Missing Observations in the Dynamic Regression Model

F.C. Palm
Th.E. Nijman

Researchmemorandum 1982-18 juni '82

Missing Observations in the Dynamic Regression Model*

F.C. Palm and Th.E. Nijman

Abstract

In this paper we consider the dynamic regression model with lagged endogenous variables and moving average disturbances, when some observations on the endogenous variable are missing. The available data are assumed to be sampled at regular intervals of length m and can be linear combinations of the realizations of the variable over a finite number of periods.

We discuss the identification of the parameters in the regression model and present several ways to obtain Maximum Likelihood (ML) estimates.

For some selected models, we evaluate the large sample variances of the ML estimates for the incomplete data and compare them with the asymptotic variances for the ML estimates when no data are missing. In this way, we get an indication of the loss of efficiency due to missing observations and of the precision of the ML estimator when the data are incomplete.

Finally, we give some results for the effects on the properties of the OLS estimator, when interpolated series are substituted for the missing observations.

* Paper to be presented at the Econometric Society European Meeting in Dublin, September 5-9, 1982.

Missing observations in the dynamic regression model

F.C. Palm and Th.E. Nijman *

April 1982
Preliminary version
Comments welcome

1. Introduction

In econometric analysis of time series, it is usually assumed that the relevant data consist of observations on the variables in the model pertaining to T subsequent time periods that are considered appropriate on a priori grounds. Attention has been drawn in the literature to the consequences of loosening these assumptions which will often not be met in applied work.

One stream of contributions is concerned with the problem of missing observations (see e.g. Dunsmuir (1981) for a survey and the references cited therein). The problems of temporal aggregation in dynamic models form another related research topic that has received increasing attention in recent years (see e.g. Sims (1971), Zellner and Montmarquette (1971), Tiao and Wei (1976), Geweke (1978) among many others).

In this paper we concentrate on the dynamic regression model with moving average disturbances when the endogenous variable is observed every m'th period, as is usually the case for stock variables, or when only a linear aggregate for the m periods, such as a flow variable measured over the m periods, is observed. Formally, we assume that the endogenous variable y_t is generated by the following regression model

$$y_t = \sum_{i=1}^{p} \rho_i y_{t-i} + \sum_{k=1}^{K} \beta_k x_{kt} + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} , \qquad (1)$$

where the ε_t's are independent normal variates with mean zero and unit varance and the x_{kt}'s are strictly exogenous variables,

* Address: Economische Faculteit, Vrije Universiteit,
Postbus 7161, 1007 MC Amsterdam, The Netherlands

i.e. x_{kt} is independent of $\varepsilon_{t'}$ for all t, t' and k.
We assume that the standard conditions for identification of the parameters in (1) are satisfied. In particular, to remove an indeterminacy, we assume that $\theta_0 > 0$. Furthermore, we assume that all x_{kt}'s are known (k=1,...,K, t=1,...,T), but that only linear combinations of the y_t's defined by

$$\tilde{y}_t = \sum_{i=0}^{A} w_i \, y_{t-i} \, , \text{ for } \quad t = m, 2m, \ldots, T, \tag{2}$$

are observed. The weights are deterministic and known. The problem of missing observations has been analyzed along two different lines. A simple formal way consists in deriving for instance quarterly data from yearly observations on the same series by minimizing some criterion function (see e.g. Boot, Feibes and Lisman (1967)). A second approach consists in specifying a model in which the missing observations are explained by other variables. The parameters of the model can be estimated, provided they are identified, and the model can be used to generate "predictions" of the missing observations. The problem of missing endogenous variables has been studied by Sargan and Drettakis (1974) for the autoregressive simultaneous equation model, by Zellner (1966), Telser (1967), Jones(1980), Harvey and Pereira (1980) and by Shaman and Tan (1981) for univariate time series models, by Harvey and Pereira(1980) for the static and dynamic regression model. The analysis of dynamic models when data are missing has also received attention in the time series literature (see e.g. Dunsmuir (1981)).

The static regression model with first order autoregressive errors and missing endogenous and exogenous variables for the same period has been analyzed by Wansbeek and Kapteyn (1981) and Kmenta (1981). When exogenous variables are missing, one usually extends the model by introducing an equation that relates the unobserved exogenous variables to other explanatory variables. For the static regression model with missing exogenous variables, the reader is referred to Anderson (1957), Dagenais (1973), Gourieroux and Montfort (1981), Hsiao (1979), Kmenta (1981) and Palm and Nijman (1981) among others. The dynamic regression model with missing observations that we analyze in this paper is not only of interest in itself. It can also be interpreted as a transfer function equation derived from a linear dynamic simultaneous equation model.

The plan of the paper is as follows. In section 2, we briefly introduce some patterns for the data transformation (2) that are particularly relevant in economic applications. In section 3, we discuss the identification of the parameters in the model. Section 4 is devoted to the computation of maximum likelihood (ML) estimates of the parameters. In section 5, we compare the asymptotic efficiency of the ML estimates for the model when data are missing with those of the parameters when the observations are available. An analysis of the efficiency of other consistent estimators compared with the ML estimator for the dynamic regression model with missing endogenous variables is currently under way. The accuracy of alternative consistent estimators for the static regression model with missing exogenous variables has been compared in Palm and Nijman (1981). In section 6, we present some results on the effect on the parameter estimates in large samples of using interpolated data as proxies for the missing endogenous variables.
Finally, in section 7 some concluding remarks are presented.

2. Relevant data transformations

In the previous section we assumed that the observations are in the form of the linear transformation (2) of the unknown data. Dropping the assumption of linearity would admittedly introduce new problems. Transformations of type (2) are somewhat restrictive however; several authors have discussed transformation patterns that do not fit into (2) (e.g. Dunsmuir and Robinson (1981), Harvey and Pereira (1980) and Tan (1979)). For economic time series, the most important cases that do not fit into (2) are perhaps the randomly missing observations and the (α, β)-sampling with $\alpha \neq 1$. In case of randomly missing observations the availability of an observation on y_t is determined by a probabilistic mechanism that is independent of the probability law according to which y_t is generated (for an interesting economic application, see e.g. Harvey et al. (1981)). The term (α, β)-sampling refers to a procedure in which the process under consideration is periodically observed for α consecutive periods and not observed for the next β consecutive periods. For the scheme in (2), $\alpha = 1$. Although the scheme (2) is restrictive, it is relevant in many economic applications. Moreover, some of the implications of the $(1, \beta)$-sampling remain valid for more general sampling schemes.

For a stock variable y_t, observations will often be available every m'th period. If the data are generated by a quarterly model and observed on an annual basis, $m = 4$ and the coefficients in (2) are

$$A = 0, \; w_0 = 1 \; . \qquad (3)$$

This set of coefficient values will be referred to as the skipped data pattern. If y_t is a flow variable, the total flow for m periods is ususally observed, so that we have

$$A = m-1 \; , \; w_i = 1, \; i = 0,1,2,\dots,m-1 \; . \qquad (4)$$

The scheme (2) is valid in other cases as well. Assume that the model (1) is formulated in first differences, that is $y_t = \Delta z_t = z_t - z_{t-1}$. If z_t is a stock variable observed every m'th period (t=m, 2m,..., T), then $z_t - z_{t-m} = \sum_{i=0}^{m-1} y_{t-i} = \tilde{y}_t$,

for $t = 2m, 3m, \dots, T$, can be obtained. In this case, (4) applies as well. If z_t is a flow variable for which every m'th sum of the last m realizations is observed, we have information on

$$\tilde{y}_t = \sum_{i=0}^{1} z_{t-i} - \sum_{i=0}^{1} z_{t-2-i} = y_t + 2y_{t-1} + y_{t-2} \, , \quad t = 4,6,\dots,T,$$

for $m = 2$, and on

$$\sum_{i=0}^{2} z_{t-i} - \sum_{i=0}^{2} z_{t-3-i} = y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4} \, , \quad t = 6,9,\dots, T,$$

for $m = 3$. These transformations of y_t, with y_t being generated by a static regression model, have been analyzed by Zellner and Montmarquette (1971). Similar patterns arise when the model (1) is formulated in second differences, $y_t = \Delta^2 z_t = z_t - 2z_{t-1} + z_{t-2}$, while we observe skipped data, z_t, $t = 2,4,\dots,T$ (assuming m=2), so that $\tilde{y}_t = z_t - 2z_{t-2} + z_{t-4} = y_t + 2y_{t-1} + y_{t-2}$, $t = 6,8,\dots,T$, can be computed.

More general weighting schemes can be obtained in a straightforward manner.

If $y_t = \Delta^k z_t$ and observations on z_t, $t=m,2m,\ldots,T$, or on $\sum_{i=0}^{m-1} z_{t-i}$, $t=m,2m,\ldots,T$, are available, the transformation will always be of type (2).

3. The identification of the model

To illustrate the nature of the identification problem in dynamic models when observations are missing, we consider a first order autoregressive - second order moving average (ARMA (1,2)) model, which is a special case of (1),

$$y_t = \rho\, y_{t-1} + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad , \tag{5}$$

with ε_t satisfying the assumptions made for (1).

Define $C_\ell = E y_t y_{t+\ell}$.

As the variable y_t is normally distributed, its distribution is determined once the variance and the autocovariances are given. The parameters of model (5) are identified if the system

$$\begin{aligned} C_0 &= \rho\, C_1 + \theta_0^2 + \theta_1^2 + \theta_2^2 + \rho\, \theta_0\theta_1 + \rho\, \theta_1\theta_2 + \rho^2\theta_0\theta_2 \\ C_1 &= \rho\, C_0 + \theta_0\theta_1 + \theta_1\theta_2 + \rho\, \theta_0\theta_2 \\ C_2 &= \rho\, C_1 + \theta_0\theta_2 \\ C_k &= \rho\, C_{k-1} \, , \quad k \geq 3 \, , \end{aligned} \tag{6}$$

can be solved uniquely for $(\rho, \theta_0, \theta_1, \theta_2)$ as a function of the second moments C_ℓ , which is the case provided $|\rho| < 1$, the roots of the moving average polynomial lie on or outside the unit circle and are different from ρ .

If however y_t is observed every second period, only every second autocovariance of y_t can be estimated. The model is identified in this case if

$$C_0 = \rho^2 C_0 + \theta_0^2 + \theta_1^2 + \theta_2^2 + 2\rho\ \theta_0\theta_1 + 2\rho\ \theta_1\theta_2 + 2\rho^2\theta_0\theta_2$$

$$C_2 = \rho^2 C_0 + \rho\ \theta_0\theta_1 + \rho\ \theta_1\theta_2 + \theta_0\theta_2 + \rho^2\theta_0\theta_2 \qquad (7)$$

$$C_k = \rho^2 C_{k-2}\ ,\quad k > 2\ ,$$

can be solved for $(\rho\ ,\ \theta_0,\theta_1,\theta_2)$.

