A simple dynamic model for limited dependent variables

Frank Gerhard

Nuffield College, Oxford OX1 1NF, UK
Center of Finance and Econometrics, University of Konstanz, Germany

frank.gerhard@nuffield.oxford.ac.uk

July 16, 2001

Abstract

A dynamic model for limited dependent variables is proposed, which estimation does not rely on simulation methods. A latent conditional mean function which is measurable with respect to past and observable information circumvents the solution of a T-dimensional integral and yields a simple and computationally parsimonious maximum likelihood estimation.

It can be shown that the latent process implied by the limited dependent autoregressive moving average model is covariance stationary. Parameter estimates of this model are shown to be consistent but inefficient estimates of the parameters of a standard latent autoregressive moving average model, for which a maximum likelihood estimator is computationally burdensome. Monte Carlo evidence is provided to assess parameter estimates based on the limited dependent ARMA given the data generation process is a standard latent ARMA. The results indicate that the asymptotic properties hold quite nicely in small samples. An application based on IBM transaction price changes from the NASDAQ demonstrates a potential use of the model suggested here.

Keywords: limited dependent variables, quantal response models, latent dynamic, ARMA process, generalised error

1 Introduction

In this article a limited dependent autoregressive moving average (LD-ARMA) process is introduced and shown to be flexible, extendible, and computationally inexpensive to estimate. Further, the likelihood function for a LD-ARMA model provides a valid quasi-likelihood for the estimation of limited dependent variable models where the latent process is an ARMA process.

We construct the limited dependent variable y_t as

$$y_t = g(m_t + e_t), \qquad \text{where } e_t \sim NID(0,1), \qquad t = 1, \ldots, T, \tag{1}$$

where m_t is the conditional mean of a latent process and where $g(\cdot)$ only needs to be a Borel measurable function. This class includes Probits, as well as ordered Probits or Tobit type models, yet, it is not limited to these classical limited dependent variable models but could be extended to a wider class of models. For a broad range of applications in economics and statistics along with different observation rules, see e.g. Maddala (1983), Cox and Snell (1989), McCullagh and Nelder (1989). Denote the information set generated by the observable limited dependent variables y_t up to time t by $\mathcal{F}_t^y = \sigma(y_t, y_{t-1}, \ldots, y_1)$ and the information generated by the errors e_t up to t by $\mathcal{F}_t^e = \sigma(e_t, e_{t-1}, \ldots, e_1)$. The two information sets will only coincide if the observation rule $g(\cdot)$ is a one-to-one function, or more formally a Borel measurable isomorphism (e.g. Davidson (1994, theorem 10.3)). In typical applications however, especially in the limited dependent case, we have that $\mathcal{F}_t^y \subset \mathcal{F}_t^e$.

The simplest illustration of this type of process is the LD-AR(1) model where the conditional mean m_t is built up recursively[1], conditioning on some initial m_0

$$m_t = \phi(m_{t-1} + c_{t-1}), \tag{2}$$

where c_t is the conditional expectation of e_t given the observable information $\mathcal{F}_t^y$ at time t, i.e.

$$c_t = \mathrm{E}\left[e_t \middle| \mathcal{F}_t^y\right] = \mathrm{E}\left[e_t \middle| m_t, y_t\right]. \tag{3}$$

[1]Note that in a standard observable AR(1) model $y_{t-1}^* = \phi y_{t-2}^* + e_{t-1} = m_{t-1} + e_{t-1}$ and therefore $y_t^* = \phi(m_{t-1} + e_{t-1}) + e_t$.

The important feature of the mean function m_t of the latent process is that it is measurable with respect to the observable information $\mathcal{F}_{t-1}^{y}$ available by construction. Thereby, the process y_t could be seen as essentially observation driven in the sense of Cox (1981). Note that the conditional expectation c_t relates to a concept known in econometrics as generalised residuals, see Gourieroux, Monfort, Renault, and Trognon (1987), or in statistics as Bayesian residuals, see Albert and Chib (1995).

Due to the measurability of the mean function m_t with respect to past observable information $\mathcal{F}_{t-1}^{y}$, the maximum likelihood estimation of the parameter ϕ is computationally inexpensive as the likelihood function of the LD-ARMA process can be directly computed using the predictive decomposition (e.g. Harvey (1990, ch. 3.5)), without recourse to simulation. Clearly, in the context of the LD-ARMA model, the conditional distribution of y_t given past observations' information $\mathcal{F}_{t-1}^{y}$ is available. In addition to the very simple case of a LD-AR(1) model just outlined, it is shown in the paper that an extension to include higher order AR terms and MA terms is easily achieved. The inclusion of exogenous regressors in the dynamic specification raises no particular problems as well as the presence of additional model parameters in the observation rule $g(\cdot)$, as it is the case in an ordered Probit with estimated thresholds. Even the inclusion of regressors in the observation rule $g(\cdot)$ is possible as long as it retains its property of Borel measurability given all available information up to t. Furthermore, it is shown that the latent process of an LD-ARMA dynamic is covariance stationary and that the autocorrelation function of the latent process is identical to the autocorrelation function of the corresponding latent ARMA process.

Apart from y_t being a dynamic process for limited dependent variables in its own right, the likelihood for ϕ turns out to be a valid quasi-likelihood in the sense of White (1982) for the parameter ρ in the more complicated process z_t, which is constructed as

$$z_t = g(\mu_t + \epsilon_t), \qquad \text{where } \epsilon_t \sim NID(0,1), \qquad t = 1, \ldots, T, \tag{4}$$

where μ_t is the conditional mean of a latent ARMA process, e.g. a latent AR(1), to match the LD-AR(1) process outlined in (1)-(2), i.e.

$$\mu_t = \rho(\mu_{t-1} + \epsilon_{t-1}), \tag{5}$$

conditioning on the initial μ_0. This latent specification yields an essentially parameter driven, or state space, model (e.g. Cox (1981) or Harvey (1989)) and has the considerable inferential drawback that the conditional mean μ_t is not measurable with respect to the available information set $\mathcal{F}^z_{t-1}$ but only with respect to the unobservable information $\mathcal{F}^\epsilon_{t-1}$. As a consequence the prediction decomposition which is readily available for one-to-one functions $g(\cdot)$ and the LD-ARMA process becomes unfeasible, since the conditional distribution of z_t given past observations' information $\mathcal{F}^z_{t-1}$ is not easily available.

Although the maximum likelihood estimator of ρ can still be formulated, its computation involves the solution of T-fold integrals, therefore there is a long tradition in statistics and econometrics of directly using sampling moments of the z_t process for inference, see Lomnicki and Zaremba (1955), Kedem (1980), Keenan (1982), Gourieroux, Monfort, Renault, and Trognon (1987), and Poirier and Ruud (1988). The wide availability of fast computing resources favored however the use of simulation methods, especially Markov Chain Monte Carlo, to overcome the inherent inferential hurdle of parameter driven dynamic models, see the general framework proposed in Chib and Greenberg (1998) and Manrique and Shephard (1998) for an emphasis on time series applications and further literature given there. Yet, all of the simulation approaches involve a considerable computational overhead. Other alternatives to the LD-ARMA model include the observation driven models by Cox and Snell (1989, chap. 2.11) and Zeger and Qaqish (1988) as well as the mixture approach suggested by Jacobs and Lewis (1978a) and Jacobs and Lewis (1978b).

