Heiss, Florian:

Nonlinear State-Space Models for Microeconometric Panel Data

Munich Discussion Paper No. 2006-24

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at https://doi.org/10.5282/ubm/epub.1157

Nonlinear State-Space Models for Microeconometric Panel Data *

Florian Heiss

University of Munich, Department of Economics,

florian.heiss@lrz.uni-muenchen.de

June 28, 2006

In applied microeconometric panel data analyses, time-constant random effects and first-order Markov chains are the most prevalent structures to account for intertemporal correlations in limited dependent variable models. An example from health economics shows that the addition of a simple autoregressive error terms leads to a more plausible and parsimonious model which also captures the dynamic features better. The computational problems encountered in the estimation of such models – and a broader class formulated in the framework of nonlinear state space models – hampers their widespread use. This paper discusses the application of different nonlinear filtering approaches developed in the time-series literature to these models and suggests that a straightforward algorithm based on sequential Gaussian quadrature can be expected to perform well in this setting. This conjecture is impressively confirmed by an extensive analysis of the example application.

*The author would like to thank Alexander Ludwig, Axel Börsch-Supan, Dan McFadden, Mike Hurd, Viktor Winschel, Joachim Winter, and David Wise for valuable discussion, comments and suggestions.

1 Introduction

Panel data provide repeated observations on the same individuals, firms, or other units over time. This allows the identification of a much richer set of effects in a more general setting than pure cross-sectional data. Many microeconometric models, especially limited dependent variable models, are inherently nonlinear. This nonlinearity complicates the analysis of panel data models, for a general discussion see for example Chamberlain (1984). In applied microeconometric research, the vast majority of nonlinear panel data models specify unobserved heterogeneity as time-constant individual effects and/or state dependence as a low-order Markov model. While the estimation of these models is fairly straightforward, they impose a quite inflexible dynamic structure on the data.

As an example, a health economics application is presented. For studying the evolution of health over time, the literature has so far focused on first-order Markov chain and random effects models. Contoyannis, Jones and Rice (2004) thoroughly discuss these approaches and their estimation. I argue that a simple ordered logit model with an AR(1) error term is theoretically more convincing. Furthermore it is more parsimonious and captures the observed intertemporal correlation pattern much better.

The widespread application of such models is hampered by the computational difficulties encountered in their estimation. This paper discusses these problems and different solutions for a class of models which includes limited dependent variable models with AR(1) errors but is much more general. It is formulated in a state space framework. This approach has a long tradition in linear time series models, see Hamilton (1994). The increase in computational power makes it also feasible for general nonlinear models which generated increased interest in the econometric time series literature, see for example Fernández-Villaverde and Rubio-Ramírez (2005).

The computational problem in evaluating the likelihood function of such models is that the unobserved state process has to be integrated out. With continuously distributed states, these integrals have to approximated numerically and their dimension is typically

proportional to the time-series dimension of the data. Unlike time series models, this data dimension is usually moderate for microeconometric analyses and asymptotic arguments are applied to the cross-sectional dimension. This makes it feasible to approximate the full multidimensional integral for example by Monte Carlo simulation or by numerical integration (Heiss and Winschel 2006).

For time-series models, various attempts have been made to break up the full integral into a sequence of lower-dimensional integrals in the spirit of the Kalman filter. Outside of economics, these nonlinear filtering approaches are widely studied e.g. in engineering (Doucet, De Freitas and Gordon, eds 2001). In the econometric time-series literature, they have been discussed e.g. by Danielsson and Richard (1993), Fernández-Villaverde and Rubio-Ramírez (2006), and Tanizaki and Mariano (1994). For a survey of these methods, see Tanizaki (2003).

Also applications with a moderate time-series dimension can profit from nonlinear filtering techniques. Loosely speaking does each reduction of dimensionality help for each method of numerical integration. This paper briefly reviews different filtering algorithms such as nonlinear particle filters (Gordon, Salmond and Smith 1993) and sequential importance sampling Tanizaki and Mariano (1994). Given the features of microeconometric models discussed here with a moderate time series dimension and a univariate latent state space, it is then argued that a similar, straightforward to implement, approach using sequential Gaussian quadrature can be expected to perform well.

Different algorithms are then implemented for the illustrative health model. While all converge to the same results as the computational effort is increased, the speed of this convergence differs dramatically. Simulation estimation with a total run time of 15 hours delivers less accurate results than the preferred sequential quadrature algorithm achieves in 11 minutes.

The paper is structured as follows. Section 2 presents the general model structure and the illustrative example. In section 3, different approaches and algorithms for the likelihood approximation are discussed and the sequential Gaussian quadrature algorithm is presented

in detail. Section 4 revisits the illustrative example, implements different algorithms and compares their performance. Section 5 concludes.

2 Microeconometric State-Space Models

This paper discusses estimation for a relatively rich set of micro-econometric models that can be represented in a state-space framework. We start out by defining the structure of and requirements on these models.

2.1 Model Specification

Suppose a sequence of dependent variables is observed over time for a number N of cross-sectional units, say individuals. All random variables involved in the model are assumed to be independent across cross-sectional units. Let T be the number of observations over time ("waves") for each cross-sectional unit. In the following discussion, assume that T is the same number for each cross-sectional unit so that we are dealing with a balanced panel. This is merely for notational convenience – unbalanced panels can easily be dealt with if the individual number of observations is random or modeled jointly.

The vectors of random variables $\mathbf{Y}_{it}$ for $i = 1, ..., N$ and $t = 1, \ldots, T$ represent dependent variables of individual i in wave t. In many applications, they are one-dimensional, but I allow for the more general case since this does not create any complications neither in the notation nor in the analysis. The vector of dependent variables may consist of discrete, continuous, or both types of random variables. They are modeled conditional on exogenous variables $\mathbf{x}_i$, unobserved states which are correlated over time a_{it} and unobserved i.i.d. error terms $\mathbf{e}_{it}$. The model is specified as

$$\mathbf{Y}_{it} = g(\mathbf{x}_i, a_{it}, \mathbf{e}_{it}; \boldsymbol{\theta}), \tag{1}$$

where $g(\cdot)$ is a general parametric function. The vectors $\mathbf{x}_i$ contain time-constant and time-varying strictly exogenous variables. In the latter case, $\mathbf{x}_i$ collects all time-specific

values. The random variables ("states") a_{it} are allowed to be continuously distributed and correlated over time in a relatively flexible way discussed in detail below. In this paper, a_{it} is assumed to be a scalar random variable, but generalizations to a higher-dimensional state-space are straightforward. The i.i.d. error terms $\mathbf{e}_{it}$ may reflect measurement errors and/or transitory influences on $\mathbf{Y}_{it}$. For notational simplicity, all model parameters are collected in the vector $\boldsymbol{\theta}$.