When θ_1 and θ_2 are known to be zero, i.e. y_t is generated by a first order autoregressive model, (7) can simply be solved for ρ^2 and θ_0^2 . Without additional a priori information, the AR (1) model is not identified, as no information on the sign of ρ is present. This finding is at variance with a conclusion by Telser (1967)[1]. Unless $\rho = 0$, the AR (1) model is locally identified, which was already implicitly shown by several authors who established the information matrix of ρ and θ_0 in this model (see Shaman and Tan (1981) and Wansbeek and Kapteyn (1981)). Another interesting conclusion that can be derived from (7) is that pure moving average (MA) models of order 1 or 2 as well as the ARMA (1,2) model are not identified. The order condition that the MA coefficients have to appear in at least $q+1$ equations of (7) for the model to be identified is not satisfied. These examples are special cases of a more general result that will be proved at the end of this section: when only skipped data are observed (no matter how many periods lie between two successive observations), an ARMA (p,q) model is not identified if $q > p$.

That an ARMA (1,1) model is locally identified if $\rho \neq 0$ can be shown by evaluating the Jacobian of the transformation of the equations for $(C_0,\ C_2,\ C_4)$ in (7) to $(\rho,\ \theta_0,\ \theta_1)$, which is $4\ \rho^2 C_2(\theta_0^2-\theta_1^2)/(1-\rho$

1) For the AR (1) model with observations for every second period, Telser states that in $\eta_t = y_t - \rho^2 y_{t-2} = \varepsilon_t + \rho\varepsilon_{t-1}$ "all powers of the roots are present" (p. 493). This is not true however for the variance of η_t as Telser implicity claims below his equation (31).

The presence of an autoregressive parameter in this example helps to identify the moving average parameter. Notice however that the ARMA (1,1) model is not globally identified, because if $(\bar{\rho}, \bar{\theta}_0, \bar{\theta}_1)$ satisfies (7), so does $(-\bar{\rho}, \bar{\theta}_0, -\bar{\theta}_1)$.

Now we assume that we observe every second period the aggregate over two periods. From (5) we derive the relationship for the temporally aggregated data

$$\bar{y}_t = \rho\bar{y}_{t-1} + \theta_0\bar{\varepsilon}_t + \theta_1\bar{\varepsilon}_{t-1} + \theta_2\bar{\varepsilon}_{t-2} , \qquad (5')$$

where $\bar{y}_t = y_t + y_{t-1}$ and $\bar{\varepsilon}_t = \varepsilon_t + \varepsilon_{t-1}$.

Define $\bar{C}_\ell = E\,\bar{y}_t\bar{y}_{t+\ell}$. For this model, the analogous expressions to (6) and (7) are

$$\bar{C}_0 = \rho\bar{C}_1 + (2+\rho)[\theta_0^2 + \theta_1^2 + \theta_2^2] + (2+2\rho+\rho^2)\ (\theta_0\theta_1+\theta_1\theta_2) + \rho(1+\rho)^2\ \theta_0\theta_2$$

$$\bar{C}_1 = \rho\bar{C}_0 + \theta_0^2 + \theta_1^2 + \theta_2^2 + (2+\rho)\ (\theta_0\theta_1+\theta_1\theta_2) + (1+\rho)^2\ \theta_0\theta_2$$

$$\bar{C}_2 = \rho\bar{C}_1 + \theta_0\theta_1 + \theta_1\theta_2 + (2+\rho)\ \theta_0\theta_2$$

$$\bar{C}_3 = \rho\bar{C}_2 + \theta_0\theta_2$$

$$\bar{C}_k = \rho\bar{C}_{k-1} \qquad k > 3 \qquad (6')$$

and

$$\bar{C}_0 = \rho^2\bar{C}_0 + 2(1+\rho)(\theta_0^2+\theta_1^2+\theta_2^2) + (2+4\rho+2\rho^2)(\theta_0\theta_1+\theta_1\theta_2) + 2\rho(1+\rho)^2\theta_0\theta_2$$

$$\bar{C}_2 = \rho^2\bar{C}_0 + \rho(\theta_0^2+\theta_1^2+\theta_2^2) + (1+\rho)^2(\theta_0\theta_1+\theta_1\theta_2) + (2+2\rho+2\rho^2+\rho^3)\ \theta_0\theta_2$$

$$\bar{C}_4 = \rho^2\bar{C}_2 + \rho\ \theta_0\theta_2$$

$$\bar{C}_{2k} = \rho^2\bar{C}_{2k-2} \qquad k > 2 \quad . \qquad (7')$$

It is immediately clear that an MA (2) model is not identified in this case either, as there are only two equations in which θ_0, θ_1, and θ_2 appear, when $\rho = 0$. The presence of the term $(1+\rho)^2\ \theta_0\theta_1$ in the second equation of (7') and of $\rho\ \theta_0\theta_2$ in the third suggest that an MA (1) model and an ARMA (1,2) model might be identified in this case. Notice also that observing sums instead of single realizations helps in identifying the parameters. The sign of ρ is determined here.

To illustrate the impact of the presence of exogenous variables for the parameter identification, we consider the model

$$y_t = \rho y_{t-1} + \beta x_{1t} + \gamma x_{2t} + \theta_0 \varepsilon_t \tag{8}$$

with ε_t being a normally distributed white noise, and we assume as in (7) that skipped data are available every second period. Substitution of the model for y_{t-1} yields a model for the observed variables

$$y_t = \rho^2 y_{t-2} + \beta x_{1t} + \beta\rho x_{1t-1} + \gamma x_{2t} + \gamma\rho x_{2t-1} + u_t \tag{9}$$

with u_t being a normally distributed white noise with mean zero and variance $\theta_0^2 + \rho^2\theta_0^2$, for $t=1,3,\ldots,T$. It should be noted that if $\beta \neq 0$ or $\gamma \neq 0$, the regression coefficients in (9) are odd functions of ρ , so that there is usually information available on the sign of ρ . The parameters ρ and γ will obviously not be identified if there is multicollinearity present in the form of $(x_{1t}+\rho x_{1t-1}) = \lambda(x_{2t}+\rho x_{2t-1})$ for $t=1,3,\ldots,T$.

Finally we discuss the identification problem for the general model (1). For this purpose, we use a transformation that has been introduced by Amemiya and Wu (1972). We write equation (1) as

$$\rho(L)y_t = \sum_{k=1}^{K} \beta_k x_{kt} + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} \quad , \tag{10}$$

where L is the lag operator and $\rho(L) = 1 - \sum_{i=1}^{p} \rho_i L^i$. Defining

$\tilde{x}_{jt} = \sum_{i=0}^{A} w_i x_{jt-i}$ and $\tilde{\varepsilon}_t = \sum_{i=0}^{A} w_i \varepsilon_{t-i}$, we have

$$\rho(L)\,\tilde{y}_t = \sum_{k=1}^{K} \beta_k \tilde{x}_{kt} + \sum_{i=0}^{q} \theta_i \tilde{\varepsilon}_{t-i} \tag{11}$$

Now let $\alpha_1,\alpha_2,\ldots,\alpha_p$ be the (possibly complex) roots of the polynomial equation $\rho(L^{-1}) = 0$. Multiplying (11) by $\prod_{i=1}^{p}(1-\alpha_i L)^{-1}\,(1-\alpha_i^m L^m)$, we get

$$\prod_{i=1}^{p}(1-\alpha_i^m L^m)\,\tilde{y}_t = \sum_{k=1}^{K}\prod_{i=1}^{p}\left(\sum_{\ell=0}^{m-1}\alpha_i^\ell L^\ell\right)\beta_k \tilde{x}_{kt} + \sum_{j=0}^{q}\prod_{i=1}^{p}\left(\sum_{\ell=0}^{m-1}\alpha_i^\ell L^\ell\right)\theta_j \tilde{\varepsilon}_{t-j} \quad . \tag{12}$$

As m is the time lag between subsequent observations on the endogenous variable and because all data on the exogenous variables are assumed to be available, (12) is an expression in observed variables. Introducing a new parametrization for notational convenience and assuming that x_{1t} is the constant term, we can write equation (12) as

$$\tilde{y}_t = \sum_{i=1}^{p}\Psi_i \tilde{y}_{t-im} + \delta_1 + \sum_{k=2}^{K}\sum_{\ell=0}^{p(m-1)}\delta_{k\ell}\tilde{x}_{kt-\ell} + u_t \tag{13}$$

where u_t is a MA disturbance term

$$u_t = \sum_{\ell=0}^{p(m-1)+q+A}\eta_\ell L^\ell \varepsilon_t$$

with η_ℓ being defined by

$$\sum_{\ell=0}^{p(m-1)+q+A}\eta_\ell L^\ell = \sum_{r=0}^{A}\sum_{j=0}^{q}\sum_{i=1}^{p}\left(\sum_{\ell=0}^{m-1}\alpha_i^\ell L^\ell\right)\theta_j L^j w_r L^r$$

and $$E\,u_t u_{t-im} = \xi_i = \sum_{j=im}^{p(m-1)+q+A}\eta_j \eta_{j-im} \quad .$$

The definition of Ψ_i and $\delta_{k\ell}$ should be clear from (12). Notice also that the parameters in (13) are all real because the α_i's will be in conjugate pairs if they are complex. Equation (13) states that the observations are generated by a dynamic regression model as well. This model is referred to as the transformed difference equation. The order of the MA disturbance in (13) denoted by mq^* satisfies $mq^* \leq (m-1)p+q+A$. It should be noted that when for one exogenous variable only skipped data are available while all other variables are observed at each time period, a transformation similar to that used above can be applied in order to obtain a dynamic regression equation in which only observed variables appear.

The conditional density function for the observed endogenous variables given the exogenous variables can be written in terms of the parameters (Ψ,δ,ξ) which are functions $(\Psi,\delta,\xi) = f(\rho,\beta,\theta)$ of the basic parameters. A sufficient condition for (global) identification of (ρ,β,θ) on a subset P of the parameter space is therefore that the parameters (Ψ,δ,ξ) are identified in $f(P)$ without the use of the restrictions on (Ψ,δ,ξ) implied by $f(\rho,\beta,\theta)$ and that the equations $(\Psi,\delta,\xi) = f(\rho,\beta,\theta)$ have a unique solution $(\rho_0,\beta_0,\theta_0)$ in P for every (Ψ,δ,ξ) in $f(P)$. A necessary condition is that $(\Psi,\delta,\xi) = f(\rho,\beta,\theta)$ has a unique solution $(\rho_0,\beta_0,\theta_0)$ for every (Ψ,δ,ξ) in $f(P)$.