The main advantage of the model proposed here is the close link it provides between observation driven and parameter driven models of limited dependent variables. First of all, the LD-AR(1) process is identical to a latent AR(1) process, if $g(\cdot)$ is one-to-one and thereby $\mathcal{F}^y_t = \mathcal{F}^e_t$. Hence in this situation model (1)-(2) would be equivalent to (4)-(5). Second, also in the general case, where $\mathcal{F}^y_t \subset \mathcal{F}^e_t$, it is shown that the autocorrelation function of $(m_t + e_t)$ and of the correspondingly specified model $(\mu_t + \epsilon_t)$ are indeed equal. Third, the unconditional variance of $(m_t + e_t)$ is bounded from above by the well-known variance of $(\mu_t + \epsilon_t)$

To assess the use of the quasi-likelihood implied by the LD-ARMA process for the latent ARMA process in practice, Monte Carlo evidence on the small sample properties of the estimator is provided. It indicates that the asymptotic properties hold quite nicely. We regard this as the most important result reported in the paper, given a computationally simple method for estimating the LD-ARMA models.

The outline of the rest of this paper is as follows. In the second section the LD-ARMA specification is introduced in the context of a Probit LD-AR(1) model, subsequently alternative observation rules, the extension of the model to the LD-ARMA(p,q) case, and the inclusion of exogenous variables are given here as well. In the third section the LD-ARMA model is characterized in particular with respect to its implied autocorrelation function. Furthermore, it is shown that the maximum likelihood parameter estimates obtained for a LD-ARMA model are indeed consistent, yet inefficient, estimates of the parameters of a latent ARMA model. A Monte Carlo study of a LD-AR(1) and a LD-MA(1) model completes the comparison. The fourth section gives a small illustration of the estimator in practice using a sample of IBM trading at the NASDAQ. The fifth section concludes.

2 Model specification

2.1 Estimation of a Probit-AR(1) model

The main advantage of a LD-ARMA dynamic given by (1)-(2) over the standard latent ARMA model given by (4)-(5) is the computationally cheap maximum likelihood estimator. Here, the maximum likelihood estimation of a very simple example of an LD-ARMA model is outlined based on the

Example 1 (Probit observation rule) *Assume the setting of Lomnicki and Zaremba (1955), i.e. a Probit observation rule $g_P(u)$ which is independent of additional parameters and defined by*

$$g_P(u) = \begin{cases} 0, & \textit{if } u < 0, \\ 1, & \textit{if } u \geq 0, \end{cases} \qquad u \in \mathbb{R} \tag{6}$$

The complementary relationship $G_P(v)$, which yields the information, i.e. a particular interval, available on the latent process by observation of the binary variable y_t is given *by*

$$G_P(v) = \begin{cases} (-\infty, 0), & \textit{if } v = 0, \\ \\ [0, \infty), & \textit{if } v = 1, \end{cases} \qquad v \in \{0, 1\}. \tag{7}$$

We choose the simple LD-AR(1) process given by (1) and (2) under the observation rule $g_P(\cdot)$ to outline the evaluation of the computational simple maximum likelihood estimator. The great advantage of the LD-ARMA specification is that the conditional expectation of the latent variable m_t is measurable with respect to the information available up to time $t-1$, $\mathcal{F}^y_{t-1}$, and thus allows to rely on a prediction error decomposition of the likelihood. The evaluation of the likelihood follows the following recursive scheme:

1. The conditional expectation of the latent variable given no available past information is assumed to equal the unconditional expectation

$$m_0 := \mathrm{E}\left[m_t\right] = 0. \tag{8}$$

2. The likelihood contribution of observation t given the probit observation rule, the Gaussian assumption on the error term and most important, the measurable mean function is

$$\mathrm{Prob}\left[y_t \left| \mathcal{F}^y_{t-1}\right.\right] = \begin{cases} \Phi(-m_t), & \text{if } y_t = 0, \\ \\ 1 - \Phi(-m_t), & \text{if } y_t = 1. \end{cases} \tag{9}$$

3. The generalized error c_t, which makes up the mean function is a (conditionally) deterministic function of the observations, concisely, of the $\mathcal{F}^y_{t-1}$ measurable mean function m_t and the current observation y_t

$$c_t = \begin{cases} \frac{-\phi(-m_t)}{\Phi(-m_t)}, & \text{if } y_t = 0, \\ \frac{\phi(-m_t)}{1-\Phi(-m_t)}, & \text{if } y_t = 1. \end{cases} \tag{10}$$

See the original paper by Gourieroux, Monfort, Renault, and Trognon (1987) for an extended discussion of generalized errors in the context of non-dynamic models.

4. Calculation of the conditional expectation of the future latent variable given the present information,

$$m_{t+1} = \phi(m_t + c_t). \tag{11}$$

5. Steps 2 through 4 are repeated for all y_t, $t = 1, \ldots, T$.

6. The likelihood $\mathfrak{L}_y$ of the observable model can be directly evaluated, using $\bar{y}_t$ which contains all observations of y_t up to t, as

$$\begin{aligned}
\mathfrak{L}_y(\bar{y}_T|\phi) &= \int_{G_P(y_1)} \int_{G_P(y_2)} \cdots \int_{G_P(y_T)} f(u_1, u_2, \ldots, u_T) du_1\, du_2 \ldots du_T \\
&= \int_{G_P(y_1)} \int_{G_P(y_2)} \cdots \int_{G_P(y_T)} f(u_1) f(u_2|\mathcal{F}_1^y) \ldots f(u_T|\mathcal{F}_{T-1}^y) du_1\, du_2 \ldots du_T \\
&= \prod_{t=1}^{T} \mathrm{Prob}\left[y_t = 1 \left|\mathcal{F}_{t-1}^y\right.\right]^{y_t} \mathrm{Prob}\left[y_t = 0 \left|\mathcal{F}_{t-1}^y\right.\right]^{(1-y_t)}.
\end{aligned} \tag{12}$$

Thus by the use of the LD-ARMA process (1) and (2) and the implied likelihood $\mathfrak{L}_y$, the quite cumbersome likelihood implied by the parameter driven model (4) and (5) can be circumvented.

2.2 An observation rule with parameters

Unlike in models which rely on the EM algorithm or on simulation methods for estimation, the introduction of parameters in the observation rule does not raise any additional problems for model specification. See Ruud (1991) for a discussion of ordered probits in the EM context.

To demonstrate this, the dynamic model is extended to the case of ordered probits.