This model can be viewed as a generalization of a random effects model. In this case, $\mathbf{a}_{i,1:T} = [a_{it} : t = 1, ..., T]$ has a degenerate joint distribution with $a_{it} = a_i$ for all $t = 1, ..., T$. In the more general case, it could for example represent an AR(1) component of the error term.

Suppose we are interested in likelihood-based estimation such as maximum likelihood or Bayesian analysis. With $P(\mathbf{y}_{i,1:t}|\mathbf{x}_i; \boldsymbol{\theta})$ denoting the joint probability mass (or probability density) of $\mathbf{Y}_{i,1:t} = [\mathbf{Y}_{is} : s = 1, ..., t]$ conditional on $\mathbf{x}_i$, evaluated at the observed values $\mathbf{y}_{i,1:t}$, the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} P(\mathbf{y}_{i,1:T}|\mathbf{x}_i; \boldsymbol{\theta}) \tag{2}$$

The computational problem in evaluating this expression mainly arises due to the presence of the latent states a_{it} in the model. Before discussing this problem and solutions in detail, the class of models is restricted in the following way. For convenience of presentation, it is understood that all expressions depend on the parameter vector $\boldsymbol{\theta}$ which is in the following left out of the notation.

Measurement

Let $P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1}, \mathbf{a}_{i,1:T})$ represent the joint probability mass (or density) of $\mathbf{Y}_{i,1:t}$ conditional on $\mathbf{x}_i$, $\mathbf{a}_{i,1:T}$, and past values of y_{it}, evaluated at the observed values $\mathbf{y}_{it}$.

Make the following conditional independence assumption.

$$P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1}, \mathbf{a}_{i,1:T}) = P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it}) \quad \forall i = 1, ..., N, t = 1, ..., T \tag{3}$$

Conditional on $\mathbf{x}_i$ and the contemporaneous value of the latent state a_{it}, the outcome probability of $\mathbf{y}_{it}$ is independent of both past and future values of the state process $\mathbf{a}_{i,1:T}$ and lagged dependent variables. The latter assumption avoids the usual initial value problems which could be dealt with with the usual approaches, see Heckman (1981), Wooldridge (2005), and Honoré and Tamer (2006). Under this assumption, all contemporaneous correlation of $\mathbf{y}_{it}$ conditional on $\mathbf{x}_i$ is generated by the sequence of latent states $\mathbf{a}_{i,1:T}$ which is correlated over time.

Assume that the i.i.d. error terms $\mathbf{e}_{it}$ can easily be integrated out of the model so that $P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it})$ is a known parametric function. It might for example follow from a typical limited dependent variable (LDV) model specification in (1) in which the unknown a_{it} enter as additional regressors analogous to LDV models with random effects.

States

For the sequence of latent states $\mathbf{a}_{i,1:T}$, assume that the marginal distribution of a_{it} is known up to a finite set of parameters included in the general parameter vector $\boldsymbol{\theta}$. In the following, these states are treated as vectors of continuous random variables. The only real difference with discrete or mixed distributions is that the numerical analysis would be less difficult. For notational simplicity assume that these marginal distributions are the same for all $i = 1, ..., N$ and $t = 1, ..., T$ and denote its p.d.f. conditional on the exogenous covariates as $f(a_{it}|\mathbf{x}_i)$. For identification of the model, it will in many cases be necessary to assume independence of $\mathbf{x}_i$ analogous to random effects models.

As noted before, states are allowed to be dependent over time. For notational and analytical convenience, assume that they are first-order Markov. Also assume that there is no feedback from the sequence of dependent variables $\mathbf{y}_{i,1:T}$. Therefore, for the conditional p.d.f.

$$f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:T}, \mathbf{a}_{i,1:t-1}) = f(a_{it}|\mathbf{x}_i, a_{i,t-1}) \tag{4}$$

This conditional distribution is again assumed to be known up to parameters. This structure allows to write the joint p.d.f. of $\mathbf{a}_{i,1:T}$ as

$$f(\mathbf{a}_{i,1:T}|\mathbf{x}_i) = f(a_{i1}|\mathbf{x}_i) \prod_{t=2}^{T} f(a_{it}|\mathbf{x}_i, a_{i,t-1}) \tag{5}$$

2.2 An Ordered Logit Model of Health with an AR(1) Error Term

One of the most frequently studied measures of individual health is the self-rated health status (SRHS). The Health and Retirement Study (HRS) asks all respondents "Would you say your health is excellent, very good, good, fair, or poor?". It is included in many other surveys with a similar wording. Despite its obvious subjectiveness, it has been found a useful and powerful measure. It maps the high-dimensional and complex concept of health into one dimension using individual perceptions and judgments. It is also a very powerful predictor of objective events such as mortality.

The data used for the empirical example is from the Health and Retirement Study (HRS) which is sponsored by the National Institute of Aging (NIA) and conducted by the University of Michigan. For the analyses presented here, I use the RAND HRS Data File (Version E). It was developed by the RAND Center for the Study of Aging with funding from the National Institute on Aging (NIA) and the Social Security Administration (SSA). The HRS contains data on different cohorts of elderly Americans. I use a sample of all cohorts with the only restriction that they are at least 50 years old at the time of the first interview. The sample includes 25,353 respondents with up to 6 observations over time each. A total of 102,233 observations are available.

Panel data analyses of SRHS are not only useful because unobserved heterogeneity of health itself, but also heterogeneity of reporting SRHS *given* health can be accounted for. Table 1 shows the distribution of SRHS in the sample. As the tabulations conditional on the previous response indicate, SRHS is highly correlated over time. Here, we focus on the question how to model this correlation. In the literature, this correlation is almost exclusively modeled as time-constant unobserved heterogeneity and/or state dependence

of SRHS. Contoyannis et al. (2004) discuss and compare these approaches. I will argue that a simple model with correlated error terms is both more plausible and fits the correlation pattern better.

Table 2 gives an impression on the intertemporal correlation pattern over a longer period of time. It shows the results of an orered logit regression of SRHS in wave 6 on a typical set of covariates plus lagged values of SRHS. Note that this is obviously only done for respondents with six observations. Due to the sampling scheme, this holds only for the original HRS cohort born between 1931 and 1941. The two most interesting results are that the coefficients of all lags (i) are all highly significantly different from zero and (ii) get smaller the further away the respective observation is from wave 6. A random effects model would imply equal predictive power of all lags which contradicts observation (ii). A first-order Markov chain model would imply no additional predictive power of waves 1 through 4 once wave 5 is controlled for which contradicts observation (i). A combination of a time-constant random effect (RE) with a first-order Markov chain model would imply predictive power of all waves with wave 5 having a higher predictive power than waves 1 through 4. But a Wald test of the hypothesis of equal predictive power of the earlier four waves is clearly rejected (test statistic $\overset{a}{\sim} \chi^2_3 = 40.05$).

