Simple Solutions to the Initial Conditions Problem in Multiple Equation Dynamic Panel Data Models with Individual Effects

Wladimir Raymond[*], Pierre Mohnen, Franz Palm
and Sybrand Schim van der Loeff

(Preliminary Draft)

February 9, 2010

Abstract

This paper generalizes Wooldridge's simple solutions to the initial conditions problems to multiple-equation dynamic models. The problem is taken care of by using, in each equation, interrelated linear projections where their unobserved components are integrated out using multiple-step Gauss-Hermite quadrature. The estimation is handled by maximum likelihood for a whole class of multiple equation dynamic models such as panel VARs and dynamic sample selection models among others. We derive a general form for the likelihood where only the core likelihood differs across models. The likelihood evaluation is straightforward and should be added as basic commands in econometric packages.

[*]University of Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands; Room A4.27; Tel.: (+31) 43 388 4962; Fax: (+31) 43 388 4874; w.raymond@maastrichtuniversity.nl

1 Introduction

One of the main difficulties that occurs when studying nonlinear dynamic models with short panel data, besides a proper treatment of the individual effects, is the so-called initial conditions problem. This problem arises when the first period of the observational data does not coincide with that of the stochastic process being modeled. In that case, the first observation depends on the dependent variable(s) in the period before the sample starts. One approach consists of ignoring the problem altogether by using exogenous initial conditions, which assumes that the dependent variable(s) at period t_0, say y_{i0}, is independent of the individual effects. This approach can be used if the panel consists of a large T, or with short T panels when one can reasonably assume that the process is observed from the start. Another approach is by Heckman (1981) and consists of specifying a reduced-form model for the initial conditions given the individual effects and the exogenous regressors. This model is often assumed to be similar to the underlying model for the remaining time periods. Then, the parameters of the initial conditions model as well as those of the model at the remaining period are jointly estimated. Similar to Heckman's is Orme's (2001) approach that consists, in a first step, of specifying and estimating a model for the initial conditions. Generalized residuals, in the spirit of Gouriéroux et al. (1987), are then computed and used as regressors in the second step estimation of the dynamic model for the remaining time periods. More recently, Wooldridge (2005a) suggests an approach that consists of specifying a distribution for the individual effects conditional on the initial conditions. The parameters of the initial conditions are not estimated as the estimation is done conditional on the initial (observed) values of the dependent variables. Wooldridge's approach is very popular in empirical work because a rather broad class of single equation dynamic models can be estimated without any additional programming effort regarding the modeling of the initial conditions. Indeed, the likelihood of such models is written such that existing commands of standard packages (e.g. Stata) that handle the estimation of these models in a static framework can be used to estimate the dynamic variants by simply adding as explanatory variables the lagged dependent variables, the dependent variables taken at the first observed period, and the time varying explanatory variables that enter the linear projection of the individual effects. Wooldridge's study, however, is completely silent when the model under study has multiple equations, even though the approach itself can be easily extended. This is where this present study comes into play.

We generalize Wooldridge's approach to multiple equation dynamic models. More specifically, we show that the initial conditions problem is taken care of by projecting in each equation the individual effects on the time varying explanatory variables and the dependent variables taken at the first observed period. The linear projections are interrelated and the unobserved components are integrated out using multiple-step Gauss-Hermite (G-H) quadrature. The estimation is handled by maximum likelihood (ML) in a rather straightforward manner for a whole class of multiple equation dynamic models such as panel vector autoregressions (VARs), dynamic bivariate probit and dynamic sample selection models among others. We derive a general form for the likelihood of such models where only the core likelihood, which is the likelihood function of the cross-sectional versions of these models, differs across models. The multiple step G-H involved in the numerical evaluation of the likelihood is shown to be implemented rather easily and should be added in econometric software (e.g. Stata) as standard commands to estimate such models. Indeed, the functions that form the core likelihood are usually well-known in the literature, and the only programming effort consists in devising a two- or -three-part loop that includes the weights and abscissas of the G-H quadrature. To keep the exposition simple, we choose to work with two-equation models, but the principle is the same for three-equation models. We illustrate our method

The remainder of the paper is as follows. Section 2 presents the class of models we consider in the analysis. The relevant conditions that these models need to satisfy, the resulting likelihood and average partial effects are also described. Section 3 discusses in more details the leading examples of this class of models. Section 4 illustrates our approach by showing an application to a dynamic model of innovation. We summarize and conclude in Section 5.

2 A class of two-equation dynamic models

Let $\mathbf{y}_{it}^* = (y_{1it}^*, y_{2it}^*)'$ denote a vector of *latent* variables for cross-sectional unit i taken at period t, with $i = 1, ..., N$ and $t = 1, ..., T$.[1] It can be expressed as a function of past *observed* outcomes $\mathbf{y}_{i,t-1}$, strictly exogenous explanatory variables $\mathbf{z}_{it}$, individual effects $\mathbf{c}_i$ and other unobserved variables summarized in the idiosyncratic errors $\mathbf{u}_{it}$. Formally, the class of two-equation dynamic models can be written as

$$\mathbf{y}_{it}^* = \boldsymbol{\rho}'\mathbf{g}(\mathbf{y}_{i,t-1}) + \boldsymbol{\gamma}'\mathbf{z}_{it} + \mathbf{c}_i + \mathbf{u}_{it}, \tag{2.1}$$

where $\mathbf{g}$ is a vector function of the lagged observed outcomes, $\mathbf{z}_{it} = (\mathbf{z}_{1it}, \mathbf{z}_{2it})'$, $\mathbf{c}_i = (c_{1i}, c_{2i})'$, $\mathbf{u}_{it} = (u_{1it}, u_{2it})'$, and $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)'$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)'$ are vectors of parameters to be estimated. Throughout the paper, we take the convention of writing vectors as boldface terms while scalars are not boldface. The following assumptions underlying equation (2.1) are made. First, unlike Wooldridge (2005a), we consider models with two variables of interest, hence the use of a two-component *vector* of individual effects. Secondly, the dynamics is of the first order conditional on $\mathbf{z}_{it}$ and $\mathbf{c}_i$, and allows for interactions with components of $\mathbf{z}_{it}$, hence the use of the function $\mathbf{g}$. For instance, we may study the persistence of product and process innovation that is industry-specific. Thirdly, the way the vectors of latent ($\mathbf{y}_{it}^*$) and observed ($\mathbf{y}_{it}$) variables are related depends on the model under study. For instance, in a dynamic bivariate probit model, $\mathbf{y}_{it} = \mathbf{1}[\mathbf{y}_{it}^* > 0]$, where $\mathbf{1}[\ ]$ is the indicator function taking on the value one if the expression between squared brackets is true, and zero otherwise.

2.1 Maximum likelihood estimation

In order to estimate the models of equation (2.1) by ML, we make the following assumptions. First, we assume that the dynamics is correctly specified conditional on $\mathbf{z}_{it}$ and $\mathbf{c}_i$, i.e. there is at most one lag of the dependent variables, and $\mathbf{z}_{it}$ is strictly exogenous conditional on $\mathbf{c}_i$. Secondly, the distribution of $\mathbf{y}_{it}$ conditional on $\mathbf{y}_{i,t-1}$, $\mathbf{z}_{it}$ and $\mathbf{c}_i$, is correctly specified. This distribution is also equal to that of $\mathbf{y}_{it}$ conditional on $\mathbf{y}_{i,t-1}, ..., \mathbf{y}_{i0}$, $\mathbf{z}_i$ and

[1] To keep the exposition simple, we use a balanced panel in the description of the general framework. In the application we use an unbalanced panel, which requires additional specifications regarding the initial conditions.