Example 2 (Ordered probit observation rule) *The observation rule $g_{OP}(u, \gamma)$ is defined using parameters γ. The latent process is mapped through a threshold function into the observable, discrete variable y_t*

$$g_{OP}(u, \mu) = \begin{cases} v_1, & \text{if} \quad u \in (-\infty; \gamma_1), \\ v_2, & \text{if} \quad u \in [\gamma_1; \gamma_2), \\ \vdots & \\ v_J, & \text{if} \quad u \in (\gamma_{J-1}; \infty), \end{cases} \tag{13}$$

$$-\infty < \gamma_1 < \gamma_2 < \ldots < \gamma_{J-1} < \infty.$$

where the variable v_j contains the distinct values y_t can take on, with $v_1 \prec v_2 \prec \ldots \prec v_J$, i.e. the different values of the dependent variable need to be ordered but not necessarily observed on a metric scale.

The form of $G_{OP}(u, \gamma)$ is just the straightforward extension of the binary case in example 1 to the present model.

Note that in this setting neither the level of the latent variable nor the scale of the latent variable are identified. The generalised error c_t follows readily by an evaluation of the conditional expectation in (3). In the context of the modified observation rule the generalized error c_t given in (10) is extended to the case of multiple categories as

$$c_t = \begin{cases} \frac{-\phi(\nu_{t,1})}{\Phi(\nu_{t,1})}, & \text{if } y_t = v_1, \\ \frac{\phi(\nu_{t,j-1})-\phi(\nu_{t,j})}{\Phi(\nu_{t,j})-\Phi(\nu_{t,j-1})}, & \text{if } y_t \in \{v_2, \ldots, v_{J-1}\}, \\ \frac{\phi(\nu_{t,J-1})}{1-\Phi(\nu_{t,J-1})}, & \text{if } y_t = v_J, \end{cases} \tag{14}$$

$$\text{with } \nu_{t,j} := \gamma_j - m_t.$$

The likelihood function of this model has the well-known form of an ordered probit with at least weakly exogenous regressors, see e.g. Maddala (1983), and is a simple extension of the probit likelihood $\mathfrak{L}_y$ given by (12). The likelihood contributions are a function of the generalized errors c_t through the conditional expectation m_t of the latent variable as in (9)

$$\text{Prob}\left[y_t \,\middle|\, \mathcal{F}^y_{t-1}\right] = \begin{cases} \Phi(\gamma_1 - m_t), & \text{if } y_t = v_1, \\ \Phi(\gamma_i - m_t) - \Phi(\gamma_{i-1} - m_t), & \text{if } y_t \in \{v_2, \ldots, v_{J-1}\}, \\ 1 - \Phi(\gamma_{J-1} - m_t), & \text{if } y_t = v_J. \end{cases} \tag{15}$$

This makes it clear that the use of the LD-ARMA dynamic suggested in (1) and (2) has a wide range of possible applications. The key feature necessary in a particular application is the evaluation of the generalized error c_t given by (3). This is however a straightforward task as long as the observation rule is conditionally deterministic given observations up to time t. It would even be possible to include regressors in the formulation of the thresholds.

2.3 Higher order dynamics

The extension to include AR(p), for $p > 1$, and MA(q) terms in the dynamic specification is described most easily in the context of an ARMA model which is cast in state space form. In order to do so coefficient matrices F, H, and the dimension of the state space r are defined as

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{r-1} & \phi_r \\ 1 & 0 & \dots & & 0 \\ 0 & 1 & 0 & \dots & \vdots \\ \vdots & & \ddots & & 0 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}, \quad H' = \begin{bmatrix} 1 & \theta_1 & \dots & \theta_r \end{bmatrix}, \quad r = \max(p, q+1), \tag{16}$$

where we have for the AR parameters $\phi_i = 0$ for $i > p$ and for the MA parameters likewise $\theta_i = 0$ for $i > q$. See e.g. Hamilton (1994, chap. 13.1). The conditional mean m_t of the latent process is just

$$m_t = H' s_t, \tag{17}$$

where an additional state process s_t is introduced. The conditional mean of the latent state s_t is defined by the recursion

$$s_t = F(s_{t-1} + u_1 c_{t-1}), \qquad \text{where } u_1' = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}, \tag{18}$$

while conditioning on some initial s_0. Thereby, an LD-ARMA(p,q) model is defined by (1) and (16)-(18). Thus, the maximum likelihood estimation of the ARMA(p,q) model proceeds almost exactly along the same lines as in the AR(1) case described initially.

2.4 The inclusion of exogenous variables

There are two ways to include exogenous regressors in the dynamic model. Using again the context of the state space model, explanatory variables w_t can be included in the mean equation of the latent process (17) to obtain

$$m_t = H' s_t + w_t' \delta, \tag{19}$$

This is equivalent to a limited dependent variables model with exogenous regressors w_t including ARMA(p,q) errors. Alternatively, regressors x_t can be included in the dynamic specification and thereby obtaining an infinite distributed lags model, see e.g. Hendry (1995). A modification of the update of the states' conditional mean in (18) yields thus

$$s_t = F(s_{t-1} + u_1 c_{t-1}) + u_1 \beta' x_t, \tag{20}$$

using regressors x_t with coefficients β. An extension of the observable information set is however necessary to defined either $\mathcal{F}_t^{y,w} = \sigma(y_t, w_{t+1}, \ldots, y_1, w_2, w_1)$ or $\mathcal{F}_t^{y,x} = \sigma(y_t, x_{t+1}, \ldots, y_1, x_2, x_1)$. The definition of the innovation term c_t in (3) is adjusted correspondingly. This makes clear that the latent state is decomposed into a weighted sum of all the past x_t and the MA already known from the ARMA(p,q) model without regressors. This flexibility is sometimes needed. A typical candidate for the inclusion as a regressor with an infinite lag structure is the observed volume per transaction in the context of an empirical market microstructure analysis. Other variables however are rendered virtually uninterpretable by a dynamic inclusion, e.g. regressors capturing a seasonality.

3 A characterization of the LD-ARMA process

3.1 Dynamic properties of the LD-ARMA process

The LD-ARMA process benefits from its close relationship to a corresponding latent ARMA process which is observed through an observation rule $g(\cdot)$. It turns out that the LD-ARMA process can be considered as a filter for data generated by the latent ARMA process. The derivation of its dynamic properties profits greatly from the fact that the parameter space for which the latent ARMA process is covariance stationary is actually well established.