I interpret these findings as an indication that the models typically used for modeling SRHS in panels such as Contoyannis et al. (2004) are not capable of capturing the correlation pattern found in the data.[1] An obvious strategy to capture the correlation pattern better would be to combine a higher-order Markov chain model of state dependence with a RE specification. But this would aggravate the initial values problem already present in the first-order Markov chain model with RE.

In a structural model, state-dependence of SRHS is actually not very convincing. While for example in a model of labor force participation lagged outcomes can causally affect

[1] Note that these models typically do not specify the lagged values as the 5-point scale SRHS measure but as four dummy variables. This does not change the conclusions from Table 2 but only makes the results harder to read.

today's outcome, this is unlikely for this application: Which of the five SRHS categories a respondent ticks in a survey won't affect future health. So in a model with state dependence and RE, the coefficients determining the state dependence can be interpreted to capture the diminishing predictive power of higher lags evident in Table 2 in a reduced-form fashion. But a structurally more plausible model would be one in which SRHS depends on current health and this underlying variable follows some random process over time with decreasing correlation. I suggest a simple model with an AR(1) error term.

Let Y_{it}^* denote a latent variable which represents a continuous representation of health. It is modeled as a function of covariates $\mathbf{x}_{it}$, an unobserved stochastic process a_{it} and an i.i.d. error term e_{it}. For simplicity, consider the linear specification

$$Y_{it}^* = \mathbf{x}_{it}\boldsymbol{\beta} + a_{it} + e_{it}. \tag{6}$$

Assume that SRHS $Y_{it} \in \{1, ..., 5\}$ is generated by a standard ordered response model.

$$Y_{it} = j \Leftrightarrow \alpha_{j-1} \leq Y_{it}^* < \alpha_j \quad \text{with } 1 \leq j \leq 5, \tag{7}$$

where $\alpha_0 = -\infty$, $\alpha_5 = \infty$, and α_1 through α_4 are unknown model parameters. In the general notation of section 2.1, equations (6) and (7) correspond to the specification of the model in (3).

In order to derive a parametric expression of conditional outcome probabilities, assume that the i.i.d. error terms e_{it} are i.i.d. with a logistic distribution. They may represent transitory health problems like a cold, general mood at the time the survey was completed or general measurement errors. This parametric assumption leads to a standard ordered logit specification except that the latent process a_{it} is present. With $\Lambda(\cdot)$ representing the logistic c.d.f., the conditional outcome probabilities in (3) can in this model be written as

$$P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it}) = \Lambda\left(\alpha_{y_{it}} - \mathbf{x}_{it}\boldsymbol{\beta} - a_{it}\right) - \Lambda\left(\alpha_{y_{it}-1} - \mathbf{x}_{it}\boldsymbol{\beta} - a_{it}\right). \tag{8}$$

To complete the model, the joint distribution of the state-space $a_i t$ has to be specified. Assume independence of $\mathbf{x}_i$ and a normal AR(1) process. With $\phi(\cdot; \mu, \sigma^2)$ denoting the normal p.d.f. with mean μ and variance σ^2, the marginal distribution is

$$f(a_{it}|\mathbf{x}_i) = \phi(a_{it}; 0, \sigma^2). \tag{9}$$

Assume the AR(1) structure

$$a_{it} = \rho a_{i,t-1} + u_{it} \tag{10}$$

where the innovations u_{it} are i.i.d. normal with zero mean and variance $(1-\rho^2)\sigma^2$. The correlation parameter $-1 \leq \rho \leq 1$ is another model parameter. This leads to a conditional distribution corresponding to (4) of

$$f(a_{it}|\mathbf{x}_i, a_{i,t-1}) = \phi(a_{it}; \rho a_{i,t-1}, (1-\rho^2)\sigma^2). \tag{11}$$

This completes the model definition discussed in general in section 2.1 with a parameter vector $\boldsymbol{\theta} = [\boldsymbol{\beta}, \alpha_1, ..., \alpha_4, \sigma, \rho]$.

A standard ordered logit model follows in the special case $\sigma = 0$ and a standard random effects ordered logit model follows in the case $\rho = 1$. The correlation between Y_{it} and Y_{is} conditional on the covariates $\mathbf{x}_i$ is $\rho^{|t-s|}$. With $0 < \rho < 1$, it can explain the significant but decreasing predictive power of lagged dependent variables in Table 2.

3 Evaluation of the Likelihood Contributions

For the evaluation of the likelihood function (2), the probabilities $P(\mathbf{y}_{i,1:T}|\mathbf{x}_i)$ have to be evaluated. Because of the presence of the latent process $\mathbf{a}_{i,1:T}$ in the conditional outcome probabilities specified in (3), this expression can in general not be evaluated directly. Instead, it can be approximated numerically as will be discussed in the remainder of this section.

3.1 Joint Simulation and Numerical Integration

The simplest approach is to integrate out the full latent process $\mathbf{a}_{i,1:T}$:

$$P(\mathbf{y}_{i,1:T}|\mathbf{x}_i) = \int \cdots \int P(\mathbf{y}_{i,1:T}|\mathbf{x}_i, \mathbf{a}_{1:T}) f(\mathbf{a}_{i,1:T}|\mathbf{x}_i)\, da_{i1} \cdots da_{iT} \tag{12}$$

All terms in this equation are known by assumption. The integral does in general not have an analytic solution. The typical approach in micro-econometrics to these kinds of problems is simulation: Given a number of R draws $[\mathbf{a}^r_{i,1:T} : r = 1, ..., R]$ from the joint distribution $f(\mathbf{a}_{i,1:T}|\mathbf{x}_i)$, the simulated probability is equal to

$$\tilde{P}^{\text{SIM}}(\mathbf{y}_{i,1:T}|\mathbf{x}_i) = \frac{1}{R} \sum_{r=1}^{R} P(\mathbf{y}_{i,1:T}|\mathbf{x}_i; a^r_{i,1:T}). \tag{13}$$

Pseudo-maximum likelihood estimators of $\boldsymbol{\theta}$ using $\tilde{P}^{\text{SIM}}(\mathbf{y}_{i,1:T}|\mathbf{x}_i)$ instead of its true value is under weak regularity conditions consistent (in N) if the number of replications R rises with N (Hajivassiliou and Ruud 1994).

It has been shown in various cases that with a given number of replications R, the accuracy of the simulated probabilities and estimators based on them can improve dramatically if instead of (pseudo-)random draws antithetic or quasi-random draws are used. This can also be expected in this setting.

Instead of simulation, deterministic numerical integration methods can be used to approximate analytically infeasible integrals. While Gaussian quadrature is know to work effectively in univariate integration problems, the integral in (12) is T-dimensional even if a_{it} is one-dimensional and a multiple thereof otherwise. The well-known product rule extension of Gaussian quadrature to multiple dimensions suffers from exponentially rising computational costs as the number of dimensions increases. Even in 3 or 4 dimensions, this "curse of dimensionality" makes this approach computationally inefficient. In higher dimensions, it quickly becomes infeasible even on modern computers.