Corollary: The MA parameters θ_j in model (1) are not identified if $q > p+(q-p+A)/m$, that is if $q > p + A/(m-1)$.

Proof: The $q+1$ θ_j's appear in $f(\rho,\beta,\theta)$ only through the q^*+1 ξ_i's. The necessary conditions can therefore not be met if $q > p+(q-p+A)/m$.

The corollary implies that the dynamic regression model (1) is not identified when only skipped data are available (transformation pattern (2) and A=0) if $q > p$. Similarly, it is not identified when only aggregates are available (transformation pattern (3) and $A=m-1$) if $q > p+1$. Of course some parameters in model (1) can be identified, although other parameters are not. Writing

$1 + \sum_{i=1}^{p} \Psi_i L^{im} = \prod_{i=1}^{p}(1-\tilde{\Psi}_i L^m)$ we can see that the ρ_i's are identified if there is sufficient a priori information to determine them uniquely from $\alpha_i^m = \tilde{\Psi}_i$.

If $K > 1$ and $q = 0$, this a priori information is not always needed as the roots of $\alpha_i^m = \tilde{\Psi}_i$ can partly be determined from the $\delta_{k\ell}$.

If $K = 0$ and $q = 0$, this a priori information may be or may not be indispensible as has been shown in the examples at the beginning of this section.

The identification of β is straightforward if (13) is identified. The identification of θ can be checked by computing the Jacobian of the transformation $\xi = \xi(\theta)$. Finally, one should notice that models with a constant term and a seasonal dummy over m periods or with two dummies cannot be identified even if the necessary condition is met, because $\prod_{i=1}^{p}(\sum_{\ell=0}^{m-1} \alpha_i^\ell L^\ell)\, \tilde{x}_{k_1 t} = \lambda \prod_{i=1}^{p}(\sum_{\ell=0}^{m-1} \alpha_i^\ell L^\ell)\, \tilde{x}_{k_2 t}$

with $k_1 \neq k_2$ in that case, that is β_{k_1} and β_{k_2} enter in (12) only as $(\beta_{k_1} + \lambda\beta_{k_2})$, hence they cannot be identified.

4. Efficient estimation of the model

One way of estimating dynamic regression models with missing data has been to use interpolated values as proxies for the missing observations and to apply standard estimation methods to the constructed data. In section 6, we shall show some results on the effects for the properties of the estimates obtained in this way. At present, we consider consistent and efficient estimation of the parameters in model (1). Simple methods to compute consistent estimates have not received much attention in the literature. Tan (1979) and Dunsmuir and Robinson (1981) are exceptions, but their methods of moment procedures cannot be used in our case: the sample moments of the process y_t cannot be estimated directly for the lags that are not a multiple of m. Dunsmuir and Robinson's frequency domain estimation method suffers from the same problem. One can estimate (Ψ,δ,ξ) in (13) by instrumental variables or by nonlinear least squares neglecting the restrictions implied by f. Then one can determine an estimate of (ρ,β,θ) that fits best with the estimate of (Ψ,δ,ξ) in some sense. The computations involved will usually not be very expensive and the estimates will be consistent if (13) is identified.

Maximum likelihood estimation of a dynamic simultaneous equation model (for the case q=0) has been discussed by Sargan and Drettakis (1974) and by Harvey and Pereira (1980) for dynamic models with MA disturbances. Sargan and Drettakis parametrize the density function of the observed variables in such a way that ML estimates of the parameters are obtained by maximizing the density with respect to the parameters and the unobserved variables. Their procedure implies of course a large number of "parameters to be estimated". Harvey and Pereira use a Kalman filter approach and write the log-likelihood function in terms of the prediction error decomposition (see also Harvey (1981)):

$$L(\tilde{y}) = -T/2m \log 2\pi - \tfrac{1}{2}\sum_{t=1}^{T/m} \log f_{tm} - \tfrac{1}{2}\sum_{t=1}^{T/m} v_{tm}^2 / f_{tm} \quad , \quad (14)$$

where $v_{tm} = \tilde{y}_{tm} - \hat{\tilde{y}}_{tm|tm-1}$, $f_{tm} = E(v_{tm}^2)$ and $\hat{\tilde{y}}_{tm|tm-1}$ is the minimum mean square estimator of $\tilde{y}_{tm}$ conditional on all endogenous variables up to $tm-m$ ($\tilde{y}_i$; $i=m,2m,\ldots,(t-1)m$) and on all exogenous variables up to tm ($\tilde{x}_{ki}$; $k=1,\ldots,K$; $i=1,\ldots,tm$). The prediction errors v_{tm} and their variances can be computed by using the Kalman filter provided the model is cast in state space form. The relevant formulae are given in e.g. Harvey (1981).
For a static regression model with first order autoregressive errors, and missing endogenous and exogenous variables, Wansbeek and Kapteyn (1981) obtain analytic expressions for the first order derivatives of the likelihood function and for the information matrix. The expressions which they derive are very useful for ML estimation of the parameters.

As the transformed difference equation (13) is a dynamic regression model that is usually subject to nonlinear restrictions, ML estimates can be obtained by computing first and second order derivatives of the log-likelihood function with respect to φ , with $\varphi = (\rho,\beta,\theta)$ and iterating the Newton-Raphson linearization (or approximations of it) of the first order conditions for a maximum

$$\hat{\hat{\varphi}} = \hat{\varphi} - P^{-1}(\hat{\varphi}) \left.\frac{\partial L}{\partial \varphi}\right|_{\varphi=\hat{\varphi}} , \tag{15}$$

where $P(\hat{\varphi})$ is a nonsingular matrix satisfying the condition

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} P(\hat{\varphi}) = \underset{=}{E} \frac{1}{T} \frac{\partial^2 L}{\partial\varphi\partial\varphi'}$$ and $\hat{\varphi}$ is a consistent

initial estimate of φ which has an appropriate limiting distribution. Iterating expression (15) until convergence yields the ML estimator, but asymptotic first order efficiency is already obtained at the second step of iteration. A matrix $P(\hat{\varphi})$ that satisfies the condition given above and the vector $\partial L/\partial\varphi$ can be computed in several ways, three of which will be briefly discussed here. First, setting $\nu = (\Psi,\delta,\xi)$ and using the result that

$$\underset{=}{E} \left[\frac{\partial L}{\partial \nu}\frac{\partial L}{\partial \nu'}\right] = -\underset{=}{E}\left[\frac{\partial^2 L}{\partial\nu\partial\nu'}\right]$$, the Hessian matrix for the restricted

model can be obtained by pre- and postmultiplication by $\partial\nu/\partial\varphi$ of the Hessian matrix for the unrestricted model

$$\underset{=}{E}\left[\frac{\partial L}{\partial\varphi}\frac{\partial L}{\partial\varphi'}\right] = \underset{=}{E}\left[\frac{\partial\nu'}{\partial\varphi}\frac{\partial L}{\partial\nu}\frac{\partial L}{\partial\nu'}\frac{\partial\nu}{\partial\varphi'}\right] = -\frac{\partial\nu'}{\partial\varphi}\underset{=}{E}\left[\frac{\partial^2 L}{\partial\nu\partial\nu'}\right]\frac{\partial\nu}{\partial\varphi'} =$$

$$-\underset{=}{E}\frac{\partial^2 L}{\partial\varphi\partial\varphi'} , \tag{16}$$

which also indicates that ML estimation is efficient in this case.

A second way consists in differentiating the log-likelihood function in prediction error decomposition form given in (14) and the equations of the Kalman filter and then substituting the observations. Finally, a third approach generalizes an idea of Tan (1979) which is also discussed in Shaman and Tan (1981). They use the missing information principle of Orchard and Woodbury (1972) which says that if $x' = (s',m')$ and $L(x)$ depends on a vector of parameters φ, the following result holds true

$$\frac{\partial L(s)}{\partial \varphi} = E_m \frac{\partial L(x|s)}{\partial \varphi} \quad . \tag{17}$$

Using (17), Tan establishes expressions for $\frac{\partial L(\hat{y})}{\partial \varphi}$ and $E\left[-\frac{\partial^2 L(\hat{y})}{\partial \varphi \partial \varphi'}\right]$, for the model with $A = K = q = 0$. Results for the case where A, K and q are non-zero have been obtained and will be given in a separate paper.

As the alternative computational procedures have the same asymptotic properties, the choice among them in applied work will depend on computational aspects. For instance, the Kalman filter approach yields the exact likelihood estimates and predictions of the missing observations. As it has to be computed for every set of parameter values, it may become computationally expensive, except in situations where there is much information available about the true parameter values. The use of the transformed equation (13) is straightforward provided the restrictions in f are not too complicated. Finally, little is known about the small sample properties of the alternative procedures, so that at present small sample consideration offer little guidance for the choice of one among the alternative asymptotically efficient procedures.

5. The loss of efficiency due to incomplete data

In this section, we shall compare the asymptotic efficiency of the ML estimator of the parameters in a dynamic regression model, when the complete sample is available, with that of the ML estimator, when some observations on the endogenous variable are missing. The relative efficiency is perhaps of limited value for practical situations, as an investigator usually does not have the choice between either using a complete sample or relying on incomplete data. However, the results that we shall present are interesting for at least two reasons.
First, they indicate for which kind of sampling and for which parameter values the loss of efficiency due to missing observations is important. This may be of interest to agencies whose task it is to collect data in that they can better appreciate which gain can be expected from a more detailed data collection.
Second, for empirical work it is important to emphasize that for some parameters the large sample precision of the ML estimator deteriorates dramatically as a result of incomplete data, whereas other parameters can be estimated fairly accurately in large but incomplete samples. Other consistent, but not fully efficient estimators will have a still larger asymptotic variance. Their relative efficiency compared with that of the ML estimator for the dynamic regression model with incomplete data is subject of further research. For the static regression model, when some data is missing, the asymptotic efficiency of alternative estimators has been investigated by Palm and Nijman (1981).

In the tables 1 to 3, we give the ratio of the asymptotic variance of the ML estimator for ρ, β, θ_0 and θ_1 with respect to their variance when skipped data, aggregates or aggregates of aggregates are observed respectively. This last term is used to designate the situation where the change of a variable z_t, $y_t = z_t - z_{t-1}$, is explained in the regression model (1), whereas one observes an aggregrate of z_t. In table 4, we give the variances of ML estimates for the complete data, so that the reader can obtain the variances of the ML estimates for the incomplete data if he wants to do so. For more details on the computation of the large sample covariance matrix, the reader is referred to appendices A and B.