We can relate the latent ACF of the LD-ARMA model to the corresponding ACF of the latent ARMA model by the following proposition:

Proposition 1 *For the autocovariance and the ACF at lag s, $\rho^*(s)$, $s > 0$, of the latent process $m_t + e_t$ implied by the LD-ARMA process for the observable process y_t defined by*

(1), (16)-(18) *and for the autocovariance and the ACF,* $\rho^{\circ}(s)$*,* $s > 0$*, of the latent ARMA process* $\mu_t + \epsilon_t$*, where*

$$\mu_t = H'\zeta_t, \qquad\qquad \zeta_t = F(\zeta_{t-1} + \epsilon_{t-1}), \tag{21}$$

we have that

1. $\mathrm{Cov}\left[m_t + e_t, m_{t-s} + e_{t-s}\right] \leq \mathrm{Cov}\left[\mu_t + \epsilon_t, \mu_{t-s} + \epsilon_{t-s}\right]$*,* $s \geq 0$*, and*

2. $\rho^*(s) = \rho^{\circ}(s)$*,* $s > 0$*,*

Proof: See appendix. □

This quite useful result characterizes the LD-ARMA model as having basically the same dynamic properties as the latent ARMA but has a lower unconditional variance of the latent process than the original latent ARMA. Based on the boundedness of the autocovariance from above of the LD-ARMA process, we can give the following proposition, which establishes the conditions for covariance stationarity of the LD-ARMA process

Proposition 2 *The latent process* $m_t + e_t$ *implied by the LD-ARMA process for the observable process* y_t *defined by* (1), (16)-(18) *is covariance stationary if the eigenvalues of* F *lie inside the unit circle.*

Proof: Follows directly from the proof of proposition 1. □

These two propositions given here establish the close relationship to latent ARMA models, especially since the sufficient conditions for a covariance stationary process match the usual assumptions in the VAR context, see e.g. Lütkepohl (1991, chap. 2.1).

3.2 The observable autocorrelation function

The term observable autocorrelation function refers to the ACF of the observable limited dependent variable y_t. Note that the two propositions given above do not involve the ACF of the observable process y_t, which might not even be defined as in the case of categorial observations, which are not measured on a metric scale. If however, the observations permit the sensible evaluation of an ACF, the following proposition establishes that the properties of the latent process carry over to the observable process.

Proposition 3 *For an observation rule $g(\cdot)$, which is a Borel measurable function and has values on a metric scale, so that the ACF of the process y_t is defined and denoted by $\rho(s)$, $s > 0$, and a latent process $(m_t + e_t)$ which is covariance stationary and has an ACF $\rho^*(s)$, then the observable process y_t is*

1. *covariance stationary, and*

2. *$|\rho(s)| \leq |\rho^*(s)|$, for all $s > 0$.*

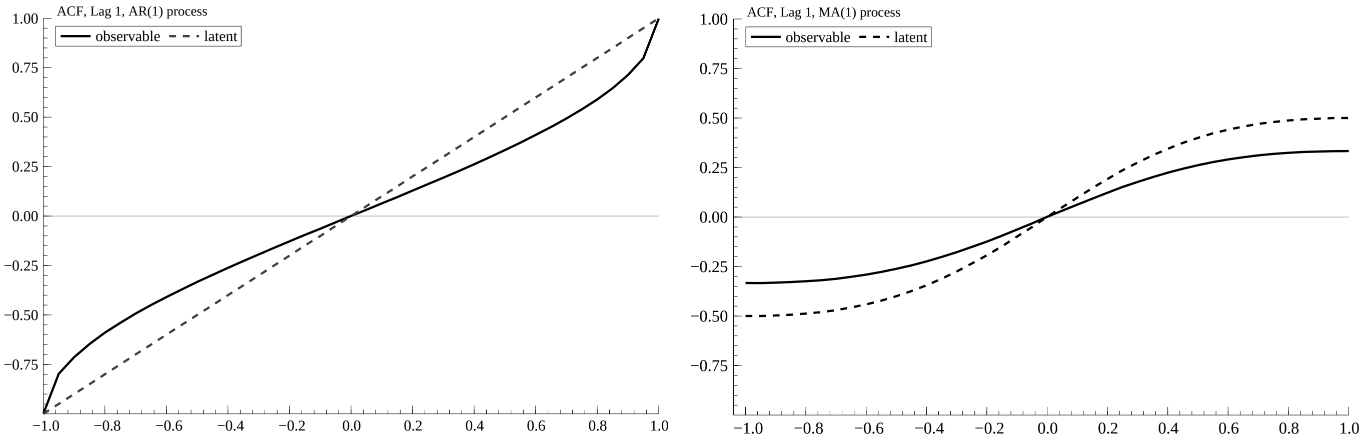
Proof: Follows directly from Granger and Newbold (1976, sec. 2) and Stone (1927, Lemma IV). □

Additionally, one example is considered where the observable ACF is defined and its relationship to the latent ACF is outlined to illustrate the scope of proposition 3. For the simple, yet fundamental, case of the Probit observation rule in example 1, the relationship between the latent ACF, $\rho^*(s)$, and the observable ACF, $\rho(s)$, is well-known, see Lomnicki and Zaremba (1955). From the properties of the bivariate Gaussian distribution the functional relationship between the observable ACF and the latent ACF is readily derived as

$$\rho(s) = \frac{2}{\pi} \arcsin \rho^*(s). \tag{22}$$

To illustrate this relationship and for future reference in the context of the Monte Carlo study, which is based on the Probit observation rule, figure 1 gives the ACF of the observable and the latent variable implied by a MA(1) and an AR(1) process at the first lag only for parameters $\phi \in [-1; 1]$ and $\theta \in [-1; 1]$. Note that the latent ACFs of the Probit-AR(1) and Probit-MA(1) are identical to their latent ARMA counterparts by virtue of proposition 1. The difference between both models is quite obvious. While the

Figure 1: **ACF of latent and observable process in a dynamic Probit** for a latent AR(1) and MA(1).



effect the Probit observation rule has on the ACF is quite limited in the AR(1) case, the effect on the ACF of the MA(1) process is considerable. While the ACF of the latent MA(1) becomes less steep for large parameters in absolute value, the ACF of the observable process is virtually flat for $|\theta| > 0.5$. This renders latent MA(1) processes with such parameters almost observationally equivalent.

3.3 The LD-ARMA process as an auxiliary model

The relationship between the latent autocorrelation functions is described by propositions 1 and 3. Here, the use of the likelihood implied by the LD-ARMA process as a valid quasi-likelihood for the latent ARMA process is examined.