Heiss and Winschel (2006) suggest to apply a different approach of extending Gaussian quadrature to multiple dimensions for integration problems such as (12). This method of

integration on sparse grids (ISG) does not suffer from the curse of dimensionality. The authors demonstrate its superior performance for a similar estimation problem and for as many as 20 dimensions. This should suffice for the typical micro-econometric model at least with a univariate latent state space. This method is similarly easy to implement as simulation. It prescribes a set of R nodes $[\mathbf{a}_{1:T}^r : r = 1, ..., R]$ and corresponding weights $[\mathbf{w}^r : r = 1, ..., R]$ according to the dimension and distribution of a_{it}. The probability is then approximated as

$$\tilde{P}^{\text{ISG}}(\mathbf{y}_{i,1:T}|\mathbf{x}_i) = \sum_{r=1}^{R} w^r P(\mathbf{y}_{i,1:T}|\mathbf{x}_i; \mathbf{a}_{1:T}^r). \tag{14}$$

The higher the time-series dimension of the data, the worse can all methods of integrating out the full sequence of latent states be expected to work. This is extremely so for the Gaussian integration based on the product rule, but also for SGI, the computational burden rises with the dimensions of integration. While the asymptotic (in R) properties of simulation estimators do not depend on the dimension, the accuracy given a finite number of replications often does, see e.g. Lee (1997).

3.2 Nonlinear filtering

Nonlinear filter techniques separate the full integral in (12) into a sequence of lower-dimensional integrals using the structure of the model. These approaches can be interpreted as a generalization of the Kalman filter to nonlinear models with possibly nonnormal disturbances. For a survey of nonlinear filtering methods, see Tanizaki (2003). Compared to time series models in engineering, finance or macroeconomics for which nonlinear filters are usually discussed, the typical microeconometric panel data model has a low-dimensional state space and a short time-series dimension.

The general idea how to decompose the integral in (12) for the model structure described in section 2.1 is the following. For a simplification of notation, denote $P(\mathbf{y}_{i1}|\mathbf{x}_i, \mathbf{y}_{i,1:0}) =$

$P(\mathbf{y}_{i1}|\mathbf{x}_i)$. By the rules of conditioning, the probabilities of interest can then in general be written as

$$P(\mathbf{y}_{i,1:T}|\mathbf{x}_i) = \prod_{t=1}^{T} P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1}). \tag{15}$$

Each of these terms $P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ are now approximated separately by some $\tilde{P}(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ and the approximated individual likelihood contributions are

$$\tilde{P}^{\text{SEQ}}(\mathbf{y}_{i,1:T}|\mathbf{x}_i) = \prod_{t=1}^{T} \tilde{P}(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1}). \tag{16}$$

Note that the expressions $P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ are nontrivial to calculate. They are complicated functions of past values because of the presence of the unobserved sequence of latent states a_{it}.

The structure of the model allows to obtain these approximations in a sequential fashion. The outcome probabilities conditional on past values can be written as

$$P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1}) = \int P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it})\, f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})\, da_{it} \tag{17}$$

by (3). This equation reflects the model assumption that all dependence of $\mathbf{y}_{it}$ conditional on $\mathbf{x}_i$ is induced by the presence of the latent state process of a_{it}. The problem of this equation is that the conditional distribution $f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ is again a complicated function of past realizations $\mathbf{y}_{i,1:t-1}$.

For $t = 1$, the conditional densities are simply equal to the initial distribution $f(a_{i1}|\mathbf{x}_i)$ which is known by the model specification. For $t > 1$, they can be expressed in a recursive fashion. Suppose that the conditional density $f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ is known so that $P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ can be calculated by (17). Then, density for the next wave $t + 1$ can be derived as follows. First, note that by the conditional independence assumption (4),

$$f(a_{it}, a_{i,t+1}|\mathbf{x}_i, \mathbf{y}_{i,1:t}) = f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t}) f(a_{i,t+1}|\mathbf{x}_i, a_{it}). \tag{18}$$

A marginalization with respect to a_{it} leads to the wanted expression of the conditional distribution for $t + 1$:

$$f(a_{i,t+1}|\mathbf{x}_i, \mathbf{y}_{i,1:t}) = \int f(a_{i,t+1}|\mathbf{x}_i, a_{it})\, f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t})\, da_{it}. \tag{19}$$

The first term in this integral is known by the model specification, the second term can be expressed in terms of known functions. Bayes' rule and the model assumption (3) imply

$$f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t}) = f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1}) \frac{P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it})}{P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})}. \tag{20}$$

A combination of (19) and (20) results in an expression for $f(a_{i,t+1}|\mathbf{x}_i, \mathbf{y}_{i,1:t})$ as a function of terms which are either known by the model specification ($P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it})$ and $f(a_{i,t+1}|\mathbf{x}_i, a_{it})$) or from the previous recursion step ($P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ and $f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$).

The computational problem lies in the fact that the integrals in (17) and (19) generally do not have an analytic solution. For the type of models discussed in this paper, a computational approach based on Gaussian quadrature to approximate these integrals is discussed in the next section.

3.3 Sequential Gaussian Quadrature

Gaussian quadrature prescribes a set of R nodes $[z_r : r = 1, ..., R]$ and corresponding weights $[w_r : r = 1, ..., R]$ for a general integration problem of the form $\int g(z)w(z)\,dz$ which depend on the weighting function $w(z)$. The approximation is then given as $\sum_{r=1}^{R} w_r g(z_r)$. It will be the exact solution of the integral if $g(z)$ is a polynomial of order $2R - 1$ or less. If the integrand is reasonably smooth and can therefore be closely approximated by a polynomial, the approximation can be expected to be very accurate. Gaussian quadrature has long been used for univariate integration problems such as random effects models, see for example Butler and Moffit (1982).

A problem with applying Gaussian quadrature directly to the integrals in (17) and (19) is that the natural weight functions $f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ have no closed-form expression and therefore the appropriate nodes and weights cannot be derived. Therefore, a reformulation of the integrals very much in the spirit of the sequential importance sampling algorithm of e.g. Tanizaki and Mariano (1994) is used here. Define a "proposal density" for which a Gaussian quadrature rule is known and which is as close to $f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ as possible.