$\mathbf{c}_i$, where $\mathbf{z}_i = (\mathbf{z}_{i1}, ..., \mathbf{z}_{iT})'$, and Wooldridge (2005a) refers to its density as "structural" density, which will be used to form what we call throughout the paper the *core likelihood*. More specifically, we form the core likelihood using the density of $(\mathbf{y}_{i1}, ..., \mathbf{y}_{iT}|\mathbf{y}_{i0}, \mathbf{z}_i, \mathbf{c}_i)$ which is then "integrated out" using the density of $(\mathbf{c}_i|\mathbf{y}_{i0}, \mathbf{z}_i)$ assumed to be correctly specified. The likelihood function of the class of dynamic models for cross-sectional unit i can be written as

$$l_i(\boldsymbol{\theta}, \boldsymbol{\delta}) = \int_{\mathbf{c}_i} \prod_{t=1}^{T} \boldsymbol{f}(\mathbf{y}_{it}|\mathbf{y}_{i,t-1}, \mathbf{z}_{it}, \mathbf{c}_i, \boldsymbol{\theta}) \boldsymbol{h}(\mathbf{c}_i|\mathbf{y}_{i0}, \mathbf{z}_i, \boldsymbol{\delta}) \mathbf{d}\mathbf{c}_i, \tag{2.2}$$

where $\boldsymbol{f}$ and $\boldsymbol{h}$ denote respectively the "structural" density and the density of the individual effects, and $\boldsymbol{\theta}$ consists of $\boldsymbol{\rho}$, γ and additional parameters of the density $\boldsymbol{f}$, and $\boldsymbol{\delta}$ consists of the parameters of the density $\boldsymbol{h}$.

A few remarks are worth noting when we consider the likelihood function written in equation (2.2). First, this likelihood can be applied to *any* dynamic random-effects model with idiosyncratic errors that are serially independent conditional on the individual effects. Secondly, modeling the individual effects conditional on the initial values of the dependent variables is what Wooldridge (2005a) refers to as "simple solutions to the initial conditions" as opposed to modeling the initial values conditional on the individual effects (Heckman, 1981). Thirdly, while Wooldridge applies this likelihood to single equation models with a *scalar* of individual effects, we apply the likelihood to two-equation models where a two-component *vector* of individual effects is considered. Hence, in practice we have to evaluate a double integral over the support of the distribution of $\mathbf{c}_i$, which is achieved by using a two-step Gauss-Hermite quadrature. The evaluation of the likelihood for the class of models of equation (2.1) is done in a similar manner where only the core likelihood differs across models. For simplicity, we use the (bivariate) normal distribution for $\boldsymbol{h}$ throughout the analysis although in principle any distribution that belongs to the exponential family can be used.

Two-step Gauss-Hermite quadrature

For any model of the class under study, we can write the density $\boldsymbol{f}$ as the product of $f_1(...|\mathbf{y}_{i0}, \mathbf{z}_i, c_{1i}, c_{2i})$ and $f_2(...|y_{2i0}, \mathbf{z}_{2i}, c_{2i})$ where f_1 denotes a function of two components of the individual effects and f_2 denotes a function of only one component of the individual effects. Similarly, we can write $\boldsymbol{h}$ as the product of $h_1(c_{1i}, c_{2i}|\mathbf{y}_{i0}, \mathbf{z}_i)$ and $h_2(c_{2i}|y_{2i0}, \mathbf{z}_{2i})$ whith h_1 and h_2 having a similar interpretation. Wooldridge's approach of handling the initial conditions problem consists in modeling $\mathbf{c}_i$ as

$$\mathbf{c}_i = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1'\mathbf{g}(\mathbf{y}_{i0}) + \boldsymbol{\alpha}_2\mathbf{z}_i + \boldsymbol{a}_i, \tag{2.3}$$

where $\boldsymbol{\alpha}_j = (\alpha_{1j}, \alpha_{2j})'$ (for $j = 0, 1, 2$) and $\boldsymbol{a}_i = (a_{1i}, a_{2i})'|\mathbf{y}_{i0}, \mathbf{z}_i \sim \boldsymbol{Normal}(\mathbf{0}, \boldsymbol{\Sigma_a})$ with $\boldsymbol{\Sigma_a} = \begin{pmatrix} \sigma_{a_1}^2 & \\ \rho_{\boldsymbol{a}}\sigma_{a_1}\sigma_{a_2} & \sigma_{a_2}^2 \end{pmatrix}$. The following remarks regarding equation (2.3) are worth making. First, if the vector of parameters $\boldsymbol{\gamma}$ contains intercepts, only the sum of those intercepts and $\boldsymbol{\alpha}_0$ is identified. Secondly, $\boldsymbol{\alpha}_1$ captures the correlation between the individual effects and the initial values of the dependent variables. A test of the significance of $\boldsymbol{\alpha}_1$ is equivalent to a test of exogeneity of the initial values. Thirdly, the function $\mathbf{g}$ enters the individual effects equation in the same way as it enters equation (2.1). In other words, if we allow for interactions between the dynamics and components of $\mathbf{z}_{it}$, we should also allow for interactions between the initial values and the same components. Fourthly, $\boldsymbol{\alpha}_2$ captures the correlation between the individual effects and the strictly exogenous regressors that have to exhibit sufficient time variation for them to be included in equation (2.3). Models that allow for such a correlation is referred to as "correlated effects" models in the literature (e.g. Chamberlain, 1980).

After substituting equation (2.3) for $\mathbf{c}_i$ into equation (2.2), the likelihood function can be written as

$$l_i(\boldsymbol{\theta}, \boldsymbol{\delta}) = \int_{a_{2i}} h_2(a_{2i}|..., \boldsymbol{\delta}) \prod_{t=1}^{T} f_2(...|..., a_{2i}, \boldsymbol{\theta}) H_2(a_{2i}|..., \boldsymbol{\theta}, \boldsymbol{\delta}) da_{2i}, \tag{2.4}$$

where H_2 denotes a function of the sole component a_{2i} of the individual effects and is given by

$$H_2(a_{2i}|..., \boldsymbol{\theta}, \boldsymbol{\delta}) = \int_{a_{1i}} h_1(a_{1i}, a_{2i}|..., \boldsymbol{\delta}) \prod_{t=1}^{T} f_1(...|..., a_{1i}, a_{2i}, \boldsymbol{\theta}) da_{1i}. \tag{2.5}$$

If $\boldsymbol{h}$ is bivariate normal, then both h_1 and h_2 involve the exponential function, which makes the use of quadrature feasible when evaluating the integrals of equations (2.4) and (2.5). The G-H quadrature states that

