Discussion Papers No. 295, March 2001
Statistics Norway, Research Department

Arvid Raknerud

A State Space Approach for Estimating VAR Models for Panel Data with Latent Dynamic Components

Abstract:
The econometric literature offers various modeling approaches for analyzing micro data in combination with time series of aggregate data. This paper discusses the estimation of a VAR model that allows unobserved heterogeneity across observation unit, as well as unobserved time-specific variables. The time-latent component is assumed to consist of a persistent and a transient term. By using a Helmert-type orthogonal transformation of the variables it is demonstrated that the likelihood function can be expressed on a state space form. The dimension of the state vector is low and independent of the time and cross section dimensions. This fact makes it convenient to employ an ECM algorithm for estimating the parameters of the model. An empirical application provides new insight into the problem of making forecasts for aggregate variables based on information from micro data.

Keywords: State space models, panel vector autoregressions, random components, latent time series, maximum likelihood, Kalman filter, Helmert transformation, aggregation, prediction.

JEL classification: C13, C15, C33, C53.

Acknowledgement: This paper is developed from Chapter 2 of my Ph.D. thesis "Dynamic Models for Panel Data: Methods and Applications", University of Oslo, 1999. Earlier versions of the paper have been presented at seminars at Stanford University, University of Oslo, and at the 1999 Econometric Society European Meeting in Santiago de Compostela. A computer program is available upon request. Extensive comments form Erik Biørn, Søren Johansen, Tor Jakob Klette, Tore Wentzel Larsen, Tore Schweder, Anders Rygh Swensen, Dag Tjøstheim, Rolf Aaberge, and seminar participants are highly appreciated.

Address: Arvid Raknerud, Statistics Norway, Research Department, P.O. Box 8131 Dep., N- 0033 Oslo, Norway. E-mail: arvid.raknerud @ssb.no

Discussion Papers	comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of
Discussion Papers are available on the Internet: http://www.ssb.no

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
N-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

Most of the literature on dynamic models for panel data with continuous variables are concerned with single-equation models. Important exceptions include Holtz-Eakin, Newey, and Rosen (1988), who analyze an instrumental variable (IV) estimator for a vector autoregressive (VAR) model, and Binder, Hsaio, and Pesaran (2000), who study inferences based on the Generalized Method of Moments (GMM) and Maximum Likelihood (ML) in a similar model. The purpose of the present paper is to analyze a dynamic model with a more general latent structure than in previous studies. The point of departure is a VAR model with a two-way latent structure with both subject-specific and time-specific effects, the latter being modelled as a latent time series consisting of a persistent and a transient component. Maximum likelihood estimators are developed with respect to computational procedures and asymptotic properties based on a state space representation of the model.

In an illustrative example, the proposed method is applied on a micro panel data set to study the relationship between labor demand and real wage in manufacturing firms for the period 1976-96. Focusing on forecasting of aggregate sectorial variables, it is first demonstrated that the in-sample predictions from an estimated micro model are similar to those from the aggregate version of the model. Next, the models are estimated on data for 1976-86, and then used to calculate one-year-ahead forecasts over the remaining out-of-sample period. The micro model performs well also out-of-sample, but the aggregate model fails dramatically. Furthermore, while the parameter estimates of the micro model are remarkably stable over time, the estimates of the aggregate model become strikingly different when the estimation period changes from 1976-96 to 1976-86. The discrepancies between the two approaches are mainly due to the role of industry-wide latent variables, and illustrate the important role of panel data for extracting commonalities across a panel of observation units. Similar observations have previously been made by macro economists; see Quah and Sargent (1993).

The main theme of this paper is to demonstrate the usefulness of state space formulations and the Kalman filter for analyzing dynamic panel data. Although state space models have been applied to panel data earlier; see Jones (1993) and Fahrmeir and Tutz (1994), they have not been explored in the context considered here, with latent structure both across observation units and along time.

So far, GMM estimation based on instrumental variables has dominated the econometric literature about dynamic panel data (see Arellano and Bover (1995) and Ahn and Schmidt (1995)). Such methods can be very inefficient in models with near unit roots, as pointed out by e.g. Blundell and Bond (1998). Moreover, the specific choices of instruments in applications tend to be somewhat arbitrary. Maximum likelihood estimators have also been studied; e.g. by Anderson and Hsiao (1982), Jones (1993), Icaza and Jones (1999), and Binder, Hsaio and Pesaran (2000), but have been less popular in

applications; apparently because of computational obstacles. An interesting recent discussion about the merits of likelihood based methods in panel data models is found in Sims (2000).

For the random components layout proposed in this paper, the situation is more complicated than for the "workhorse" panel VAR model with fixed time effects and random individual effects. Nevertheless, it will be shown that likelihood methods can be attractive in this context, both from a theoretical and a computational point of view. In particular, it is demonstrated that a Helmert-type orthogonal transformation of the variables permits us to concentrate the likelihood function into a simple state space form. The dimension of the state vector is low and independent of the time and cross section dimensions. Thus there is no "curse of dimensionality" associated with the method when N and/or T is large. An iterative ECM algorithm (see Meng and Rubin (1993)) is proposed for computing the maximum likelihood estimator, and its convergence and asymptotic properties are assessed.

The rest of this paper is organized as follows: Section 2 presents and motivates the model; Section 3 establishes an implicit representation of the likelihood function in terms of a state space model; Section 4 presents the ECM algorithm which is used to estimate the model and discusses initialization, implementation of rank restrictions on covariance matrices, and initial conditions; Section 5 applies the model on a micro panel data set and discusses forecasting and aggregation; Section 6 establishes consistency and asymptotic normality of the ML estimator when T is fixed and N goes to infinity and provides Monte Carlo evidence about finite sample properties, including cases with large T; Section 7 concludes.

2 The Model

The starting point of the analysis is a vector autoregressive VAR(k) model with a two-way random components structure:

$$y_t^i = \sum_{j=1}^{k} \Pi_j y_{t-j}^i + \sum_{j=0}^{m} \Lambda_j x_{t-j}^i + f_t + v^i + \varepsilon_t^i. \tag{1}$$

For $i = 1, ..., N$ and $t = 1, ..., T$; y_t^i is a $p \times 1$ vector of endogenous variables on subject i at time t; x_t^i is a $q \times 1$ vector of exogenous variables, and f_t is a latent time series which is composed of a pure random walk μ_t and a transient component δ_t:

$$f_t = \mu_0 + \Gamma \mu_t + \Upsilon \delta_t, \tag{2}$$

where

$$\begin{aligned} \mu_t &= \begin{cases} 0 & t=1 \\ \mu_{t-1}+\eta_t & t=2,..,T \end{cases} \\ \eta_t &\sim \mathcal{IN}_p(0, I_p) \\ \delta_t &\sim \mathcal{IN}_p(0, I_p) \end{aligned} \tag{3}$$

(where 0 denotes a vector of zeros of appropriate dimension and I_p is the identity matrix of order p). Finally, ε_t^i and v^i are p-dimensional standard normal variates:

$$\varepsilon_t^i \sim \mathcal{IN}_p(0, \Sigma) \tag{4}$$

$$v^i \sim \mathcal{IN}_p(0, \Omega), \tag{5}$$

with ε_t^i, v^j, μ_s and δ_u being independent of each other for every (i, j, t, s, u).

The model contains the following unknown parameters: matrices of regression coefficients Π_j and Λ_j, with dimension $p \times p$ and $p \times q$, respectively; an intercept vector μ_0; lower triangular loading matrices Γ and Υ, and a random effects covariance matrix Ω. Since any initial value μ_1 can be absorbed into the intercept μ_0 in equation (2), $\mu_1 = 0$ is chosen as an identifying restriction. Because

$$Var(f_t) = (t-1)\Gamma\Gamma' + \Upsilon\Upsilon',$$

Γ and Υ are identified as the unique Cholesky factors in a covariance matrix decomposition of the time latent variable f_t, provided these matrices are positive definite. Rank restrictions will be considered in Section 4.5.

Two interesting special cases emerge if we set *(i)* $\Omega \equiv 0$ – no individual effects, or *(ii)* $\Gamma = \Upsilon \equiv 0$ – no stochastic time effects. In the first case, maximum likelihood estimation can be carried out quite straightforwardly, since the model then is easily seen to have a state space representation with state variables (μ_t, δ_t) and measurement-type equation (1). In the latter case, the likelihood function has an attractive explicit form, as will be demonstrated in Section 3.

Let $\Pi = [\Pi_1, ..., \Pi_k]$ and $\Lambda = [\Lambda_0, ..., \Lambda_m]$. We partition the vector θ of unknown parameters as $\theta = (\beta, \phi)$, where $\beta = (vec(\Pi)', vec(\Lambda)', vech(\Sigma)', vech(\Omega)')$ and $\phi = (\mu_0', vech(\Gamma)', vech(\Upsilon)')$. Thus ϕ contains the parameters of the latent time series, while β contains all other parameters. Clearly, we cannot estimate μ_0, Γ and Υ consistently with fixed T (however, if $\Upsilon = 0$ the intercept vector μ_0 is identifiable from cross sectional data alone).

The non-stationarity of f_t does not affect large-N asymptotics, as is demonstrated in Section 6. For large-N *and* large-T asymptotics, which would allow us to make inference about ϕ, the situation is much more complicated. Conventional limit theorems do not allow multi-index asymptotics, and

the relatively few existing results in this field (see Phillips and Moon (2000) for an overview) are not applicable to the case with both serial and cross-sectional dependence. However, the Monte Carlo evidence presented in Section 6 are encouraging and suggest that conventional normal-theory approximations may be valid also in situations with large N and T.

To motivate equation (1)-(2), it is instructive to look at the aggregate version of the model. If we denote by y_t and x_t the mean of y_t^i and x_t^i over an "infinite" number of fixed observation units, we obtain

$$y_t = \mu_0 + \sum_{j=1}^{k} \Pi_j y_{t-j} + \sum_{j=0}^{m} \Lambda_j x_{t-j} + \Gamma \mu_t + e_t, \tag{6}$$

where $e_t \sim \mathcal{IN}(0, \Upsilon\Upsilon')$. Hence, it is seen that if $\Lambda_j = 0$ for all j and $\Gamma = 0$, the VAR(k) model is obtained. On the other hand, if $\Gamma \neq 0$, the stochastic trend μ_t can be interpreted as an intercept correction (see Clements and Hendry (1996)). Another interesting case emerges if $\Lambda_j = \Pi_j = 0$ for all j – the so-called local level model (see Harvey (1989)). Thus, by imposing different parameter constraints it is seen that a number of commonly applied time series models are obtained through aggregation of the model.

In the literature about dynamic panel data much effort have been made to specify proper initial conditions (see Anderson and Hsiao (1982) and Blundell and Bond (1998)). When deriving the likelihood function in the next section, we will condition on the initial observation vectors $Y_0^i = \{y_0^i, .., y_{1-k}^i\}$ and $X_0^i = \{x_0^i, .., x_{1-m}^i\}$, and assume that these are ancillary statistics with respect to θ (see Johansen (1995)), and independent of the individual effect v^i. However, the latter assumption may be unrealistic and is relaxed in Section 4.6.

Let $F_t = vec[f_1, ..., f_t]$ and $H_{t-1}^i = \{(x_s^i, y_s^i), s < t\}$. *Exogeneity* of x_t^i with respect to θ requires that x_t^i is conditionally independent of v^i and F_t given H_{t-1}^i, and independent of θ in distribution (cf. the definition of sequential exogeneity in Gourieroux and Montfort (1995)). If for any random variables X and Y, we let $g(Y|X)$ be generic notation for the conditional density of Y given X, we can state this requirement formally as

$$g(x_t^i, y_t^i | H_{t-1}^i, F_t, v^i; \theta) = g(y_t^i | x_t^i, H_{t-1}^i, F_t, v^i; \theta) g(x_t^i | H_{t-1}^i). \tag{7}$$

If (Y_0^i, X_0^i) are exogenous w.r.t. θ, and $L(\theta)$ is the likelihood function, we have by (7)

$$\begin{aligned} L(\theta) &= \int \left\{ \prod_{i=1}^{N} \int \left[\prod_{t=1}^{T} g(x_t^i, y_t^i | H_{t-1}^i, F_t, v^i; \theta) \right] g(v^i; \theta) dv^i \right\} \prod_{t=1}^{T} g(f_t | F_{t-1}; \theta) dF_T \\ &\times \prod_{i=1}^{N} g(Y_0^i, X_0^i) \\ &= \int \left\{ \prod_{i=1}^{N} \int \left[\prod_{t=1}^{T} g(y_t^i | x_t^i, H_{t-1}^i, F_t, v^i; \theta) \right] g(v^i; \theta) dv^i \right\} \prod_{t=1}^{T} g(f_t | F_{t-1}; \theta) dF_T \qquad (8) \\ &\times \prod_{i=1}^{N} \prod_{t=1}^{T} g(x_t^i | H_{t-1}^i) \times \prod_{i=1}^{N} g(Y_0^i, X_0^i). \qquad (9) \end{aligned}$$

Estimation based on the *partial likelihood* − the factor (8) only − is therefore fully efficient provided *(i)* $g(x_t^i | H_{t-1}^i)$ and $g(Y_0^i, X_0^i)$ contain no unknown parameters or, if they do, these parameters and θ are variation free, and *(ii)* (Y_0^i, X_0^i) and v^i are independent. However, the latter restriction can be relaxed. In Section 4.6 we assume instead that (Y_0^i, v^i) have a joint normal distribution.

Let Z_t^i denote the sum of the latent variables in (1):

$$\begin{aligned} Z_t^i &= f_t + v^i + \varepsilon_t^i \\ &= y_t^i - \sum_{j=1}^{k} \Pi_j y_{t-j}^i - \sum_{j=0}^{m} \Lambda_j x_{t-j}^i. \end{aligned} \qquad (10)$$

We see that Z_t^i has a rather complicated, but identifiable, covariance structure in the time- and cross section dimensions:

$$\mathrm{Cov}(Z_s^i, Z_t^j) = \begin{cases} (t-1)\Gamma\Gamma' + \Upsilon\Upsilon' + \Omega + \Sigma & t = s,\ i = j \\ ((s \wedge t) - 1)\,\Gamma\Gamma' + \Omega & t \neq s,\ i = j \\ (t-1)\Gamma\Gamma' + \Upsilon\Upsilon' & t = s,\ i \neq j \\ ((s \wedge t) - 1)\,\Gamma\Gamma' & t \neq s,\ i \neq j. \end{cases} \qquad (11)$$

3 The Likelihood Function

The purpose of this section is to derive a representation of the likelihood function which is convenient for estimation purposes. Of course, the likelihood function corresponding to (1)-(5) is that of a multivariate normal distribution with covariance structure (11). Unfortunately, this is a pTN- dimensional distribution, which is a highly complicated function of the unknown parameters. Furthermore, there is no simple factorization of the density into conditionally independent terms due to the two-way random components structure. Instead, an indirect representation of the likelihood function is proposed. It will be shown that a particular orthogonal transformation of the variables enables us to represent the likelihood function implicitly on a state space form with a $4p$-dimensional state vector, regardless of N and T. In Section 4, this state space form is used when developing an algorithm for maximizing the likelihood function.

