{Preliminary – please do not quote}

IDENTIFICATION OF COEFFICIENTS IN A QUADRATIC MOVING AVERAGE PROCESS USING THE GENERALIZED METHOD OF MOMENTS

Richard A. Ashley
Department of Economics
Virginia Tech
Blacksburg, VA 24061
(540) 231 6220
ashleyr@vt.edu

Douglas M. Patterson
Department of Finance
Virginia Tech
Blacksburg, VA 24061
(540) 231 5737
amex@vt.edu

June 21, 2002

Abstract

The output of a causal, stable, time-invariant nonlinear filter can be approximately represented by the linear and quadratic terms of a finite parameter Volterra series expansion. We call this representation the "quadratic nonlinear MA model" since it is the logical extension of the usual linear MA process. Where the actual generating mechanism for the data is fairly smooth, this quadratic MA model should provide a better approximation to the true dynamics than the two-state threshold autoregression and Markov switching models usually considered.

As with linear MA processes, the nonlinear MA model coefficients can be estimated via least squares fitting, but it is essential to begin with a reasonably parsimonious model identification and non-arbitrary preliminary estimates for the parameters. In linear ARMA modeling these are derived from the sample correlogram and the sample partial correlogram, but these tools are confounded by nonlinearity in the generating mechanism. Here we obtain analytic expressions for the second and third order moments – the autocovariances and third order cumulants – of a quadratic MA process driven by i.i.d. symmetric innovations. These expressions allow us to identify the significant coefficients in the process by using GMM to obtain preliminary coefficient estimates and their concomitant estimated standard errors. The utility of the method for specifying nonlinear time series models is illustrated using artificially generated data.

1. Introduction

Nonlinear serial dependence in the mean of a time series is typically modeled within a "switching regressions" framework in which the observed realization at time t is taken to be generated by one of several (usually two) linear AR(p) processes.[1] Which process is operative may be determined by the value of a lagged value of the observed series – as in the TAR or STAR processes studied by Tong(1983), Chan and Tong (1986), and Teräsverta, et al. (1992, 1994) – or it may be stochastically determined, as in the Markov switching processes studied by Hamilton (1989) and later authors. These specifications are sufficiently flexible as to provide good sample fits to a number of economic and financial time series, but recent evidence in Ashley and Patterson (2000) illustrates how several such switching models for U.S. real output can fit the sample data fairly well, yet fail to capture the nonlinear serial dependence in this time series.

Here we propose a method for identifying the important coefficients in an alternative parametric framework for modeling nonlinear serial dependence in time series. This framework is based on the general Volterra expansion for a time series process. The usual linear MA(q) process can be viewed as the first term in this expansion; by retaining a sufficiently large number of terms in the expansion an extremely broad class of nonlinear time series processes can be accurately represented. Where a time series varies fairly smoothly in its nonlinear behavior – rather than shifting from one distinct linear process to another – the first few terms in this expansion will approximate the true generating mechanism far better than can even an elaborate

[1] We focus here on parametric modeling methods because we are more sanguine as to the feasibility of applying such methods to macroeconomic and financial data sets in which the sample length has been restricted so as to make reasonably credible the assumption of a stable relationship. Granger and Teräsvirta (1993, Chapter 7) review alternative – nonparametric and semiparametric – approaches; they also review several parametric approaches – e.g., bilinear models – which cannot be classified as "switching regressions."

switching regression model. Conversely, where a process actually does switch between several distinct regimes, only an impractically elaborate Volterra expansion will be adequate. In this sense, the framework proposed here can be viewed as a complement to existing methods based on the regime-switching paradigm.

The Volterra expansion is described in Section 2 below. There we use the second order expansion – what one might call a "quadratic moving average" model – to explicitly demonstrate how nonlinear serial dependence confounds the information in the usual autocorrelation function. In Section 3 we describe an identification procedure which uses the generalized method of moments (GMM) method to obtain preliminary estimates of the coefficients in the second order expansion. The utility of this procedure is illustrated in Section 4 with an application to artificially generated data and directions for further research in this area are outlined in Section 5.