The tables need a short explanation. The true values of the paramaters are $\rho \varepsilon \{-.8,0,.8\}$, $\beta = 1$, $\sigma_\varepsilon^2 = \theta_0^2 = 1$, $\theta_1 = .6$ for the case that results are given in the tables, otherwise β and/or θ_1 are zero. The values of the coefficient of determination R^2 which are close to those frequently observed in applied work with economic time series, have been used to determine the parameters of the process for x_t (see appendix C). When x_t is generated by a stationary process, R^2 is defined as the variance of $(1 - \rho L)^{-1} x_t$ divided by the variance of y_t for the complete sample. When x_t is nonstationary, the R^2 is evaluated at $T = 30$. For the exogenous variable, several alternative processes are considered :

$x_t = v_t$ is a white noise with mean zero and variance σ_v^2, denoted by WHI;

$x_t = .9x_{t-1} + v_t$, denoted by AR;

$x_t = c$, a constant denoted by CON;

$x_t = t$, a trend denoted by TRE;

$x_t = x_{t-1} + v_t$, a random walk denoted by RWA;

$x_t = x_{t-1} + v_t + \mu$, a random walk with drift denoted by RWD, with $\mu/\sigma_v = 1$.

The number of periods m is 2 and 4 respectively and is reported in the first column of tables 1 to 4. The variances in the tables have been computed as the diagonal elements of the inverse of the information matrix (see appendix A).

The reader should have a look at the tables. For ρ, θ_0 and θ_1, a large relative efficiency of the ML estimator for complete data is caused by the large variance of the ML estimator for incomplete data. For β, the relative efficiency of the former is sometimes important, although the variance for the estimator based on incomplete data seems to be reasonable. Notice also that for skipped data, a nonstationary process for x_t seems to imply a large relative efficiency of the ML estimator for ρ and β when data is incomplete. For aggregates, this happens when x_t is a constant. By NID we indicate that the parameter is not identified.

To conclude, the results in the tables 1 to 4 give an indication about the loss of precision in parameter estimates and about the order of magnitude of the variance of the ML estimator when observations are missing.

Table 1. Relative efficiency of the ML estimator for complete data compared with that for incomplete data (skipped data)

			rel. eff. of ρ_{ML}			rel. eff. of β_{ML}			rel. eff. of $\theta_{0_{ML}}$			rel. eff. of $\theta_{1_{ML}}$		
m	R^2	x_t	ρ: -.8	0	0.8	ρ: -.8	0	0.8	ρ: -.8	0	0.8	ρ: -.8	0	0.8
2			1.28	NID	1.28				2.22	NID	2.22			
2	0.70	WHI	1.21	2.86	1.21	2.07	2.00	2.07	2.06	2.00	2.06			
2	0.95	WHI	1.18	2.11	1.18	2.10	2.00	2.10	2.01	2.00	2.01			
2	0.70	AR	1.25	6.51	1.20	9.45	4.56	1.18	2.11	2.00	2.11			
2	0.95	AR	1.22	2.55	1.14	6.98	2.43	1.19	2.02	2.00	2.02			
2	0.70	CON	1.28	NID	1.28	25.04	NID	1.20	2.22	NID	2.22			
2	0.95	CON	1.28	NID	1.28	5.05	NID	1.27	2.22	NID	2.22			
4			2.20	NID	2.20				6.34	NID	6.34			
4	0.70	WHI	1.79	5.71	1.79	4.67	4.00	4.67	4.57	4.00	4.57			
4	0.95	WHI	1.69	4.21	1.68	4.85	4.00	4.85	4.09	4.00	4.09			
4	0.70	AR	1.83	13.02	1.73	17.01	9.12	1.62	5.01	4.00	4.96			
4	0.95	AR	1.60	5.11	1.46	12.56	4.85	1.57	4.20	4.00	4.18			
4	0.70	CON	2.20	NID	2.20	26.81	NID	1.87	6.34	NID	6.34			
4	0.95	CON	2.20	NID	2.20	6.05	NID	2.16	6.34	NID	6.34			
2			2.83	NID	2.21				2.24	NID	278.32	4.27	NID	255.45
2	0.70	WHI	1.45	NID	1.53	1.74	NID	5.29	2.16	NID	190.78	3.06	NID	174.98
2	0.95	WHI	1.44	NID	1.27	1.74	NID	4.98	2.16	NID	157.38	3.11	NID	144.21
2	0.70	AR	1.47	NID	1.82	3.22	NID	1.59	2.16	NID	218.90	3.01	NID	201.32
2	0.95	AR	1.45	NID	1.47	3.22	NID	1.66	2.16	NID	166.36	3.10	NID	152.76
2	0.70	CON	2.83	NID	2.21	2.90	NID	1.62	2.24	NID	278.32	4.27	NID	255.45
2	0.95	CON	2.83	NID	2.21	2.45	NID	2.09	2.24	NID	278.32	4.27	NID	255.45
2	0.70	TRE	1.28	2051.78	1.26	1.28	2051.78	1.26	2.22	2.00	2.20			
2	0.95	TRE	1.32	86.62	1.16	1.32	86.62	1.16	2.22	2.00	2.13			
2	0.70	RWA	1.24	15.29	1.21	1.24	15.29	1.21	2.16	2.00	2.12			
2	0.95	RWA	1.16	3.63	1.14	1.16	3.63	1.14	2.05	2.00	2.03			
2	0.70	RWD	1.26	187.82	1.23	1.26	187.82	1.23	2.21	2.00	2.17			
2	0.95	RWD	1.32	24.94	1.12	1.32	24.94	1.12	2.16	2.00	2.06			

Table 2. Relative efficiency of the ML estimator for complete data compared with that for incomplete temporally aggregated data

			rel. eff. of ρ_{ML}			rel. eff. of β_{ML}			rel. eff. of $\theta_{0_{ML}}$			rel. eff. of $\theta_{1_{ML}}$		
m	R^2	x_t	ρ: −.8	0	0.8	ρ: −.8	0	0.8	ρ: −.8	0	0.8	ρ: −.8	0	0.8
2			36.56	8.00	1.25				3.69	6.00	2.25			
2	0.70	WHI	4.80	3.33	1.17	2.08	2.58	2.08	2.07	2.50	2.07			
2	0.95	WHI	3.66	2.76	1.15	2.08	2.66	2.11	2.01	2.07	2.01			
2	0.70	AR	6.33	4.23	1.17	3.58	3.13	1.14	2.15	3.47	2.12			
2	0.95	AR	3.74	2.45	1.10	2.32	2.34	1.13	2.02	2.27	2.03			
2	0.70	CON	36.56	8.00	1.25	21.98	5.90	1.17	3.69	6.00	2.25			
2	0.95	CON	36.56	8.00	1.25	29.33	7.65	1.24	3.69	6.00	2.25			
4			81.13	64.00	2.00				4.58	76.00	6.96			
4	0.70	WHI	6.65	12.31	1.68	4.02	8.95	4.87	4.01	8.15	4.74			
4	0.95	WHI	5.01	9.55	1.58	4.02	9.10	5.11	4.00	4.54	4.12			
4	0.70	AR	9.32	12.58	1.55	5.03	8.97	1.44	4.03	13.81	5.19			
4	0.95	AR	5.35	5.61	1.30	3.88	5.33	1.39	4.00	5.37	4.23			
4	0.70	CON	81.13	64.00	2.00	50.10	45.10	1.70	4.58	76.00	6.96			
4	0.95	CON	81.13	64.00	2.00	67.32	50.85	1.96	4.58	76.00	6.96			
2			853.59	NID	1.79				893.41	NID	255.85	116.13	NID	211.71
2	0.70	WHI	20.96	5.24	1.43	4.06	4.77	5.62	185.04	170.11	208.45	237.19	145.11	167.24
2	0.95	WHI	20.66	4.62	1.26	4.05	4.57	5.31	184.78	150.45	174.44	252.62	135.24	146.27
2	0.70	AR	23.20	8.37	1.62	5.03	6.39	1.52	185.45	245.49	219.91	219.43	198.95	183.30
2	0.95	AR	22.47	5.69	1.44	5.03	5.47	1.63	184.83	159.71	181.69	249.84	141.47	152.38
2	0.70	CON	853.59	NID	1.79	621.33	NID	1.43	893.41	NID	255.85	116.18	NID	211.71
2	0.95	CON	853.59	NID	1.79	705.34	NID	1.72	893.41	NID	255.85	116.18	NID	211.71

Table 3. Relative efficiency of the ML estimator for the complete data compared with that for incompletely observed aggregates of aggregates

			rel. eff. of ρ_{ML}			rel. eff. of β_{ML}			rel. eff. of $\theta_{1_{ML}}$		
m	R^2	x_t	ρ: -.8	0	0.8	ρ: -.8	0	0.8	ρ: -.8	0	0.8
2			3237.04	8.26	1.26				595.69	7.67	2.29
2	0.70	WHI	28.38	3.76	1.15	3.82	2.90	2.10	3.56	2.77	2.07
2	0.95	WHI	20.96	3.14	1.12	3.83	3.02	2.13	2.19	2.11	2.01
2	0.70	AR	21.48	4.28	1.15	10.64	3.16	1.12	4.05	4.04	2.13
2	0.95	AR	11.73	2.45	1.07	7.87	2.35	1.12	2.25	2.37	2.02
2	0.70	CON	3237.04	8.26	1.26	1919.33	6.08	1.15	595.69	7.67	2.29
2	0.95	CON	3237.04	8.26	1.26	2592.53	7.99	1.24	595.69	7.67	2.29