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)dx \simeq \sum_{m=1}^{M} w_m f(a_m), \tag{2.6}$$

where w_m and a_m are respectively the weights and abscissas of the G-H quadrature, the tables of which are formulated in mathematical textbooks, and M is the total number of integration points. The two-step G-H quadrature works as follows. In the first step, we apply the approximation of equation (2.6) to equation (2.5) by making the appropriate variable change. In the second step, we replace the approximated expression of $H_2(a_{2i}|..., \boldsymbol{\theta}, \boldsymbol{\delta})$ into equation (2.4) and apply again the approximation of equation (2.6). In the case of the bivariate normal distribution, the expressions of $h_1(a_{1i}, a_{2i}|..., \boldsymbol{\delta})$ and $h_2(a_{2i}|..., \boldsymbol{\delta})$ are given by

$$h_1(a_{1i}, a_{2i}|..., \boldsymbol{\delta}) = k e^{\frac{-a_{1i}^2}{2\sigma_{a_1}^2(1-\rho_{\mathbf{a}}^2)}} e^{\frac{\rho_{\mathbf{a}}^2 \frac{a_{1i}}{\sigma_{a_1}} \frac{a_{2i}}{\sigma_{a_2}}}{(1-\rho_{\mathbf{a}}^2)}}; \; h_2(a_{2i}|..., \boldsymbol{\delta}) = e^{\frac{-a_{2i}^2}{2\sigma_{a_2}^2(1-\rho_{\mathbf{a}}^2)}}, \tag{2.7}$$

where $k = (2\pi\sigma_{a_1}\sigma_{a_2})^{-1}\left(1-\rho_{\mathbf{a}}^2\right)^{\frac{-1}{2}}$. The appropriate variable change to make in each step of the G-H integration is $x_{ji} = \frac{a_{ji}}{\sigma_{a_j}\sqrt{2(1-\rho_{\mathbf{a}}^2)}}$ $(j = 1, 2)$ so that $da_{ji} = dx_{ji}\sigma_{a_j}\sqrt{2(1-\rho_{\mathbf{a}}^2)}$. The two-step G-H evaluation of the likelihood function yields

$$l_i(\boldsymbol{\theta}, \boldsymbol{\delta}) \simeq \frac{\sqrt{1-\rho_{\mathbf{a}}^2}}{\pi} \sum_{p=1}^{P} w_p \prod_{t=1}^{T} f_2(...|..., a_p, \boldsymbol{\theta}, \boldsymbol{\delta}) \sum_{m=1}^{M} w_m e^{2\rho_{\mathbf{a}} a_p a_m} \prod_{t=1}^{T} f_1(...|..., a_p, a_m, \boldsymbol{\theta}, \boldsymbol{\delta}),^{2} \tag{2.8}$$

where w_m, w_p, a_m and a_p are respectively the weights and abscissas of the first- and second-stage G-H integration with M and P being the first- and second-stage total number of integration points.[3] The exact expressions involving the G-H abscissas at which f_1 and f_2 are operated stem from the variable change and are equal to $a_m\sigma_{a_1}\sqrt{2(1-\rho_{\mathbf{a}}^2)}$ and $a_p\sigma_{a_2}\sqrt{2(1-\rho_{\mathbf{a}}^2)}$. In practice, we maximize

$$l(\boldsymbol{\theta}, \boldsymbol{\delta}) = \sum_{i=1}^{N} \ln l_i(\boldsymbol{\theta}, \boldsymbol{\delta}) \tag{2.9}$$

[2]The parameters $\boldsymbol{\delta}$ enters now the conditioning set of f_1 and f_2 because of the variable change.

[3]See Raymond (2007, Chapter 3) for more details on the derivation of the two-step G-H integration.

using standard numerical techniques (e.g. Newton-Raphson) to obtain estimates of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$. The standard errors are obtained using standard methods like the Hessian matrix or the outer product of gradients.

The strength of this approach is that it handles in a single framework the estimation of a whole class of dynamic panel data models. In other words, equation (2.8) represents the individual likelihood of all these models where only f_1 and f_2 vary across models. Furthermore, the implementation of the likelihood (equation (2.8)) can be done rather easily and should be added in econometric software (e.g. Stata) as standard commands to estimate such models. Indeed, the functions f_1 and f_2 are usually well-known in the literature since their product form the core likelihood which has a similar expression to that of the likelihood of such models taken in a cross-sectional framework. The only programming effort consists in devising a two-part loop that includes the weights and abscissas of the G-H quadrature. The larger M and P, the more accurate the ML estimates but also the more time-consuming the maximization of the likelihood. Computation time also depends on the tractability of the core likelihood, the number of explanatory variables and the choice of the initial values for the optimization algorithm. For instance, other things being equal (e.g. given M and P), it is more time-consuming to estimate a dynamic bivariate probit whose core likelihood contains a bivariate normal cdf than a bivariate VAR(1) whose core likelihood contains a bivariate normal pdf.

2.2 Average partial effects

In the dynamic models of equation (2.1), the estimated parameter $\widehat{\boldsymbol{\theta}}$ usually gives only the direction and significance of the effect of the explanatory variables on the dependent variables. In order to obtain the magnitude of the effect of explanatory variable, say z_j, on the dependent variables $\mathbf{y}_t$, we need to compute marginal effects also known as *partial effects* defined as

$$\mathbf{m}(\mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c}) = \partial \mathbf{E}(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c}) / \partial z_j, \tag{2.10}$$

where we must choose values of interest (e.g. the mean) for $\mathbf{y}_{t-1}, \mathbf{z}_t$, and $\mathbf{c}$. For instance, in a dynamic bivariate probit, $\mathbf{E}(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c}) = P(y_{1t} = 1, y_{2t} = 1 | \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c})$, and the partial effect of z_j is on the joint probability of success for both dependent variables conditional on interesting values of the regressors and the individual effects. One approach consists in filling values for the individual effects $\mathbf{c}$ in the calculation of the partial effects. For instance, if $\mathbf{c}_i$ is independent of $\mathbf{y}_{i0}$ and $\mathbf{z}_{it}$ so that $\mathbf{E}(\mathbf{c}_i|\mathbf{y}_{i0}, \mathbf{z}_{it}) = \mathbf{0}$, we could substitute $\mathbf{0}$ for $\mathbf{c}$ in equation (2.10). However, if $\mathbf{c}_i$ has a continuous distribution, as in our case, calculating the partial effects at $\mathbf{c} = \mathbf{0}$ may be representative of only a small percentage of the population (Chamberlain, 1984). Alternatively, if $\mathbf{c}_i$ depends on $\mathbf{y}_{i0}$ and $\mathbf{z}_{it}$ as in equation (2.3) so that $\mathbf{E}(\mathbf{c}_i|\mathbf{y}_{i0}, \mathbf{z}_i) = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1'\mathbf{g}(\mathbf{y}_{i0}) + \boldsymbol{\alpha}_2\mathbf{z}_i$, we could replace $\mathbf{c}$ by the latter expression in equation (2.10).[4] Wooldridge (2002) argues that, since $\mathbf{c}$ is unobserved, it does not have any meaningful units of measurement. Thus, replacing $\mathbf{c}$ by any value in the calculation of the partial effects is not meaningful. Wooldridge (2002) suggests averaging the partial effects over the distribution of the individual effects to obtain *average partial effects* (APEs), i.e.

$$\boldsymbol{\mu}(\mathbf{y}_{t-1}, \mathbf{z}_t) = \mathbf{E}_{\mathbf{c}}[\mathbf{m}(\mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c})], \tag{2.11}$$

where $\mathbf{E}_{\mathbf{c}}$ denotes the expectation of the partial effects with respect to the distribution of the individual effects $\mathbf{c}$. The partial effects (eq. (2.10)) and the average partial effects (eq. (2.11)) quantify the *joint* effect of variable z_j on the expected value of both variables of interest y_{1t} and y_{2t}. Furthermore, the partial effects are functions of both components of the individual effects. Hence, the expectation in equation (2.11) involves a double integral which is to be evaluated over the support of the (bivariate) distribution of $\mathbf{c}$. The evaluation can be done in a similar manner to that of the likelihood of equation (2.2), as described in subsection 2.1, where the core likelihood $\boldsymbol{f}$ is to be replaced by the partial effects $\mathbf{m}$.