2. The Volterra Expansion and the Quadratic Nonlinear MA Model

Sandberg (1983) shows that , where $\{x_t\}$ is strictly stationary with finite higher moments, it can be represented as the multi-order convolution of a set of causal, stable, time-invariant filters with i.i.d. noise:

$$
\begin{aligned}
x_t = \; & h_o + \sum_{n=1}^{\infty} h_1(n)\, e_{t-n} + \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} h_2(n,m)\, e_{t-n}\, e_{t-m-n} \\
& + \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} h_3(n,m,k)\, e_{t-n}\, e_{t-m-n}\, e_{t-k-m-n} + \dots + e_t
\end{aligned}
\qquad (1)
$$

where $e_t \sim$ i.i.d $(0, \sigma^2)$ and the functions $h_i(n, m, k, \dots)$ are called the Volterra kernels of the filter. Here we will restrict attention to approximations to this underlying process which are quadratic and have finite memory L:

$$
x_t = \alpha + \sum_{n=1}^{L} h(n)\, e_{t-n} + \sum_{n=0}^{L} \sum_{m=0}^{L} a(n,m)\, e_{t-n}\, e_{t-m-n} + e_t \qquad (2)
$$

Data generated by this process, denoted the “quadratic nonlinear MA model of order L” below, exhibits asymmetry whenever any of the a(n,0) coefficients are non-zero. While there is no guarantee that this moving average process is invertible in general, it *must* be invertible if the quadratic nonlinear MA model is an adequate approximation to the true generating mechanism. Thus, invertibility is solely a numerical issue here, and that only because the optimization algorithm used to estimate (2) will necessarily examine parameter values well away from the optimum; this issue is discussed in Section 3 below.

Where the underlying process is smoothly related to its own past – as opposed to shifting from one discrete state to another – the quadratic nonlinear MA model may often provide a good approximation to the actual process generating the data.

Data generated by this quadratic nonlinear MA process may be serially correlated or they may not. In fact, in this setting the autocovariance function for x_t is quite misleading because it confounds the linear and the nonlinear serial dependence in x_t. More explicitly, letting σ^2 denote the variance of the innovation e_t and letting μ_k denote its kth moment for $k > 2$, the autocovariance function for x_t generated by the quadratic nonlinear MA process is

$$\begin{aligned} c_{xx}(\tau) = E(x_t x_{t-\tau}) &= \sigma^2 \sum_{n=0}^{L} h(n)h(n+\tau) + \mu_3 \sum_{n=0}^{L} \left[h(n)a(n+\tau,0) + h(n+\tau)a(n,0)\right] \\ &\quad + \sigma^4 \sum_{n=0}^{L}\sum_{m=0}^{L} \left[1 - \delta(m)(\mu_4 \sigma^{-4} - 2)\right] a(n,m)a(n+\tau,m) \end{aligned} \tag{3}$$

where $\delta(m)$ is the usual kronecker delta function.

Thus, x_t might depend quite strongly on its own past – i.e., some or all of h(1) ... h(L) and a(1,1) ... a(L,L) are substantially non-zero – yet x_t might appear to be serially uncorrelated – i.e., $|c_{xx}(1)|$ ... $|c_{xx}(L)|$ are small. For example, x_t generated from

$$x_t = a(1,m)\, e_{t-1}\, e_{t-1-m} + a(2,m+1)\, e_{t-2}\, e_{t-2-m+1} + e_t \tag{4}$$

are serially uncorrelated but still serially dependent.

Alternatively, some or all of $|c_{xx}(1)|$ ... $|c_{xx}(L)|$ might be large – implying substantial linear serial dependence in x_t to the unwary – entirely because of nonlinear serial dependence associated with substantial a(m,n) coefficients. Thus, for example, x_t generated from

$$x_t = a(1,m)\, e_{t-1}\, e_{t-1-m} + a(2,m)\, e_{t-2}\, e_{t-2-m} + e_t \tag{5}$$

will have the same population correlogram as that of a (linear) MA(1) process.[2]

In general, we can conclude that nonlinear serial dependence – nonzero values for some a(m,n) in the present instance – can severely distort the shape of the population autocovariance function. This distortion greatly diminishes the usefulness of the correlogram as a tool for identifying the linear dependence {h(1) ... h(L)} in a time series. Moreover, since the probability limits of the estimated parameters in a linear model for prewhitening x_t are functions of the autocovariances, this result also implies that the coefficients in a linear pre-whitening model will

[2]Note that $x_t\, x_{t-1}$ equals $a(1,m)a(2,m)(e_{t-2})^2\,(e_{t-2-m})^2$ plus terms with expectation zero, whereas $E\{x_t\, x_{t-k}\}$ is zero for all k exceeding 1.

be inconsistently estimated. Thus, pre-whitening is valid only under a null hypothesis of (at most) linear serial dependence.