Table 4. Large sample variance of the ML estimator for the complete data

			variance of ρ_{ML}			variance of β_{ML}			variance of $\theta_{0_{ML}}$			variance of $\theta_{1_{ML}}$		
m	R^2	x_t	ρ: -.8	0	0.8	ρ: -.8	0	0.8	ρ: -.8	0	0.8	ρ: -.8	0	0.8
1			0.36	1.00	0.36				0.50	0.50	0.50			
1	0.70	WHI	0.11	0.30	0.11	0.43	0.43	0.43	0.50	0.50	0.50			
1	0.95	WHI	0.02	0.05	0.02	0.05	0.05	0.05	0.50	0.50	0.50			
1	0.70	AR	0.19	0.69	0.19	0.12	0.99	4.57	0.50	0.50	0.50			
1	0.95	AR	0.04	0.22	0.04	0.02	0.23	0.76	0.50	0.50	0.50			
1	0.70	CON	0.36	1.00	0.36	0.16	1.43	12.69	0.50	0.50	0.50			
1	0.95	CON	0.36	1.00	0.36	0.12	1.05	9.31	0.50	0.50	0.50			
1			2.42	2.78	0.40				0.50	0.50	0.50	4.48	1.96	0.90
1	0.70	WHI	0.04	0.37	0.18	0.33	0.41	0.30	0.50	0.50	0.50	0.88	0.97	0.85
1	0.95	WHI	0.01	0.05	0.04	0.04	0.05	0.04	0.50	0.50	0.50	0.83	0.84	0.83
1	0.70	AR	0.09	1.01	0.26	0.15	1.35	7.42	0.50	0.50	0.50	0.95	1.24	0.87
1	0.95	AR	0.01	0.18	0.07	0.02	0.19	1.21	0.50	0.50	0.50	0.84	0.89	0.83
1	0.70	CON	2.42	2.78	0.40	0.88	3.86	19.58	0.50	0.50	0.50	4.48	1.96	0.90
1	0.95	CON	2.42	2.78	0.40	0.78	2.91	11.07	0.50	0.50	0.50	4.48	1.96	0.90
1	0.70	TRE	0.36	1.00	0.34	0.11	1.00	8.41	0.50	0.50	0.50			
1	0.95	TRE	0.36	0.99	0.23	0.11	0.99	5.72	0.50	0.50	0.50			
1	0.70	RWA	0.27	0.87	0.21	0.08	0.87	5.30	0.50	0.50	0.50			
1	0.95	RWA	0.09	0.45	0.35	0.03	0.45	1.35	0.50	0.50	0.50			
1	0.70	RWD	0.34	0.99	0.29	0.11	0.99	7.19	0.50	0.50	0.50			
1	0.95	RWD	0.26	0.93	0.12	0.08	0.93	2.95	0.50	0.50	0.50			

6. The use of interpolated data

An approach that has been used quite often in empirical work to solve the problem of missing observations consists in first interpolating the missing values in such a way that the resulting series is plausible according to some criterion and is in agreement with the observed values of the series and then using the constructed series as realizations for the missing observations. Usually, this kind of procedure will yield inconsistent parameter estimates. However, it has the advantage of being straightforward to apply. In this section we shall present some results on the magnitude of the parameter inconsistency for some selected models. An interpolation method that has been applied on a large scale is that proposed by Boot, Feibes and Lisman (1967). Generalizing their method to other cases than the observation of aggregates, one obtains the interpolated series as the solution to the following optimization problem

$$\min_{\hat{y}_t} \sum_{t=1}^{T} [(1-L)^d \hat{y}_t]^2 \tag{18}$$

$$\text{subject to } \tilde{\hat{y}}_t = \tilde{y}_t \text{ , for } t=1, 1+m, \ldots, T,$$

and d being a priori given. Boot, Feibes and Lisman suggest using $d=1$ or $d=2$. The procedure reflects the fact that many economic time series are smooth and that the constructed series should have that property too. A more sophisticated smoothing method has been proposed and applied by Somermeyer et al. (1976). A review of various other methods of interpolation is given by Gelauff and Harkema (1977). Note that interpolation methods that use related series (see e.g. Chow and Lin (1971, 1976), and Ginsburgh (1973)) are not well suited to the problem at hand as the model for the missing data is dynamic.

For many interpolated methods, including that by Boot, Feibes and Lisman, the interpolated series can be written as a linear transformation $\hat{y} = R\,y$ of the realizations, where $\hat{y} = (\hat{y}_1, \hat{y}_2, \ldots \hat{y}_T)'$, $y = (y_1, y_2, \ldots y_T)'$ and R is a matrix of constants that does not depend on the observations $\tilde{y}_t$. Assume that $q = 0$ in equation (1) and that one estimates the parameters of equation (1) by ordinary least squares using the constructed series $\hat{y}_t$. Denote the resulting estimates of ρ and β by $\dot{\rho}$ and $\dot{\beta}$. Furthermore define R_i as the matrix obtained by deleting the first $(\rho-i)$ and the last i rows of R.

We can express the OLS estimator as

$$\begin{bmatrix} \hat{\rho} \\ \dot{\beta} \end{bmatrix} = \begin{bmatrix} y' R_1' R_1 y & \cdots & y' R_1' R_p y & y' R_1' X \\ \vdots & & \vdots & \\ y' R_p R_1 y & \cdots & y' R_p' R_p y & y' R_p' X \\ X' R_1 y & \cdots & X' R_p y & X' X \end{bmatrix}^{-1} \begin{bmatrix} \hat{y}' R_1' R_0 y \\ \vdots \\ y' R_p' R_0 y \\ X' R_0 y \end{bmatrix} . \quad (19)$$

If R is a block-Toeplitz matrix with a limited number of non-zero elements, that is

$$R = \begin{bmatrix} A_0 & A_1 & \cdots & A_p & 0 \cdots\cdots 0 \\ A_{-1} & A_0 & A_1 \cdots & A_p & 0 \cdots 0 \\ \cdots & & & & \\ 0 \cdots 0 & A_{-p} & \cdots & A_{-1} & A_0 \end{bmatrix} , \quad (20)$$

we can compute the probability limit of the OLS estimator defined in (19). Thereby, we use the property that products of matrices of the type defined in (20) are again block-Toeplitz matrices. Note that the matrix R for the procedure of Boot, Feibes and Lisman is not of the form (20) but can be very closely approximated by such a matrix, so that the probability limit of the OLS estimator in (19) can be obtained straightforwardly by replacing the cross-products in (19) by their second order moments. For instance, when $m=2$ and $\hat{y}_{2t} = \hat{y}_{2t-1} = \frac{1}{2} \tilde{y}_{2t} = \frac{1}{2} (y_{2t} + y_{2t-1})$, $t = 1, 2, \ldots$, we can write the first element of the matrix of cross-products in (19), i.e. the sum of squares of the interpolated series lagged one period, as

$$\begin{aligned} \sum_{t=2}^{T} \hat{y}_{t-1}^2 &= \sum_{t=1}^{T/2} (\hat{y}_{2t-1}^2 + \hat{y}_{2t}^2) - \hat{y}_T^2 \\ &= \frac{1}{2} \sum_{t=1}^{T/2} (y_{2t-1}^2 + 2y_{2t}\, y_{2t-1} + y_{2t}^2) - \hat{y}_T^2 . \end{aligned} \quad (21)$$

When y_t is nonstationary, the expectation of the first r.h.s. term of (21) can be obtained along the lines of appendix B. and that of the second r.h.s. term in (21) can readily be obtained as a function of T.

In tables 5 and 6 we report the probability limits of $\hat{\rho}$ and $\dot{\beta}$ for four selected models. For the first three models, we have assumed that aggregates over m periods are observed and that the investigator assigns a fraction $1/m$ of the observed value to each of the m periods. For the last model, the weights w_i have been obtained using the method proposed by Boot, Feibes and Lisman (1967) assuming $d=2$ and $m=4$. For the exogenous variable

x_t, we consider the processes that have been used in section 5. In particular we assume that x_t is a white noise, a first order autoregressive process, a constant term, a linear trend or a random walk respectively. These assumptions imply that y_t is generated by an AR(1) model with autoregressive coefficient ρ or by an ARMA (2,1) model with autoregressive coefficients equal to ρ and .9 and by an ARIMA (1,1,1)model. The R^2 takes the values .70 and .95 respectively, (see appendix C), whereas $\beta=1$, $\theta_0=1$ and $\theta_1=1$. For ρ we choose the values $\rho \varepsilon \{-.8, -.4, 0, .4, .8\}$. For the details we refer to the preceding section.

At this point we would like to emphasize that the method by Boot, Feibes and Lisman (1967) has not been designed for situations, where the endogenous variable has a negative autoregressive coefficient, and that therefore, we should not draw strong conclusions from the figures in table 6 for case 2 when $\rho = -.8$ or when $\rho = 0$ and x_t is a white noise or a constant. From tables 5 and 6, it is obvious that the probability limits differ substantially from the true values of ρ and β. When $\rho = .8$, the probability limit of $\dot{\rho}$ is reasonably close to the true value, except for some nonstationary models when the method by Boot et al. is used. For most of the models with nonpositive ρ, the OLS method largely overestimates ρ, whereas β is usually underestimated, except when x_t is nonstationary and $\rho = .8$. The figures slightly improve when R^2 increases. When x_t is generated by a trend or a random walk with drift, the probability limit of $\dot{\rho}$ hardly varies with the true value of ρ, when the missing data are interpolated using the method by Boot et al. As an overall conclusion, we cannot recommend using OLS on interpolated data, if the aim is to estimate the parameters of a dynamic regression model from incomplete observations. Whether this conclusion holds true for the predictions generated by a model that has been estimated from interpolated data using OLS is subject to further research.

Table 5. The probability limit of the OLS estimator when using interpolated data for the missing observations.

	m	R^2	x_t	$\rho = -.8$ $\hat{\rho}$	$\hat{\beta}$	$\rho = -.4$ $\hat{\rho}$	$\hat{\beta}$	$\rho = 0$ $\hat{\rho}$	$\hat{\beta}$	$\rho = .4$ $\hat{\rho}$	$\hat{\beta}$	$\rho = .8$ $\hat{\rho}$	$\hat{\beta}$
$w_i = \frac{1}{2}$	2	0.70	WHI	0.32	0.22	0.28	0.33	0.36	0.41	0.54	0.45	0.82	0.49
	2	0.95	WHI	0.26	0.23	0.21	0.35	0.30	0.43	0.50	0.47	0.81	0.53
	2	0.70	AR	0.17	0.44	0.25	0.52	0.37	0.61	0.56	0.72	0.83	0.85
	2	0.95	AR	0.02	0.52	0.07	0.64	0.23	0.74	0.48	0.84	0.81	0.92
	2	0.70	CON	0.46	0.30	0.44	0.49	0.50	0.50	0.64	0.60	0.86	0.69
	2	0.95	CON	0.46	0.30	0.44	0.40	0.50	0.50	0.54	0.60	0.86	0.59
	2	0.70	TRE	0.43	0.32	0.43	0.41	0.49	0.51	0.63	0.61	0.85	0.75
	2	0.95	TRE	0.24	0.42	0.35	0.46	0.44	0.56	0.59	0.68	0.81	0.96
	2	0.70	RWA	0.26	0.41	0.34	0.47	0.44	0.56	0.60	0.67	0.83	0.83
	2	0.95	RWA	0.04	0.53	0.13	0.62	0.23	0.72	0.51	0.82	0.81	0.95
	2	0.70	RWD	0.35	0.36	0.42	0.41	0.43	0.51	0.63	0.62	0.84	0.82
	2	0.95	RWD	0.00	0.55	0.31	0.49	0.42	0.57	0.57	0.71	0.78	1.18
$w_i = \frac{1}{3}$	3	0.70	WHI	0.45	0.17	0.53	0.17	0.57	0.21	0.56	0.26	0.85	0.32
	3	0.95	WHI	0.43	0.17	0.49	0.17	0.53	0.22	0.63	0.27	0.84	0.32
	3	0.70	AR	0.38	0.32	0.44	0.37	0.51	0.46	0.63	0.59	0.84	0.78
	3	0.95	AR	0.33	0.34	0.33	0.44	0.39	0.57	0.53	0.72	0.81	0.88
	3	0.70	CON	0.49	0.28	0.62	0.27	0.67	0.33	0.73	0.45	0.88	0.63
	3	0.95	CON	0.49	0.28	0.62	0.27	0.67	0.33	0.73	0.45	0.88	0.63
	3	0.70	TRE	0.45	0.31	0.59	0.29	0.55	0.35	0.72	0.47	0.86	0.69
	3	0.95	TRE	0.22	0.43	0.41	0.42	0.52	0.48	0.63	0.62	0.79	1.05
	3	0.70	RWA	0.40	0.33	0.51	0.35	0.58	0.42	0.57	0.54	0.84	0.73
	3	0.95	RWA	0.32	0.38	0.35	0.46	0.42	0.58	0.56	0.74	0.81	0.94
	3	0.70	RWD	0.39	0.34	0.58	0.30	0.64	0.36	0.71	0.48	0.84	0.31
	3	0.95	RWD	0.11	0.50	0.37	0.45	0.51	0.49	0.60	0.66	0.76	1.21