3.1 The Helmert Transform

Define the orthonormal $T \times T$ Helmert-type matrix

$$W = [w_1, .., w_T],$$

with

$$w_t = \begin{cases} \sqrt{\frac{t}{t+1}} \times \left[-\frac{1}{t}, ..., -\frac{1}{t}, 1, 0, .., 0\right]' & t = 1, .., T-1 \\ \left[\frac{1}{\sqrt{T}}, ..., \frac{1}{\sqrt{T}}\right]' & t = T, \end{cases} \tag{12}$$

where the 1 is in the $t+1$'th position of w_t. Furthermore, define $p \times T$ matrices

$$Z^i = [Z_1^i, ..., Z_T^i]$$

$$f = [f_1, ..., f_T]$$

$$E^i = [\varepsilon_1^i, ..., \varepsilon_T^i].$$

Since (10) holds for $t = 1, ..., T$, we have for $i = 1, ..., N$:

$$\begin{aligned} (Z^i - f) &= [v^i, ..., v^i]_{(p \times T)} + E^i \\ &= \left[0, ..., 0, \sqrt{T} v^i\right] W' + E^i. \end{aligned} \tag{13}$$

Postmultiplying (13) by W gives an equivalent formulation

$$Z^i W = fW + \left[0, ..., 0, \sqrt{T} v^i\right] + \zeta^i, \tag{14}$$

with

$$\zeta^i = E^i W.$$

Since W is orthonormal, and ε_t^i and ε_s^i are normal and independent when $s \neq t$, the distribution of ζ^i is the same as the distribution of E^i (see Anderson (1984), p. 68). Thus, by (4),

$$\zeta_t^i = E^i w_t$$

is mean zero Gaussian with

$$E(\zeta_t^i \zeta_s^{i\,\prime}) = \begin{cases} 0 & t \neq s \\ \Sigma & t = s. \end{cases} \tag{15}$$

Define

$$z_t^i = \begin{cases} Z^i w_t & t = 1, ..., T-1 \\ T^{-1/2} Z^i w_T & t = T. \end{cases} \tag{16}$$

Then it follows from (14)-(15) that

$$z_t^i | (f, v^i) \backsim \begin{cases} \mathcal{N}_p(f w_t, \Sigma) & t = 1, .., T-1 \\ \mathcal{N}_p(T^{-1/2} f w_T + v^i, T^{-1}\Sigma) & t = T, \end{cases} \tag{17}$$

where z_t^i and z_s^i are conditionally independent given f and v^i whenever $s \neq t$. We can use (17) to obtain the conditional density of $z_t^i|f$. By the rule of iterated expectation, $E(z_t^i|f) = E(E(z_t^i|f, v^i))$ and $Var(z_t^i|f) = E(Var(z_t^i|f, v^i)) + Var(E(z_t^i|f, v^i))$. Hence, by normality, (5), and (17)

$$z_t^i|f \backsim \begin{cases} \mathcal{N}_p(fw_t, \Sigma) & t = 1, .., T-1 \\ \mathcal{N}_p(T^{-1/2} f w_T, \Omega + T^{-1}\Sigma) & t = T. \end{cases} \tag{18}$$

Conditional on f, the z_t^i are independent random variables for $i = 1, ..., N$ and $t = 1, ..., T$.

From (18) and (3), the *marginal* probability density of z_t^i is implicitly given by the following equations:

$$\begin{aligned} z_t^i &= \begin{cases} fw_t + \xi_t^i & t = 1, ..., T-1 \\ T^{-1/2} f w_T + \xi_T^i & t = T \end{cases} \\ f_t &= \mu_0 + \Gamma \mu_t + \Upsilon \delta_t \\ \mu_t &= \begin{cases} 0 & t = 1 \\ \mu_{t-1} + \eta_t & t = 2, ..., T \end{cases} \\ \eta_t &\backsim \mathcal{IN}_p(0, I_p) \\ \delta_t &\backsim \mathcal{IN}_p(0, I_p) \\ \xi_t^i &\backsim \mathcal{IN}_p(0, \Xi_t), \end{aligned} \tag{19}$$

where

$$\Xi_t = \begin{cases} \Sigma & t = 1, ..., T-1 \\ \Omega + T^{-1}\Sigma & t = T. \end{cases} \tag{20}$$

The relation between z_t^i and the observed variables y_t^i is given by (10), (12), and (16), and thus involves the unknown parameters β. z_t^i is therefore a latent variable.

Equations (19) provide an implicit form of the likelihood function of the data $Y = \{y_t^i\}_{t,i}$ for all (i, t), although it is not immediately clear that this is a particularly useful representation for estimation purposes. However, it is shown in the next sub-section that the choice of transformation matrix W leads to a simple state space representation in terms of the variables μ_t, δ_t, and z_t^i.

If W instead was chosen as the matrix with last column equal to $T^{-\frac{1}{2}} w_T$ and t'th column equal to $\iota_t - T^{-\frac{1}{2}} w_T$ $(t < T)$, where ι_t is the t'th column of the identity matrix, the transformation in (14) would correspond to be the familiar "within" and "between" transformations: The last column would contain the individual means of the variables and the other columns the deviations from the means. However, this transformation would not be orthogonal.

3.2 A State Space Representation

Equations (19) bear a resemblance to a state space form. Moreover

$$fw_t = \begin{cases} \sqrt{\frac{t}{t+1}}(f_{t+1} - \frac{1}{t}\sum_{s=1}^{t} f_s) & t = 1, .., T-1 \\ \frac{1}{\sqrt{T}} \sum_{s=1}^{T} f_s & t = T. \end{cases}$$

We see that each of the transformed variables fw_t for $t < T$ only involve $f_1, .., f_{t+1}$ – and in a very appealing way: fw_t is (except for a scale factor) the difference between f_{t+1} and the mean of $f_1, .., f_t$. In fact, the variable fw_t can be written as a linear combination of VAR(1) processes. To see this, define

$$\alpha_t = \begin{cases} \begin{bmatrix} 0' & \delta_1' & 0' & 0' \end{bmatrix}' & t = 0 \\ \begin{bmatrix} \mu_{t+1}' & \delta_{t+1}' & \sum_{s=1}^{t} \mu_s' & \sum_{s=1}^{t} \mu_s' \end{bmatrix}' & t = 1, ..., T. \end{cases}$$

Then, from (19)

$$\begin{array}{ll} z_t^i = \tau_t + G_t \alpha_t + \xi_t^i & t = 1, ..., T \\ \alpha_t = F \alpha_{t-1} + \omega_t & t = 1, ..., T, \end{array} \tag{21}$$

where

$$\begin{aligned} \tau_t &= \begin{cases} \mu_0 & t = T \\ 0 & \text{else} \end{cases} \\ G_t &= \begin{cases} \begin{bmatrix} \sqrt{\frac{t}{t+1}}\Gamma & \sqrt{\frac{t}{t+1}}\Upsilon & -\sqrt{\frac{t}{t(t+1)}}\Gamma & -\sqrt{\frac{t}{t(t+1)}}\Upsilon \end{bmatrix} & t = 1, ..., T-1 \\ \begin{bmatrix} 0 & 0 & T^{-1}\Gamma & T^{-1}\Lambda \end{bmatrix} & t = T \end{cases} \\ F &= \begin{bmatrix} I_p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ I_p & 0 & I_p & 0 \\ 0 & I_p & 0 & I_p \end{bmatrix} \\ \omega_t &= \begin{bmatrix} \eta_{t+1} \\ \delta_{t+1} \\ 0 \\ 0 \end{bmatrix} \text{ for } t = 1, ..., T. \end{aligned} \tag{22}$$

Note that, by (3), ω_t is *i.i.d.* normal, with mean zero and singular covariance matrix

$$E(\omega_t \omega_t') \equiv Q = \begin{bmatrix} I_p & 0 & 0 & 0 \\ 0 & I_p & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ for } t = 1, ..., T.$$

The importance of (21) is that, regardless of T and N, the dimension of the state vector α_t is only $4p$, where p is the number of equations.

To obtain a state space representation for the complete set of N subjects, we only need to stack observation- and coefficient matrices:

$$\begin{array}{ll} Z_t^* = \tau_t^* + G_t^* \alpha_t + \xi_t^* & t = 1, ..., T \\ \alpha_t = F \alpha_{t-1} + \omega_t & t = 1, ..., T, \end{array} \tag{23}$$

where

$$\begin{aligned} Z_t^* &= (z_t^{1\prime}, .., z_t^{N\prime})' \\ \tau_t^* &= (\tau_t', .., \tau_t')' \\ G_t^* &= (G_t', ..., G_t')' \\ \xi_t^* &= (\xi_t^{1\prime}, .., \xi_t^{N\prime})'. \end{aligned}$$

Thus, by (20),

$$E(\xi_t^*\xi_t^{*\prime}) \equiv \Sigma_t = \begin{cases} I_N \otimes \Sigma & t = 1, ..., T-1 \\ I_N \otimes (\Omega + T^{-1}\Sigma) & t = T. \end{cases} \tag{24}$$

To formulate the z_t^i in (21) in terms of the observable variables y_t^i, we must examine the relationship between z_t^i and y_t^i. We start by defining a "moving average" operator. For any time series χ_t^i, define

$$\chi_{t;s}^i = \frac{1}{s}\sum_{v=0}^{s-1} \chi_{t-v}^i. \tag{25}$$

$\chi_{t;s}^i$ is thus the mean of the latest s observations up until time t. Next, define

$$\begin{aligned}
Y_{1t}^i &= \begin{cases} \sqrt{\frac{t}{t+1}}(y_{t+1}^i - y_{t;t}^i) & t = 1, ..., T-1 \\ y_{T;T}^i & t = T \end{cases} \\
Y_{0t}^i &= \begin{cases} \sqrt{\frac{t}{t+1}}\left[(y_t^i - y_{t-1;t}^i)', ..., (y_{t-k+1}^i - y_{t-k;t}^i)'\right]' & t = 1, ..., T-1 \\ [y_{T-1;T}^{i\prime}, ..., y_{T-k;T}^{i\prime}]' & t = T \end{cases} \\
X_{0t}^i &= \begin{cases} \sqrt{\frac{t}{t+1}}\left[(x_{t+1}^i - x_{t;t}^i)', ..., (x_{t-m+1}^i - x_{t-m;t}^i)'\right]' & t = 1, ..., T-1 \\ [x_{T;T}^{i\prime}, ..., x_{T-m;T}^{i\prime}]' & t = T. \end{cases}
\end{aligned} \tag{26}$$

From (10), (16) and (26) we can rewrite z_t^i as

$$z_t^i = Y_{1t}^i - \Pi Y_{0t}^i - \Lambda X_{0t}^i \quad \text{for } t = 1, ..., T. \tag{27}$$

Since z_t^i depends on β, we will use the notation $z_t^i(\beta)$ to mean the left side of (27) as a function of β.

Above a state space model with α_t as state vector and z_t^i as "observation" vector has been established. Since z_t^i depends on the unknown parameters β, the state space form cannot be used to maximize the likelihood function directly as in Harvey (1989); Ch. 4. Instead I will propose an indirect maximization method, which is similar to that described in Fahrmeir and Tutz (1994); Ch. 8. This method is the subject of the next section.

4 Estimation

The main tool in the estimation of the model (1)-(5) based on the state space formulation (23), is a generalization of the EM (Expectation Maximization) algorithm, called the ECM (Expectation Conditional Maximization) algorithm by Meng and Rubin (1993). The EM algorithm was originally developed by Dempster, Laird and Rubin (1977) as a tool for estimating models with incomplete or missing data, when the likelihood of the complete data has a simple explicit form. In our case, y_t^i is the observed (incomplete) data and the state vector α_t is "missing." Conditional on $F = vec[f_1, ..., f_T]$, the explicit likelihood of the y_t^i is a relatively simple multivariate normal distribution given by (18).

But the likelihood of the "incomplete" data is complex, since we have to integrate out the high-dimensional vector F, leading to a likelihood function which is a highly non-linear function of the unknown parameters.

4.1 The EM and ECM Algorithm.

For general random vectors z and α, let z denote the incomplete data and α the missing data. Furthermore, let $g(z, \alpha; \theta)$ be their joint density (i.e. the "complete data" density), and $g(\alpha|\, z; \theta)$ the conditional density of α given z. The ML estimator, $\widehat{\theta}$, is the maximum of the log-likelihood $L(\theta)$ of the observed data, where

$$L(\theta) = \ln g(z; \theta). \tag{28}$$

Since

$$g(z; \theta) = \frac{g(z, \alpha; \theta)}{g(\alpha|\, z; \theta)},$$

(28) can be rewritten as

$$L(\theta) = \ln g(z, \alpha; \theta) - \ln g(\alpha|\, z; \theta). \tag{29}$$

Taking the expectation on both sides in (29) with respect to the conditional density of α given z, evaluated at an arbitrary parameter value θ', gives:

$$L(\theta) = M(\theta|\theta') - H(\theta|\, \theta'), \tag{30}$$

where

$$M(\theta|\theta') = E\{\ln g(z, \alpha; \theta)|z; \theta'\}$$
$$H(\theta|\theta') = E\{\ln g(\alpha|z; \theta)|z; \theta'\}$$

It is shown in Wu (1983) that the following algorithm will converge to a stationary point of the likelihood function under quite general conditions:

Let $\theta^{(1)}$ be given. For $m = 1, 2, ...$

(i) E-step: Compute $M(\theta|\, \theta^{(m)})$.

(ii) M-step: Set $\theta^{(m+1)} = \arg\max_{\theta} M(\theta|\, \theta^{(m)})$.

(iii) Set $m = m + 1$, and go to *(i)*.

Since by Kullback's inequality $H(\theta|\theta^{(m)}) \leq H(\theta^{(m)}|\theta^{(m)})$ for all θ, it is easy to verify that $\{L(\theta^{(m)})\}$ is an increasing sequence of likelihood values.

It will be more convenient for us to replace the maximization in the M-step by a *conditional maximization* (CM) procedure: For a fixed partition of the parameters, the maximization is partial; that is, with respect to one group of parameters at a time, keeping the remaining parameters fixed at their current values. This partial maximization procedure guarantees that $M(\theta^{(m+1)}|\theta^{(m)}) > M(\theta^{(m)}|\theta^{(m)})$ and therefore preserves the ascent property, and thus stability, of the EM algorithm. The convergence might, however, be slower.

The EM (ECM) algorithm does not require calculation of the log-likelihood $L(\theta)$ – only of the function $M(\theta|\theta^{(m)})$. An important property of the algorithm is that

$$\frac{\partial L(\theta^{(m)})}{\partial \theta} = \left.\frac{\partial M(\theta|\theta^{(m)})}{\partial \theta}\right|_{\theta=\theta^{(m)}}, \tag{31}$$

which follows from the fact that $\theta^{(m)}$ is the maximizer of $H(\theta|\theta^{(m)})$, and hence a stationary point. The Hessian of $L(\theta)$ at $\widehat{\theta}$ can therefore be obtained by numerical differentiation of $\left.\frac{\partial M(\theta|\widehat{\theta})}{\partial \theta}\right|_{\theta=\widehat{\theta}}$. This result is important, because it yields a computationally simple estimator of the covariance matrix of $\widehat{\theta}$. In my experience, it also yields more accurate covariance estimates than the SECM algorithm proposed in Meng and Rubin (1992).

4.2 The E-Step

To perform the E-step of the ECM algorithm, we have to evaluate

$$\begin{aligned} M(\theta|\,\theta^{(m)}) = &-\frac{N(T-1)}{2}\ln|\Sigma| - \frac{N}{2}\ln|\tilde{\Omega}| \\ &- \frac{1}{2}E\left\{\sum_{i=1}^{N}\sum_{t=1}^{T-1}\left(z_t^i(\beta) - G_t(\Gamma,\Upsilon)\alpha_t\right)'\Sigma^{-1}\left(z_t^i(\beta) - G_t(\Gamma,\Upsilon)\alpha_t\right)\,|\,Y;\theta^{(m)}\right\} \\ &- \frac{1}{2}E\left\{\sum_{i=1}^{N}\left(z_T^i(\beta) - \mu_0 - G_T(\Gamma,\Upsilon)\alpha_T\right)'\tilde{\Omega}^{-1}\left(z_T^i(\beta) - \mu_0 - G_T(\Gamma,\Upsilon)\alpha_T\right)|\,Y;\theta^{(m)}\right\}, \end{aligned} \tag{32}$$

where $\tilde{\Omega} = \Omega + T^{-1}\Sigma$, $G_t(\Gamma,\Upsilon)$ denotes G_t as a function of Γ and Υ (see 22), and we have ignored uninteresting constants. In equation (32), the expectation is with respect to the latent variables $(\alpha_1,...,\alpha_T)$, conditional on the data Y, and with θ evaluated at $\theta^{(m)}$.