In principle the coefficients in the quadratic nonlinear MA model can be estimated via least squares fitting, but the number of coefficients to be estimated quickly becomes unwieldy as L rises; moreover, this model is quite nonlinear in the coefficients because the e_t's themselves depend on the coefficients.[3] Consequently, it is essential to initially obtain preliminary parameter estimates and a parsimonious representation of the series. A practical method for doing so is discussed in the next Section.

[3] Indeed, this property can make even the estimation of linear MA models troublesome.

3. Preliminary Estimates of the Quadratic Nonlinear MA Model Using GMM

Nowadays the generalized method of moments (GMM) is most commonly used to estimate parametric models based on orthogonality conditions which are either implied by theory or by least squares minimization itself. Our application of GMM here is more fundamental: we obtain consistent preliminary estimates of the $(L+1)^2 + 1$ coefficients in the quadratic nonlinear MA model of order L,[4]

$$x_t = \alpha + \sum_{n=1}^{L} h(n)\, e_{t-n} + \sum_{n=1}^{L}\sum_{m=0}^{L} a(n,m)\, e_{t-n}\, e_{t-m-n} + e_t \qquad e_t \sim \mathrm{NIID}[0,\sigma^2] \tag{6}$$

(and estimated standard errors for these estimates) by matching the non-zero first, second, and third order moments of x_t as accurately as possible to the analogous sample moments obtained from the data. The second moment of the innovations is also matched, so as to estimate σ^2.[5]

The discrepancies between these (L+1)(L+2) + 2 population and sample moments can be written as the vector $\bar{m}$ defined in Table 1. In this notation the parameter estimates are obtained as argmin$\{\bar{m}^t W^{-1} \bar{m}\}$ where W is an (L+1)(L+2) + 2 square weighting matrix.[6] Note that this

---

[4]There are $(L+1)^2 + 1$ coefficients to estimate if the martingale terms involving the contemporaneous innovation – i.e., the coefficients a(0,0) ... a(0,L) – are omitted from the model and taking into account the fact that σ^2 must be estimated. See Robinson (1977) for a detailed treatment of the issues involved in estimating equation 6 for the special case where only h(1) and the first order martingale {a(0,1)} terms are non-zero.

[5]The normality assumption on e_t can be replaced by an assumption that the innovations are i.i.d. and symmetrically distributed, in which case two more parameters (the fourth and sixth contemporaneous moments of the innovations) must be estimated and two more moments are matched.

[6]The last element of $\bar{m}$ involves the estimated innovations, $e_1 \ldots e_T$, calculated based on the sample data and the current set of parameter estimates. If the quadratic nonlinear MA model is a reasonable approximation to the

model is over-identified since the number of moment conditions exceeds the number of parameters.

As usual in implementing GMM, initial consistent (but relatively inefficient) parameter estimates are obtained by setting W equal to the identity matrix. These initial parameter estimates are then used to estimate the optimal weighting matrix, W_{GMM}, which is the asymptotic variance-covariance matrix of $\bar{m}$. Ordinarily this matrix can be obtained from (1/T) times the matrix of cross products of the T terms making up each component of $\bar{m}$, but that is not appropriate here since these T terms are not serially independent. Consequently, this asymptotic variance-covariance matrix is estimated by using equation six (and the parameter estimates) to generate 1000 T-samples (x_1 ... x_T), yielding 1000 realizations of the vector $\bar{m}$.[7]

An estimated t ratio for each coefficient estimate is then obtained using the usual GMM estimator of the asymptotic variance-covariance matrix of the parameter estimates, $[G^t (W_{GMM})^{-1}G]^{-1}$, where G contains the partial derivatives of each component of $\bar{m}$ with respect to each parameter.[8]

---

actual generating mechanism, then it can be expected to be invertible for parameter values close to the optimum, but the optimization algorithm examines parameter value combinations well away from the optimum. Where such combinations lead to non-invertible models, the e_t calculation blows up. In such cases the objective function is penalized in an amount proportional to $(e_{t-1})^2$ and the recursion is re-started, re-setting e_{t-1} ... e_{t-L} to zero.

[7]This is not as computationally burdensome as it sounds – it typically takes well under a minute.

[8]All of these gradients are obtained analytically except those of the last component of $\bar{m}$, which must be calculated numerically (using Ridder's method) since the $\hat{e}_t$ are only available numerically.