For simplicity, I assume that this is the case for the marginal distribution $f(a_{it}|\mathbf{x}_i)$ and use it as the "proposal density". Define the ratio of the two densities as

$$q_{it}(a_{it}) = \frac{f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})}{f(a_{it}|\mathbf{x}_i)}. \tag{21}$$

With this definition, rewrite (17) as

$$P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1}) = \int q_{it}(a_{it})\, P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it})\, f(a_{it}|\mathbf{x}_i)\, da_{it} \tag{22}$$

and combine (19) and (20) to obtain

$$f(a_{i,t+1}|\mathbf{x}_i, \mathbf{y}_{i,1:t}) = \int q_{it}(a_{it})\, f(a_{i,t+1}|\mathbf{x}_i, a_{it}) \frac{P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it})}{P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})}\, f(a_{it}|\mathbf{x}_i)\, da_{it}. \tag{23}$$

This gives a recursion of the "importance weights":

$$q_{i,t+1}(a_{i,t+1}) = \int q_{it}(a_{it}) \frac{f(a_{i,t+1}|\mathbf{x}_i, a_{it})}{f(a_{i,t+1}|\mathbf{x}_i)} \frac{P(\mathbf{y}_{it}|\mathbf{x}_i, a_{it})}{P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})}\, f(a_{it}|\mathbf{x}_i)\, da_{it}. \tag{24}$$

Both integrals in (22) and (24) can now be sequentially approximated by Gaussian quadrature with quadrature nodes $[a^r : r = 1, ..., R]$ and weights $[w^r : r = 1, ..., R]$ appropriate for $f(a_{it}|\mathbf{x}_i)$. Initialize $q_{i1}^r = 1$ for all $r = 1, ..., R$ and for all $t = 1, ..., T$ do the following calculations:

1. Approximate $P(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ as

$$\tilde{P}(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1}) = \sum_{r=1}^{R} q_{it}^r\, P(\mathbf{y}_{it}|\mathbf{x}_i, a^r) w^r \tag{25}$$

2. For all $s = 1, ..., R$, approximate $q_{i,t+1}(a^r)$ as

$$q_{i,t+1}^{*s} = \sum_{r=1}^{R} \frac{f(a^s|\mathbf{x}_i, a^r)}{f(a^s|\mathbf{x}_i)} \frac{q_{it}^r\, P(\mathbf{y}_{it}|\mathbf{x}_i, a^r)}{\tilde{P}(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})} w^r. \tag{26}$$

3.4 Other Nonlinear Filtering Approaches

There are various other approaches to the approximations of the time-specific probabilities $\tilde{P}(\mathbf{y}_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ in (16). In the following, different approaches known in the time series literature are briefly discussed. For a more extensive overview, see Tanizaki (2003).

Very closely related to the sequential is the sequential importance sampling (SIS) approach of Tanizaki and Mariano (1994) and Tanizaki (1999). It uses the same transformations as (21) through (24) but uses Monte Carlo simulation instead of numerical integration to approximate the involved integrals. At least in the case of a one-dimensional state-space discussed in this paper, Gaussian quadrature can be expected to be more accurate in most cases. Furthermore, it uses the same nodes for all observations which can save computational costs.

Both SGQ and the SIS share the problem that the involved importance weights $q_{it}(a_{it}) = \frac{f(a_{it}|\mathbf{x}_i,\mathbf{y}_{i,1:t-1})}{f(a_{it}|\mathbf{x}_i)}$ generally diverge as $t \to \infty$ since more and more information is accumulated in $\mathbf{y}_{i,1:t-1}$. For the moderate time series dimension typical for microeconometric panel data models and with asymptotics in the cross-sectional dimension N with a fixed T, this problem is less problematic than in pure time-series models with asymptotics in T.

For the estimation of long time series, the nonlinear particle filter (NPF) has proved to work efficiently, see Doucet et al., eds (2001) or – for econometric time series – Fernández-Villaverde and Rubio-Ramírez (2005, 2006). Instead of the transformations in (21) through (24), the particle filter generates draws from $f(a_{it}|\mathbf{x}_i, \mathbf{y}_{i,1:t-1})$ by a sequential resampling scheme. While it avoids the problem of diverging importance weights as $T \to \infty$, the involved resampling creates additional noise compared to SIS and is computationally cumbersome. Furthermore, the simulated likelihood contributions are not smooth in the parameters. This impedes gradient-based maximization algorithms for the likelihood function.

In section 2.2, SGQ, SIS, and NPF are compared to the joint simulation and numerical integration methods discussed in section 3.1 for a typical micro-econometric model. For the discussion of other approaches, the reader is referred to the literature, see for example the survey of Tanizaki (2003).

4 Results

For the SRHS model described in section 2.2, this section implements and compares the following algorithms for evaluating the likelihood function discussed above:

- **Random Simulation:** Simulation of the full sequence of the latent state space discussed in section 3.1 using a standard random number generator for drawing from the joint distribution.
- **Antithetic Simulation:** The same algorithm but using Modified Latin Hypercube Sequences (MLHS) instead of random draws. These are straightforward to implement and work effectively in the context of likelihood approximation, see Hess, Train and Polak (2006) for details.
- **Sparse grids integration:** Numerical integration on sparse grids as discussed in section 3.1. The algorithm for generating nodes and weights is given in Heiss and Winschel (2006).[2]
- **Nonlinear particle filter:** the standard nonlinear particle filter (Fernández-Villaverde and Rubio-Ramírez 2006) is implemented with MLHS for the initial state and innovations to improve the performance.
- **Sequential importance sampling:** The algorithm of Tanizaki (1999) except that a fixed grid of nodes instead of antithetic draws are used. For the univariate state space, this proved to be most successful.
- **Sequential Gaussian quadrature** as detailed in section 3.3. Pseudo-Code is shown in the appendix.

Each algorithm is implemented for a different number of nodes R at which all likelihood contributions are to be evaluated. As $R \to \infty$, all methods should converge to the true

[2] Code in Matlab and Stata for generating nodes and weights for integration on sparse grids can be requested from the author.

likelihood. The question is how fast they do and what computational costs each algorithm needs to achieve accurate results. First, the approximate values of the log likelihood function at a fixed parameter vector is calculated and compared to its limiting value. Second, each method is used to estimate the parameter vector by maximizing the approximated likelihood function.

4.1 Approximation of the likelihood function

Figure 1 shows the approximated log likelihood value at a fixed parameter vector for the different algorithms with the numbers of calculations R on the abscissa – note the logarithmic scaling. As expected, all algorithms converge to the same value, but the speed of convergence differs dramatically. Random simulation of the whole sequence converges the slowest and even with 5000 replications, there is still a notable difference to the limiting value. For small R, the approximation is severely biased downwards. This is due to the fact that while outcome probabilities are simulated without bias, the concave log transformation creates downward bias by Jensen's inequality. Antithetic simulation with MLHS performs better requiring roughly half as many evaluations to achieve a comparable accuracy. Sparse grids integration converges notably faster.