Table 6. The probability limit of the OLS estimator when using interpolated data for the missing observations.

	m	R^2	x_t	$\rho = -.8$		$\rho = -.4$		$\rho = 0$		$\rho = .4$		$\rho = .8$	
				$\dot{\rho}$	$\dot{\beta}$	$\dot{\rho}$	$\dot{\beta}$	$\dot{\rho}$	$\dot{\beta}$	$\dot{\rho}$	$\dot{\beta}$	$\dot{\rho}$	$\dot{\beta}$
$w_i = \frac{1}{4}$	4	0.70	WHI	0.68	0.06	0.67	0.09	0.69	0.12	0.73	0.17	0.87	0.23
	4	0.95	WHI	0.66	0.06	0.64	0.10	0.65	0.13	0.71	0.17	0.86	0.23
	4	0.70	AR	0.54	0.22	0.56	0.28	0.60	0.36	0.68	0.50	0.85	0.73
	4	0.95	AR	0.51	0.24	0.50	0.32	0.51	0.44	0.59	0.62	0.82	0.83
	4	0.70	CON	0.72	0.16	0.73	0.19	0.75	0.25	0.79	0.35	0.89	0.54
	4	0.95	CON	0.72	0.16	0.73	0.19	0.75	0.25	0.79	0.35	0.89	0.54
	4	0.70	TRE	0.56	0.25	0.66	0.24	0.71	0.29	0.76	0.40	0.86	0.57
	4	0.95	TRE	0.12	0.49	0.37	0.45	0.51	0.49	0.62	0.64	0.77	1.15
	4	0.70	RWA	0.55	0.25	0.60	0.28	0.65	0.35	0.72	0.47	0.85	0.75
	4	0.95	RWA	0.49	0.28	0.49	0.37	0.52	0.48	0.60	0.66	0.81	0.93
	4	0.70	RWD	0.42	0.32	0.64	0.25	0.71	0.29	0.75	0.41	0.83	0.33
	4	0.95	RWD	0.12	0.49	0.36	0.46	0.51	0.49	0.59	0.67	0.73	1.34
w_i : Boot et al.	4	0.70	WHI	0.86	0.02	0.86	0.03	0.86	0.06	0.87	0.11	0.93	0.20
	4	0.95	WHI	0.86	0.02	0.85	0.03	0.85	0.06	0.86	0.12	0.93	0.20
	4	0.70	AR	0.78	0.11	0.77	0.16	0.78	0.22	0.81	0.34	0.90	0.60
	4	0.95	AR	0.75	0.13	0.71	0.19	0.68	0.30	0.71	0.48	0.85	0.75
	4	0.70	CON	0.88	0.07	0.88	0.09	0.89	0.11	0.91	0.15	0.96	0.22
	4	0.95	CON	0.88	0.07	0.88	0.09	0.89	0.11	0.91	0.15	0.96	0.22
	4	0.70	TRE	-0.01	0.56	0.00	0.71	0.02	0.98	0.03	1.61	0.05	4.77
	4	0.95	TRE	-0.01	0.56	-0.01	0.72	-0.01	1.01	-0.00	1.68	-0.00	5.00
	4	0.70	RWA	0.65	0.19	0.72	0.28	0.76	0.24	0.81	0.32	0.87	0.65
	4	0.95	RWA	0.55	0.25	0.54	0.32	0.56	0.44	0.63	0.62	0.79	1.13
	4	0.70	RWD	-0.01	0.56	0.00	0.71	0.02	0.98	0.03	1.62	0.02	4.92
	4	0.95	RWD	-0.01	0.56	-0.01	0.72	-0.01	1.01	-0.01	1.68	-0.00	5.01

7. Some tentative conclusions

In this paper we have discussed the problems arising in a dynamic regression model when some realizations of the endogenous variable are not observed. After a presentation of the different schemes in which information on the endogenous variable may be available, we consider the identification of the parameters in the regression model with incomplete data and present several ways to obtain ML estimates of these parameters provided they are identified. Results on the loss of efficiency due to an incomplete data set are given for several models. Finally, we analyze the effects on the probability limit of the OLS estimator when the data is completed through interpolation. In the light of the results presented in section 5 and given the choice among several procedures to obtain ML estimates, we should like to advise the investigator at this time to use the ML method to estimate the parameters of the dynamic regression model with incompletely observed endogenous variable. Each of the procedures to obtain ML estimates has specific computational advantages. Nevertheless, as has been illustrated in section 4, the investigator should not be surprised that some of the coefficients in the model cannot be determined very accurately from the sample information.

Among the problems that remain to be analyzed, there are three questions that will receive more attention in the future. The first question concerns the possible loss of efficiency implied by the use of alternative consistent but not fully efficient estimation methods that are easier to implement than the ML method. Secondly, for empirical work it is important to know what the effects are on the forecasting performance of the dynamic regression model, when a specific method is used to solve the problem of incomplete observations. Third, the problem of missing exogenous variables in a dynamic regression model will be analyzed.

Appendix A

In this appendix, we briefly outline how the asymptotic covariance matrix for the ML estimator of the unrestricted parameters in the transformed equation (13) have been obtained. The large sample covariances have been used to compute the results in tables 1 to 4 in section 5.

We consider a first order dynamic regression model with one exogenous variable x_t and assume that observations are available at every m-th period.

$$\tilde{y}_t = \rho \, \tilde{y}_{t-1} + \tilde{x}_t' \beta + \tilde{\varepsilon}_t \qquad \text{(A.1a)}$$

or after substitution for $\tilde{y}_{t-1}$

$$\tilde{y}_t = \rho^m \, \tilde{y}_{t-m} + \sum_{i=0}^{m-1} \tilde{x}_{t-i} \, \beta\rho^i + \sum_{i=0}^{m-1} \tilde{\varepsilon}_t \, \rho^i \quad . \qquad \text{(A.1b)}$$

For a sample of T realizations, equation (A.1a) can be written as

$$\underset{(T\times T)}{A} \; \underset{(T\times 1)}{\tilde{y}} = \underset{(T\times T)}{C} \; \underset{(T\times 1)}{\tilde{x}} + \underset{(T\times 1)}{\tilde{\varepsilon}} \quad , \qquad \text{(A.2a)}$$

where $A = \begin{bmatrix} 1 & 0 & 0 & \cdots & & 0 \\ -\rho & 1 & 0 & \cdots & & \\ 0 & -\rho & 1 & 0 & \cdots & \\ \vdots & & & & & \\ 0 & \cdots & & & -\rho & 1 \end{bmatrix}$, $C = I_T \otimes \beta$ and $E \, \tilde{\varepsilon} \tilde{\varepsilon}' = \Omega$.

For the T/m observations, the model is

$$R \, \tilde{y} = R \, A^{-1} \, C \, \tilde{x} + R \, A^{-1} \, \tilde{\varepsilon}, \qquad \text{(A.2b)}$$

where R is a selection matrix

$$\underset{T/m \times T}{R} = \begin{bmatrix} 1 & \overbrace{0 \quad 0}^{m-1} & 0 & & \cdots & 0 \\ 0 & \cdots \quad 0 & 1 & 0 & \cdots & 0 \\ & & & & & \\ 0 & & & & & 1 \end{bmatrix}$$

Alternatively, equation (A.1b) can be written as

$$R\,\tilde{y} = R\,\tilde{y}_{-m}\,\psi + R\,C\,S\,\tilde{x} + v \quad , \tag{A.3}$$

$$\text{where } \underset{T\times T}{S} = \begin{bmatrix} 1 & 0 & \dots & & & 0 \\ \rho & 1 & 0 & \dots & & 0 \\ \rho^2 & \rho & 1 & 0 & \dots & \\ \vdots & & & & & \vdots \\ 0 & \dots & 0 & \rho^{m-1} & \dots\ \rho & 1 \end{bmatrix} , \quad \Psi = \rho^m \quad \text{and} \quad v = R \sum_{i=0}^{m-1} \rho^i\, \tilde{\varepsilon}_{-i}$$

with $E\,v\,v' = R\,\Sigma\,R' = (H)$, that is a Toeplitz-matrix as $\sum_{i=0}^{m-1} \rho^i\, \tilde{\varepsilon}_{-i}$ can be represented as a finite MA model. We parametrize (H) in terms of the covariances $\xi_i = E\,v_t\,v_{t-im}$.

Equation (A.3) can also be written as

$$R\,\tilde{y} = D^{-1}\,R\,C\,S\,\tilde{x} + D^{-1}\,v \tag{A.4}$$

$$\text{with } D = \begin{bmatrix} 1 & 0 & \dots & & 0 \\ -\rho^m & 1 & 0 & & \\ 0 & -\rho^m & 1 & & \\ \vdots & & \ddots & \ddots & \\ 0 & & & -\rho^m & 1 \end{bmatrix} .$$

When v is normally distributed, the Hessian-matrix of the log-likelihood function L can be computed as follows:

$$- E\,\frac{\partial^2 L}{\partial \psi^2} = E\,\tilde{y}'_{-m}\,R'\,(H)^{-1}\,R\,\tilde{y}_{-m} \tag{A.5}$$

and using (A.2b)

$$= \operatorname{tr}\,\{A'^{-1}\,R'\,(H)^{-1}\,R\,A^{-1}\,\Omega\} + E[\tilde{x}'_{-m}\,C'A'^{-1}\,R'\,(H)^{-1}\,R\,A\,C\,\tilde{x}_{-m}] \tag{A.6}$$

and using (A.4)

$$= \operatorname{tr}\,\{D'^{-1}\,(H)^{-1}\,D\,(H)\} + E[\tilde{x}'_{-m}\,S'\,C'\,R'\,D'^{-1}\,(H)^{-1}\,D^{-1}\,R\,C\,S\,\tilde{x}_{-m}] \, . \tag{A.7}$$

Combining (A.6) and (A.7) yields

$$- E \frac{\partial^2 L}{\partial \psi^2} = \mathrm{tr}\{D'^{-1} Ⓗ^{-1} D Ⓗ\} + E[\tilde{x} C' A'^{-1} L^{m'} R' Ⓗ^{-1} R L^m A C \tilde{x}] , \tag{A.8}$$

where the T×T backshift matrix L^m , defined as

$$L^m = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \\ 0 & \dots & 0 \\ & & 0\dots0 \\ I_{T-m} & & 0\dots0 \end{bmatrix} \} \, m$$

implying for instance that $\tilde{y}_{-m} = L^m \tilde{y}$, has been used.