We consider the scores of both models with respect to the parameter of the latent dynamic ρ in a simple AR(1) and LD-AR(1) model in the context of a Probit observation rule. From the structure of the arguments it will be obvious that the limitation to the latter model eases the exposition, yet does not limit the validity of the approach for a more general model. We define the unobservable process $z_t^* = \mu_t + \epsilon_t$ and denote its likelihood by $\mathfrak{L}_z^*(\bar{z}_T^*|\rho)$, where $\bar{z}_t^*$ collects all observations of the dependent variable z_t^* up to time t. The score of the unobservable latent AR(1) model is just

$$\frac{\partial \log \mathfrak{L}_z^*(\bar{z}_T^*|\rho)}{\partial \rho} = -\sum_{t=1}^{T} \epsilon_t \frac{\partial \mu_t}{\partial \rho}, \tag{23}$$

$$= -\sum_{t=1}^{T} \sum_{i=0}^{t-1} (i+1)\rho^i \epsilon_{t-1-i}\epsilon_t. \tag{24}$$

The score of the observable model for z_t follows from a straightforward application of the EM algorithm of Dempster, Laird, and Rubin (1977) and is given by the conditional expectation of the latent score given the observable information $\mathcal{F}_T^z$

$$\frac{\partial \log \mathfrak{L}_z(\bar{z}_T|\rho)}{\partial \rho} = \mathrm{E}\left[\left. \frac{\partial \log \mathfrak{L}_z^*(\bar{z}_T^*|\rho)}{\partial \rho} \right| \mathcal{F}_T^z \right] \tag{25}$$

$$= -\sum_{t=1}^{T} \sum_{i=0}^{t-1} (i+1)\rho^i \mathrm{E}\left[\epsilon_{t-1-i} | \mathcal{F}_T^z \right] \mathrm{E}\left[\epsilon_t | \mathcal{F}_T^z \right] \tag{26}$$

At this point it is obvious, that the use of the EM algorithm of Dempster, Laird, and Rubin (1977) does not solve the inferential problem of this parameter driven model. The evaluation of $\mathrm{E}\left[\epsilon_t | \mathcal{F}_T^z \right]$ is computationally just as involved as the direct maximization of the likelihood of the latent ARMA model, since it involves the computation of T-fold integrals as well. See also the discussion of the EM algorithm and extensions in Ruud (1991).

The score of the LD-ARMA model on the other hand, has a much simpler structure. It is directly derived from the likelihood $\mathfrak{L}_y$ in (12) to obtain

$$\frac{\partial \log \mathfrak{L}_y(\bar{y}_T|\phi)}{\partial \phi} = -\sum_{t=1}^{T} \sum_{i=0}^{t-1} (i+1)\phi^i c_{t-1-i} c_t, . \tag{27}$$

Thus, in the LD-ARMA model the identification condition for the parameters of the latent dynamic boils down to the uncorrelatedness of generalised errors c_t. Note that this result is independent from the particular form of the observation rule $g(u)$, given the usual regularity conditions. Here, the conditional expectation of the latent error c_t can be thought of as an extension of the generalised errors introduced by Gourieroux, Monfort, Renault, and Trognon (1987) to dynamic models.

To actually compare the latent ARMA and the LD-ARMA model, assume that the data generating process (DGP) z_t is of the latent AR(1) form with parameter ρ. If one estimates, however, the parameter ϕ of a LD-AR(1) based on observations z_t, the question is whether the estimate $\hat{\phi}$ is a consistent estimate of the parameter ρ of the DGP. In the given context, the conditional expectation of the latent variable $\tilde{m}_t$ is given by

$$\tilde{m}_t = \phi \sum_{i=0}^{t-1} \phi^i \tilde{c}_{t-1-i}, \tag{28}$$

$$\text{where } \tilde{c}_t = \mathrm{E}\left[e_t \middle| \mathcal{F}_t^z \right]. \tag{29}$$

The score of the model is similar to the score in (27) under the LD-AR(1) DGP. Here, however, the generalized errors are evaluated on the basis of the observations $\bar{z}_T$, thus, the generalized errors c_t are replaced by $\tilde{c}_t$ yielding

$$\frac{\partial \log \mathfrak{L}_y(\bar{z}_T|\phi)}{\partial \phi} = -\sum_{t=1}^{T} \sum_{i=0}^{t-1} (i+1)\phi^i \tilde{c}_{t-1-i}\tilde{c}_t. \tag{30}$$

If one reformulates the score of the latent AR model in terms of the generalized error $\tilde{c}_t$ and an error ν_t the consistency of the quasi maximum likelihood estimation of ρ based on the simple likelihood $\mathfrak{L}_y$ will become obvious.

$$\begin{aligned}
&\frac{\partial \log \mathfrak{L}_z(\bar{z}_T|\rho)}{\partial \rho} = \\
&-\sum_{t=1}^{T} \sum_{i=0}^{t-1} (i+1)\rho^i (\mathrm{E}\left[\epsilon_t \middle| \mathcal{F}_t^z \right] + \nu_t)(\mathrm{E}\left[\epsilon_{t-1-i} \middle| \mathcal{F}_{t-1-i}^z \right] + \nu_{t-1-i}), \\
&= -\sum_{t=1}^{T} \sum_{i=0}^{t-1} (i+1)\rho^i (\mathrm{E}\left[\epsilon_t \middle| \mathcal{F}_t^z \right] \mathrm{E}\left[\epsilon_{t-1-i} \middle| \mathcal{F}_{t-1-i}^z \right] + \nu_t \nu_{t-1-i} + \nu_t \mathrm{E}\left[\epsilon_{t-1-i} \middle| \mathcal{F}_{t-1-i}^z \right] + \mathrm{E}\left[\epsilon_t \middle| \mathcal{F}_t^z \right] \nu_{t-1-i}),
\end{aligned} \tag{31}$$

with $\nu_t := \mathrm{E}\left[\epsilon_t \middle| \mathcal{F}_T^z \right] - \mathrm{E}\left[\epsilon_t \middle| \mathcal{F}_t^z \right]$.

Now, we can show, that the difference between the LD-ARMA model (30) and the latent ARMA model (31) boils down to three additional terms being present in the score of the latter. Due to the i.i.d. nature of errors and the fact that $\mathcal{F}_t^z \subseteq \mathcal{F}_T^z$, the unconditional expectation taken with respect to the observations generated by the original process of each the four terms in (31) is zero, i.e.

$$\begin{aligned}
\mathrm{E}_{\bar{z}_T}\left[\mathrm{E}\left[\epsilon_t \middle| \mathcal{F}_t^z \right] \mathrm{E}\left[\epsilon_{t-1-i} \middle| \mathcal{F}_{t-1-i}^z \right] \right] = \mathrm{E}_{\bar{z}_T}\left[\mathrm{E}\left[\epsilon_t \middle| \mathcal{F}_t^z \right] \nu_{t-1-i} \right] = \\
\mathrm{E}_{\bar{z}_T}\left[\nu_t \mathrm{E}\left[\epsilon_{t-1-i} \middle| \mathcal{F}_{t-1-i}^z \right] \right] = \mathrm{E}_{\bar{z}_T}\left[\nu_t \nu_{t-1-i} \right] = 0
\end{aligned} \tag{32}$$

This opens a different perspective to the estimation problem as, the original model could be interpreted as a GMM estimator based on the four moment restrictions outlined in (32). The alternative estimator, however, relies only on a subset of these moment conditions, namely the first one. See the surveys of Newey and McFadden (1994) and Wooldridge (1994) for an extended discussion of GMM estimators and their relationship to ML estimators. From the GMM perspective of this ML estimation problem it is apparent that the LD-ARMA model yields indeed a consistent estimator of the parameter of the dynamic in the original model. The intuition behind this can be found in the fact that the reduced information set $\mathcal{F}_t^z$ used to form an expectation of the error term ϵ_t is a subset of the full information set $\mathcal{F}_T^z$. This is driven by the i.i.d. nature of the error term, as the incremental information contained in $\mathcal{F}_T^z$ but not contained in $\mathcal{F}_t^z$ is uncorrelated with $\mathcal{F}_t^z$.