Another type of partial effects involves the conditional distribution of one dependent variable given the other one, i.e.

$$m_{1|2}(y_{2t}, \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c}) = \frac{\partial E(y_{1t}|y_{2t}, \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c})}{\partial z_j}, \; m_{2|1}(y_{1t}, \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c}) = \frac{\partial E(y_{2t}|y_{1t}, \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c})}{\partial z_j}. \tag{2.12}$$

Both *conditional* partial effects are functions of the two components of the individual effects. Thus, the resulting conditional APEs are obtained by averaging the conditional

[4] In practice, we would replace $\boldsymbol{\alpha}_j$ $(j = 0, 1, 2)$ by their estimates $\widehat{\boldsymbol{\alpha}}_j$.

partial effects over the distribution of the individual effects, i.e.

$$\mu_{1|2}(y_{2t}, \mathbf{y}_{t-1}, \mathbf{z}_t) = E_{\mathbf{c}}[m_{1|2}(y_{2t}, \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c})],\ \mu_{2|1}(y_{1t}, \mathbf{y}_{t-1}, \mathbf{z}_t) = E_{\mathbf{c}}[m_{2|1}(y_{1t}, \mathbf{y}_{t-1}, \mathbf{z}_t, \mathbf{c})]. \tag{2.13}$$

The conditional APEs also involve double integration which can be evaluated as for the likelihood and the joint APEs.

Finally, *marginal* partial effects can be obtained for each equation as

$$m_1(y_{1t-1}, \mathbf{z}_{1t}, c_1) = \frac{\partial E(y_{1t}|y_{1t-1}, \mathbf{z}_{1t}, c_1)}{\partial z_j},\ m_2(y_{2t-1}, \mathbf{z}_{2t}, c_2) = \frac{\partial E(y_{2t}|y_{2t-1}, \mathbf{z}_{2t}, c_2)}{\partial z_j},$$

where each m is a function of only one component of the individual effects. The corresponding marginal APEs are obtained in a similar fashion to that of univariate APEs except that the estimated coefficients used in the derivation of the former take account of the cross-equation correlation through among others $\mathbf{c}$, while those used in the latter ignore this correlation. The marginal APEs can then be computed by the general method described in Wooldridge (2005b) and applied to single equation panel data models in Wooldridge (2005a).

The relevance of the three types of APEs depends on the model that is estimated. For instance, all three types of APEs can be of interest in a dynamic bivariate probit, while only the conditional and the marginal APEs are of relevance in a dynamic type 2 tobit

3 Examples

In this section, we study several leading examples of two-equation dynamic models of the class described in equation (2.1), namely the dynamic bivariate probit, the dynamic bivariate (type 1) tobit and two dynamic sample selection models also referred to as type 2 and type 3 tobit by Amemiya (1984). While these models are frequently used in a cross-sectional framework, and the one-equation counterparts of some of them are frequently used in a cross-sectional and panel data framework (e.g. univariate type 1 tobit), their use in the context of panel data remains to date rather scarce. This is in part due to the presence, in both equations, of two different components of the individual effects that need to be integrated out with respect to their bivariate distribution, which makes the estimation of such models fairly difficult in the context of panel data. One approach that simplifies the estimation procedure is to reduce the dimension of the integration. This can be achieved by using a common factor specification with two different factor loadings. Thus, integrating out the individual effects is done with respect to a univariate distribution (see e.g. Flaig and Stadler, 1994), which makes the use of the quadrature approach described in Butler and Moffitt (1982) or the adaptive quadrature approach described in Rabe-Hesketh et al. (2005) possible. While the common factor approach simplifies the estimation procedure, it imposes the rather strong restriction of a one correlation between the individual effects of the two equations. However, it is more interesting to allow for two different components of the individual effects and estimate their cross-equation correlation. In that case, neither the procedure of Butler and Moffitt (1982) nor that of Rabe-Hesketh et al. (2005) can be used, as both procedures only account for a single factor of individual effects. This is where our approach comes into play. As mentioned earlier, in order to apply the approach of Section 2, the models are assumed to be autoregressive of first order, and their idiosyncratic errors are assumed to be serially independent conditional on the regressors and the individual effects.

For each model, we show the expressions of f_1 and f_2 that form the core likelihood (cf. Table 1). These expressions enter equation (2.8) evaluated at $a_m \sigma_{a_1} \sqrt{2(1-\rho_{\mathbf{a}}^2)}$ and $a_p \sigma_{a_2} \sqrt{2(1-\rho_{\mathbf{a}}^2)}$ where a_m and a_p are the abscissas of the G-H quadrature. We also derive for each model the APEs.

3.1 Dynamic bivariate probit

Let

$$
\begin{aligned}
y_{1it} &= \mathbf{1}[\rho_{11} y_{1i,t-1} + \rho_{12} y_{2i,t-1} + \boldsymbol{\gamma}_1' \mathbf{z}_{1it} + c_{1i} + u_{1it} > 0], \\
y_{2it} &= \mathbf{1}[\rho_{21} y_{1i,t-1} + \rho_{22} y_{2i,t-1} + \boldsymbol{\gamma}_2' \mathbf{z}_{2it} + c_{2i} + u_{2it} > 0],
\end{aligned}
\tag{3.1}
$$

where we use a similar notation to that of equation (2.1),[5] and $\mathbf{1}[\ ]$ denotes the indicator function. Under the assumptions of subsection 2.1 and if $\mathbf{u}_{it} | \mathbf{y}_{i,t-1}, \mathbf{z}_{it}, \mathbf{c}_i \sim$

[5]If we used the notation in terms of latent variables, they would equal the expressions in squared brackets. For simplicity, we drop $\mathbf{g}$ from the notation in the examples. However, we could still use some function of the autoregressive terms, e.g. interactions with the exogenous regressors, provided that $\mathbf{g}$ appears in both equation (2.1) and (2.3).

