. These models are known as projection pursuit regression (PPR) models due to an algorithm developped by Friedman and Stützle (1981).