Since $M(\theta|\,\theta^{(m)})$ is quadratic in $(\alpha_1,...,\alpha_T)$, to evaluate the expectations in (32) we only need to calculate the conditional moments

$$\begin{aligned} a_{t|T} &= E\{\alpha_t\,|\,Y;\theta^{(m)}\} \\ V_{t\,|T} &= E\{(\alpha_t - a_{t|T})(\alpha_t - a_{t|T})'\,|\,Y;\theta^{(m)}\}. \end{aligned} \tag{33}$$

When $\theta = \theta^{(m)}$, we can calculate $z_t^i(\beta^{(m)})$ from (27). The state space form (23) can then be used with z_t^i replaced by $z_t^i(\beta^{(m)})$ to derive (33) by means of the Kalman filter. Following the exposition in

Fahrmeir and Tutz (1994), p. 264, the filtering recursions can be described by the following algorithm:

Kalman filtering:

$$a_{0|0} = 0$$

$$V_{0|0} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & I_p & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

For $t = 1, ..., T$:

$$a_{t|t-1} = F a_{t-1|t-1}$$

$$V_{t|t-1} = F V_{t-1|t-1} F' + Q$$

$$K_t = V_{t|t-1} G_t^{*\prime} [G_t^* V_{t|t-1} G_t^{*\prime} + \Sigma_t]^{-1}$$

$$a_{t|t} = a_{t|t-1} + K_t (Z_t^* - \tau_t^* - G_t^* a_{t|t-1})$$

$$V_{t|t} = V_{t|t-1} - K_t G_t^* V_{t|t-1},$$

where all parameters are evaluated at $\theta = \theta^{(m)}$. Note that no inversions of high dimensional matrices are needed, since, from a well-known matrix inversion lemma (see Anderson and Moore (1979), p. 138),

$$[G_t^* V_{t|t-1} G_t^{*\prime} + \Sigma_t]^{-1} = \Sigma_t^{-1} - \Sigma_t^{-1} G_t^* (V_{t|t-1}^{-1} + G_t^{*\prime} \Sigma_t^{-1} G_t^*)^{-1} G_t^{*\prime} \Sigma_t^{-1}, \tag{34}$$

with

$$\Sigma_t^{-1} = \begin{cases} I_N \otimes \Sigma^{-1} & t = 1, ..., T-1 \\ I_N \otimes (\Omega + T^{-1}\Sigma)^{-1} & t = T. \end{cases}$$

The required conditional expectations $a_{t|T}$ and variances $V_{t|T}$ are obtained in subsequent backward smoothing recursions (see Fahrmeir and Tutz (1994) p. 265):

Kalman smoothing:

For $t = T, ..., 2$:

$$a_{t-1|T} = a_{t-1|t-1} + B_t (a_{t|T} - a_{t|t-1})$$

$$V_{t-1|T} = V_{t-1|t-1} + B_t (V_{t|T} - V_{t|t-1}) B_t',$$

where

$$B_t = V_{t-1|t-1} F' V_{t|t-1}^{-1}.$$

4.3 The CM-Step

In the m'th CM-step, after $a_{t|T}$ and $V_{t|T}$ have been evaluated in the preceding E-step, we update θ to obtain $\theta^{(m+1)}$. However, maximization of $M(\theta|\theta^{(m)})$ would require iterative methods. To avoid

iterations, we partition θ such that a closed form solution for the optimum with respect to each subset in the partition of θ can be found when the remaining parameters are fixed at its current value. Define $\theta_1 = (\mu_0', vec(\Pi)', vec(\Lambda)', vech(\Gamma)', vech(\Upsilon)')$ and $\theta_2 = (vech(\Sigma)', vech(\Omega)')$. Thus $\theta = (\theta_1, \theta_2)$ is a partition of θ. Let $M(\theta_1, \theta_2|\theta) \equiv M(\theta|\theta)$. The conditional maximizations then consist in finding:

$$\theta_1^{(m+1)} = \arg\max_{\theta_1} M(\theta_1, \theta_2^{(m)}|\theta^{(m)}) \tag{35}$$

$$\theta_2^{(m+1)} = \arg\max_{\theta_2} M(\theta_1^{(m+1)}, \theta_2|\theta^{(m)}). \tag{36}$$

Conditional maximization w.r.t. θ_1 in (35) is equivalent to maximization of the quadratic function

$$\begin{aligned} q(\theta_1) = &-\sum_{i=1}^{N}\sum_{t=1}^{T-1}\left(Y_{1t}^i - \Pi Y_{0t}^i - \Lambda X_{0t}^i - G_t(\Gamma,\Upsilon)a_{t|T}\right)' \Sigma^{(m)^{-1}} \left(Y_{1t}^i - \Pi Y_{0t}^i - \Lambda X_{0t}^i - G_t(\Gamma,\Upsilon)a_{t|T}\right) \\ &-\sum_{i=1}^{N}\left(Y_{1T}^i - \mu_0 - \Pi Y_{0T}^i - \Lambda X_{0T}^i - G_T(\Gamma,\Upsilon)a_{T|T}\right)' \widetilde{\Omega}^{(m)^{-1}} \left(Y_{1T}^i - \mu_0 - \Pi Y_{0T}^i - \Lambda X_{0T}^i - G_T(\Gamma,\Upsilon)a_{T|T}\right) \\ &-\sum_{i=1}^{N}\sum_{t=1}^{T-1} \text{tr } \Sigma^{(m)^{-1}} G_t(\Gamma,\Upsilon)V_t G_t(\Gamma,\Upsilon)' + \sum_{i=1}^{N} \text{tr } \widetilde{\Omega}^{(m)^{-1}} G_T(\Gamma,\Upsilon)V_T G_T(\Gamma,\Upsilon)'. \end{aligned} \tag{37}$$

By setting the derivative of $q(\theta_1)$ equal to zero, we get linear 1. order conditions. The updating of θ_1 in (35) is therefore trivial.

Conditional maximization with respect to $\theta_2 = (vech(\Sigma)', vech(\Omega)')$ in (36) is equivalent to maximization of

$$\begin{aligned} r(\theta_2) = &-N(T-1)\ln|\Sigma| - N\ln|\,\Omega + T^{-1}\Sigma| - \sum_{i=1}^{N}\sum_{t=1}^{T-1} \text{tr } \Sigma^{-1}\left(R_t^i R_t^{i\prime} + G_t^{(m+1)} V_t G_t^{(m+1)\prime}\right) \\ &-\sum_{i=1}^{N} \text{tr } (\Omega + T^{-1}\Sigma)^{-1}\left(S^i S^{i\prime} + G_T^{(m+1)} V_T G_T^{(m+1)\prime}\right) \end{aligned} \tag{38}$$

where

$$\begin{aligned} G_t^{(m+1)} &= G_t(\Gamma^{(m+1)}, \Upsilon^{(m+1)}) \\ R_t^i &= Y_{1t}^i - \Pi^{(m+1)} Y_{0t}^i - \Lambda^{(m+1)} X_{0t}^i - G_t^{(m+1)} a_{t|T} \\ S^i &= Y_{1T}^i - \mu_0^{(m+1)} - \Pi^{(m+1)} Y_{0T}^i - \Lambda^{(m+1)} X_{0T}^i - G_T^{(m+1)} a_{T|T}. \end{aligned}$$

Hence

$$\Sigma^{(m+1)} = \frac{1}{N(T-1)} \sum_{i=1}^{N}\sum_{t=1}^{T-1}\left(R_t^i R_t^{i\,\prime} + G_t^{(m+1)} V_{t|T}\, G_t^{(m+1)\prime}\right) \tag{39}$$

$$\Omega^{(m+1)} = \frac{1}{N}\sum_{i=1}^{N}\left(S^i S^{i\,\prime} + G_T^{(m+1)}\, V_{T|T}\, G_T^{(m+1)\prime}\right) - \frac{1}{T}\Sigma^{(m+1)}. \tag{40}$$

Unfortunately, $\Omega^{(m+1)}$ may not be positive definite. This potential problem is addressed in Section 4.5 by reparametrizing Σ and Ω in terms of their Cholesky factors, and then maximizing $r(\theta_2)$ by iterative methods.

We can sum up the ECM algorithm as follows:

Let $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)})$ be given. For $m = 1, 2, ...$

(i) The E-step: Run the Kalman filter with θ evaluated at $\theta^{(m)}$ and compute $a_{t|T}$, $V_{t|T}$ and B_t.

(ii) The CM-step: Set

$$\theta_1^{(m+1)} = \arg\max_{\theta_1} \; M(\theta_1, \theta_2^{(m)} | \theta^{(m)})$$
$$\theta_2^{(m+1)} = \arg\max_{\theta_2} \; M(\theta_1^{(m+1)}, \theta_2 | \theta^{(m)}).$$

(iii) Set $m = m + 1$, and go to *(i)*.

The general convergence properties of the EM and ECM algorithm are studied in Wu (1983) and Meng and Rubin (1993). Under regularity conditions (5)-(10) in Wu (1983), Theorem 3 in Meng and Rubin (1993) states that the ECM algorithm converges to a stationary point provided the partition of the parameter space is space filling, which follows in our case from the fact that cyclical maximizations over a fixed partition of the parameter vector is always space filling (see Meng (1992) for a proof). Wu (1983) shows that his conditions (5)-(8) and (10) are always satisfied for curved exponential families (see Barndorff-Nielsen and Cox (1994)), and therefore in our case because of normality. The remaining regularity condition to be checked is his condition (9):

$$\Omega_\theta^{(1)} = \{\theta \in \Theta : L(\theta) \geq L(\theta^{(1)}) \text{ is compact and in the interior of } \Theta\}, \tag{41}$$

where Θ is the parameter space. In fact, the compactness condition (41) is required for any numerical algorithm to ensure convergence to a stationary point. However, it is well-known that (41) could fail in latent variables models when $\Theta^{(1)}$ includes points where some of the covariance matrices are singular; i.e. either Ω, Γ or Υ have reduced rank. To remedy this problem, one may impose rank restrictions in such a way that the remaining free parameters satisfy (41). (See also Harvey (1989); Ch. 4 and 5, for a discussion of a similar problem in the context of structural time series models.) Fortunately, using (31) it is always possible to cheque whether a limit point of the ECM algorithm is a stationary point or not. Moreover, by numerical differentiation of the gradient vector it is also possible to find out whether the point is a local maxima.

4.4 Initialization of the ECM Algorithm

To ensure that the ECM algorithm gives a consistent estimator of the parameters β in the case of finite-T asymptotics, we can use an IV-based estimator to initiate the algorithm. This method requires that the true parameter β_0 is an interior point of the parameter space Ω_β, and that the roots of the autoregressive polynomial $I_p - \sum_{j=1}^{k} \Pi_j z^j$ are outside the unit circle (see Binder, Hsaio and Pesaran (2000)). Of course, consistent estimation of the ϕ-parameters is not an issue as long as we hold T constant. Hence initialization of ϕ will not be considered here.

To obtain an initial estimate of β, we difference (1) to get

$$\Delta y_t^i - \sum_{j=1}^{k} \Pi_j \Delta y_{t-j}^i - \sum_{j=0}^{m} \Lambda_j \Delta x_{t-j}^i - \Delta f_t = \Delta \varepsilon_t^i \text{ for } t = 2, ..., T,$$

where $\Delta y_t^i = y_t^i - y_{t-1}^i$, etc. We then have the following orthogonality conditions for $i = 1, ..., N$:

$$\begin{array}{ll} E\{(\varepsilon_{t+1}^i - \varepsilon_t^i) y_{t-l}^i{}'\} = 0 & t = 1, .., T-1;\ l = 1, .., t+k-1 \\ E\{(\varepsilon_{t+1}^i - \varepsilon_t^i) x_{t-l}^i{}'\} = 0 & t = 1, .., T-1;\ l = 1, .., t+m-1. \end{array}$$

These are identical to the orthogonality conditions used by Holtz-Eakin et al. (1988). Furthermore, using the notation in (25), we have

$$E\left\{ y_{T;T}^i - \sum_{j=1}^{k} \Pi_j y_{T-j;T}^i - \sum_{j=0}^{m} \Lambda_j x_{T-j;T}^i - f_{T;T} \right\} = 0 \quad \text{for } i = 1, ..., N.$$

The GMM method proposed in Holtz-Eakin et al. (1988) is applicable if we treat the realization of f as a matrix of fixed effects. The method yields estimates $(\Pi^{IV}, \Lambda^{IV}, f^{IV})$, say, which is consistent as $N \to \infty$ with fixed f. The rest of the parameters in β can be initiated as follows:

$$\Sigma^{IV} = \frac{1}{N(T-1)} \sum_{i}^{N} \sum_{t=1}^{T-1} (Y_{1t}^i - \Pi^{IV} Y_{0t}^i - \Lambda^{IV} X_{0t}^i - f^{IV} w_t)(Y_{1t}^i - \Pi^{IV} Y_{0t}^i - \Lambda^{IV} X_{0t}^i - f^{IV} w_t)'$$

$$\Omega^{IV} = \frac{1}{N} \sum_{i}^{N} (Y_{1T}^i - \Pi^{IV} Y_{0T}^i - \Lambda^{IV} X_{0t}^i - T^{-\frac{1}{2}} f^{IV} w_T)(Y_{1T}^i - \Pi^{IV} Y_{0T}^i - \Lambda^{IV} X_{0t}^i - T^{-\frac{1}{2}} f^{IV} w_T)' - \frac{1}{T} \Sigma^{IV}.$$

Hence $\beta^{IV} = (\Pi^{IV}, \Lambda^{IV}, \Sigma^{IV}, \Omega^{IV})$ is a consistent estimator of β with fixed f. By dominated convergence it is also a consistent estimator of β when T (but not f) is fixed. In my experience, β^{IV} tends to be far away from the optimum of the likelihood function $L(\theta)$. To improve the initial value, I propose to use β^{IV} as a starting point for an iterative ascent method for maximizing the conditional likelihood (18). Then any limit point β^* of this method will be a consistent initiator of the ECM algorithm. In my experience, β^* is typically close to the ML estimator $\widehat{\beta}$ (see also the proof of Lemma 3, Appendix A).

4.5 Parameter Restrictions

In this sub-section we will briefly consider implementation of parameter restrictions. Such restrictions are important for several reasons. First, the full model may contain too many parameters relative to the information in the data. This can be seen as a practical identification problem, causing slow convergence of the estimation algorithm and large standard errors. Second, in some analyses parameter restrictions, such as e.g. cointegration, may have a theoretical foundation. Third, as mentioned above, the ECM algorithm described in Section 4 does not ensure that the estimate of the covariance matrix Ω is positive definite.

We shall first consider the case where $\Lambda_j = \Pi_j = 0$ for every j. Model (1)-(2) then reduces to a panel data version of the local level model

$$y_t^i = \mu_0 + \Gamma\mu_t + \Upsilon\delta_t + v^i + \varepsilon_t^i \tag{42}$$

$$\mu_t = \mu_{t-1} + \eta_t \tag{43}$$

Assume that Γ is a $p \times r$ matrix of rank $r < p$, with positive diagonal elements and zeros above the diagonal, while μ_t is an r-dimensional random walk. Then there exists a $p \times (p-r)$ matrix υ of full rank, such that $\upsilon'\Gamma = 0$. From (42) it follows that $\upsilon' y_t^i = \upsilon'\mu_0 + \upsilon'\Upsilon\delta_t + \upsilon' v^i + \upsilon'\varepsilon_t^i$ is a $(p-r)$-vector of stationary variables; i.e. the columns of υ are cointegrating vectors. Nyblom and Harvey (2000) present tests for the rank of Γ in the aggregate version of the model (42)-(43). These results may be useful also in the present context given that T is large: one may first determine the rank of Γ based on aggregate data, and then estimate the unknown parameters from panel data. For the latter task, no adjustment of the algorithm presented in Section 4 is needed, except that the number of columns in Γ is r rather than p.

In applications it is of equal importance to consider rank and definiteness restrictions on the estimator of the random effects covariance matrix Ω. In my experience, the conditional maximization formula (40) may yield non-definite estimates $\widehat{\Omega}$. To remedy this problem, we can reparametrize Ω and Σ in terms of their Cholesky factors:

$$\begin{aligned} \Omega &= UU' \\ \Sigma &= LL' \end{aligned} \tag{44}$$

for lower triangular matrices U and L. This reparametrization complicates the CM-step, since we no longer get closed form solutions when maximizing (jointly) with respect to U and L. Fortunately, analytic expressions for the derivatives of the objective function (38) with respect to U and L are available (see Lütkepohl (1996)), and efficient quasi-Newton algorithms can be used to solve this sub-

problem. It is sometimes also desirable to implement rank restriction on U, so that U is a $p \times r$ matrix with rank r, $r < p$, and with zeros above the diagonal. Technically, such restrictions are straightforward to implement. (See Section 5 for an application.)

4.6 Initial Conditions

Now the initial condition assumption that $Y_0^i = \{y_0^i, .., y_{1-k}^i\}$ is independent of the subject-specific intercept v^i will be relaxed. This independence assumption may be unrealistic because of the following reason: Since v^i determines the level of y_t^i for $t \geq 1$, the level of the initial value Y_0^i may also depend of v^i. A natural way to introduce such dependence in our model is to assume that v^i and Y_0^i have a joint normal distribution and that $v^i|Y_0^i$ is conditionally independent of Y_0^j for $j \neq i$ (see Sims (2000)). Hence, imposing the identifying restriction that v^i has *unconditional* mean equal to zero, we get

$$v^i|Y_0^i \backsim \mathcal{N}(\sum_{j=1}^{k} B_j(y_{1-j}^i - \widetilde{\mu}_{1-j}), \Omega), \tag{45}$$

where $\widetilde{\mu}_{1-j}$ is the unconditional expectation of y_{1-j}^i; B_j are regression coefficients; and Ω is the *conditional* covariance matrix given Y_0^i derived from the joint distribution of (v^i, Y_0^i).