Similarly, for $\delta_i = \beta \rho^i$,

$$- E \frac{\partial^2 L}{\partial \psi \, \partial \delta_i} = E[\tilde{y}'_{-m} R' Ⓗ^{-1} R \tilde{x}_{-i}] = E[\tilde{x}' C' A'^{-1} L^{m'} R Ⓗ R L^i \tilde{x}] \tag{A.9}$$

and

$$- E \frac{\partial^2 L}{\partial \delta_i \, \partial \delta_j} = E[\tilde{x}' L^{i'} R' Ⓗ^{-1} R L^j \tilde{x}] . \tag{A.10}$$

For the parameters in $Ⓗ = R' \Sigma R$, with $\Sigma = E(\sum_{i=0}^{m-1} \rho^i \tilde{\varepsilon}_{-i})(\sum_{i=0}^{m-1} \rho^i \tilde{\varepsilon}_{-i})'$,

we have

$$- E \frac{\partial^2 L}{\partial \psi \, \partial \xi_i} = E[v' D'^{-1} L^{m'} Ⓗ^{-1} \frac{\partial Ⓗ}{\partial \xi_i} Ⓗ^{-1} v] = \mathrm{tr}\{D'^{-1} L^{m'} Ⓗ^{-1} \frac{\partial Ⓗ}{\partial \xi_i} \} \tag{A.11}$$

$$E \frac{\partial^2 L}{\partial \delta_i \, \partial \xi_j} = 0 \tag{A.12}$$

and

$$- E \frac{\partial^2 L}{\partial \xi_i \, \partial \xi_j} = \tfrac{1}{2} \mathrm{tr}\{Ⓗ^{-1} \frac{\partial Ⓗ}{\partial \xi_i} Ⓗ^{-1} \frac{\partial Ⓗ}{\partial \xi_j} \} . \tag{A.13}$$

We compute approximate inverses by writing the matrix in Toeplitz-form and truncating it.

Appendix B The expectation of the sample moments for y_t and x_t

In order to compute the asymptotic covariance matrix for the ML estimator, we need expressions for the expectation of the sample moments for the variables y_t and x_t . We assume that x_t is non-stationary (when x_t is stationary, the derivation is much easier) and generated by $x_t = x_{t-1} + \mu + v_t$, with v_t being independently distributed as $N(0, \sigma_v^2)$ and independent of past x_t's . We now compute the second order moments for x_t and y_t .

$$\gamma_{22}(k,\ell) = E \sum_{t=0}^{T/n} x_{tn+k}\, x_{tn+\ell}$$

$$\gamma_{12}(k,\ell) = E \sum_{t=0}^{T/n} y_{tn+k}\, x_{tn+\ell}$$

$$\gamma^*_{22}(k,\ell) = E \sum_{t=0}^{T/n} x^*_{tn+k}\, x^*_{tn+\ell}\ , \text{ where } x^*_{tn+k} = \sum_{i=0}^{tn+k} \rho^i\, x_{tn+k-i}\ ,$$

$$\gamma_{11}(k,\ell) = E \sum_{t=0}^{T/n} y_{tn+k}\, y_{tn+\ell}\ . \tag{B.1}$$

For the sake of convenience, we delete the indices k and ℓ . As $\sum_{i=0}^{T/n} \rho^i\, i^a$ is $O(T^0)$ for finite a , we ignore the terms of this form. When divided by T , they become negligible in large samples. In the sequel, "$\doteq$" indicates that the equality holds except for terms of order $O(T^0)$. First we consider the case where $\sigma_v^2 = 0$ and $\mu = 1$. Then x_t is a deterministic linear trend and γ_{ij} can be solved as a difference equation.
For $\sigma_v^2 = 0$ and $\mu = 1$ we denote the second moments by $\overline{\gamma}_{ij}$.

$$\overline{\gamma}_{22} \doteq \sum_{t=0}^{T/n} (nt+k)(nt+\ell) = \frac{1}{3n} T^3 + \frac{1}{2}\left(1+\frac{\ell k}{n}\right) T^2$$
$$+ \left[\frac{n}{6} + \frac{1}{2}(\ell+k) + \frac{\ell k}{n} \right] T\ . \tag{B.2}$$

For the cross moments, we have

$$\overline{\gamma}_{12} \doteq \sum_{t=0}^{T/n} \left[\frac{nt+k}{1-\rho} - \frac{\rho}{(1-\rho)^2} \right] (nt+\ell)$$
$$\doteq \frac{\overline{\gamma}_{22}}{1-\rho} - \frac{\rho}{2n(1-\rho)^2} T^2 - \frac{\rho}{1-\rho^2}\left(\frac{1}{2} + \frac{\ell}{n}\right) T\ . \tag{B.3}$$

$$\bar{\gamma}^*_{22} \doteq \sum_{t=0}^{T/n} \left[\frac{nt+k}{1-\rho} - \frac{\rho}{(1-\rho)^2} \right] \left[\frac{nt+\ell}{1-\rho} - \frac{\rho}{(1-\rho)^2} \right]$$

$$\doteq \frac{\bar{\gamma}_{12}}{1-\rho} - \frac{\rho}{2n(1-\rho)^3} T^2 - \left[\frac{\rho}{(1-\rho)^3} \left(\frac{1}{2} + \frac{k}{n}\right) + \frac{\rho^2}{(1-\rho)^4} \right] T \tag{B.4}$$

and

$$\bar{\gamma}_{11} \doteq \bar{\gamma}^*_{22} + \frac{\sigma_\varepsilon^2 \rho^{|\ell-k|}}{n(1-\rho^2)} T \quad . \tag{B.5}$$

Now, we consider the general case where $\mu \neq 1$ and $\sigma_v^2 \neq 0$. The process for x_t and y_t can be written as

$$x_t = \mu t + \sum_{s=1}^{t} v_s + x_0$$

$$y_t = \sum_{i=0}^{t} \rho^i (x_{t-i} + \varepsilon_{t-i}) \quad , \tag{B.6}$$

where y_{-1} and ε_0 are assumed to be zero.

For y_t, we have

$$y_t = \sum_{i=0}^{t} \rho^i \mu(t-i) + \sum_{i=0}^{t} \sum_{s=1}^{t-i} \rho^i v_s + \sum_{i=0}^{t} \rho^i \varepsilon_{t-i}$$

$$= \frac{\mu t(1-\rho^{t+1})}{1-\rho} - \frac{\mu[\rho - (t+1)\rho^{t+1} + t\rho^{t+2}]}{(1-\rho)^2}$$

$$+ \sum_{s=1}^{t} \sum_{i=0}^{t-s} \rho^i v_s + \sum_{i=0}^{t} \rho^i \varepsilon_{t-i} \quad . \tag{B.7}$$

As $\sum_{i=0}^{T/n} \rho^i i^a$ is $O(T^0)$ for finite a, we can use for y_t

$$y_t \doteq \frac{\mu t}{1-\rho} - \frac{\mu\rho}{(1-\rho)^2} + \sum_{s=1}^{t} \frac{1-\rho^{t-s+1}}{1-\rho} v_s + \sum_{i=0}^{t} \rho^i \varepsilon_{t-i} \tag{B.8}$$

to derive the cross-moments

$$\gamma_{22} = \mu^2 \bar{\gamma}_{22} + \frac{\sigma_v^2}{2n} T^2 + \left(\frac{1}{2} + \frac{r}{n}\right) \sigma_v^2 T \quad ,$$

where $r = \min(k,\ell)$,

$$\gamma_{12} = \mu^2 \bar{\gamma}_{12} + \frac{\sigma_v^2}{2(1-\rho)n} T^2 + \frac{\sigma_v^2}{1-\rho} \left(\frac{1}{2} + \frac{r}{n}\right) T$$

$$- \frac{\rho^{k-r+1} \sigma_v^2 T}{(1-\rho)^2 n} ,$$

and finally

$$\gamma_{11} = \mu^2 \bar{\gamma}^*_{22} + \frac{\sigma_\varepsilon^2 \rho^{|\ell-k|}}{n(1-\rho^2)} T + \frac{\sigma_v^2}{1-\rho^2} C \quad , \text{ with}$$

$$C = \frac{T^2}{2n} + \left\{ \frac{1}{2} + \frac{r}{n} - \frac{\rho^{\ell-r+1}}{(1-\rho)n} - \frac{\rho^{k-r+1}}{(1-\rho)n} + \frac{\rho^{k+\ell-2r+2}}{(1-\rho^2)n} \right\} T \qquad (B.9)$$

Appendix C The use of R^2 in choosing the parameters of the model

In this appendix, we give the expressions that relate the parameters of the model to the theoretical coefficient of determination R^2 . Plausible values for R^2 have been selected and used in these relationships to determine the parameter values for the models in sections 5 and 6 .
Consider the following dynamic regression model

$$y_t = \rho\, y_{t-1} + x_t\, \beta + \varepsilon_t \quad \text{with} \quad \beta = 1 , \quad \varepsilon_t \sim IN\,(0, \sigma_\varepsilon^2) , \tag{C.1}$$

and define R^2 as

$$R^2 = \text{var}\,(z_t) \,/\, \text{var}\,(y_t) , \text{ where } z_t = \frac{1}{1-\rho L}\, x_t . \tag{C.2}$$