Table 1: Core likelihood expressions of the 2-equation dynamic panel data models†

Model		$f_1(\ldots\vert\ldots,a_p,a_m,\boldsymbol{\theta},\boldsymbol{\delta})$	$f_2(\ldots\vert\ldots,a_p,\boldsymbol{\theta},\boldsymbol{\delta})$
1) Dynamic bivariate probit			
	$\rho_{\mathbf{u}}\neq 0$	$\Phi_2\big[(2y_{1it}-1)(M_{1it}+a_m\ldots),(2y_{2it}-1)\big(M_{2it}+a_p\ldots\big),(2y_{1it}-1)(2y_{2it}-1)\rho_{\mathbf{u}}\big]$	1
	$\rho_{\mathbf{u}}=0$	$\Phi_1[(2y_{1it}-1)(M_{1it}+a_m\ldots)]$	$\Phi_1[(2y_{2it}-1)(M_{2it}+a_p\ldots)]$
2) Dynamic bivariate tobit			
	$\rho_{\mathbf{u}}\neq 0$	$\left[\Phi_1\left(\frac{-M_{2it}-a_p\cdots-\rho_{\mathbf{u}}\frac{\sigma_{u_2}}{\sigma_{u_1}}(y_{1it}-M_{1it}-a_m\cdots)}{\sigma_{u_2}\sqrt{1-\rho_{\mathbf{u}}^2}}\right)\frac{\phi_1\left(\frac{y_{1it}-M_{1it}-a_m\cdots}{\sigma_{u_1}}\right)}{\sigma_{u_1}}\right]^{d_{1it}(1-d_{2it})}$ $\left[\kappa\phi_1\left(\frac{y_{1it}-M_{1it}-a_m\cdots-\rho_{\mathbf{u}}\frac{\sigma_{u_1}}{\sigma_{u_2}}(y_{2it}-M_{2it}-a_p\cdots)}{\sigma_{u_1}\sqrt{1-\rho_{\mathbf{u}}^2}}\right)\right]^{d_{1it}d_{2it}}$ $\left[\Phi_1\left(\frac{-M_{1it}-a_m\cdots-\rho_{\mathbf{u}}\frac{\sigma_{u_1}}{\sigma_{u_2}}(y_{2it}-M_{2it}-a_p\cdots)}{\sigma_{u_1}\sqrt{1-\rho_{\mathbf{u}}^2}}\right)\right]^{(1-d_{1it})d_{2it}}$ $\left[\Phi_2\left(\frac{-M_{1it}-a_m\cdots}{\sigma_{u_1}},\frac{-M_{2it}-a_p\cdots}{\sigma_{u_2}},\rho_{\mathbf{u}}\right)\right]^{(1-d_{1it})(1-d_{2it})}$	$\left[\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_p\cdots}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{d_{1it}d_{2it}}\left[\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_p\cdots}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{(1-d_{1it})d_{2it}}$
	$\rho_{\mathbf{u}}=0$	$\left[\frac{\phi_1\left(\frac{y_{1it}-M_{1it}-a_m\cdots}{\sigma_{u_1}}\right)}{\sigma_{u_1}}\right]^{d_{1it}(1-d_{2it})}\left[\kappa_{\vert\rho_{\mathbf{u}}=0}\phi_1\left(\frac{y_{1it}-M_{1it}-a_m\cdots}{\sigma_{u_1}}\right)\right]^{d_{1it}d_{2it}}$ $\left[\Phi_1\left(\frac{-M_{1it}-a_m\cdots}{\sigma_{u_1}}\right)\right]^{(1-d_{1it})d_{2it}}\left[\Phi_1\left(\frac{-M_{1it}-a_m\cdots}{\sigma_{u_1}}\right)\right]^{(1-d_{1it})(1-d_{2it})}$	$\left[\Phi_1\left(\frac{-M_{2it}-a_p\cdots}{\sigma_{u_2}}\right)\right]^{d_{1it}(1-d_{2it})}\left[\Phi_1\left(\frac{-M_{2it}-a_p\cdots}{\sigma_{u_2}}\right)\right]^{(1-d_{1it})(1-d_{2it})}$ $\left[\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_p\cdots}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{d_{1it}d_{2it}}\left[\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_p\cdots}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{(1-d_{1it})d_{2it}}$
3) Dynamic type 2 tobit			
	$\rho_{\mathbf{u}}\neq 0$	$\left[\Phi_1\left(\frac{M_{1it}+a_m\cdots+\frac{\rho_{\mathbf{u}}}{\sigma_{u_2}}(y_{2it}-M_{2it}-a_p\cdots)}{\sqrt{1-\rho_{\mathbf{u}}^2}}\right)\right]^{y_{1it}}[\Phi_1(-M_{1it}-a_m\ldots)]^{1-y_{1it}}$	$\left[\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_p\cdots}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{y_{1it}}$
	$\rho_{\mathbf{u}}=0$	$[\Phi_1(M_{1it}+a_m\ldots)]^{y_{1it}}[\Phi_1(-M_{1it}-a_m\ldots)]^{1-y_{1it}}$	$\left[\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_p\cdots}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{y_{1it}}$
4) Dynamic type 3 tobit			
	$\rho_{\mathbf{u}}\neq 0$	$\left[\kappa\phi_1\left(\frac{y_{1it}-M_{1it}-a_m-\rho_{\mathbf{u}}\frac{\sigma_{u_1}}{\sigma_{u_2}}(y_{2it}-M_{2it}-a_p)}{\sigma_{u_1}\sqrt{1-\rho_{\mathbf{u}}^2}}\right)\right]^{d_{1it}}\left[\Phi_1\left(\frac{-M_{1it}-a_m}{\sigma_{u_1}}\right)\right]^{1-d_{1it}}$	$\left[\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_p}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{d_{1it}}$
	$\rho_{\mathbf{u}}=0$	$\left[\kappa_{\vert\rho_{\mathbf{u}}=0}\phi_1\left(\frac{y_{1it}-M_{1it}-a_m}{\sigma_{u_1}}\right)\right]^{d_{1it}}\left[\Phi_1\left(\frac{-M_{1it}-a_m}{\sigma_{u_1}}\right)\right]^{1-d_{1it}}$	$\left[\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_p}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{d_{1it}}$

†Notes: $a_m\ldots=a_m\sigma_{a_1}\sqrt{2(1-\rho_{\mathbf{a}}^2)}$; $a_p\ldots=a_p\sigma_{a_2}\sqrt{2(1-\rho_{\mathbf{a}}^2)}$; ϕ_1 and Φ_1 are the univariate standard normal pdf and cdf; Φ_2 is the bivariate standard normal cdf; $\kappa=(1/\sigma_{u_1})(1-\rho_{\mathbf{u}}^2)^{-1/2}$ and $\kappa_{\vert\rho_{\mathbf{u}}=0}=1/\sigma_{u_1}$. y_{1it} and y_{2it} are binary in model 1 and y_{1it} is binary in model 3; y_{1it} and y_{2it} are continuous in a certain interval in models 2 and 4, so is y_{2it} in model 3; $d_{1it}=\mathbf{1}[y_{1it}>0]$ and $d_{2it}=\mathbf{1}[y_{2it}>0]$ in models 2 and 4.