To derive the likelihood under (45), we only need to redefine z_T^i in (16) as

$$z_T^i = T^{-1/2} Z^i w_T - \sum_{j=1}^{k} B_j(y_{1-j}^i - \widetilde{\mu}_{1-j}),$$

where $\widetilde{\mu}_{1-j}$ can be estimated from the sample mean of y_{1-j}^i. Hence,

$$z_T^i|(f, v^i, Y_0^i) \backsim \mathcal{N}(T^{-1/2} f w_T + v^i - \sum_{j=1}^{k} B_j(y_{1-j}^i - \widetilde{\mu}_{1-j}), T^{-1}\Sigma),$$

and

$$z_T^i|\, f, Y_0 \backsim \mathcal{N}(T^{-1/2} f w_T, T^{-1}\Sigma + \Omega).$$

The state space- and likelihood representation of the model are unaltered, except that the definition of z_T^i has changed. The necessary modifications of the CM-step of the ECM algorithm are straightforward.

5 Application: A New Look at the Aggregation Problem

The model and estimation procedure outlined in the previous sections will now be applied on a real data set. The data consist of a balanced sample of 111 Norwegian manufacturing firms from the sector *mineral products* observed during the period 1976-96. At issue is the determinants of the wage and employment process at the plant level. We shall analyze a bivariate model ($p = 2$), with

$$y_t^i = (\,\text{log-real wage},\ \text{log-employment})'$$

as endogenous variables, and with no exogenous variables. The application serves two purposes: to illustrate the ECM algorithm and to assess the merits of the micro model with regard to making forecasts about *aggregate* variables – and compare with predictions from an (exact) aggregate model.

The aggregation problem has a long history in econometrics. Classical contributions include Theil (1954) and Grunfeld and Griliches (1960). Aggregation of non-linear structural models have been treated for example in Caballero and Engel (1999) and Caballero, Engel and Haltiwanger (1995), while Pesaran (1999) studies aggregation in autoregressive models. Although it is outside the scope of this paper to analyze the problems of aggregation in depth, this section will provide some new insight into this topic.

We shall look at aggregate variables of the type $y_t = \frac{1}{N}\sum_{i=1}^{N} y_t^i$, where y_t^i is assumed generated by the model (1)-(5). Specifically, we will study the problem of predicting y_{t+1} from two different information sets: *(i)* the aggregate information set $\mathcal{F}_t^a = \{y_0, ..., y_{t-1}\}$; and *(ii)* the disaggregate information set $\mathcal{F}_t^d = \{y_0^i, .., y_t^i;\, i = 1, ..., N\}$ consisting of all the micro variables up until time t.

We will analyze a VAR(1) specification of the model. Thus $k = 1$, $T = 20$, and $N = 111$. This specification is not rejected when testing against the alternative $k = 2$ using a Wald-test. For known parameters θ, assuming that $N^{-1}\sum_{i=1}^{N} v^i \approx 0$, the optimal predictors in the mean squared error sense are the conditional expectations (see Lehman (1983), p. 54):

$$\begin{aligned} \widehat{y}_{t+1}^a &= \mu_0 + \Pi y_t + \Gamma E\{\mu_{t+1}|\mathcal{F}_t^a;\theta\} \\ \widehat{y}_{t+1}^d &= \mu_0 + \Pi y_{t-1} + \Gamma E\{\mu_{t+1}|\mathcal{F}_t^d;\theta\}, \end{aligned}$$

where $E\{\mu_{t+1}|\mathcal{F}_t^a;\theta\}$ and $E\{\mu_{t+1}|\mathcal{F}_t^d;\theta\}$ are outputs from the Kalman filter. Note that there may be a substantial discrepancy between the two predictors, because $\mathcal{F}_t^a$ and $\mathcal{F}_t^d$ contain different information about μ_t. Undoubtedly, given that the model is correctly specified and θ is known, $\widehat{y}_{t+1}^d$ is the most accurate predictor, being based on the largest information set.

In practice, θ is not known and must be replaced by estimates $\widehat{\theta}^d$ and $\widehat{\theta}^a$ based on panel data and aggregate data, respectively. The prediction functions which can be applied are therefore the estimated relations:

$$\begin{aligned} \widehat{y}_{t+1}^a &= \widehat{\mu}_0^a + \widehat{\Pi}^a y_t + \widehat{\Gamma}^a E\{\mu_t|\mathcal{F}_t^a;\widehat{\theta}^a\} \\ \widehat{y}_{t+1}^d &= \widehat{\mu}_0^d + \widehat{\Pi}^d y_{t-1} + \widehat{\Gamma}^d E\{\mu_t|\mathcal{F}_t^d;\widehat{\theta}^d\}. \end{aligned}$$

Due to specification errors and estimation uncertainty, systematic differences between $\widehat{\theta}^d$ and $\widehat{\theta}^a$ are likely to occur even in samples with large N and large T. Moreover, identification problems could arise because the aggregate model cannot distinguish between idiosyncractic shocks ε_t^i and industry-wide (common) shocks δ_t and η_t. For example, if the variables are $I(1)$ one would expect that from

aggregate data it is nearly impossible to determine whether the stochastic trend is due to accumulation of common shocks, or to a unit root in the autoregressive polynomial.

An important aspect of the aggregation problem is that of parameter stability. Is there any evidence that either $\widehat{\theta}^d$ or $\widehat{\theta}^a$ is the more stable estimator as new time periods are included in the sample used to estimate θ? In fact, Grunfeld and Griliches (1960) argue that in the presence of specification errors, aggregation of economic variables may reduce such errors. However, some evidence to the contrary will be presented below.

Let us now turn to the implementation of the micro model. To estimate its parameters, the ECM algorithm was written as a GAUSS program. The algorithm was iterated until

$$\max_j \left| \frac{\partial L(\theta^{(m)})}{\partial \theta_j} \right| = \max_j \left| \frac{\partial M(\theta|\theta^{(m)})}{\partial \theta_j} \right|_{\theta=\theta^{(m)}} < 0.1,$$

where θ_j is the j'th component of θ. Convergence was obtained after approximately 750 iterations, with $\max_j |\theta_j^{(m+1)} - \theta_j^{(m)}| < 0.00001$. This took about 20 minutes on a Pentium 750 Mhz PC, using the maximizer of the fixed effect likelihood to initiate the algorithm (See Section 4.4). I found that the point of convergence was insensitive to the choice of starting point.

The estimates of the micro model using *(i)* all the data, and *(ii)* data for 1976-86 only, are shown in Table 1, column 2 and 4, respectively. Standard deviation estimates (squared roots of the inverse negative Hessian of the log-likelihood function) are reported in parenthesis.

As described in Section 4.5, the covariance matrices were reparametrized in terms of Cholesky factors, i.e. $\Sigma = LL'$ and $\Omega = UU'$. The reparametrization slowed down the speed of each CM-step, since (U, L) had to be updated iteratively (see (38)), using a quasi-Newton algorithm. But after approximately 50 ECM iterations, the updating of U and L were quick; the effect of superlinear convergence for the quasi-Newton method was then already at full force. The estimates of both $U^{(2,2)}$ and $\Upsilon^{(2,2)}$ (i.e. component $(2,2)$ in U and Υ, respectively) both tended to zero, indicating that the rank of Ω as well as the number of components in δ_t is 1. Fortunately, the same estimates were obtained regardless of whether these rank restrictions were imposed a priori or not.

Estimates for the aggregate model are reported in Table 1, column 3 (full sample) and column 5 (partial sample). The error term in the aggregate model is:

$$e_t = \Upsilon \delta_t + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_t^i.$$

Note that the Cholesky factor L in the covariance matrix of ε_t^i and the loading matrix Υ of δ_t cannot be identified from aggregate data. In Table 1; column 3 and 5, the estimated Cholesky factor in the covariance matrix of e_t is reported in the Υ-rows. However, the micro estimates (column 2 and 4) and macro estimates (column 3 and 5) are not directly comparable for these parameters.

Although the loading matrices Γ and Υ are difficult to estimate when $T = 20$ (see Section 6 for some Monte Carlo evidence), it appears from Table 1, column 2, that the matrix components $\Gamma^{(1,1)}$, $\Gamma^{(2,2)}$ and $\Upsilon^{(1,1)}$ are significant in the micro model. In particular, the sector-wide stochastic trend μ_t seems to be an important determinant of y_t^i. On the other hand, turning to the estimated aggregate relations, we find contradictory evidence: the estimate of Γ reported in column 3 is zero. On the other hand, the largest eigenvalue of $\widehat{\Pi}^a$ is 0.98. Hence, also from aggregate data we find evidence that shocks have a high degree of persistence.

While there is a striking similarity in the estimates of the micro relations using *(i)* all the data, and *(ii)* data for 1976-86, only, the opposite is true for the aggregate model. Figure 1 offers an explanation of this: After 1986 there is a structural break in the data, as the industry enters a long recession period with declining employment. Moreover, the recession appears to influence real wages to a lesser extent than employment, causing a change in the pattern of co-movements of employment and real wage at the aggregate level. The parameters of the aggregate model are clearly not invariant to this break, while the parameters of the micro model are remarkably stable, indicating that the break is appropriately accounted for by the latent trend μ_t. The ability of the micro model to identify industry wide and idiosyncractic shocks seems to be the basis for parameter stability. As we shall see, this has severe consequences for forecasting.

To compare predictions from the two models, the measure R^2 defined as

$$R^2 = 1 - \frac{tr(Var(y_t - \widehat{y}_t))}{tr(Var(y_t))}$$

was calculated, where $\widehat{y}_t = \widehat{y}_t^a$ and $\widehat{y}_t = \widehat{y}_t^d$ for the aggregate and disaggregate model, respectively. The in-sample predictions from the micro and aggregate model are in agreement with previous studies (see Grunfeld and Griliches (1960)): the macro model fairs slightly better, with $R^2 = .88$ versus $R^2 = .86$ for the micro model. This is not surprising, as the aggregate model contains much more parameters per data point. On the other hand, when R^2 is calculated for 1987-96, with estimates obtained from the 1976-86 data, the results become markedly different. In the micro model, $R^2 = .76$, but in the aggregate model a dramatic prediction failure occurs, with $R^2 = -.03$. The variance of the prediction errors is larger than the variance of y_t itself! The complete failure of the aggregate model out-of-sample is also evident from Figure 2.

The application has revealed two striking results:*(i)* the parameter estimates of the micro model remain largely invariant despite a distinct structural break in the data generating process after 1986, and *(ii)* the out-of-sample forecasts from the micro model fit the data well, while the aggregate model fails dramatically. The lesson from this seems to be that the Kalman filter utilizes micro data in an efficient way to identify persistent industry-wide shocks as opposed to firm-specific shocks.

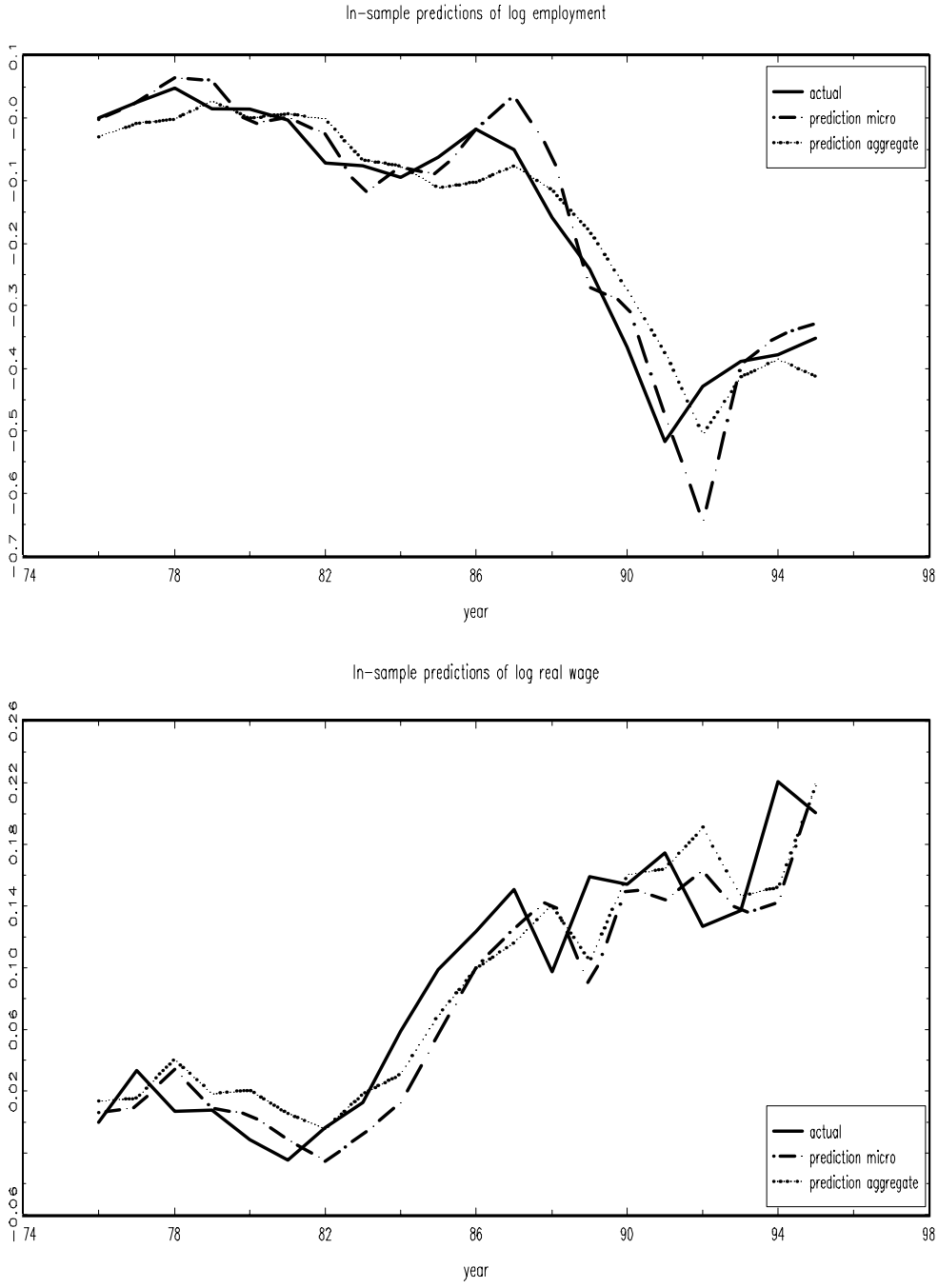

Figure 1: In-sample predictions.

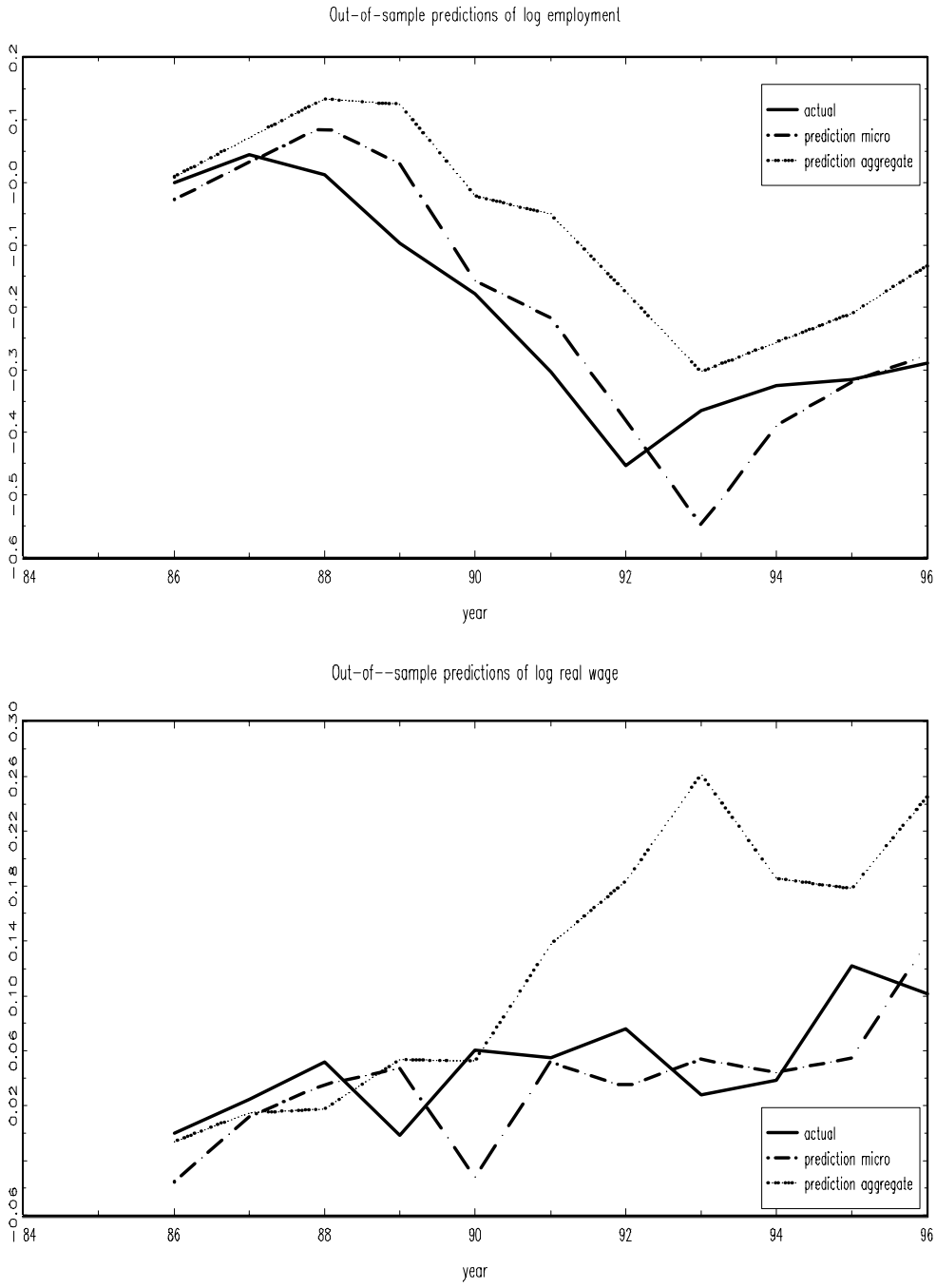

Figure 2: Out-of-sample predictions.