1) We assume that $x_t = \varphi\, x_{t-1} + v_t$ with $|\varphi| < 1$ and $v_t \sim IN\,(0, \sigma_v^2)$ and independent of ε_t and past x_t's . The available z_t is generated by a second order autoregressive process. Its variance is

$$\text{var}\,(z_t) = A\, \sigma_v^2 , \tag{C.3}$$

where $A = \dfrac{(1+\rho\varphi)}{(1-\rho\varphi)\,[(1+\rho\varphi)^2 - (\rho+\Psi)^2]}$,

and $\quad R_2 = A\, \sigma_v^2 \,/\, [\, A\, \sigma_v^2 + B\, \sigma_\varepsilon^2 \,]$, with $B = 1/(1-\rho^2)$,

so that

$$\sigma_v^2 = \frac{R^2}{(1-R^2)}\, \frac{B}{A}\, \sigma_\varepsilon^2 \quad \text{and} \quad \sigma_x^2 = \frac{R^2}{(1-R^2)}\, \frac{B}{A}\, \frac{\sigma_\varepsilon^2}{(1-\varphi^2)} . \tag{C.4}$$

When x_t is assumed to be white noise (i.e. $\varphi = 0$) , we have

$$\sigma_x^2 = R^2 \,/\, (1-R^2) \tag{C.5}$$

2) When x_t is constant, i.e. $x_t = \mu$, we can determine μ from R^2

$$R^2 = \underset{T\to\infty}{\text{plim}} \frac{\sum_{t=1}^{T} (\frac{x_t}{1-\rho L})^2}{\sum_{t=1}^{T} y_t^2} = \frac{\frac{\mu^2}{(1-\rho)^2}}{\frac{\mu^2}{(1-\rho)^2} + \frac{\sigma_\varepsilon^2}{1-\rho^2}}$$

$$\mu = \left[\frac{R^2}{1-R^2} \frac{(1-\rho)}{(1+\rho)} \sigma_\varepsilon^2 \right]^{\frac{1}{2}} . \tag{C.6}$$

3) For non-stationary x_t , $n = 1$ and $\sigma_v^2 \neq 0$, we have

$$R^2 = \frac{\gamma_{11}(0,0) - \frac{\sigma_\varepsilon^2}{1-\rho^2} T}{\gamma_{11}(0,0)} , \tag{C.7}$$

where $\gamma_{11}(0,0)$ is obtained from (B.9) by setting $k = \ell = 0$.

When $\sigma_v^2 = 0$, it can readily be seen using (B.5) and (B.9) that (C.7) specializes to

$$R^2 = \frac{\mu^2(\overline{\gamma}_{11}(0,0) - \sigma_\varepsilon^2\, T/1-\rho^2)}{\mu^2(\overline{\gamma}_{11}(0,0) - \sigma_\varepsilon^2\, T/1-\rho^2) + \frac{\sigma_\varepsilon^2\, T}{1-\rho^2}} ,$$

from which μ is easily obtained as

$$\mu = \left[\frac{R^2\, \sigma_\varepsilon^2\, T/1-\rho^2}{(1-R^2)\; (\overline{\gamma}_{11}(0,0) - \frac{\sigma_\varepsilon^2\, T}{1-\rho^2})} \right]^{\frac{1}{2}} . \tag{C.8}$$

When $\mu = 0$, we have from (B.9) and (C.7) that

$$\sigma_v^2 = \frac{\sigma_\varepsilon^2\, T}{C} \frac{R^2}{1-R^2} . \tag{C.9}$$

Finally, when the ratio $\mu^2/\sigma_v^2 = \alpha$ is given, we get from (B.9) and (C.7)

$$\mu = \left[\frac{R^2}{1-R^2} \frac{\sigma_\varepsilon^2\, T}{(1-\rho^2)\, \overline{\gamma}_{22}^* + C/\alpha} \right]^{\frac{1}{2}} . \tag{C.10}$$

References

Amemiya, T., and R.Y. Wu (1972), "The effect of aggregation on prediction in the autoregressive model", Journal of the American Statistical Association, 67, 628-632.

Anderson, T.W. (1957), "Maximum likelihood estimates for a multivariate normal distribution when some observations are missing", Journal of the American Statistical Association, 52, 200-203.

Boot, J.C.G., Feibes, W., and J.H.C. Lisman (1967), "Further methods of derivation of quarterly figures from annual data", Applied Statistics, 16, 65-75.

Chow, G.C., and A. Lin (1971),"Best linear unbiased interpolating distribution and extrapolation of time series by related series", The Review of Economics and Statistics, 53, 372-375.

Chow, G.C., and A. Lin (1976), "Best linear unbiased estimation of missing observations in an economic time series", Journal of the American Statistical Association, 71, 719-721.

Dagenais, M.G. (1973), "The use of incomplete observations in multiple regression analysis, a generalized least squares approach", Journal of Econometrics, 1, 317-328.

Dunsmuir, W. (1981), "Estimation for stationary time series when data are irregularly spaced or missing", in Findley, D.F. (ed.), Applied Time Series II, Academic Press, New York.

Dunsmuir, W., and P.M. Robinson (1981), "Estimation of time series models in the presence of missing data", Journal of the American Statistical Association, 76, 560-568.

Gelauff, G.M.M., and R. Harkema (1977), "Estimating quarterly models with partly missing quarterly observations",Erasmus University Rotterdam, Report 7724.

Geweke, J. (1978), "Temporal aggregation in the multiple regression model", Econometrica, 46, 643-661.

Ginsburgh, V.A. (1973), "A further note on the derivation of quarterly figures consistent with annual data", Applied Statistics, 22, 368-374.

Gourieroux, Ch., and A. Monfort (1981), "On the problem of missing data in linear models, Review of Economic Studies, 48, 579-586.

Harvey, A.C., and P.L.V. Pereira (1980), "The estimation of dynamic models with missing observations", Paper presented at the Econometric Society World Meeting in Aix-en-Provence.

Harvey, A.C. (1981), The Econometric Analysis of Time Series, Philip Allan Publ. Ltd., Oxford.

Harvey, A.C., McKenzie, C.R., Blake, D.P.C., and M.J. Desai (1981), "Data revisions in the U.K.", Paper presented at the ASA-CENSUS-NBER Conference on Applied Time Series Analysis of Economic Data held in Washington, D.C., October 1981.

Hsiao, Cheng (1979), "Linear regression using both temporally aggregated and temporally disaggregated data", Journal of Econometrics, 10, 203-252.

Jones, R.H. (1980), "Maximum likelihood fitting of ARMA models to time series with missing observations", Technometrics, 22, 389-395.

Kmenta, J. (1981), "On the problem of missing measurements in the estimation of economic relationships", in Charatsis E.G. (ed.), Proceedings of the Econometric Society Meeting 1979, Amsterdam, North-Holland Publish. Comp. .

Orchard, T., and M.A. Woodbury (1972), "A missing information principle: theory and applications", Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, 1, 697-715.

Palm, F.C., and Th.E. Nijman (1981), "Linear regression using both temporally aggregated and temporally disaggregated data", Free University , Amsterdam, Research Memorandum, forthcoming in Journal of Econometrics.

Sargan, J.D., and E.G. Drettakis (1974),"Missing data in an autoregressive model", International Economic Review, 15, 39-58.

Shaman, P., and S. B. Tan (1981), "The mean square error of one-step-ahead prediction for an autoregressive process when the coefficients are estimated with missing data", University of Pennsylvania, manuscript.

Sims, C.A. (1971), "Discrete approximations to continuous time distributed lags in econometrics", Econometrica, 39, 545-563.

Somermeyer, W.H., Jansen, R., and A.S. Louter (1976), "Estimating quarterly values of annually known variables in quarterly relationships", Journal of the American Statistical Association, 71, 588-595.

Tan, S.B. (1979), "Maximum likelihood estimation in autoregressive processes with missing observations", Ph.D. Thesis, University of Pittsburgh.

Telser, L. (1967), "Discrete samples and moving sums in stationary stochastic processes", Journal of the American Statistical Association, 62, 484-499.

Tiao, G.C., and W.S. Wei (1976), "Effect of temporal aggregation on the dynamic relationship of two time series variables", Biometrika, 63, 513-523.

Wansbeek, T., and A. Kapteyn (1981), "Maximum likelihood estimation in a linear model with serially correlated errors when observations are missing", Central Bureau of Statistics, The Netherlands, manuscript.

Zellner, A. (1966), "On the analysis of first order autoregressive models with incomplete data", International Economic Review, 7, 72-76.

Zellner, A., and C. Montmarquette (1971), "A study of some aspects of temporal aggregation problems in econometric analyses", Review of Economics and Statistics, 53, 335-342.

1979-1	A.A. Schreuder-Sunderman, Frans Blommaert en Hein Schreuder,	Werknemers en Soc. Jaarverslag.
1979-2	H. Schreuder,	De maatschappelijke verantwoordelijkheid van Ondernemingen.
1979-3	P. Nijkamp en P. Rietveld,	Multilevel Multi-objective Models in a Multiregional System.
1979-4	J. Arntzen, G. Bornmalm-Jardelöw and P. Nijkamp,	Duality, Segmentation and Dynamics on a Regional Labour Market.
1979-5	P. Nijkamp, A. Soffer	Soft Multicriteria Decision Models for Urban Renewal Plans.
1979-6	drs. A.J. Mathot	A Model of choosing a car with or without a credit.
1979-7	H. Blommestein, P. Nijkamp en W. van Veenendaal,	Shopping Perceptions and preferences: A multi-dimensional Attractivenerss Analysis of Consumer and Entrepreneurial Attitudes.
1979-8	H.J.Blommestein and F.C. Palm	The Aggregate Demand for Money in the Netherlands- a new look at a study of the Bank of the Netherlands.
1980-1 Jan.	P. Nijkamp and H. Voogd	New Multicriteria Methods for Physical Planning by Means of Multidimensional Scaling Techniques.
1980-2	Hidde P. Smit	Medium- and Long-Term Models for the Escap-Region - A review of existing models and a proposal for a new model system -
1980-3 april	P.v. Dijck en H. Verbruggen	Productive Employment in Developing Countries' Exporting Industries
1980-4	P. Nijkamp en L. Hordijk	Integrated Approaches to Regional Development Models; A survey of some Western European Models
1980-5	P. Nijkamp	Soft Econometric Models. An analysis of Regional Income Determinants
1980-6	P. Nijkamp en F. van Dijk	Analysis of Conflicts in Dynamical Environ-mental Systems via Catastrophe Theory.
1980-7 juni	E. Vogelvang	A short term econometric model for the consumer demand of roasted coffee in the Netherlands
1980-8	N. van Hulst	De effectiviteit van de geleide loonpolitiek in Nederland.
1980-9 Oct. 1980	P. Nijkamp	A survey of Dutch integrated Energy-Environmental-Economic Policy Models
1980-10 Oct. 1980	P. Nijkamp	Perspectives for Urban analyses and policies
1980-11 Oct. 1980	P. Nijkamp	New developments in multidimensional geographical data and policy analysis
1980-12	F.C. Palm, E. Vogelvang en D.A. Kodde	Efficient Estimation of the Geometric Distributed Lag Model; Some Monte Carlo Results on Small Sample Properties