Coming to the sequential algorithms, the nonlinear particle filter performs better than the joint algorithms with $R < 100$ but it is still far away from the limiting value and converges slower than sparse grids integration with a higher number $R > 100$. Compared to this, the sequential importance sampling algorithm with a fixed grid of nodes is very successful. With $R = 200$ the results are hardly different from the limiting value. By far the fastest algorithm is sequential Gaussian quadrature. With only $R = 20$ replications, the results are practically indistinguishable from the limit $R \to \infty$.

The number of evaluations of the conditional outcome probabilities R is not the only determinant of computational costs. The additional required calculations are

- The random simulation, antithetic simulation, and nonlinear particle filter require a large number of random numbers. With $R = 5000$, $102233 * 5000 \approx 500$ million

random numbers have to be generated. With 8 bytes storage for each number, this corresponds to roughly 4GB. This is beyond the RAM capacity of most modern personal computers, so the numbers have to be sequentially generated for each likelihood evaluation.

- The resampling step of the nonlinear particle filter is computationally costly.
- The updating of the importance weights of the sequential importance sampling and Gaussian quadrature algorithms is not too expensive with low R but rise quadratically with R since it involves the calculation of R weighted sums over R elements.

To provide a different comparison, Figure 2 shows the same results as Table 1 but with the total time the implemented methods needed for each likelihood evaluation instead of the number of function evaluations. The methods were implemented in Matlab and run on a Pentium 4 PC with 3GHz. Of course, these results depend on how efficiently the different algorithms were coded. While the author tried to do a good job for all algorithms, these results should not be interpreted too literally. As can be seen, the long run times for the resampling step make the nonlinear particle filter less competitive than when just considering the number of evaluations R. The random simulation now performs better than the MLHS because the generation of random numbers is considerably faster. The sequential Gaussian quadrature run very fast, so its advantage over the other methods is at least as pronounced as in Figure 1.

For an gradient-based maximization algorithm with numerical gradients, the likelihood function of the model has to be evaluated 17 times. Assuming 10 iterative steps for the maximization, the total number of likelihood calculations is 170. This translates into a total computing time of about 15 hours for random simulation with $R = 5000$ and 11 minutes for sequential quadrature with $R = 20$ with the latter approach clearly delivering more accurate results.

4.2 Parameter estimation

The ultimate goal of the approximated likelihood functions is to base parameter estimation on them. Intuitively, a better approximation of the likelihood function *ceteris paribus* leads to better estimates based on it. For the different algorithms and accuracy levels discussed above, the model parameters can for example be estimated by maximization of the approximated likelihood. As seen above, the sequential Gaussian quadrature algorithm seems to perform very well with only $R = 20$ function evaluations. To be on the safe side, this algorithm with $R = 50$ is declared as a "reference algorithm". Table 3 shows the QML estimates obtained by this algorithm. Notably, the estimated standard deviation of the latent state process σ is large compared to the standard logistic i.i.d. error term e_{it} which has a standard deviation normalized to $\pi/\sqrt{3} \approx 1.82$. The correlation parameter ρ is large but highly significantly smaller than unity.

Figure 3 shows the estimates of these two most interesting parameters that drive the intertemporal correlation pattern using the different algorithms and number of evaluations R. The qualitative picture is the same as for the likelihood values. While with 20 evaluations, the sequential Gaussian quadrature algorithm has reached its limiting value, the other methods need considerably more computations with random simulation performing worst. The estimated standard deviation σ seems to be downward biased with the simulation methods, while the correlation parameter is upward biased.

A measure of overall deviations of the estimated parameters from their limiting values is shown in Figure 4. It shows the LR test statistic for the null hypothesis that all parameters are equal to the estimates obtained by the different algorithms, where the statistic is calculated for the "reference algorithm". The broad message of this graph is the same as from the previous figures: all methods converge to the parameters obtained by the "reference algorithm" so that the test statistic approaches zero. Sequential quadrature does so extremely faster than the other algorithms.

4.3 Performance of the model

How well is this extremely simple and parsimonious model able to capture the intertemporal correlation patterns observed in Table 2? Table 4 approaches this question. In the first column, the parameter estimates from the descriptive regression on covariates and five lags of SRHS are repeated from Table 2. The other three columns show results from simulated data set. Given the SRHS model, the parameter estimates and the actual covariates in the data, 100 different data set were simulated. The same descriptive regression is then repeated for each of the simulated data sets. The table shows shows the mean and the 95% confidence interval over the 100 repititions.

Most of the parameters are well in line with the original estimates. The coefficients of lagged SRHS are all highly significant and decrease over time. Only the coefficient of the most recent lag (SRHS wave 5) is significantly higher in the original estimates than with the simulated data. This might be an indication that the model specification can be refined, for example the AR(1) process is too simple. Or the correlation parameters differ across the population so that some interactions might help. Remember that the model is estimated on the full sample, whereas the results in Table 4 are obviously for the subsample with six observations. This subsample only contains the original (relatively young) HRS cohort. Overall, the model does a very good job in replicating the intertemporal correlation pattern – much better than e.g. a first-order markov chain model with a random effect which would imply equal coefficients for SRHS in wave 1 through 4.

The general model structure discussed in section 2.1 would also easily allow more elaborate models. An obvious extension would be to add mortality to the measurement model. This would allow a straightforward and model-consistent treatment of the obvious dynamic selection effect through mortality. Another straightforward generalization would be to specify the unobserved state process in continuous time. The only required change would be in the transition equation (10). This would allow to easily take care of the fact that the time between surveys, and therefore probably also the correlation between adja-

cent measures, differ considerably in the HRS. For the discussion of such issues, see Heiss, Börsch-Supan, Hurd and Wise (2006).

5 Conclusions

This paper discusses the numerical approximation of the likelihood for a certain class of nonlinear panel data models including limited dependent variable models with AR(1) error terms. The computational difficulties arise because the likelihood function involves multiple integrals without analytic solutions. While methods for multiple numerical integration are available, their accuracy decreases with a rising dimensionality if the computational effort is held constant. Equivalently, the computational costs for a given accuracy increase with a rising dimensionality.

This paper discusses how these models allow to split the multiple integrals into several integrals with lower dimensions using nonlinear filtering algorithms. In the examples discussed here, the integrals become one-dimensional. Since these integrals are approximated accurately with relatively low computational costs, the overall approximation can be expected to perform better than the "brute force" approach to approximate the joint integral.

There are several approaches to actually implement the sequential evaluation of the likelihood function. In engineering where most of these methods were developed and in the econometric time series literature where they receive increased attention, the number of time periods is high compared to the typical microeconometric panel data. This affects the relative advantages and disadvantages of the algorithms. I suggest an approach that is plausibly very powerful for moderate time series dimensions. It is based on Gaussian quadrature for one-dimensional problems. This allows very precise approximations with little computational effort.