	Complete Data: 76-96				Incomplete Data: 76-86	
Param.	Micro model		Aggr. Model		Micro Model	Aggr. Model
	Est.	S.d.	Est	S.d.	Estimates	Estimates
$\mu_0^{(1)}$	2.93	.12	1.23	.98	2.87	1.97
$\mu_0^{(2)}$	-1.24	.15	2.75	1.73	-1.14	-4.16
$\Pi^{(1,1)}$	.43	.02	.81	.16	.43	.85
$\Pi^{(2,1)}$	.26	.03	-.42	.27	.24	.80
$\Pi^{(1,2)}$	.04	.01	-.06	.12	.06	-.40
$\Pi^{(2,2)}$	.97	.01	.82	.16	.96	.96
$\Gamma^{(1,1)}$	.020	.007	.000	–	.021	.000
$\Gamma^{(2,1)}$	.009	.025	.000	–	.009	.000
$\Gamma^{(2,2)}$	.053	.020	.000	–	.012	.000
$\Upsilon^{(1,1)}$	.016	.008	.03	.005	.002	.016
$\Upsilon^{(2,1)}$	-.011	.023	-.008	.013	-.008	-.007
$\Upsilon^{(2,2)}$	–	–	.057	.009	–	-.023
$U^{(1,1)}$	-.076	.008	–		-.09	–
$U^{(2,1)}$	.037	.009	–		.055	–
$L^{(1,1)}$	.180	.003	–		.157	–
$L^{(2,1)}$	-.053	.005	–		-.042	–
$L^{(2,2)}$	.243	.004	–		.200	–
R^2	.862		.887			

Table 1: Estimates.

6 Inference

Let $\widehat{\theta}_N = (\widehat{\beta}_N, \widehat{\phi}_N)$ be the estimator of θ generated by the ECM algorithm described in Section 4. In this section consistency and asymptotic normality of $\widehat{\beta}_N$ are established for the case when $N \to \infty$ and T is fixed. This type of (semi-) asymptotics is often the most relevant in applications, as panel data sets typically have small T relative to N – although there are important situations where this is not the case, e.g. if he units are countries or industries or we have other types of aggregate data. Monte Carlo results are presented to evaluate the finite sample properties of both $\widehat{\beta}_N$ and $\widehat{\phi}_N$.

When T is fixed, consistent estimation of ϕ is not an issue. Thus, the semi-asymptotic limit theorems provided in this section are not helpful for making inferences about the parameters of the latent time series. However, well-known result for non-stationary state space models (see Harvey (1989); Ch. 4) can be used to establish lage-T asymptotics for $\widehat{\phi}_N$ when N is fixed. The Monte Carlo results indicate that these results can be extended to cases where both T and N become large, but I have not been able to establish asymptotic theorems for this case. Certainly, that is a challenging and important topic for future work.

6.1 Asymptotic Results

We start by introducing some new notation: $l^i(\beta, F)$ is the log of the conditional density of $(z_1^i, .., z_T^i)$ given F (defined in (18));

$$l^i(\beta, F) = -\frac{T-1}{2}\ln|\Sigma| - \frac{1}{2}\ln|\Omega + T^{-1}\Sigma| - \tfrac{1}{2}S^i + \text{constant}, \tag{46}$$

where

$$\begin{aligned} S^i = &\sum_{t=1}^{T-1}(z_t^i - fw_t)'\Sigma^{-1}(z_t^i - fw_t) \\ &+ (z_T^i - T^{-1/2}fw_T)'(\Omega + T^{-1}\Sigma)^{-1}(z_T^i - T^{-1/2}fw_T); \end{aligned} \tag{47}$$

β_N^* is the preliminary estimator described in Section 4.4; β^0 is the true parameter; $\phi_N^* \in \Omega_\phi$ is some arbitrary initial estimate of ϕ; and $L(\beta, \phi) \equiv L(\theta)$ is the log-likelihood function based on unit $i = 1, ..., N$.

Condition 1 *(i) $(\beta, \phi) \in \Omega_\beta \times \Omega_\phi$, a compact and convex subset of $\mathcal{R}^n$, with β_0 being an interior point of Ω_β. (ii) All roots of the autoregressive polynomial $I_p - \sum_{j=1}^k \Pi_j z^j$ are outside the unit circle.*

The assumption that $\Theta = \Omega_\beta \times \Omega_\phi$ is compact may be restrictive when no natural compactification of the parameter space exists. In particular, it means that the maximizer in each CM-step may not be an interior point. Nevertheless, the solutions of the conditional maximizations (35)-(36) are still well-defined (but may not coincide with the solution of the 1.order conditions).

Define the conditional information matrix

$$I = -E\left\{ \begin{bmatrix} l^i_{\beta\beta}(\beta^0, F^0) & l^i_{\beta F}(\beta^0, F^0) \\ l^i_{F\beta}(\beta^0, F^0) & l^i_{FF}(\beta^0, F^0) \end{bmatrix} | F^0 \right\}, \tag{48}$$

that is, conditional on $F = F^0$, and the negative Hessian of the log-likelihood function

$$\widehat{H} = \frac{-1}{N}\begin{bmatrix} L_{\beta\beta}(\widehat{\beta}_N, \widehat{\phi}_N) & L_{\beta\phi}(\widehat{\beta}_N, \widehat{\phi}_N) \\ L_{\phi\beta}(\widehat{\beta}_N, \widehat{\phi}_N) & L_{\phi\phi}(\widehat{\beta}_N, \widehat{\phi}_N) \end{bmatrix}.$$

The main results of this section are stated as Proposition 1 and 2. Proofs are provided in Appendix A and B, respectively.

Proposition 1 *(Consistency) Let $\widehat{\theta}_N = (\widehat{\beta}_N, \widehat{\phi}_N)$ be any limit point of the ECM algorithm. Then $\widehat{\beta}_N$ is a consistent estimator of β^0.*

The compactness of Θ implies that a limit point $\widehat{\theta}_N$ exists. Any limit point $\widehat{\beta}_N$ is consistent for β^0. To ensure that a limit point of the ECM algorithm is consistent, the estimator used to initiate the ECM algorithm must also be consistent. Condition 1 *(ii)* ensures that the initiator proposed in Section 4.4 satisfies this requirement.

Proposition 2 *(Asymptotic normality) Conditional on* $F = F^0$, $\sqrt{N}(\widehat{\beta}_N - \beta^0) \xrightarrow{D} N(0, I^{\beta\beta})$ *where* $I^{\beta\beta}$ *is the* β*-block of the inverse information matrix* I^{-1}*. If with probability tending to one,* $\widehat{\theta}_N$ *is a (local) maximizer of* $L(\theta)$ *in the interior of* Θ*, a consistent estimator of* $I^{\beta\beta}$ *is* $\widehat{H}^{\beta\beta}$*, where* $\widehat{H}^{\beta\beta}$ *is the* β*-block of* $\widehat{H}^{-1}$*.*

Proposition 2 is conditional on $F = F^0$, meaning that F is random ex ante, but fixed at its actual realization F^0 in repeated samples as $N \to \infty$. On the other hand, if F is re-drawn in repeated samples (with T fixed), the asymptotic distribution of $\sqrt{N}(\widehat{\beta}_N - \beta^0)$ will be $\mathcal{N}(0, E\{I^{\beta\beta}\})$, where the expectation is with respect to the marginal distribution of F.

An interesting question is whether $(N\widehat{H})^{-1}$ provides an approximate covariance matrix estimator for $(\widehat{\beta}_N, \widehat{\phi}_N)$ when both N and T are (moderately) large, and thus can be used to make inference about ϕ. Some Monte Carlo evidence of this is provided in the next subsection.

6.2 Monte Carlo Results

Monte Carlo simulations are very costly in terms of CPU time, especially when both N and T become large (e.g. ≥ 80). Therefore, data were simulated from one model only. To have realistic parameter values, the estimated micro model reported in Section 5 – a bivariate VAR(1) model with no exogenous covariates – was used to simulate all the data. In the simulations, $N = 20, 40, 80$ and 160 was combined with $T = 20, 40, 80$ and 160, to produce 16 different designs (i.e. combinations of N and T).

In each Monte Carlo simulation, a sample of N starting values were drawn from the empirical distribution of initial values (see Section 5) with replacement. Then, N random effects v^i, NT idiosyncractic effects ε_t^i, and time effects μ_t and δ_t were simulated using a random number generator. Thus, the latent time series was re-drawn in each Monte Carlo sample (with N and T fixed). ML estimates $\widehat{\theta}_m$ for $m = 1, ..., M$ were calculated by the ECM algorithm, where M is the total number of Monte Carlo replications (given N and T). Results for the two types of parameters $\widehat{\beta}$ and $\widehat{\phi}$ are reported in Table 2 and 3, respectively, where *mean* and *s.d.* are the mean and standard deviation of $\widehat{\theta}_m$ in the Monte Carlo sample, and $\widehat{s.d}$ is the mean of the standard deviation estimates based on the inverse Hessian of the negative log-likelihood (i.e. $(N\widehat{H})^{-1}$). Because M is finite, the reported *mean*, *s.d.* and $\widehat{s.d.}$ all are subject to sampling errors. In particular, referring to the 90% confidence level, the last digit of each entry in the tables has a margin of error of *at most* ± 1.5 point for the β-parameters (Table 2) and ± 2 points for the ϕ-parameters (Table 3). A different M was chosen for each design to meet these error bounds. A further substantial reduction in the margin of error would require implementation of the estimation algorithm in a much more efficient programming language than GAUSS, preferably Fortran – a task which I have not attempted to undertake for the present

paper.

From Table 2, it is clear that the asymptotic limit theorems provided in the previous subsection are valid in the present context. Even in the smallest sample with $N = 20$ and $T = 20$, the bias of $\widehat{\beta}$ is very small. Furthermore, there is excellent agreement between the actual standard deviations of the estimators (*s.d.*) and the standard deviation estimates obtained from $(N\widehat{H})^{-1}$ ($\widehat{s.d}$). It is noticeable that both these results hold along the diagonal path with $T = N$, despite unpublished results cited in Phillips and Moon (2000) regarding dynamic panel regression models where the ML estimator of the autoregressive parameter has a negative asymptotic bias along the diagonal path. We notice that when N and T both are doubled, the standard deviations are cut in half, indicating that $\widehat{\beta}$ converges to β^0 at the rate of $\sqrt{NT}$.

As anticipated, the bias and standard deviation of $\widehat{\phi}$ are much larger. For example, with $T = 20$ and $N = 160$, there is a substantial bias in the Γ-parameters, particularly for $\Gamma^{(2,2)}$ which has a downward bias of more than 40%. However, for $T \geq 40$ the situation improves substantially – especially when N is large – both in terms of bias and in terms of agreement between *s.d.* and $\widehat{s.d}$. We also notice from Table 3 that going from $N = T = 80$ to $N = T = 160$ cuts standard deviations roughly in half, while the biases vanish, indicating that along the diagonal path $\widehat{\phi}$ is $\sqrt{NT}$-consistent. However, a much more careful study is required before we can reach a definitive conclusion.

Parameter (β):			$\Pi^{(1,1)}$	$\Pi^{(2,1)}$	$\Pi^{(1,2)}$	$\Pi^{(2,2)}$	$U^{(1,1)}$	$U^{(2,1)}$	$L^{(1,1)}$	$L^{(2,1)}$	$L^{(2,2)}$
True Value (β^0):			.430	.256	.036	.967	-.076	.038	.179	-.053	.243
T=20	N=20	*mean*	.428	.255	.037	.963	-.071	.034	.178	-.053	.241
		s.d.	.051	.063	.016	.014	.018	.019	.007	.012	.009
		$\widehat{s.d.}$	.050	.066	.014	.014	.017	.020	.007	.013	.009
	N=40	*mean*	.429	.255	.036	.966	-.075	.037	.179	-.053	.242
		s.d.	.033	.047	.011	.010	.012	.015	.005	.010	.006
		$\widehat{s.d.}$	.035	.045	.009	.009	.012	.014	.005	.009	.006
	N=80	*mean*	.425	.260	.036	.966	-.076	.039	.179	-.053	.242
		s.d.	.026	.034	.008	.007	.009	.011	.003	.006	.004
		$\widehat{s.d.}$	.025	.033	.007	.007	.009	.010	.003	.006	.004
	N=160	*mean*	.431	.260	.035	.967	-.076	.034	.179	-.053	.243
		s.d.	.017	.020	.005	.005	.006	.007	.003	.005	.002
		$\widehat{s.d.}$	.017	.023	.005	.005	.006	.007	.002	.005	.003
T=40	N=20	*mean*	.424	.252	.035	.965	-.075	.034	.179	-.053	.240
		s.d.	.048	.065	.017	.012	.016	.019	.006	.012	.008
		$\widehat{s.d.}$	.050	.066	.014	.014	.018	.020	.007	.013	.009
	N=40	*mean*	.431	.256	.036	.966	-.075	.037	.179	-.053	.242
		s.d.	.024	.033	.007	.007	.010	.010	.003	.006	.004
		$\widehat{s.d.}$	.024	.033	.007	.007	.009	.010	.003	.006	.004
	N=80	*mean*	.430	.255	.036	.967	-.075	.037	.179	-.053	.243
		s.d.	.015	.022	.005	.004	.007	.006	.002	.004	.003
		$\widehat{s.d.}$	.017	.022	.005	.005	.007	.007	.002	.004	.003
	N=160	*mean*	.431	.253	.036	.967	-.075	.038	.179	-.053	.243
		s.d.	.010	.017	.003	.003	.004	.005	.002	.003	.002
		$\widehat{s.d.}$	.012	.016	.004	.003	.005	.005	.002	.003	.002
T=80	N=20	*mean*	.431	.258	.036	.965	-.075	.037	.180	-.053	.242
		s.d.	.026	.029	.005	.007	.014	.012	.003	.007	.005
		$\widehat{s.d.}$	.023	.033	.005	.006	.014	.011	.003	.006	.004
	N=40	*mean*	.430	.256	.036	.967	-.076	.037	.179	-.053	.243
		s.d.	.016	.022	.004	.004	.008	.007	.002	.004	.003
		$\widehat{s.d.}$	.016	.023	.004	.004	.009	.007	.002	.004	.003
	N=80	*mean*	.432	.253	.035	.967	-.075	.037	.179	-.053	.243
		s.d.	.012	.017	.003	.003	.006	.005	.002	.003	.002
		$\widehat{s.d.}$	.012	.016	.003	.003	.007	.005	.002	.003	.002
	N=160	mean	.432	.251	.036	.967	-.075	.037	.179	-.053	.243
		s.d.	.009	.013	.002	.002	.004	.004	.001	.002	.002
		$\widehat{s.d.}$	.008	.011	.002	.002	.005	.004	.001	.002	.002
T=160	N=20	mean	.433	.253	.036	.967	-.074	.038	.179	-.053	.242
		s.d.	.014	.003	.003	.004	.010	.009	.002	.004	.003
		$\widehat{s.d.}$	.016	.003	.003	.004	.012	.009	.002	.004	.003
	N=40	mean	.430	.255	.036	.967	-.074	.038	.179	-.054	.242
		s.d.	.012	.017	.002	.003	.009	.007	.002	.003	.002
		$\widehat{s.d.}$	.011	.016	.002	.003	.009	.006	.002	.003	.002
	N=80	mean	.435	.253	.036	.967	-.075	.037	.178	-.052	.243
		s.d.	.010	.016	.002	.003	.007	.005	.001	.003	.002
		$\widehat{s.d.}$	.008	.011	.002	.002	.006	.004	.001	.002	.002
	N=160	*mean*	.431	.256	.036	.967	-.076	.037	.179	-.053	.243
		s.d.	.007	.006	.001	.001	.005	.003	.001	.002	.001
		$\widehat{s.d.}$	.006	.008	.001	.001	.004	.003	.001	.002	.001

Table 2: Monte Carlo Results for β.