$\boldsymbol{Normal}(\mathbf{0}, \mathbf{\Sigma_u})$ with $\mathbf{\Sigma_u} = \begin{pmatrix} 1 & \\ \rho_{\mathbf{u}} & 1 \end{pmatrix}$, the core likelihood is given by

$$l_i(\boldsymbol{\theta}, \boldsymbol{\delta}|\mathbf{c}) = \prod_{t=1}^{T} \Phi_2[(2y_{1it} - 1)(M_{1it} + a_{1i}), (2y_{2it} - 1)(M_{2it} + a_{2i}), (2y_{1it} - 1)(2y_{2it} - 1)\rho_{\mathbf{u}}], \tag{3.2}$$

where Φ_2 denotes the bivariate standard normal cdf, and M_{1it} and M_{2it} are defined as

$$\begin{aligned} M_{1it} &\equiv \rho_{11}y_{1i,t-1} + \rho_{12}y_{2i,t-1} + \boldsymbol{\gamma}_1'\mathbf{z}_{1it} + \alpha_{10} + \alpha_{11}y_{1i0} + \alpha_{12}\mathbf{z}_{1i}, \\ M_{2it} &\equiv \rho_{21}y_{2i,t-1} + \rho_{22}y_{2i,t-1} + \boldsymbol{\gamma}_2'\mathbf{z}_{2it} + \alpha_{20} + \alpha_{21}y_{2i0} + \alpha_{22}\mathbf{z}_{2i}. \end{aligned} \tag{3.3}$$

In the general case where $\rho_{\mathbf{u}} \neq 0$, no simple factorization exists for Φ_2. It results that f_1 is equal to the expression of Φ_2 in equation (3.2) and f_2 is equal to 1. In the special case $\rho_{\mathbf{u}} = 0$, Φ_2 can be factorized into $\Phi_1[(2y_{1it} - 1)(M_{1it} + a_{1i})]\Phi_1[(2y_{2it} - 1)(M_{2it} + a_{2i})]$, where Φ_1 denotes the univariate standard normal cdf, so that f_1 and f_2 equal respectively the first and second component of the factorization.

3.2 Dynamic bivariate tobit

Let

$$\begin{aligned} y_{1it} &= \max[0, \rho_{11}y_{1i,t-1} + \rho_{12}y_{2i,t-1} + \boldsymbol{\gamma}_1'\mathbf{z}_{1it} + c_{1i} + u_{1it}], \\ y_{2it} &= \max[0, \rho_{21}y_{1i,t-1} + \rho_{22}y_{2i,t-1} + \boldsymbol{\gamma}_2'\mathbf{z}_{2it} + c_{2i} + u_{2it}], \end{aligned} \tag{3.4}$$

where $\mathbf{u}_{it}|\mathbf{y}_{i,t-1}, \mathbf{z}_{it}, \mathbf{c}_i \sim \boldsymbol{Normal}(\mathbf{0}, \mathbf{\Sigma_u})$ with $\mathbf{\Sigma_u} = \begin{pmatrix} \sigma_{u_1}^2 & \\ \rho_{\mathbf{u}}\sigma_{u_1}\sigma_{u_2} & \sigma_{u_2}^2 \end{pmatrix}$. As Wooldridge (2002; 2005a) notes, this model applies to *corner solution* outcomes as opposed to *censored data*. The main difference is that the lagged *observed* dependent variables appear on the right-hand sides of equation (3.4), while the lagged *latent* variables would appear as regressors in models suitable for censored data.[6] The core likelihood of this model is given

[6] Even though corner solution outcomes and censored data are conceptually different, the static variant of the model applies similarly in both situtations (see Wooldridge, 2002, chapter 16).

by

$$l_i(\boldsymbol{\theta}, \boldsymbol{\delta}|\mathbf{c}) = \prod_{t=1}^{T} \left[\kappa\phi_1\left(\frac{y_{1it} - M_{1it} - a_{1i} - \rho_{\mathbf{u}} \frac{\sigma_{u_1}}{\sigma_{u_2}} (y_{2it} - M_{2it} - a_{2i})}{\sigma_{u_1}\sqrt{1-\rho_{\mathbf{u}}^2}} \right) \frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_{2i}}{\sigma_{u_2}}\right)}{\sigma_{u_2}} \right]^{d_{1it}d_{2it}}$$
$$\left[\Phi_1\left(\frac{-M_{2it} - a_{2i} - \rho_{\mathbf{u}} \frac{\sigma_{u_2}}{\sigma_{u1}} (y_{1it} - M_{1it} - a_{1i})}{\sigma_{u_2}\sqrt{1-\rho_{\mathbf{u}}^2}} \right) \frac{\phi_1\left(\frac{y_{1it}-M_{1it}-a_{1i}}{\sigma_{u_1}}\right)}{\sigma_{u_1}} \right]^{d_{1it}(1-d_{2it})}$$
$$\left[\Phi_1\left(\frac{-M_{1it} - a_{1i} - \rho_{\mathbf{u}} \frac{\sigma_{u_1}}{\sigma_{u_2}} (y_{2it} - M_{2it} - a_{2i})}{\sigma_{u_1}\sqrt{1-\rho_{\mathbf{u}}^2}} \right) \frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_{2i}}{\sigma_{u_2}}\right)}{\sigma_{u_2}} \right]^{(1-d_{1it})d_{2it}}$$
$$\left[\Phi_2\left(\frac{-M_{1it} - a_{1i}}{\sigma_{u_1}}, \frac{-M_{2it} - a_{2i}}{\sigma_{u_2}}, \rho_{\mathbf{u}} \right) \right]^{(1-d_{1it})(1-d_{2it})}, \tag{3.5}$$

where $\kappa = \frac{1}{\sigma_{u_1}\sqrt{1-\rho_{\mathbf{u}}^2}}$, ϕ_1 denotes the univariate standard normal pdf, M_{1it} and M_{2it} are defined in equation (3.3) and $d_{1it} = \mathbf{1}[y_{1it} > 0]$ and $d_{2it} = \mathbf{1}[y_{2it} > 0]$. It results from equation (3.5) that f_1 is equal to the product of standard normal pdfs and cdfs that are operated at a_{1i} and a_{2i}, and f_2 is the product of such functions that are operated only at a_{2i}. As with the bivariate probit, if we assume that the cross-equation correlation operates only through the individual effects, i.e. $\rho_{\mathbf{u}} = 0$, then equation (3.5) can be factorized further.

3.3 Dynamic sample selection models

We consider in this analysis the two most frequently used sample selection models, namely the type 2 and type 3 tobit.