A second conclusion which can be drawn from the GMM interpretation is that the alternative estimator is an inefficient version of the original model, as three possible moment restrictions were not used in estimation, see e.g. Newey and McFadden (1994) for an extended discussion. Thus, the alternative estimator is a consistent but inefficient estimator of the dynamic parameter in the original model, which has the considerable advantage of being easy to evaluate and straightforward to extend to higher order dynamics and alternative observation rules.

3.4 Small sample evidence

To obtain some evidence on the small sample performance of the estimator based on the LD dynamic a small Monte Carlo study is performed for the Probit observation rule given in example 1 in conjunction with the latent AR(1) outlined in the introduction (4)-(5). Parameter estimates are obtained from the quasi-likelihood implied by the LD-AR(1) model characterized by the alternative conditional mean of the latent variable given by (1)-(2). The only parameter of the DGP, ϕ, is drawn from a uniform distribution over the interval $[-0.95; 0.95]$ for each of the $N = 10000$ replications. The errors are drawn from the standard normal distribution. The experiment is carried out for sample sizes $T \in \{50, 100, 200, 1000\}$. Descriptive statistics of the difference between true parameter and estimate, $\phi_i - \hat{\phi}_i$, $i = 1, \ldots, N$, are reported in table 1. The small sample properties match the expectation build from the asymptotic results, i.e. the variance decreases over an increasing sample size and likewise do the interquantile ranges (q75 - q25). The results

Table 1: **Probit with latent AR(1)** Descriptive statistics of $\phi_i - \hat{\phi}_i$ in a Monte Carlo study.

sample	mean	median	variance	skewness	kurtosis	q01	q25	q75	q99
$T = 50$	-3.4e-3	-3.6e-3	3.3e-2	4.8e-2	3.7	-0.45	-0.12	0.11	0.47
$T = 100$	1.9e-3	1.1e-3	1.7e-2	6.1e-2	3.7	-0.31	-0.08	0.08	0.33
$T = 200$	3.7e-5	7.2e-4	8.2e-3	-9.8e-3	3.7	-0.23	-0.06	0.06	0.22
$T = 1000$	-1.3e-4	-7.8e-5	1.7e-3	3.1e-3	3.4	-0.10	-0.03	0.03	0.10

indicate that even a moderately sized sample of 50 observations is quite sufficient to obtain reasonable results.

To gain more insight in the consequences the observation rule bears for the estimation task, the results of the Monte Carlo experiment are scrutinized with respect to the parameter ϕ of the model. The parameter space is subdivided into ten categories, $j = 1, \ldots, 10$, where the category j is an interval c_j of size 0.1, $c_j \in \{(-1; -0.9], (-0.9; -0.8], \ldots, (0.9; 1)\}$. Note that the border categories c_1 and c_{10} have an effective size of 0.05 due to the support of the random variable the parameter is drawn from. Descriptive statistics of the difference between the true parameter and the estimate, $\phi_i - \hat{\phi}_i$, $i = 1, \ldots, N$, are depicted in figure 2 in the form of Box plots for each interval c_j. It is quite interesting to observe that the small bias the estimator shows for a sample size of $T = 50$ diminishes substantially once a sample size of $T = 1000$ is reached. The small sample bias shown in the left hand figure of 2 indicates that the simple estimator tends to underestimate the absolute size of the parameter, particularly for larger parameter values.

To further explore the properties, of the proposed model a LD-MA(1) is estimated for a DGP process which is of the latent MA(1) form. The setup of the experiment is analogue to the AR(1) experiment. Results are reported in table 2, where descriptive statistics of the difference between true parameter and estimate, $\theta_i - \hat{\theta}_i$, $i = 1, \ldots, N$ are given. Results differ significantly from the LD-AR(1) results reported in table 1. Especially, the variance and the interquantile range of the difference decreases at a much slower rate compared to the former model and an increasing kurtosis over an increasing

Figure 2: **Probit with latent AR(1)** Box plots of $\phi_i - \hat{\phi}_i$ for ten size categories of the parameter ϕ_i in a Monte Carlo study.

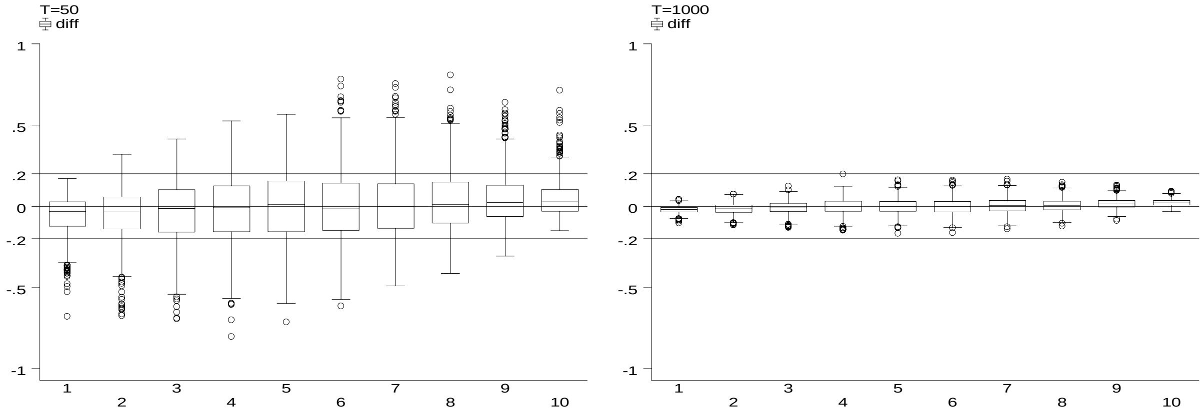


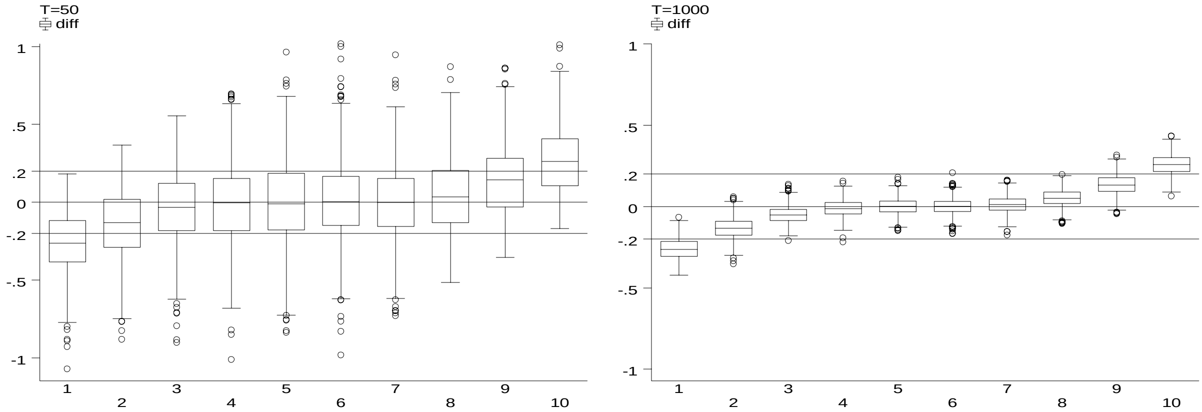
Table 2: **Probit with latent MA(1)** Descriptive statistics of $\theta_i - \hat{\theta}_i$ in a Monte Carlo study.