In an application, the panel data modeling of self-rated health status is discussed. It it argued that a simple ordered logit model with an AR(1) error term is more plausible that

the typically specified random effects and/or first-order Markov models. It is also more parsimonious and yet captures the observed intertemporal corralation pattern better. For the estimation of this example model, different algorithms are implemented. The proposed sequential quadrature method dramatically outperforms the typically used approach of joint simulation and also other nonlinear filters. While sequential Gaussian quadrature needs only about 20 function evaluations R for an accurate parameter estimation, the joint simulation still suffers from bias with $R = 5000$. The method is also easily implemented and needs only moderate additional computations.

Appendix: Pseudo-Code for sequential Gaussian quadrature

The sequential Gaussian quadrature algorithm was discussed in section 3.3. In the following, pseudo-code for the implementation of the SRHS model is presented for the convenience of the reader.[3]

1. Preparations:

   - Fix a number of replications R.
   - Obtain R nodes and weights for Gaussian quadrature and store them in the Rx1 vectors `a` and `w`, respectively.
   - For the updating, relative densities are required often. Since they do not change, calculate them once and reuse them every time: Generate a RxR "transition matrix" `m`, where
     `m(r,c)` $= \frac{\phi\left(\texttt{a(r)};\texttt{rho}\cdot\texttt{a(c)},(1-\texttt{rho}^2)\cdot\texttt{sigma}^2\right)}{\phi\left(\texttt{a(r)};0,\texttt{sigma}^2\right)}$ represents $\frac{f(\texttt{a(r)}|\texttt{x}_i,\texttt{a(c)})}{f(\texttt{a(r)}|\texttt{x}_i)}$.

2. For each cross-sectional unit $i = 1, ..., N$

   a) Initialize the vector `q` as a Rx1 vector of ones.

   b) For each wave $t = 1, ..., T$ (T may differ across i):

      - Calculate the Rx1 vector of weighted conditional probabilities `qp` as `qp(r)` $=$ `q(r)` $\cdot P($`y(i,t)`$|$`x(i,t)`$,$ `sigma` $\cdot$ `a(r)`$)$.
      - Approximate the likelihood contribution according to (25) as `L(i,t)` $=$ `qp`$'$`w`.
      - Update the importance weights according to (26) as `q` $= \frac{1}{\texttt{L(i,t)}}($`m*qp`$)'$`w` with "`*`" denoting matrix multiplication.

[3] All calculations for this paper were done in Matlab. The actual code can be requested from the author.

References

Butler, J. S., Moffit, Robert 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* **50**(3): 761–764.

Chamberlain, Gary 1984. Panel data. In *Handbook of Econometrics,* ed. Zvi Griliches and Michael D. Intriligator, vol. II. Amsterdam, New-York: Elsevier pp. 1247–1318.

Contoyannis, Paul, Jones, Andrew M, Rice, Nigel 2004. The dynamics of health in the british household panel survey. *Journal of Applied Econometrics* **19**: 473 – 503. Mimeo, University of York.

Danielsson, Jon, Richard, Jean-François 1993. Accelerated gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics* **8**: S153–S173.

Doucet, Arnaud, Nando De Freitas, Neil Gordon, eds 2001 *Sequential Monte Carlo Methods in Practice.* New York: Springer Verlag.

Fernández-Villaverde, Jesús, Rubio-Ramírez, Juan F. 2005. Estimating dynamic equilibrium economies: Linear versus nonlinear likelihood. *Journal of Applied Econometrics* **20**: 891–910.

___ 2006. Estimating dynamic equilibrium economies: Linear versus nonlinear likelihood. Technical Report, NBER Technical Working Paper 321.

Gordon, Neil J., Salmond, D. J., Smith, A. F. M. 1993. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F* **140**(2): 107–113.

Hajivassiliou, Vassilis A., Ruud, Paul A. 1994. Classical estimation methods for LDV models using simulation. In *Handbook of Econometrics Vol. IV,* ed. Robert F. Engle and Daniel L. McFadden. New-York: Elsevier pp. 2383–2441.

Hamilton, James D. 1994. State-space models. In *Handbook of Econometrics Volume 4,* ed. R. Engle and D. McFadden. North-Holland chapter 50.

Heckman, James J. 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process. In *Structural Analysis of Discrete Data and Econometric Applications,* ed. Charles F. Manski and Daniel McFadden. Cambridge, Mass.: MIT Press pp. 179–195.

Heiss, Florian, Börsch-Supan, Axel, Hurd, Michael, Wise, David 2006. Pathways to disability: Predicting health trajectories. In *Perspectives on the Economics of Aging,* ed. David Wise. University of Chicago Press. forthcoming.

Heiss, Florian, Winschel, Viktor 2006. Estimation with numerical integration on sparse grids. Technical Report, Department of Economics Discussion paper No. 2006-15, University of Munich.

Hess, Stephane, Train, Kenneth E., Polak, John W. 2006. On the use of a modified latin hypercube sampling (mlhs) method in the estimation of a mixed logit model for vehicle choice. *Transportation Research Part B* **40**: 147–163.

Honoré, Bo E., Tamer, Elie 2006. Bounds on parameters in panel dynamic discrete choice models. *Econometrica* **74**: 611–629.

Lee, Lung-Fei 1997. Simulated maximum likelihood estimation of dynamic discrete choice statistical models: Some monte carlo results. *Journal of Econometrics* **82**: 1–35.

Tanizaki, H., Mariano, R.S. 1994. Prediction, filtering and smoothing in non-linear and non-normal cases using monte carlo integration. *Journal of Applied Econometrics* **9**(2): 163–179.

Tanizaki, Hisashi 1999. Nonlinear and nonnormal filter using importance sampling: Antithetic monte-carlo integration. *Communications in Statistics, Simulation and Computation* **28**: 463–486.

____ 2003. Nonlinear and non-gaussian state-space modeling with monte carlo techniques: A survey and comparative study. In *Handbook of Statistics,* ed. D.N. Shanbhag and C.R. Rao, vol. 21. Elsevier pp. 871–929.

Wooldridge, Jeffrey M. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* **20**: 39–54.