The dynamic type 2 tobit can be described as

$$\begin{aligned} y_{1it} &= \mathbf{1}[\rho_{11}y_{1i,t-1} + \rho_{12}y_{2i,t-1} + \boldsymbol{\gamma}_1'\mathbf{z}_{1it} + c_{1i} + u_{1it} > 0] \\ y_{2it} &= y_{1it}[\rho_{21}y_{1i,t-1} + \rho_{22}y_{2i,t-1} + \boldsymbol{\gamma}_2'\mathbf{z}_{2it} + c_{2i} + u_{2it}], \end{aligned} \tag{3.6}$$

where $\mathbf{u}_{it}|\mathbf{y}_{i,t-1}, \mathbf{z}_{it}, \mathbf{c}_i \sim \boldsymbol{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u}})$ with $\boldsymbol{\Sigma}_{\mathbf{u}} = \begin{pmatrix} 1 & \\ \rho_{\mathbf{u}}\sigma_{u_2} & \sigma_{u_2}^2 \end{pmatrix}$. Like the dynamic bivariate tobit, this model applies to corner solution outcomes as the lagged observed dependent variables appear on the right-hand sides of equation (3.6). We can write the core likelihood of the model as

$$l_i(\boldsymbol{\theta},\boldsymbol{\delta}|\mathbf{c}) = \prod_{t=1}^{T}\left[\Phi_1\left(\frac{M_{1it}+a_{1i}+\frac{\rho_{\mathbf{u}}}{\sigma_{u_2}}(y_{2it}-M_{2it}-a_{2i})}{\sqrt{1-\rho_{\mathbf{u}}^2}}\right)\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_{2i}}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{y_{1it}}$$
$$[\Phi_1(-M_{1it}-a_{1i})]^{1-y_{1it}}, \tag{3.7}$$

where the notation stands as before. We can easily see from equation (3.7) that f_1 is equal to the product of the two univariate cdfs, and f_2 is equal to the expression of the univariate pdf.

The dynamic type 3 tobit can be written as

$$\begin{aligned} y_{1it} &= \max[0, \rho_{11}y_{1i,t-1}+\rho_{12}y_{2i,t-1}+\boldsymbol{\gamma}_1'\mathbf{z}_{1it}+c_{1i}+u_{1it}] \\ y_{2it} &= d_{1it}[\rho_{21}y_{1i,t-1}+\rho_{22}y_{2i,t-1}+\boldsymbol{\gamma}_2'\mathbf{z}_{2it}+c_{2i}+u_{2it}], \end{aligned} \tag{3.8}$$

where d_{1it} is defined as in the dynamic bivariate tobit model and $\mathbf{u}_{it}|\mathbf{y}_{i,t-1},\mathbf{z}_{it},\mathbf{c}_i \sim$ $\boldsymbol{Normal}(\mathbf{0},\boldsymbol{\Sigma}_{\mathbf{u}})$ with $\boldsymbol{\Sigma}_{\mathbf{u}} = \begin{pmatrix} \sigma_{u_1}^2 & \\ \rho_{\mathbf{u}}\sigma_{u_1}\sigma_{u_2} & \sigma_{u_2}^2 \end{pmatrix}$. The core likelihood is obtained as

$$l_i(\boldsymbol{\theta},\boldsymbol{\delta}|\mathbf{c}) = \prod_{t=1}^{T}\left[\kappa\phi_1\left(\frac{y_{1it}-M_{1it}-a_{1i}-\rho_{\mathbf{u}}\frac{\sigma_{u_1}}{\sigma_{u_2}}(y_{2it}-M_{2it}-a_{2i})}{\sigma_{u_1}\sqrt{1-\rho_{\mathbf{u}}^2}}\right)\frac{\phi_1\left(\frac{y_{2it}-M_{2it}-a_{2i}}{\sigma_{u_2}}\right)}{\sigma_{u_2}}\right]^{d_{1it}}$$
$$\left[\Phi_1\left(\frac{-M_{1it}-a_{1i}}{\sigma_{u_1}}\right)\right]^{1-d_{1it}} \tag{3.9}$$

with the same notation as before. It results that f_1 is the product of the univariate cdf and the univariate pdf that is evaluated at both a_{1i} and a_{2i}, and f_2 is the expression of the univariate pdf evaluated only at a_{2i}.

4 Application

In order to illustrate our approach, we estimate the dynamics of innovation using a two-equation panel VAR of the first order. The two dependent variables are the intensity of innovation input defined as the ratio of total innovation expenditures over total (domestic) sales, and the intensity of innovation output defined as the share of total sales accounted for by sales of new or significantly improved products.[7] Total innovation expenditures comprise, in addition to R&D, the purchase of rights and licenses to use external technology and the purchase of advanced machinery and computer hardware devoted to the implementation of product and process innovations. We take a logit transformation of both dependent variables so as to make them lie with the whole set of real numbers.[8]

The data are collected by the *Centraal Bureau voor de Statistiek* (CBS) and stem from five waves of the Dutch Community Innovation Survey, namely CIS 2 (1994-1996), CIS 2.5 (1996-1998), CIS 3 (1998-2000), CIS 3.5 (2000-2002) and CIS 4 (2002-2004), merged with data from the Production Survey (PS). The enterprises of this application belong to the manufacturing sector (NACE 15.1-37.2), and have at least ten employees and positive sales at the end of each period covered by the innovation survey. We only include firms for which total innovation expenditures count for no more than 50% of total sales, the full set of variables included in the panel VAR are observed, and total innovation expenditures and the share of innovative sales are positive at all time periods.[9]

4.1 Model specification

The explanatory variables used to explain innovation input intensity and the share of innovative are listed as follows. Employment, as a measure of firm (absolute) size and market share, as a measure of relative size, are chosen on the grounds of Schumpeter (1942) hypothesis. Employment is given by the number of employees (in log) of the firm, market share is defined as the ratio of the sales of an enterprise over the total sales (in log) of the 3-digit industry it belongs to.[10] The variables of technology push and demand pull are the share of firms within a 3-digit industry for which process- and product-oriented effects of innovation are deemed very important on a 0-3 Likert scale. The process-oriented effects include improvement of flexibility of production, increase of capacity of production and reduction of labor costs, materials and energy per unit output, and the product-oriented effects include increase of range of goods, improvement of quality of

[7] New or significantly improved products are so defined only with respect to other products of the firm, and not necessarily with respect to other products of the market.

[8] The share of innovative sales may take on the value one, e.g. for newly-established firms, in the sample. The ones are replaced by some ϵ between the second largest value of the variable and one.

[9] See Raymond et al. (2009) for more details on the data.

[10] Total sales of a 3-digit industry is obtained by adding up the sales of all the firms in our sample that belong to that industry after multiplying them by the appropriate raising factor.

goods, market share increase and new market entrance. Technology push and demand are included according to Schumpeter's (1934) and Schmookler's (1966) tradition respectively. Three dummy variables (D_{Q_2}-D_{Q_4}) of distance to technological frontier are included in the specification. They take on the value one if the technology gap variable, defined as the difference between the largest (labor) productivity (within each 3-digit industry) and the productivity of the firm, lies respectively between the first and second quartile, the second and the third quartile, and above the third quartile. The dummy variable D_{Q_1}, which takes on the value one if the technology gap lies below or at the first quartile, is used as the reference. Finally, three dummy variables of industry category defined according to the OECD (2007) technology based classification, and three time dummy variables are included in the specification, with the low-tech category and the time period 1996-1998 being the references.

Table 2 shows descriptive statistics of the dependent and explanatory variables of the VAR. For instance, we read that total innovation expenditures represent on average 3.5% of total sales, with a median of 1.8%. Innovative sales represent on average 30% (a median of 22%), the firms in our sample have on average 272 employees and have an average market share of 0.73%. The percentage of firms for which process-oriented and product-oriented effects of innovation are important is 36% and 64% respectively.