Parameter (ϕ):			$\mu_0^{(1)}$	$\mu_0^{(2)}$	$\Gamma^{(1,1)}$	$\Gamma^{(2,1)}$	$\Gamma^{(2,2)}$	$\Upsilon^{(1,1)}$	$\Upsilon^{(2,1)}$
True Value (ϕ^0):			2.93	-1.24	.020	.009	.053	.016	-.011
T=20	N=20	*mean*	2.95	-1.22	.012	.015	.017	.016	-.008
		s.d.	.27	.33	.010	.034	.024	.013	.028
		$\widehat{s.d.}$	.27	.34	.008	.030	.027	.020	.029
	N=40	*mean*	2.94	-1.23	.014	.011	.030	.016	-.008
		s.d.	.18	.24	.009	.026	.023	.011	.024
		$\widehat{s.d.}$	.19	.23	.008	.020	.028	.021	.034
	N=80	*mean*	2.96	-1.26	.016	.011	.032	.014	-.008
		s.d.	.14	.18	.010	.025	.024	.009	.022
		$\widehat{s.d.}$	.13	.17	.007	.026	.017	.011	.018
	N=160	*mean*	2.93	-1.26	.015	.011	.030	.016	-.009
		s.d.	.09	.11	.008	.031	.023	.009	.017
		$\widehat{s.d.}$	.09	.12	.006	.021	.016	.008	.018
T=40	N=20	*mean*	2.97	-1.21	.014	.016	.017	.015	-.011
		s.d.	.27	.34	.010	.035	.023	.013	.024
		$\widehat{s.d.}$	.27	.34	.009	.022	.024	.021	.031
	N=40	*mean*	2.94	-1.23	.016	.009	.043	.016	-.008
		s.d.	.13	.17	.007	.016	.015	.009	.020
		$\widehat{s.d.}$	.17	.18	.006	.015	.018	.010	.018
	N=80	*mean*	2.94	-1.23	.018	.009	.044	.015	-.011
		s.d.	.09	.11	.006	.015	.014	.007	.014
		$\widehat{s.d.}$	.09	.12	.006	.016	.011	.008	.013
	N=160	*mean*	2.93	-1.22	.018	.010	.046	.015	-.009
		s.d.	.06	.10	.006	.018	.010	.007	.011
		$\widehat{s.d.}$	.06	.09	.005	.017	.010	.008	.014
T=80	N=20	*mean*	2.93	-1.24	.019	.011	.046	.017	-.010
		s.d.	.14	.16	.006	.012	.014	.009	.019
		$\widehat{s.d.}$	.12	.17	.005	.012	.010	.012	.020
	N=40	*mean*	2.94	-1.24	.018	.008	.050	.016	-.009
		s.d.	.09	.12	.004	.009	.008	.007	.014
		$\widehat{s.d.}$	.09	.12	.005	.013	.010	.008	.015
	N=80	*mean*	2.92	-1.22	.019	.010	.048	.017	-.011
		s.d.	.06	.09	.004	.013	.009	.005	.011
		$\widehat{s.d.}$	.06	.09	.004	.011	.007	.005	.010
	N=160	*mean*	2.93	-1.22	.019	.010	.048	.016	-.012
		s.d.	.05	.07	.004	.008	.007	.005	.009
		$\widehat{s.d.}$	.05	.06	.004	.012	.007	.005	.010
T=160	N=20	*mean*	2.92	-1.23	.019	.010	.049	.016	-.009
		s.d.	.08	.12	.003	.009	.006	.009	.017
		$\widehat{s.d.}$	.09	.13	.004	.013	.007	.008	.014
	N=40	*mean*	2.93	-1.24	.020	.007	.050	.014	-.011
		s.d.	.07	.10	.003	.007	.005	.006	.011
		$\widehat{s.d.}$	.06	.09	.003	.010	.005	.007	.012
	N=80	*mean*	2.91	-1.22	.019	.008	.051	.016	-.011
		s.d.	.06	.09	.003	.006	.005	.005	.010
		$\widehat{s.d.}$	.05	.06	.003	.009	.005	.004	.009
	N=160	*mean*	2.93	-1.24	.020	.009	.052	.016	-.012
		s.d.	.04	.04	.002	.005	.003	.003	.006
		$\widehat{s.d.}$	.03	.05	.003	.009	005	.003	.007

Table 3: Monte Carlo Results for ϕ.

7 Conclusions

The model considered in this paper generalizes some of the most popular and commonly applied dynamic panel data models in the econometric literature. In particular, by allowing random time effects a feature of potential interest in many econometric applications is incorporated, for example in the context of combining micro data and aggregate data for forecasting purposes. Contrary to the "common wisdom" (expressed by e.g. Matyas (1996), p. 64), it has been shown that likelihood methods can be powerful tools for statistical inference in very complicated random components models, and yet computationally attractive. Maximum likelihood estimation of models with a one-way error structure is an almost trivial special case of the approach.

Many important problems remain, however: The paper throughout assumed a balanced design with no missing observations, but most economic panel data sets lack both these properties. The above model framework can be extended to allow certain types of unbalanced panel data sets, e.g. rotating panels (see Biørn (1981)). This is a topic which will be addressed in future work. Another challenging issue for future work is the assessment of finite sample and asymptotic properties of estimators when both N and T becomes large, e.g. along a diagonal path.

Appendix

A. Proof of Proposition 1

To prove this proposition, a preliminary result is needed:

Lemma 3 *For any sufficiently small* $\delta > 0$, $P(\sup_{|\beta-\beta^0|=\delta,\ \phi\in\Omega_\phi} L(\beta,\phi) > L(\beta_N^*,\phi_N^*)) \to 0$.

Proof. It is sufficient to prove the lemma with "frozen" $F = F^0$, since the unconditional case follows from the dominated convergence theorem. The main idea of the proof is to show that asymptotically the exact likelihood is equivalent to the fixed-effect profile likelihood, i.e. where F is considered as a fixed effect which is "maximized out" of the likelihood.

Let $P(\phi, F)$ be the marginal density of F as a function of (ϕ, F). We then have

$$
\begin{aligned}
L(\theta) &= \ln \int \exp(\sum_{i=1}^{N} l^i(\beta, F)) P(\phi, F) dF \\
&= \ln \int \exp(\sum_{i=1}^{N} l^i(\beta, F) + \ln P(\phi, F)) dF \\
&= \ln \int \exp(Q(\beta, F) + \ln P(\phi, F)) dF \\
&= \ln \int \exp(S(\theta, F)) dF, \qquad (49)
\end{aligned}
$$

where $Q(\beta, F) = \sum_{i=1}^{N} l^i(\beta, F)$ and $S(\theta, F) = Q(\beta, F) + \ln P(\phi, F)$. Let

$$
F_N(\theta) = \arg\max_F S(\theta, F).
$$

Because $S(\theta, F)$ is quadratic in F (see 47), we have

$$
S(\theta, F) = S(\theta, F_N(\theta)) - \frac{1}{2}(F - F_N(\theta))'(-H(\theta))(F - F_N(\theta)), \qquad (50)
$$

where $H(\theta)$ is the Hessian:

$$
H(\theta) = N\nabla^2 l^i(\beta, F_N(\theta)) + \nabla^2 \ln P(\phi, F_N(\theta)), \qquad (51)
$$

with $\triangledown^2$ denoting the second derivative w.r.t. F. Note that $H(\theta)$ does not depend on the data, nor on $F_N(\theta)$.

The formulation (50) allows us to solve the integral in (49):

$$
L(\theta) = Q(\beta, F_N(\theta)) + \ln P(\phi, F_N(\theta)) - \frac{1}{2} \ln |-H(\theta)| \qquad (52)
$$

– ignoring uninteresting constants. We will study the limit behavior of $\frac{1}{N} L(\theta)$ as $N \to \infty$. Starting with the first term in (52), we get

$$
\frac{1}{N} Q(\beta^0, F_N(\theta^0)) = \frac{1}{N} \sum_{i=1}^{N} l^i(\beta^0, F_N(\theta^0)) \xrightarrow{P} E\{l^i(\beta^0, F^0) | F^0\}, \qquad (53)
$$

where F^0 is the actual realization of F. The proof of (53) follows immediately from the proof of (67) below.

Define the conditional profile likelihood

$$
Q^P(\beta) = \max_F Q(\beta, F). \qquad (54)
$$

If we regard the realization F^0 as a vector of fixed effect parameters, the maximizer of $Q(\beta^0, F)$ with respect to F is a weighted least square estimator for F^0 in a linear model, and is therefore consistent. A fortiori

$$
\frac{1}{N} Q^P(\beta^0) \xrightarrow{P} E\{l^i(\beta^0, F^0) | F^0\}. \qquad (55)
$$

Conditional on $F = F^0$, $l^i(\beta^0, F^0)$ are *i.i.d.* random variables. It is therefore possible to generalize a classical result (see Lehman (1983), p.430) to the profile likelihood $Q^P(\beta^0)$:

$$P(\sup_{|\beta-\beta^0|=\delta} \frac{1}{N}Q^P(\beta) - \frac{1}{N}Q^P(\beta^0) > -\varepsilon) \to 0 \tag{56}$$

for sufficiently small δ and ε. To show (56), a second order Taylor expansion of $\frac{1}{N}Q^P(\beta)$ is sufficient. We have

$$\begin{aligned}
\frac{1}{N}\frac{\partial Q^P(\beta^0)}{\partial \beta} &= \frac{\partial Q(\beta^0, F^0)}{\partial \beta} + o_p(1) \\
\frac{1}{N}\frac{\partial^2 Q^P(\beta^0)}{\partial \beta \partial \beta'} &= \frac{1}{N}\frac{\partial^2 Q(\beta^0, F^0)}{\partial \beta \partial \beta'} - \frac{1}{N}\frac{\partial^2 Q(\beta^0, F^0)}{\partial \beta \partial F'}\left[\frac{1}{N}\frac{\partial^2 Q(\beta^0, F^0)}{\partial F \partial F'}\right]^{-1}\frac{1}{N}\frac{\partial^2 Q(\beta^0, F^0)}{\partial F \partial \beta'} + o_p(1).
\end{aligned} \tag{57}$$

By standard arguments, $\frac{1}{N}\frac{\partial Q(\beta^0, F^0)}{\partial \beta} \xrightarrow{P} 0$, while the first order term in the expansion of $\frac{1}{N}\frac{\partial^2 Q^P(\beta^0)}{\partial \beta \partial \beta'}$ tends to a positive definite matrix. The remainders in (57) are due to the argmax being $o_p(1)$ away from F^0. Provided we can show that the third order terms in the expansion of $\frac{1}{N}Q^P(\beta)$ are uniformly bounded in probability over the set $\{\beta : |\beta - \beta^0| \leq \delta\}$, equation (56) follows from the arguments in Lehman (1983), p. 430 (with $Q^P(\beta)$ in place of $L(\theta)$). But this is true because of continuity of third derivatives, and because the argmax in (54) is a regression estimator which is the mean of N *i.i.d.* terms which are continuous functions of β.

We have from (53) and (55)

$$\frac{1}{N}Q(\beta^0, F_N(\theta^0)) = \frac{1}{N}Q^P(\beta^0) + o_p(1). \tag{58}$$

Define

$$M = N_\delta(\beta^0) \times \Omega_\phi,$$

where $N_\delta(\beta^0)$ is a closed sphere in Ω_β with radius δ and centre at β^0.

We shall now investigate the second term in (52). We will see that *(i)* $\frac{1}{N}\ln P(\phi, F_N(\theta))$ converges to zero in probability evaluated at $\theta = \theta^0$, and *(ii)* that this term, for fixed N, is uniformly upwards bounded on M by a constant that goes to zero as $N \to \infty$. Part *(i)* follows from $F_N(\theta^0) \xrightarrow{P} F^0$, which implies

$$\frac{1}{N}\ln P(\phi^0, F_N(\theta^0)) \xrightarrow{P} 0. \tag{59}$$

Regarding *(ii)*, $\ln P(\phi, F_N(\theta))$ depends on $F_N(\theta)$ only through a non-positive quadratic form. Because of the smoothness of $\ln P(\phi, F_N(\theta))$ in ϕ, and compactness of M:

$$\frac{1}{N}\sup_{\theta \in M} \ln P(\phi, F_N(\theta)) \leq \frac{C}{N} \to 0 \tag{60}$$

for some constant $C < \infty$.

Finally, consider the last term in (52). We have

$$\begin{aligned}\frac{1}{N}\ln|-H(\theta)| &= \frac{1}{N}\ln|\frac{-H(\theta)}{N}| + \frac{pT}{N}\ln N \\ &= \frac{1}{N}\ln|\nabla^2 l^i(\beta, F_N(\theta)) + \frac{1}{N}\nabla^2 \ln P(\phi, F_N(\theta))| \\ &+ \frac{pT}{N}\ln N \end{aligned} \tag{61}$$

(recall that $H(\theta)$ has dimension pT). Since $\frac{pT}{N}\ln N \to 0$

$$\frac{1}{N}\ln|-H(\theta^0)| \xrightarrow{P} 0. \tag{62}$$

Due to compactness of M, smoothness of the functions involved, and the fact that the Hessian $H(\theta)$ does not depend on the data, nor on $F_N(\theta)$, $|\nabla^2 l^i(\beta, F_N(\theta)) + \frac{1}{N}\nabla^2 \ln P(\phi, F_N(\theta))|$ is bounded on M. Hence

$$\sup_{\theta\in M}\left|\frac{1}{N}\ln|-H(\theta)|\right| \xrightarrow{P} 0. \tag{63}$$

By (52), (60), and (63)

$$\begin{aligned}\sup_{|\beta-\beta^0|=\delta,\phi e\Omega_\phi}\frac{1}{N}L(\theta) &\le \frac{1}{N}\sup_{|\beta-\beta^0|=\delta,\phi e\Omega_\phi} Q(\beta, F_N(\theta)) + r_N \\ &\le \frac{1}{N}\sup_{|\beta-\beta^0|=\delta} Q^P(\beta) + r_N, \end{aligned} \tag{64}$$

where $r_N \xrightarrow{P} 0$. By (52), (58), (59) and (62)

$$\frac{1}{N}L(\theta^0) = \frac{1}{N}Q^P(\beta^0) + o_p(1). \tag{65}$$

We shall now see that the limit behavior of $\frac{1}{N}L(\theta_N^*)$ is the same as that of $\frac{1}{N}L(\theta^0)$. From (52):

$$\begin{aligned}\frac{1}{N}L(\theta_N^*) &= \frac{1}{N}Q(\beta_N^*, F_N(\theta_N^*)) + \frac{1}{N}\ln P(\phi_N^*, F_N(\theta_N^*)) \\ &- \frac{1}{2N}\ln|-H(\theta_N^*)|. \end{aligned} \tag{66}$$

By almost identical arguments as above, the second and third terms in (66) vanish, whereas

$$\frac{1}{N}Q(\beta_N^*, F_N(\theta_N^*)) \xrightarrow{P} E\{l^i(\beta^0, F^0)|F^0\} \tag{67}$$

(to be proved below) and therefore

$$\frac{1}{N}L(\theta_N^*) = \frac{1}{N}L(\theta^0) + o_p(1). \tag{68}$$

The proof of (67) goes as follows: $F_N(\theta_N^*)$ satisfies the 1. order condition

$$\begin{aligned}\frac{1}{N}\left.\frac{\partial S(\theta_N^*, F)}{\partial F}\right|_{F=F_N(\theta_N^*)} &= \frac{1}{N}\sum_{i=1}^N \left.\frac{\partial l^i(\beta_N^*, F)}{\partial F}\right|_{F=F_N(\theta_N^*)} \\ &+ \frac{1}{N}\left.\frac{\partial \ln P(\phi_N^*, F)}{\partial F}\right|_{F=F_N(\theta_N^*)} \\ &= 0. \end{aligned} \tag{69}$$

As $\beta_N^* \xrightarrow{P} \beta^0$,

$$\frac{1}{N}\frac{\partial S(\theta_N^*, F)}{\partial F} \xrightarrow{P} E\left\{\frac{\partial l^i(\beta^0, F)}{\partial F}|F^0\right\}. \tag{70}$$

Hence, the first order condition for maximizing $S(\beta_N^*, F)$ w.r.t. F converges pointwise (for given F) to the first order condition for maximizing $E\{l^i(\beta^0, F)|F^0\}$. Because of the linearity of equation (69), the argmax also converges (see Lehman (1983), Lemma 6.4.1, p.432). Hence $F_N(\theta_N^*) \xrightarrow{P} F^0$. If we replace θ_N^* by θ^0 in (69) and (70), we get a proof of (53) as well.