sample	mean	median	variance	skewness	kurtosis	q01	q25	q75	q99
$T = 50$	-7.1e-4	-4.7e-3	7.0e-2	3.5e-2	3.1	-0.61	-0.18	0.18	0.61
$T = 100$	3.1e-3	3.8e-3	4.1e-2	-4.0e-2	3.0	-0.48	-0.13	0.14	0.47
$T = 200$	6.9e-4	1.3e-3	2.7e-2	4.6e-3	3.1	-0.39	-0.11	0.11	0.40
$T = 1000$	9.1e-4	8.6e-4	1.8e-2	-4.3e-3	3.5	-0.33	-0.07	0.07	0.33

sample size indicates that a closer examination of the results is in place.

Figure 3 reveals the nature of this unexpected behaviour. It shows that for in absolute value smaller parameters the bias is indeed small and the interquartile range decreases as in the LD-AR(1) case. Yet, for larger parameters in absolute value the underestimation is quite substantial and does only decrease slightly for an increasing sample size. This behaviour seems to indicate a serious deficiency of the LD-ARMA estimator of the latent ARMA parameters, yet, the lesson learned from the ACF of the latent and the observable model given in 3.2 is that the latent and the observable ACF can deviate substantially depending on the observation rule. This actually helps to resolve the considerable bias reported in figure 3, which is due to the information loss incurred by the Probit observation rule. Note, however, that this is not attributed to the LD-ARMA process but is a problem of latent MA models as such, and thus a problem simulation based approaches would have to struggle with as well. Further experiments, not reported

Figure 3: **Probit with latent MA(1)** Box plots of $\theta_i - \hat{\theta}_i$ for ten size categories of the parameter θ_i in a Monte Carlo study.



here, have revealed, that the bias goes indeed away, if the information loss imposed by the observation rule is reduced, e.g. by considering an ordered probit with a latent MA.

4 An empirical illustration

To illustrate the working of the LD-ARMA model in practice a simple empirical study over the absolute value of transaction-to-transaction price changes is reported in table 3. Asymptotic t-statistics are given in parentheses. They were evaluated using the sandwich estimator of the parameters' covariance matrix suggested by White (1982). The data is extracted from the TAQ data set provided by the NYSE. Here, all the $13,421$ transactions for IBM carried out at the NASDAQ in September 2000 are employed. The tick size for IBM at the NASDAQ was at this time 1/16. Thus, categories employed were chosen as follows. Category one contains the transactions which were not associated with a price change. Categories two and three, capture price changes of 2/16 and 3/16, respectively. Category four contains all price changes equal or larger than 4/16. All in absolute value. Several specifications up to an LD-ARMA(3,3) model are assessed in this study. For each estimate the t-Statistic is reported and the Schwartz information criterion (BIC) and the Portmanteau type statistic (ξ_{20}) suggested by Gourieroux, Monfort, and Trognon (1985) including 20 lags are given for each specification. Note that an increase in the number of lags for the test statistic, i.e. using ξ_{40} instead, tends to shift the p-values towards 1, thus decreasing the evidence for misspecification. For the sake of brevity, this is however not reported in table 3. Overall the LD-ARMA(2,2) specification seems appropriate, showing

Table 3: **Estimation results for IBM trading at NASDAQ in September 2000.**

	μ_1	μ_2	μ_3	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3	BIC	ξ_{20}	p-value
LD-ARMA(1,0)	-0.13	0.77	1.33	0.25						-16064	976.90	0.00
	(10.28)	(53.42)	(78.19)	(22.26)								
LD-ARMA(0,1)	-0.13	0.76	1.32				0.19			-16139	1476.80	0.00
	(11.41)	(57.36)	(82.17)				(21.92)					
LD-ARMA(1,1)	-0.11	0.82	1.38	0.94			-0.79			-15717	29.69	0.08
	(3.79)	(25.91)	(42.14)	(129.96)			(54.54)					
LD-ARMA(2,2)	-0.11	0.82	1.39	1.63	-0.64		-1.46	0.50		-15715	10.99	0.95
	(3.24)	(22.80)	(37.33)	(14.36)	(5.91)		(12.23)	(5.03)				
LD-ARMA(3,3)	-0.11	0.82	1.39	0.85	0.60	-0.48	-0.67	-0.60	0.37	-15724	10.95	0.95
	(3.24)	(22.83)	(37.38)	(4.35)	(2.41)	(3.15)	(3.41)	(2.67)	(2.79)			

the best BIC and no misspecification. The additional lags of an LD-ARMA(3,3) yield no additional information, yet the LD-ARMA(1,1) is barely rejected due to the serial correlation indicated by the test.

5 Conclusion

In this paper a new dynamic for limited dependent variable models is proposed, which is tailored to cope with the information loss incurred by the imposition of an observation rule on a latent process. The LD-ARMA model circumvents by the reformulation of the mean function of the latent process the multiple integral problem, which would usually necessitate the use of simulation methods. It turns out that the LD-ARMA models are applicable to a wide range of observation rules opening thereby a broad range of applications. Further it is shown that the latent process implied by this specification is covariance stationary and has the same autocorrelation function as the corresponding latent ARMA model. The main difference is that the variance of the latent process of the LD-ARMA model is bounded from above by the variance of the corresponding latent model. The main result in this context is that the likelihood of the LD-ARMA model serves as a valid simple to evaluate quasi-likelihood of the latent ARMA. It is demonstrated in a Monte Carlo study that the quasi-likelihood has indeed favourable small sample properties.

Further extensions include the formulation of multivariate models, including further limited dependent processes as well as standard observable processes. The latter is easily achieved as the modified mean process boils down to a standard mean process, when the latent variable is observed.

Acknowledgements

Simulation and estimation of the models was done using *Ox*, the matrix programming language created by Doornik (1998). Software for simulation and estimation is available from the author in form of the *Ox* class *DynLD*. Descriptive statistics and box plots were generated in *Stata*. I am grateful to Nikolaus Hautsch, Winfried Pohlmeier, Claudia Czado for very helpful discussions and I am greatly indebted to Neil Shephard who helped me clarify the ideas put forward here. The remaining faults and weaknesses of exposition are of course my own responsibility. The research was supported by the Center of Finance and Econometrics (CoFE, `http://cofe.uni-konstanz.de`) and the UK's Economic and Social Research Council through the grant 'Econometrics of trade-by-trade price dynamics', which is coded R00023839.