Table 1: Distribution of SRHS

	poor	fair	good	very good	excellent
Frequency	10,099	19,579	30,811	27,665	14,079
Percent	9.9	19.2	30.1	27.1	13.8
By previous SRHS [%]:					
poor	56.9	30.5	9.5	2.4	0.8
fair	16.6	48.0	26.5	7.3	1.6
good	4.4	19.1	49.9	22.4	4.2
very good	1.8	6.7	27.7	50.6	13.3
excellent	1.0	3.2	12.9	33.9	48.9

Table 2: Ordered Logit of SRHS in wave 6 on past SRHS

age	-0.0188	(0.006)**
female	0.0823	(0.046)+
high school	0.1861	(0.065)**
some college	0.1762	(0.076)*
college degree+	0.3308	(0.078)**
nonwhite	-0.1198	(0.063)+
SRHS wave 5	0.9847	(0.039)**
SRHS wave 4	0.5175	(0.038)**
SRHS wave 3	0.3251	(0.036)**
SRHS wave 2	0.2034	(0.035)**
SRHS wave 1	0.2390	(0.032)**
Observations	7173	
Log likelihood	-7663.4	

Robust SE in parantheses, $+ : p < 0.10, * : p < 0.05, ** : p < 0.01$

Table 3: Parameter estimates (sequ. quadrature with $R = 50$)

age splines: 50+	−0.1069	(0.0048)**
age splines: 60+	0.0624	(0.0076)**
age splines: 70+	−0.0485	(0.0082)**
age splines: 80+	−0.0074	(0.0122)
age splines: 90+	0.0912	(0.0306)**
female	0.0828	(0.0360)**
nonwhite	−1.0119	(0.0465)**
high school	1.2658	(0.0380)**
some college	1.8584	(0.0474)**
college degree+	2.7922	(0.0509)**
Latent states a_{it}: SD σ	2.8764	(0.0276)**
Latent states a_{it}: corr. ρ	0.9439	(0.0128)**
Individuals	25,353	
Observations	102,233	
Log likelihood	-128,311.0	

Robust SE in parantheses, $** : p < 0.01$

Table 4: Correlation patterns: ordered logit on original and simulated data

	original	simulated data		
	data	mean	2.5%	97.5%
age	-0.0188	-0.0036	-0.0140	0.0083
female	0.0823	0.0214	-0.0726	0.1138
high school	0.1861	0.1590	0.0294	0.2853
some college	0.1762	0.2310	0.0938	0.4135
college degree+	0.3308	0.3642	0.2209	0.5177
nonwhite	-0.1198	-0.1286	-0.2601	-0.0196
SRHS wave 5	0.9847	0.7832	0.7206	0.8481
SRHS wave 4	0.5175	0.4874	0.4255	0.5567
SRHS wave 3	0.3251	0.3161	0.2470	0.3795
SRHS wave 2	0.2034	0.2112	0.1394	0.2804
SRHS wave 1	0.2390	0.1676	0.0937	0.2206

Figure 1: Approximate log likelihood at fixed parameter vector

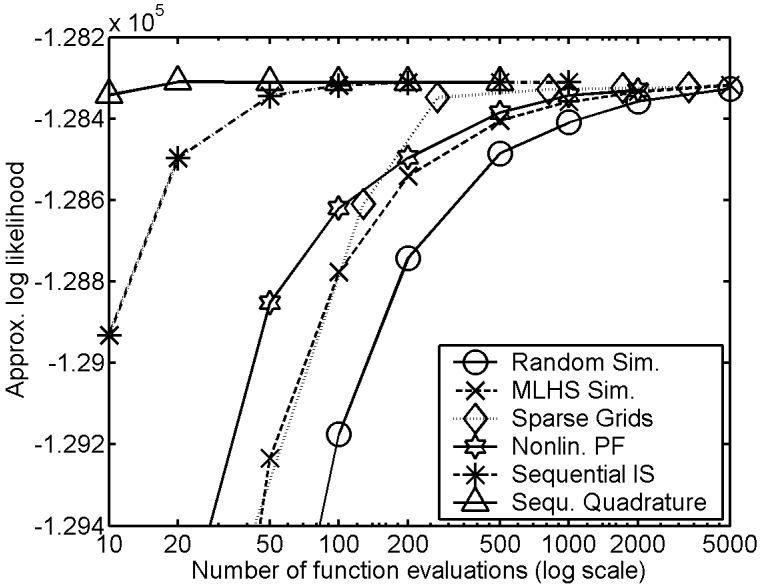

Figure 2: Approximation of the likelihood function by computational costs

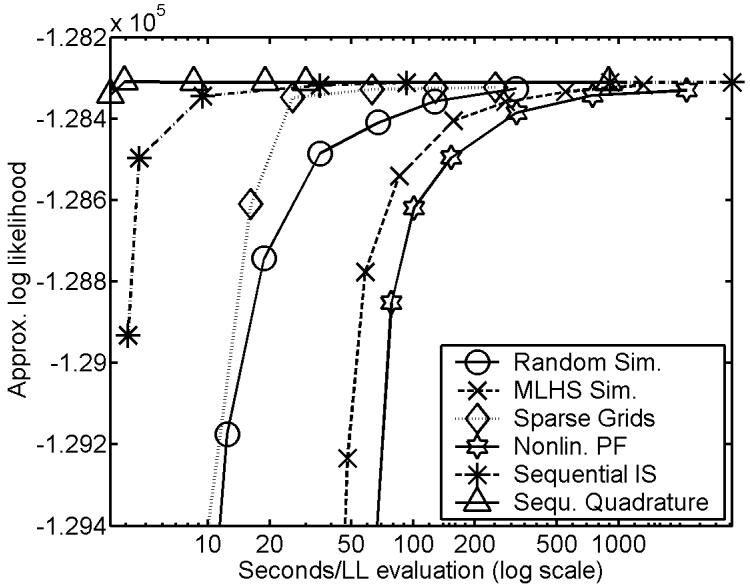

Figure 3: Results: Estimated Parameters σ and ρ

(a) σ (b) ρ

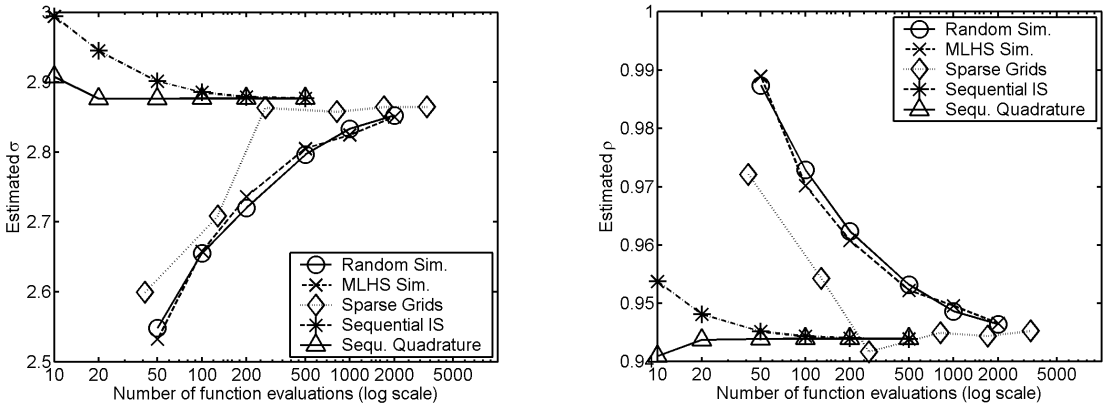

Figure 4: Results: LR statistic for estimated parameters