From (65) and (68)

$$\frac{1}{N}L(\theta_N^*) = \frac{1}{N}Q^P(\beta^0) + \varepsilon_N, \tag{71}$$

where $\varepsilon_N \xrightarrow{P} 0$. In particular,

$$r_N - \varepsilon_N = o_p(1), \tag{72}$$

where r_N is defined in (64). For a sufficiently small $\varepsilon > 0$, and conditional on F^0, we then have

$$\begin{aligned}
&P(\sup_{|\beta-\beta^0|=\delta, \phi e \Omega_\phi} \frac{1}{N}L(\theta) - \frac{1}{N}L(\theta_N^*) > 0) \\
&= P(\sup_{|\beta-\beta^0|=\delta, \phi e \Omega_\phi} \frac{1}{N}L(\theta) - \frac{1}{N}L(\theta_N^*) > 0 \cap [-\varepsilon + r_N - \varepsilon_N > 0]) \\
&+ P\left(\sup_{|\beta-\beta^0|=\delta, \phi e \Omega_\phi} \frac{1}{N}L(\theta) - \frac{1}{N}L(\theta_N^*) > 0 \cap [-\varepsilon + r_N - \varepsilon_N \leq 0]\right) \\
&\leq P(r_N - \varepsilon_N > \varepsilon) \\
&+ P\left(\sup_{|\beta-\beta^0|=\delta, \phi e \Omega_\phi} \frac{1}{N}L(\theta) - \frac{1}{N}L(\theta_N^*) > -\varepsilon + r_N - \varepsilon_N\right) \\
&= P(r_N - \varepsilon_N > \varepsilon) \\
&+ P\left(\sup_{|\beta-\beta^0|=\delta, \phi e \Omega_\phi} \frac{1}{N}L(\theta) - r_N - (\frac{1}{N}L(_N^*) - \varepsilon_N) > -\varepsilon\right) \\
&\leq P(r_N - \varepsilon_N > \varepsilon) \\
&+ P\left(\sup_{|\beta-\beta^0|=\delta} \frac{1}{N}Q^P(\beta) - \frac{1}{N}Q^P(\beta^0) > -\varepsilon\right) \text{ (by (64) and (71)} \\
&\to 0 \text{ (by (56) and (72)).}
\end{aligned}$$

■

PROOF OF PROPOSITION 1: Consistency of $\widehat{\beta}_N$ will now be proved. Assume, on the contrary, that $\widehat{\beta}_N$ is not consistent. Then there exists $\delta > 0$ such that

$$\limsup_N P(\widehat{\beta}_N \notin N_\delta(\beta^0)) > 0, \tag{73}$$

where $N_\delta(\beta^0)$ is a closed sphere in Ω_β with radius δ and centre at β^0. Each time $\beta^{(1)} = \beta_N^* \in N_\delta(\beta^0)^\circ$; the interior of $N_\delta(\beta^0)$, which happens with probability tending to one, and $\widehat{\beta}_N \notin N_\delta(\beta^0)$, there exist $\beta^{(m)}$ and $\beta^{(m+1)}$, such that $\beta^{(m)} \in N_\delta(\beta^0)^\circ$ and $\beta^{(m+1)} \notin N_\delta(\beta^0)^\circ$. At the start of the m'th CM step $\theta^{(m)} = (\theta_1^{(m)}, \theta_2^{(m)})$ is given. In the first phase of this CM step $q(\theta_1)$ is maximized (see (37)), and we obtain an intermediate point $\theta^{(m+1/2)} = (\theta_1^{(m+1)}, \theta_2^{(m)})$. In the second phase, we obtain $\theta^{(m+1)} = (\theta_1^{(m+1)}, \theta_2^{(m+1)})$ by maximizing $r(\theta_2)$ (see (38)). If the intermediate point $\theta^{(m+1/2)}$ is outside $N_\delta(\beta^0)^\circ$, but $\theta^{(m)}$ is inside, consider the straight line from $\theta_1^{(m)}$ to $\theta_1^{(m+1)}$, and define $h(c) = q(\theta_1^{(m)} + cs)$ where $s = \theta_1^{(m+1)} - \theta_1^{(m)}$. Because $q(\theta_1)$ is concave with a maximizer at $\theta_1^{(m+1)}$, $h(c)$ is concave with a maximizer at $c = 1$. In particular, $h'(c) \geq 0$ for all $c \in [0, \overline{c}]$, where $\overline{c} \leq 1$ is defined by $\overline{\theta} = (\theta_1^{(m)} + \overline{c}s, \theta_2^{(m)})$, $\overline{\theta} = (\overline{\beta}, \overline{\phi})$, and $|\overline{\beta} - \beta^0| = \delta$.

On the other hand, if $\theta^{(m+1/2)}$ is inside $N_\delta(\beta^0)^\circ$ and $\theta^{(m+1)}$ is outside, we consider the function $r(\theta_2)$. This function is concave in the parameters $\varphi = (vech(\Sigma^{-1})', vech(\widetilde{\Omega}^{-1})')$, where $\theta_2 = g(\varphi)$ is one-to-one, with $\theta_2^{(m)} = g(\varphi^{(m)})$ and $\theta_2^{(m+1)}) = g(\varphi^{(m+1)})$. Define $h(c) = r(g(\varphi^{(m)} + cs))$ where $s = \varphi^{(m+1)} - \varphi^{(m)}$. Since $h(c)$ is concave with a maximizer at $c = 1$, $h'(c) \geq 0$ for all $c \in [0, \overline{c}]$, where $\overline{\theta} = (\theta_1^{(m+1)}, g(\varphi_2^{(m)} + \overline{c}s))$, $\overline{\theta} = (\overline{\beta}, \overline{\phi})$ and $|\overline{\beta} - \beta^0| = \delta$ defines $\overline{c}$.

In either case, there is a trajectory connecting both conditional maxima generated in the m'th CM step, such that $M(\theta|\theta^{(m)})$ is increasing along this trajectory as long as we are inside $N_\delta(\beta^0)$. Let $\overline{\theta}_N = (\bar{\beta}_N, \bar{\phi}_N)$ be the point where the trajectory first crosses $N_\delta(\beta^0)$. Thus $|\bar{\beta}_N - \beta^0| = \delta$, and

$$\begin{aligned} L(\bar{\beta}_N, \bar{\phi}_N) &= M(\overline{\theta}|\theta^{(m)}) - H(\overline{\theta}|\theta^{(m)}) \\ &> M(\theta^{(m)}|\theta^{(m)}) - H(\theta^{(m)}|\theta^{(m)}) \\ &= L(\theta^{(m)}) \\ &\geq L(\beta_N^*, \phi_N^*). \end{aligned}$$

Define a new estimator

$$(\tilde{\beta}_N, \tilde{\phi}_N) = \begin{cases} (\bar{\beta}_N, \bar{\phi}_N) & \text{if } \widehat{\beta}_N \notin N_\delta(\beta^0) \text{ and } \beta_N^* \in N_\delta(\beta^0)^\circ \\ (\beta_N^*, \phi_N^*) & \text{else.} \end{cases}$$

If $\widehat{\beta}_N$ is inconsistent

$$0 < \limsup_N P\left(L((\tilde{\beta}_N, \tilde{\phi}_N)) > L(\beta_N^*, \phi_N^*)\right) \leq \limsup_N P\left(\sup_{|\beta - \beta^0| = \delta,\ \phi \in \Omega_\phi} L(\beta, \phi) > L(\beta_N^*, \phi_N^*)\right),$$

but this is impossible, because of Lemma 3. ■

B. Proof of Proposition 2

We need two preliminary results. The first is a corollary from the convergence theorem of Wu (1983), about the generalized EM (GEM) algorithm. Since ECM is a special case of GEM (Meng and Rubin (1993), Theorem 1), we have:

Theorem 4 *(Wu, 1983). Let a solution set $\mathcal{L}$ be given, and let $\theta^{(m)} \in \Theta$ be a ECM sequence generated by the mapping $\theta^{(m+1)} \in \nu(\theta^{(m)})$. Suppose that (i) ν is a closed point-to-set map over the complement of $\mathcal{L}$, and (ii) $L(\theta^{(m+1)}) > L(\theta^{(m)})$ for all $\theta^{(m)} \notin \mathcal{L}$. Then all limit points of $\{\theta^{(m)}\}$ are in $\mathcal{L}$, and $L(\theta^{(m)})$ converges monotonically to $L(\widehat{\theta})$ for some $\widehat{\theta} \in \mathcal{L}$.*

The next result characterizes the limit points of the ECM algorithm.

Proposition 5 *If $\widehat{\theta} = (\widehat{\beta}, \widehat{\phi})$ is a limit point of the ECM algorithm, then with probability tending to one $\frac{\partial L(\widehat{\beta},\widehat{\phi})}{\partial \beta} = 0$ and $L(\beta^{(m)}, \phi^{(m)})$ converges monotonically towards $L(\widehat{\beta}, \widehat{\phi})$. If $\widehat{\phi}$ is an interior point of Ω_ϕ, then $\frac{\partial L(\widehat{\beta},\widehat{\phi})}{\partial \phi} = 0$.*

Proof. We will use the convergence theorem of Wu with $\mathcal{L} = \{(\widehat{\beta}, \widehat{\phi}) : \frac{\partial L(\widehat{\beta},\widehat{\phi})}{\partial \beta} = 0 \ \mathit{\&} \ \frac{\partial L(\widehat{\beta},\widehat{\phi})}{\partial \phi} = 0$ *if $\widehat{\phi}$ is an interior point of Ω_ϕ*$\}$. (i) is easily seen to be satisfied because $M(\theta|\theta^{(m)})$ is continuous in θ and $\theta^{(m)}$, and Θ is compact. It remains to show (ii): Assume that $\theta^{(m)} \notin \mathcal{L}$ and $\beta^{(m)} \in N_\delta(\beta^0)$ for all m. Then, *either* $\frac{\partial L(\beta^{(m)},\phi^{(m)})}{\partial \beta} = \frac{\partial \theta}{\partial \beta} \frac{\partial M(\theta|\theta^{(m)})}{\partial \theta}\Big|_{\theta=\theta^{(m)}} \neq 0$ for an interior point $\beta^{(m)}$ of Ω_β *or* $\frac{\partial L(\beta^{(m)},\phi^{(m)})}{\partial \phi} = \frac{\partial \theta}{\partial \phi} \frac{\partial M(\theta|\theta^{(m)})}{\partial \theta}\Big|_{\theta=\theta^{(m)}} \neq 0$ for an interior point $\phi^{(m)}$ of Ω_ϕ. In the first case $M(\theta|\theta^{(m)})$ can be increased during the m'th CM step by changing β. In the second case, $M(\theta|\theta^{(m)})$ can be increased by changing ϕ. Hence $L(\theta^{(m+1)}) > L(\theta^{(m)})$. The conclusion now follows because $\beta^{(m)} \in N_\delta(\beta^0)$ with probability tending to one by the arguments used in the proof of Proposition 1. ■

PROOF OF PROPOSITION 2.

Let $\widehat{\beta}_N = \beta^0 + \frac{\widehat{s}}{\sqrt{N}}$, $F = F^0 + \frac{t}{\sqrt{N}}$, and $\widehat{t} = E\{t|Y;\hat{\theta}\}$. By Proposition 5, $\frac{\partial L(\widehat{\beta}_N,\widehat{\phi}_N)}{\partial \beta} = \frac{\partial \theta}{\partial \beta} \frac{\partial M(\theta|\widehat{\theta}_N)}{\partial \theta}\Big|_{\theta=\widehat{\theta}_N} = 0$ with probability tending to one. Hence

$$\begin{aligned} & E\left\{\frac{1}{\sqrt{N}} Q_\beta(\beta^0, F^0) + \frac{1}{N} Q_{\beta\beta}(\beta^0, F^0)\widehat{s} + \frac{1}{N} Q_{\beta F}(\beta^0, F^0)t + o_p(1)|Y;\hat{\theta}\right\} \\ & = \frac{1}{\sqrt{N}} Q_\beta(\beta^0, F^0) + \frac{1}{N} Q_{\beta\beta}(\beta^0, F^0)\widehat{s} + \frac{1}{N} Q_{\beta F}(\beta^0, F^0)\widehat{t} + o_p(1) \\ & = 0, \end{aligned} \tag{74}$$

where the remainder vanishes on a set with probability tending to one by consistency and the arguments used in the proof of Lemma 3. The conditional expectation $\widehat{t}$ is the stationary point on the

"complete data" log-likelihood given $\widehat{\theta}$. Therefore

$$\frac{1}{\sqrt{N}}Q_F(\beta^0,F^0)+\frac{1}{N}Q_{F\beta}(\beta^0,F^0)\widehat{s}+\frac{1}{N}Q_{FF}(\beta^0,F^0)\widehat{t}+o_p(1)=0. \tag{75}$$

Then, from (74) and (75),

$$\left[\begin{array}{c}\widehat{s}\\ \widehat{t}\end{array}\right]=J^{-1}\frac{1}{\sqrt{N}}\left[\begin{array}{c}Q_\beta(\beta^0,F^0)\\ Q_F(\beta^0,F^0)\end{array}\right]+o_{p(1)},$$

where

$$J=-\frac{1}{N}\left[\begin{array}{cc}Q_{\beta\beta}(\beta^0,F^0) & Q_{\beta F}(\beta^0,F^0)\\ Q_{F\beta}(\beta^0,F^0) & Q_{FF}(\beta^0,F^0)\end{array}\right]. \tag{76}$$

Since $Q(\beta^0,F^0)=\sum_{i=1}^N l^i(\beta^0,F^0)$,

$$J\xrightarrow{P}-E\left\{\left[\begin{array}{cc}l^i_{\beta\beta}(\beta^0,F^0) & l^i_{\beta F}(\beta^0,F^0)\\ l^i_{F\beta}(\beta^0,F^0) & l^i_{FF}(\beta^0,F^0)\end{array}\right]|F^0\right\}\equiv I.$$

By a standard result

$$Var\left\{\left[\begin{array}{c}l^i_\beta(\beta^0,F^0)\\ l^i_F(\beta^0,F^0)\end{array}\right]|F^0\right\}=I,$$

so it follows from the central limit theorem that, for $F=F^0$,

$$\left[\begin{array}{c}\sqrt{N}(\widehat{\beta}_N-\beta^0)\\ \sqrt{N}(\widehat{F}_N-F^0)\end{array}\right]\overset{D}{\Rightarrow}N(0,I^{-1}),$$

where $\widehat{F}_N=F^0+\frac{\widehat{t}}{\sqrt{N}}$The first part of the proposition thus follows.

To prove the second part, we first note that

$$\begin{aligned}(J^{\beta\beta})^{-1}&=\frac{-1}{N}(Q_{\beta\beta}(\beta^0,F^0)-Q_{\beta F}(\beta^0,F^0)Q_{FF}^{-1}(\beta^0,F^0)Q_{F\beta}(\beta^0,F^0))\\&=(\frac{-1}{N}Q^P_{\beta\beta}(\beta^0))+o_p(1),\end{aligned} \tag{77}$$

where the second equation follows from the envelope theorem and from differentiating the first order condition for maximizing $Q(\beta^0,F)$ with respect to F (see Barndorff-Nielsen and Cox (1994)). The remainder is due to the maximizer being $o_p(1)$ away from F^0. Define

$$\tilde{L}(\beta,\phi,F)=Q(\beta,F)+\ln P(\phi,F)-\frac{1}{2}\ln|-H(\theta)| \tag{78}$$

$$L^*(\beta,F)=\max_\phi\tilde{L}(\beta,\phi,F) \tag{79}$$

$$L^P(\beta)=\max_F L^*(\beta,F). \tag{80}$$

By reversing the order of maximization in (79)-(80), while noticing that $F_N(\theta)$ is the maximizer of (78), we see that

$$L^P(\beta)=\max_\phi L(\beta,\phi).$$

– i.e. the log-likelihood profile. By (79)-(80),

$$\frac{-1}{N}L^P_{\beta\beta}(\beta^0) = \frac{-1}{N}(L^*_{\beta\beta}(\beta^0, F^0) - L^*_{\beta F}(\beta^0, F^0)L^*_{FF}(\beta^0, F^0)^{-1}L^*_{F\beta}(\beta^0, F^0)) + o_p(1),$$

and by (78)-(79) and direct calculations

$$\begin{aligned}
\frac{1}{N}L^*_{\beta\beta}(\beta^0, F^0) &= \frac{1}{N}Q_{\beta\beta}(\beta^0, F^0) + o_p(1) \\
\frac{1}{N}L^*_{\beta F}(\beta^0, F^0) &= \frac{1}{N}Q_{\beta F}(\beta^0, F^0) + o_p(1) \\
\frac{1}{N}L^*_{FF}(\beta^0, F^0) &= \frac{1}{N}Q_{FF}(\beta^0, F^0) + o_p(1).
\end{aligned}$$

Hence, using (77)

$$(J^{\beta\beta})^{-1} = \frac{-1}{N}L^P_{\beta\beta}(\beta^0) + o_p(1).$$

A consistent estimator of $I^{\beta\beta}$ is therefore $\widehat{H}^{\beta\beta}$, since $\widehat{H}^{\beta\beta}$ equals $\frac{-1}{N}L^P_{\beta\beta}(\widehat{\beta}_N)^{-1}$, and $\widehat{\beta}_N$ is a consistent estimator of β^0.