Proofs

Proof of proposition (1): The mean of the latent process is evaluated from the MA form of the states' mean function (18) and the transformation of the latent state in (17)

$$\mathrm{E}\left[m_t + e_t\right] = \mathrm{E}\left[H'\Psi_t s_0 + H'\sum_{i=0}^{t-1}\Psi_i e_1 c_{t-i} + u_1 e_t\right] = 0, \tag{33}$$

since $s_0 = 0$ and $\mathrm{E}\left[c_{t-i}\right] = 0$ from the law of iterated expectations.

The autocovariance of the latent process associated with the LD-ARMA model is evaluated from (33) and

$$\mathrm{E}\left[(m_t + e_t)(m_{t-s} + e_{t-s})\right] = H'\mathrm{E}\left[s_t s'_{t-s}\right]H, \tag{34}$$

since the cross term vanish. The covariance of the state process is derived from the usual Yule-Walker equations, see e.g. Lütkepohl (1991, chap. 2.1). To ease notation, we define $\Gamma(l) := \mathrm{E}\left[s_t s'_{t-l}\right]$. By multiplying (18) with s'_{t-1} and taking expectations we obtain

$$\Gamma(1) = \mathrm{E}\left[F s_{t-1} s'_{t-1} + F u_1 c_t s'_{t-1}\right] = F\Gamma(0). \tag{35}$$

This uses again the fact that $\mathrm{E}\left[c_t s_{t-1}\right] = 0$. The covariance follows from a multiplication of (18) with s_t' and taking again expectations.

$$\Gamma(0) = F\Gamma(1)' + \Sigma_c, \tag{36}$$

where $\Sigma_c := F u_1 u_1' F' \mathrm{E}\left[c_t^2\right]$ and again $\mathrm{E}\left[c_t s_{t-1}\right] = 0$ was used. Inserting (35) into (36) and solving for $\Gamma(0)$ yields

$$\mathrm{vec}\Gamma(0) = (I - F \otimes F)^{-1} \mathrm{vec}\Sigma_c. \tag{37}$$

The autocovariance function is therefore almost identical to the corresponding latent ARMA model, see e.g. Lütkepohl (1991, chap. 2.1). The difference boils down to the unconditional expectation of the squared innovation term driving either process. These are however related by

$$\mathrm{E}\left[c_t^2\right] = \mathrm{E}\left[\mathrm{E}\left[e_t | \mathcal{F}_t^y\right]^2\right] \leq \mathrm{E}\left[e_t^2\right] = \mathrm{E}\left[\mathrm{E}\left[e_t | \mathcal{F}_t^e\right]^2\right], \tag{38}$$

$$\text{since } \mathcal{F}_t^y \subseteq \mathcal{F}_t^e, \tag{39}$$

see e.g. Davidson (1994, theorem 10.27), which proves no. 1. To prove no. 2, define $R_y(s)$ as the matrix valued ACF at lag s of the state process s_t in the dynamic of y_t by

$$R_y(s) := D^{-1}\Gamma(s)D^{-1}, \qquad DD = diag(\Gamma(0)). \tag{40}$$

From this it is evident, that the factor $\mathrm{E}\left[c_t^2\right]$ cancels out and that $R_y(s)$ is equal to the analogue ACF matrix in the model of z_t, i.e. $R_z(s)$. Therefore we have that $\rho^*(s) = \rho^\circ(s)$.□

References

Albert, J., and S. Chib (1995): "Bayesian Residual Analysis for Binary Response Regression Models," *Biometrika*, 82, 747–759.

Chib, S., and E. Greenberg (1998): "Analysis of Multivariate Probit Models," *Biometrika*, 85, 347–361.

Cox, D. R. (1981): "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics*, 8, 93–115.

Cox, D. R., and E. J. Snell (1989): *Analysis of Binary Data*. Chapman and Hall, New York, 2nd. edn.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford, New York.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society Series B*, 39, 1–38.

Doornik, J. A. (1998): *Ox: Object Oriented Matrix Programming, 2.0.* Timberlake, London.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987): "Generalized Residuals," *Journal of Econometrics*, 34, 5–32.

Gourieroux, C., A. Monfort, and A. Trognon (1985): "A General Approach to Serial Correlation," *Econometric Theory*, 1, 315–340.

Granger, C. W. J., and P. Newbold (1976): "Forecasting Transformed Series," *Journal of the Royal Statistical Society, Series B*, 38, 189–203.

Hamilton, J. D. (1994): *Time Series Analysis.* Princeton University Press.

Harvey, A. C. (1989): *Forecasting Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

——— (1990): *The Econometric Analysis of Time Series.* Philip Allan, Hemel Hempstead, 2nd. edn.

Hendry, D. F. (1995): *Dynamic Econometrics.* Oxford University Press, Oxford.

Jacobs, P. A., and P. A. Lewis (1978a): "Discrete Time Series Generated by Mixtures I: Correlational and Runs Properties," *Journal of the Royal Statistical Society, Series B*, 40, 94–105.

——— (1978b): "Discrete Time Series Generated by Mixtures II: Asymptotic Properties," *Journal of the Royal Statistical Society, Series B*, 40, 222–228.

Kedem, B. (1980): "Estimation of the Parameters in Stationary Autoregressive Processes After Hard Limiting," *Journal of the American Statistical Association*, 75, 146–153.

Keenan, D. M. (1982): "A Time Series Analysis of Binary Data," *Journal of the American Statistical Association*, 77, 816–821.

Lomnicki, Z. A., and S. K. Zaremba (1955): "Some Applications of Zero-One Processes," *Journal of the Royal Statistical Society, Series B*, 17, 243–255.

Lütkepohl, H. (1991): *Introduction to Multiple Time Series Analysis.* Springer, New York.

Maddala, G. S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge.

Manrique, A., and N. Shephard (1998): "Simulation-Based Likelihood Inference for Limited Dependent Processes," *Econometrics Journal*, 1, C174–C202.

McCullagh, P., and J. A. Nelder (1989): *Generalized Linear Models.* Chapman and Hall, London.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, chap. 36, pp. 2111–2245. Elsevier Science.

Poirier, D. L., and P. A. Ruud (1988): "Probit with Dependent Observations," *Review of Economic Studies*, 55, 593–614.

Ruud, P. A. (1991): "Extension of Estimation Methods Using the EM Algorithm," *Journal of Econometrics*, 49, 305–341.

Stone, M. H. (1927): "Developments in Hermite Polynomials," *The Annals of Mathematics*, 29, 1–13.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

Wooldridge, J. M. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, chap. 45, pp. 2639–2738. Elsevier Science B.V.

Zeger, S. L., and B. Qaqish (1988): "Markov Regression Models for Time Series: A Quasi-Likelihood Approach," *Biometrics*, 44, 1019–1031.