Table 1

Definition of $\bar{m}$ vector components

Component	Definition[9]
1	$E\{x_t\} - (1/T)\,\Sigma\, x_t$
2	$E\{x_t\, x_{t-1}\} - (1/T)\,\Sigma\, x_t\, x_{t-1}$
(etc.)	
(L+1)	$E\{x_t\, x_{t-L}\} - (1/T)\,\Sigma\, x_t\, x_{t-L}$
(L+1) + 1	$E\{x_t\, x_t\, x_t\} - (1/T)\,\Sigma\, x_t\, x_t\, x_t$
(L+1) + 2	$E\{x_t\, x_t\, x_{t-1}\} - (1/T)\,\Sigma\, x_t\, x_t\, x_{t-1}$
(etc.)	
(L+1) + (L+1)	$E\{x_t\, x_t\, x_{t-L}\} - (1/T)\,\Sigma\, x_t\, x_t\, x_{t-L}$
(L+1) + (L+1) + 1	$E\{x_t\, x_{t-1}\, x_{t-2}\} - (1/T)\,\Sigma\, x_t\, x_{t-1}\, x_{t-2}$
(L+1) + (L+1) + 2	$E\{x_t\, x_{t-1}\, x_{t-3}\} - (1/T)\,\Sigma\, x_t\, x_{t-1}\, x_{t-3}$
(etc.)	
(L+1) (L+2)	$E\{x_t\, x_{t-L}\, x_{t-2L}\} - (1/T)\,\Sigma\, x_t\, x_{t-L}\, x_{t-2L}$
(L+1) (L+2) + 1	$E\{(x_t)^2\} - (1/T)\,\Sigma\, (x_t)^2$
(L+1) (L+2) + 2	$E\{(e_t)^2\} - (1/T)\,\Sigma\, \hat{e}_t^2$

[9] All sums run from t = 1 to t = T; expectations are derived analytically from equation 6 (after a great deal of tedious algebra) based on the current estimates of α, h(1) ... h(L), a(1,1) ... a(L,L), and σ^2. The $\hat{e}_t$ are the innovations implied by the sample data, given these current parameter estimates and $\hat{e}_{t-i}$ set to zero for $i \in [0,2L-1]$.

4. An Illustrative Example Using Artificially Generated Data

To illustrate the usefulness of the method, T observations were generated from the quadratic nonlinear MA model:

$$x_t = 2.0 + .2e_{t-1} + .5e_{t-1}e_{t-2} + .3e_{t-1}e_{t-3} + e_t \qquad e_t \sim \mathrm{NIID}[0, 1.69] \qquad (7)$$

Applying the GMM method described above to obtain preliminary parameter estimates with T = 500 yields the estimates given in Table 2. At this sample length the algorithm correctly identifies all three terms {h(1), a(1,1), and a(1,2)} as statistically significant, although it does underestimate their values. No coefficients are falsely identified as significant, but note that a modest number of such "false positives" would be entirely inconsequential since this is only a preliminary identification step: such coefficients will be eliminated when the model is estimated using least squares fitting.

Failing to identify a coefficient that does belong in the model is more serious. Table 3 summarizes how the a(1,1) and a(1,2) parameter estimates (and their associated estimated t ratios) vary with T. The identification procedure appears to work reasonably well for $T \geq 400$ with this particular generating process. The complete set of estimates is given in Table 4 for the T = 250 case. Examining these, it is evident that one still would have identified the right model in this case – evidently (as one might expect) it is the GMM standard error estimates that are becoming unusable at T = 250, not the parameter estimates.

Table 2

Preliminary GMM Estimates – T = 500

$$x_t = 2.0 + .2e_{t-1} + .5e_{t-1}e_{t-2} + .3e_{t-1}e_{t-3} + e_t \qquad e_t \sim \mathrm{NIID}[0, 1.69]$$

	coef.	t
α	2.1272	12.65
h(1)	0.1564	2.53
h(2)	0.0508	0.73
h(3)	-0.0433	-0.56
h(4)	-0.0049	-0.08
a(1. 0)	0.0362	0.93
a(1. 1)	0.2430	2.56
a(1. 2)	0.1365	2.14
a(1. 3)	0.0060	0.11
a(1. 4)	0.0058	0.07
a(2. 0)	-0.0421	-1.09
a(2. 1)	0.0383	0.60
a(2. 2)	-0.0574	-0.97

	coef.	t
a(2. 3)	-0.0222	-0.27
a(2. 4)	0.0406	0.52
a(3. 0)	0.0029	0.08
a(3. 1)	-0.0122	-0.22
a(3. 2)	0.0017	0.02
a(3. 3)	-0.0257	-0.36
a(3. 4)	0.0246	0.34
a(4. 0)	-0.0165	-0.44
a(4. 1)	-0.0192	-0.20
a(4. 2)	0.0961	1.10
a(4. 3)	0.0221	0.29
a(4. 4)	0.0828	1.13