■

References

Ahn, S.C. and Schmidt, P. (1995). "Efficient Estimation of Models for Panel Data". *Journal of Econometrics*, 68, 5-27.

Anderson, B., Moore, L. (1979). *Optimal Filtering.* Englewood Cliffs, Prentice Hall., N.J.

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis.* John Wiley, New York.

Anderson, T.W. and Hsiao, C. (1982). "Formulation and Estimation of Dynamic Models using Panel Data". *Journal of Econometrics*, 18, 578-606.

Arellano, M. and Bover, O. (1995). "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, 68, 29-51.

Barndorff-Nielsen, O.E., and Cox, D. R. (1994). *Inference and Asymptotics.* Chapman and Hall, London.

Binder, M., C. Hsaio, and M. H. Pesaran (2000). "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration." *DAE Working Paper 0003*, University of Cambridge.

Biørn, E. (1981). "Estimating Economic Relations from Incomplete Cross-Section/Time-Series Data", *Journal of Econometrics*, 16, 221-36.

Blundell, R. and Bond, S. (1998). "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics*, 87, 115-143.

Caballero, R.J. and E. Engel (1999). "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S, s) Approach". *Econometrica*, 67, 783-826.

Caballero, R.J, E. Engel, and J.C. Haltiwanger (1995). "Plant-Level Adjustment and Aggregate Investment Dynamics." *Brookings Papers on Economic Activity*, 2, 1-39.

Clements, M.P and D.F. Hendry (1996). "Intercept Corrections and Structural Change."*Journal of Applied Econometrics*, 11, 475-94.

Dempster, A.P., Laird,N., and Rubin, D.B. (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society* B39, 1-38.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer, New York.

Gourieroux, C. and Montfort, A. (1995). *Statistics and Econometric Models*. Cambridge University Press, Cambridge.

Grunfeld, Y. and Zvi Griliches (1960). Is Aggregation Necessarily Bad. The Review of Economics and Statistics, XLII (1), 21-33.

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

Holtz-Eakin, D., Newey,W., and Rosen, H.S. (1988). "Estimating Vector Autoregressions with Panel Data." *Econometrica*, 56, 1371-1395

Icaza, G. and R.H. Jones (1999). "A State Space EM Algorithm for Longitudinal Data." *Journal of Time Series Analysis,* 20, 537-50.

Johansen, S. (1995). "The Role of Ancillarity in Inference for Non-stationary Variables." *Economic Journal*. 105, 302-20.

Jones, R.H. (1993). *Longitudinal data with serial correlation: A state space approach*. Chapman and Hall, London.

Lehman, E.L. (1983). *Theory of Point Estimation*. John Wiley, New York.

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley, New York.

Matyas, L (1996). "Error Components Models" in *The Econometrics of Panel Data*, Matyas, M and Sevestre, P. (eds.), Kluwer, Dordrecht.

Meng, X. L. and Rubin, D.B. (1992). "Recent Extensions to the EM Algorithm". *Bayesian Statistics*, 4, 307-320.

Meng, X. L. and Rubin, D.B. (1993). "Maximum Likelihood Estimation via the ECM Algorithm: A General Framework". *Biometrika*, 80, 267-78.

Nyblom J. and A.C. Harvey (2000). "Tests of Common Stochastic Trends." *Econometric Theory*, 16, 176-99.

Pesaran, M.H. (1999). "On Aggregation of Linear Dynamic Models." Mimeo. University of Cambridge.

Phillips, P.C.B. and H.R. Moon (2000). "Nonstationary Panle Data Analysis: An overview of Some Recent Developments." Econometric Reviews, 19, 263-86.

Quah, D.and Sargent, T.J.(1993). "A Dynamic Index Model for Large Cross Sections", in *Business cycles, indicators, and forecasting,* Stock, J.H. and Watson, M.W. (eds), University of Chicago Press, London.

Sims, C. (2000). "Using a likelihood perspective to sharpen econometric discourse Three examples." *Journal of Econometrics*, 95, 443-62.

Theil, H. (1954). *Linear Aggregation of Economic Relations.* Amsterdam.

Wu, C. F. Jeff (1983). "On the Convergence Properties of the EM algorithm". *The Annals of Statistics*, 11,1, 95-103.

Recent publications in the series Discussion Papers

204 A. Rygh Swensen (1997): Change in Regime and Markov Models

205 K. Nyborg and I. Spangen (1997): Cost-Benefit Analysis and the Democratic Ideal

206 L. Belsby and J.F. Bjørnstad (1997): Modelling and Estimation Methods for Household Size in the Presence of Nonresponse: Applied to The Norwegian Consumer Expenditure Survey

207 K.O. Aarbu and T.O. Thoresen (1997): The Norwegian Tax Reform; Distributional Effects and the High-Income Response

208 T. Hægeland and T.J. Klette (1997): Do Higher Wages Reflect Higher Productivity? Education, Gender and Experience Premiums in a Matched Plant-Worker Data Set

209 J. Gjerde, S. Grepperud and S. Kverndokk (1998): Optimate Climate Policy under the Possibility of a Catastrophe

210 T. Eika and K.A. Magnussen (1998): Did Norway Gain from the 1979-85 Oil Price Shock?

211 K.O. Aarbu and J.K. MacKie-Mason (1998): Why Some Corporations Pay More Tax than Necessary

212 R. Aaberge (1998): UMP Unbiased Tests for Multiparameter Testing Problems with Restricted Alternatives

213 M. Søberg (1998): "EPA's New Emissions Trading Mechanism: A Laboratory Evaluation" – A Comment

214 K. Nyborg (1998): Non-Verifiable Emissions, Voluntary Agreements, and Emission Taxes

215 H. C. Bjørnland (1998): Economic Fluctuations in a Small Open Economy - Real versus Nominal Stocks

216 L.C. Zhang (1998): Post-Stratification and Calibration – A Synthesis

217 R. Aaberge and Y. Zhu: The Pattern of Household Savings during a Hyperinflation. The Case of Urban China in the Late 1980s

218 L.C. Zhang (1998): A Note on Post-stratification when Analyzing Binary Survey Data Subject to Non-response

219 T.J. Klette and J. Møen (1998): From Growth Theory to Technology Policy – Coordination Problems in Theory and Practice

220 L.C. Zhang (1998): Synthetic Graphical Modelling of Binary Panel Data Subject to Non-response. A Case Study

221 J.K. Dagsvik (1998): Choice among Lotteries when Preferences are Stochastic

222 J.K. Dagsvik (1998): Nonparametric Identification of Discrete Choice Models

223 L. Brubakk and J.K. Dagsvik (1998): Consumer Demand and Unobservable Product Attributes

224 E. Holmøy (1998): Potential Aggregate Gains from Improved Sectoral Efficiency: General Equilibrium Estimates for Norway

225 B.H. Vatne and J.K. Dagsvik (1998): Estimation of Generalized Extreme Value Models by a Max-spectral Representation

226 I. Svendsen (1998): Rational Expectations in Price Setting – Tests Based on Norwegian Export Prices

227 K.A. Brekke and R.B. Howarth (1998): The Social Contingency of Wants: Implications for Growth and the Environment

228 K.A. Brekke and E. Moxnes (1998): Do Models Improve Fishery Management? Empirical Evidence from a Experimental Study

229 J.K. Dagsvik, Yu Zhu and R. Aaberge (1998): A Framework for Empirical Modelling of Consumer Demand with Latent Quality Attributes

230 R. Aaberge, U. Colombino and S. Strøm (1998): Social Evaluation of Individual Welfare Effects from Income Taxation: Empirical Evidence Based on Italian Data for Married Couples

231 R. Nesbakken (1998): Residential Energy Consumption for Space Heating in Norwegian Households. A Discrete-Continuous Choice Approach

232 R. Nesbakken (1998): Price Sensitivity of Residential Energy Consumption in Norway

233 M. Søberg (1998): Uncertainty and International Negotiations on Tradable Quota Treaties

234 J.K. Dagsvik and L. Brubakk: Price Indexes for Elementary Aggregates Derived from Behavioral Assumptions

235 E. Biørn, K.-G. Lindquist and T. Skjerpen (1998): Random Coefficients and Unbalanced Panels: An Application on Data from Norwegian Chemical Plants

236 K. Ibenholt (1998): Material Accounting in a Macroeconomic Framework. Forecast of waste generated in manufacturing industries in Norway

237 K-G. Lindquist (1998): The Response by the Norwegian Aluminium Industry to Changing Market Structure

238 J.K. Dagsvik, A.S. Flaatten and H. Brunborg: A Behavioral Two-Sex Model

239 K.A. Brekke, R.B. Howarth and K. Nyborg (1998): Are there Social Limits to Growth?

240 R. B. Howarth and K. A. Brekke (1998): Status Preferences and Economic Growth

241 H. Medin, K. Nyborg and I. Bateman (1998): The Assumption of Equal Marginal Utility of Income: How Much Does it Matter?

242 B. Bye (1998): Labour Market Rigidities and Environmental Tax Reforms: Welfare Effects of Different Regimes

243 B.E. Naug (1999): Modelling the Demand for Imports and Domestic Output

244 J. Sexton and A. R. Swensen (1999): ECM-algorithms that converge at the rate of EM

245 E. Berg, S. Kverndokk and K.E. Rosendahl (1999): Optimal Oil Exploration under Climate Treaties

246 J.K. Dagsvik and B.H. Vatne (1999): Is the Distribution of Income Compatible with a Stable Distribution?

247 R. Johansen and J.K. Dagsvik (1999): The Dynamics of a Behavioral Two-Sex Demographic Model

248 M. Søberg (1999): Asymmetric information and international tradable quota treaties. An experimental evaluation

249 S. Grepperud, H. Wiig and F.A. Aune (1999): Maize Trade Liberalization vs. Fertilizer Subsidies in Tanzania: A CGE Model Analysis with Endogenous Soil Fertility

250 K.A. Brekke and Nils Chr. Stenseth (1999): A Bio-Economic Approach to the study of Pastoralism, Famine and Cycles. Changes in ecological dynamics resulting from changes in socio-political factors

251 T. Fæhn and E. Holmøy (1999): Welfare Effects of Trade Liberalisation in Distorted Economies. A Dynamic General Equilibrium Assessment for Norway

252 R. Aaberge (1999): Sampling Errors and Cross-Country Comparisons of Income Inequality

253 I. Svendsen (1999): Female labour participation rates in Norway – trends and cycles

254 A. Langørgen and R. Aaberge: A Structural Approach for Measuring Fiscal Disparities

255 B. Halvorsen and B.M. Larsen (1999): Changes in the Pattern of Household Electricity Demand over Time

256 P. Boug (1999): The Demand for Labour and the Lucas Critique. Evidence from Norwegian Manufacturing

257 M. Rege (1999): Social Norms and Private Provision of Public Goods: Endogenous Peer Groups

258 L. Lindholt (1999): Beyond Kyoto: CO_2 permit prices and the markets for fossil fuels

259 R. Bjørnstad and R. Nymoen (1999): Wage and Profitability: Norwegian Manufacturing 1967-1998

260 T.O. Thoresen and K.O. Aarbu (1999): Income Responses to Tax Changes – Evidence from the Norwegian Tax Reform

261 B. Bye and K. Nyborg (1999): The Welfare Effects of Carbon Policies: Grandfathered Quotas versus Differentiated Taxes

262 T. Kornstad and T.O. Thoresen (1999): Means-testing the Child Benefit

263 M. Rønsen and M. Sundström (1999): Public Policies and the Employment Dynamics among new Mothers – A Comparison of Finland, Norway and Sweden

264 J.K. Dagsvik (2000): Multinomial Choice and Selectivity

265 Y. Li (2000): Modeling the Choice of Working when the Set of Job Opportunities is Latent

266 E. Holmøy and T. Hægeland (2000): Aggregate Productivity and Heterogeneous Firms

267 S. Kverndokk, L. Lindholt and K.E. Rosendahl (2000): Stabilisation of CO_2 concentrations: Mitigation scenarios using the Petro model

268 E. Biørn, K-G. Lindquist and T. Skjerpen (2000): Micro Data On Capital Inputs: Attempts to Reconcile Stock and Flow Information

269 I. Aslaksen and C. Koren (2000): Child Care in the Welfare State. A critique of the Rosen model

270 R. Bjørnstad (2000): The Effect of Skill Mismatch on Wages in a small open Economy with Centralized Wage Setting: The Norwegian Case

271 R. Aaberge (2000): Ranking Intersecting Lorenz Curves

272 J.E. Roemer, R. Aaberge , U. Colombino, J, Fritzell, S.P. Jenkins, I. Marx, M. Page, E. Pommer, J. Ruiz-Castillo, M. Jesus SanSegundo, T. Tranaes, G.G.Wagner and I. Zubiri (2000): To what Extent do Fiscal Regimes Equalize Opportunities for Income Acquisition Among citizens?

273 I. Thomsen and L.-C. Zhang (2000): The Effect of Using Administrative Registers in Economic Short Term Statistics: The Norwegian Labour Force Survey as a Case Study

274 I. Thomsen, L.-C. Zhang and J. Sexton (2000): Markov Chain Generated Profile Likelihood Inference under Generalized Proportional to Size Non-ignorable Non-response

275 A. Bruvoll and H. Medin (2000): Factoring the environmental Kuznets curve. Evidence from Norway

276 I. Aslaksen, T. Wennemo and R. Aaberge (2000): "Birds of a feather flock together". The Impact of Choice of Spouse on Family Labor Income Inequality

277 I. Aslaksen and K.A. Brekke (2000): Valuation of Social Capital and Environmental Externalities

278 H. Dale-Olsen and D. Rønningen (2000): The Importance of Definitions of Data and Observation Frequencies for Job and Worker Flows - Norwegian Experiences 1996-1997

279 K. Nyborg and M. Rege (2000): The Evolution of Considerate Smoking Behavior

280 M. Søberg (2000): Imperfect competition, sequential auctions, and emissions trading: An experimental evaluation

281 L. Lindholt (2000): On Natural Resource Rent and the Wealth of a Nation. A Study Based on National Accounts in Norway 1930-95

282 M. Rege (2000): Networking Strategy: Cooperate Today in Order to Meet a Cooperator Tomorrow

283 P. Boug, Å. Cappelen and A.R. Swensen (2000): Expectations in Export Price Formation: Tests using Cointegrated VAR Models

284 E. Fjærli and R. Aaberge (2000): Tax Reforms, Dividend Policy and Trends in Income Inequality: Empirical Evidence based on Norwegian Data

285 L.-C. Zhang (2000): On dispersion preserving estimation of the mean of a binary variable from small areas

286 F.R. Aune, T. Bye and T.A. Johnsen (2000): Gas power generation in Norway: Good or bad for the climate? Revised version

287 A. Benedictow (2000): An Econometric Analysis of Exports of Metals: Product Differentiation and Limited Output Capacity

288 A. Langørgen (2000): Revealed Standards for Distributing Public Home-Care on Clients

289 T. Skjerpen and A.R. Swensen (2000): Testing for long-run homogeneity in the Linear Almost Ideal Demand System. An application on Norwegian quarterly data for non-durables

290 K.A. Brekke, S. Kverndokk and K. Nyborg (2000): An Economic Model of Moral Motivation

291 A. Raknerud and R. Golombek: Exit Dynamics with Rational Expectations

292 E. Biørn, K-G. Lindquist and T. Skjerpen (2000): Heterogeneity in Returns to Scale: A Random Coefficient Analysis with Unbalanced Panel Data

293 K-G. Lindquist and T. Skjerpen (2000): Explaining the change in skill structure of labour demand in Norwegian manufacturing

294 K. R. Wangen and E. Biørn (2001): Individual Hetero-geneity and Price Responses in Tobacco Consumption: A Two-Commodity Analysis of Unbalanced Panel Data

295 A. Raknerud (2000): A State Space Approach for Estimating VAR Models for Panel Data with Latent Dynamic Components