Table 3

Preliminary GMM Estimates for a(1, 1) and a(1, 2) for Various Values of T

$$x_t = 2.0 + .2e_{t-1} + .5e_{t-1}e_{t-2} + .3e_{t-1}e_{t-3} + e_t \qquad e_t \sim \mathrm{NIID}[0, 1.69]$$

T	a(1, 1)	$t_{1.1}$	a(1, 2)	$t_{1,2}$
250	.440	1.24	.198	0.85
400	.251	1.96	.175	1.97
500	.243	2.56	.136	2.14
1000	.307	3.91	.183	3.27

Table 4

Preliminary GMM Estimates – T = 250

$$x_t = 2.0 + .2e_{t-1} + .5e_{t-1}e_{t-2} + .3e_{t-1}e_{t-3} + e_t \qquad e_t \sim \mathrm{NIID}[0, 1.69]$$

	coef.	t
α	1.9554	8.05
h(1)	0.1530	1.02
h(2)	-0.1234	-0.78
h(3)	-0.0582	-0.34
h(4)	0.0967	0.79
a(1. 0)	-.0304	-0.25
a(1. 1)	0.4396	1.24
a(1. 2)	0.1978	0.85
a(1. 3)	-0.1217	-0.90
a(1. 4)	-0.0482	-0.31
a(2. 0)	-0.0026	-0.03
a(2. 1)	-0.0304	-0.25
a(2. 2)	-0.0288	-0.19

	coef.	t
a(2. 3)	0.1186	0.57
a(2. 4)	0.0020	0.01
a(3. 0)	-0.0418	-0.41
a(3. 1)	-0.0435	-0.35
a(3. 2)	0.0232	0.14
a(3. 3)	0.0354	0.25
a(3. 4)	0.0243	0.15
a(4. 0)	0.0513	0.58
a(4. 1)	-0.0358	-0.22
a(4. 2)	-0.0280	-0.16
a(4. 3)	0.0588	0.37
a(4. 4)	0.0054	0.03

5. Future Directions

Currently we are working on applying the method to identify quadratic nonlinear MA models for real data – e.g., the change in the monthly index of industrial production and daily common stock returns. We will examine the postsample forecasting performance of models for these series which have been identified using our framework and fit (with parameter updating) using nonlinear least squares fitting.[10]

In future work we plan to relax the assumption of gaussian innovations made above. The analytic calculation of the population moments and gradients only requires that the innovations be symmetrically (and serially independently) distributed, with given fourth and sixth moments. Assuming gaussianity thus corresponds to setting the fourth moment to $3\sigma^2$ and the sixth moment to $15\sigma^2$. These two moments can be instead be estimated along with the other parameters, but it seems likely that rather large samples will be needed for reliable estimation of a sixth moment. Consequently – and since the principal interest in considering non-gaussian innovations is to incorporate leptokurtosis – our plan is to parameterize the innovations as Student's t with k degrees of freedom and estimate k. Additional extensions which we are working on include exploring the possibility of going on the cubic nonlinear MA model and extending the framework to the identification of multivariate nonlinear MA models.

[10] We are also exploring the impact on forecasting effectiveness of including the martingale terms in the preliminary identification and perhaps in the model estimation steps.

References

Ashley, R. and D.M. Patterson (2001) "Nonlinear Model Specification/Diagnostics: Insights from a Battery of Nonlinearity Tests," Economics Department Working Paper #E99-05.

Chan, K.S. and H. Tong (1986) "On Estimating Thresholds in Autoregressive Models," *Journal of Time Series Analysis 7*, 178-190.

Granger, C.W.J. and T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships* Oxford University Press: Oxford.

Hamilton, J. (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica 57*, 357-384.

Patterson, D.M. and R. Ashley (2000) *A Nonlinear Time Series Workshop,* Kluwer:Norwell.

Robinson, P.M. (1977) "The Estimation of a Nonlinear Moving Average Model," *Stochastic Processes and their Applications 5*, 81-90.

Sandberg, I. W. (1983) "Expansions for Discrete-Time Nonlinear Systems," *Circuits, Systems, and Signal Processing 3(2)*, 180-192.

Teräsverta, T. and H. Anderson (1992) "Characterising Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models," *Journal of Applied Econometrics 7*, 119-36.

Teräsverta, T. (1994) "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association 89*, 208-218.

Tong, H. (1983) "Threshold Models in Non-linear Time Series Analysis," *Lecture Notes in Statistics, No. 21,* Springer: Heidelberg.