Table 2: Descriptive statistics of the dependent and explanatory variables: Estimation sample

Variable	Mean	Std. Dev.	Q_1	Median	Q_3
Innov. input intensity	0.035	0.052	0.007	0.018	0.039
Share of innov. sales	0.302	0.236	0.100	0.221	0.400
Employment	272	956	63	110	216
Market share (%)	0.733	2.312	0.044	0.126	0.489
Techpush	0.359	0.084	0.281	0.367	0.403
Demand pull	0.635	0.085	0.587	0.646	0.677

4.2 Estimation

Using the same notation as in Sections 2 and 3, we write the bivariate VAR(1) as

$$
\begin{aligned}
y_{1it} &= \rho_{11} y_{1i,t-1} + \rho_{12} y_{2i,t-1} + \boldsymbol{\gamma}_1' \mathbf{z}_{1it} + c_{1i} + u_{1it}, \\
y_{2it} &= \rho_{21} y_{1i,t-1} + \rho_{22} y_{2i,t-1} + \boldsymbol{\gamma}_2' \mathbf{z}_{2it} + c_{2i} + u_{2it}.
\end{aligned} \tag{4.1}
$$

If the assumptions of subsection 2.1 are satisfied and $\mathbf{u}_{it}|\mathbf{y}_{i,t-1}, \mathbf{z}_{it}, \mathbf{c}_i \sim \boldsymbol{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{u})$ with $\boldsymbol{\Sigma}_\mathbf{u} = \begin{pmatrix} \sigma^2_{u_1} & \\ \rho_\mathbf{u}\sigma_{u_1}\sigma_{u_2} & \sigma^2_{u_2} \end{pmatrix}$, the core likelihood of the bivariate VAR can be written, after replacing $\mathbf{c}_i$ by equation (2.3), as

$$l_i(\boldsymbol{\theta}, \boldsymbol{\delta}|\mathbf{c}) = \prod_{t=1}^{T} \frac{1}{\sigma_{u_1}\sqrt{1-\rho^2_\mathbf{u}}}\phi_1\left(\frac{y_{1it} - M_{1it} - a_{1i} - \rho_\mathbf{u}\frac{\sigma_{u_1}}{\sigma_{u_2}}(y_{2it} - M_{2it} - a_{2i})}{\sigma_{u_1}\sqrt{1-\rho^2_\mathbf{u}}}\right) \frac{1}{\sigma_{u_2}}\phi_1\left(\frac{y_{2it} - M_{2it} - a_{2i}}{\sigma_{u_2}}\right) \tag{4.2}$$

where the notation stands as before. It can be easily seen that f_1 and f_2 are equal to the first and second univariate pdf respectively.

4.3 Results

Table 3 shows the ML estimation results of the VAR model of innovation. There is evidence of dynamics in the process of innovation in the sense the all four coefficients of the lagged dependent variables are positive and statistically significant even after controlling for individual effects and conditioning out the initial conditions. Employment and market share do not seem to matter in either equation. Technology push has a negative and significant effect in both innovation input and innovation output, while demand pull has a positive and significant role only in innovation output. The results seem to indicate that the farther away the firm from the technological frontier, the larger innovation expenditures so as to catch up with those that are closer to the frontier. This is in accordance with the literature on growth convergence. The share of innovative sales is, however, is more important for firms that are very close to the frontier even though the evidence is somewhat mixed. Finally, unobserved heterogeneity plays a significant role in the innovation process, at least in innovation output, as shown by the highly significant coefficient σ_{a_2}, and both equations are significantly related as shown by $\rho_{\boldsymbol{a}}$ and $\rho_\mathbf{u}$.

5 Conclusion

We have generalized in this study (Wooldridge, 2005a)'s approach to solve the initial conditions problem in multiple equation dynamic models. We have shown that the likelihood

Table 3: ML estimates of a panel VAR with individual effects: Unbalanced panel from Dutch manufacturing, CIS 2, CIS 2.5, CIS 3, CIS 3.5 and CIS 4[‡]

Variable	Coefficient	(Std. Err.)
	Individual effects	
	Innovation input intensity (in logit)	
(Innov. input intensity)$_{-1}$	0.329**	(0.023)
(Share of innov. sales)$_{-1}$	0.065**	(0.015)
(Employment)$_{-1}$ (in log)	-0.046	(0.029)
(Market share)$_{-1}$ (in log)	0.019	(0.019)
(Technology push)$_{-1}$	-1.038**	(0.332)
(Demand pull)$_{-1}$	0.169	(0.321)
Distance to frontier		
D_{Q_2}	0.047	(0.062)
D_{Q_3}	0.278**	(0.067)
D_{Q_4}	0.641**	(0.076)
(Innov. input intensity)$_0$	0.071**	(0.016)
	Share of innovative sales (in logit)	
(Innov. input intensity)$_{-1}$	0.132**	(0.024)
(Share of innov. sales)$_{-1}$	0.243**	(0.026)
(Employment)$_{-1}$ (in log)	0.034	(0.038)
(Market share)$_{-1}$ (in log)	0.019	(0.025)
(Technology push)$_{-1}$	-0.986*	(0.428)
(Demand pull)$_{-1}$	1.020*	(0.412)
Distance to frontier		
D_{Q_2}	-0.166*	(0.077)
D_{Q_3}	-0.144†	(0.083)
D_{Q_4}	-0.127	(0.094)
(Share of innov. sales)$_0$	0.052**	(0.019)
	Extra parameters	
Individual effects		
σ_{a_1}	0.980	(12.973)
σ_{a_2}	2.205**	(1.253)
$\rho_{\boldsymbol{a}}$	0.949**	(0.002)
Idiosyncratic errors		
σ_{u_1}	1.111**	(0.020)
σ_{u_2}	1.271**	(0.043)
$\rho_{\mathbf{u}}$	0.120**	(0.022)
# observations	2262	
Log-likelihood	-10456.415	

[‡]Three dummies of category of industry, three time dummies and an intercept are included in each equation of the model.

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

function can be written and evaluated in a general manner for a wide class of dynamic models. This has been achieved by using the multivariate normal distribution properties of the individual effects which enables us to use multiple-step Gauss-Hermite quadrature.

We have argued that evaluation of the likelihood of such models is rather straightforward, and should be programmed into basic econometric package (e.g. Stata) commands. We have given leading examples of models that belong the mentioned class, and illustrated our approach using a dynamic model (VAR) of innovation.

In order to evaluate the performance of our approach, we plan to carry out in the near future a Monte Carlo analysis comparing our approach to other numerical methods such as simulated maximum likelihood or Markov Chain Monte Carlo. We also plan to draw a table similar to Table 1 where the three types of average partial effects (when relevant) are to be reported.

References

Amemiya, T. (1984), "Tobit models: A survey," *Journal of Econometrics,* 24, 3–61.

Butler, J. S., and Moffitt, R. (1982), "A computationally efficient quadrature procedure for the one-factor multinomial probit model," *Econometrica,* 50, 761–764.

Chamberlain, G. (1980), "Analysis of covariance with qualitative data," *Review of Economic Studies,* 47, 225–238.

——— (1984), "Panel data," In *Handbook of Econometrics,* ed. Z. Griliches and M. Intriligator, vol. 2, Amsterdam: North Holland: Elsevier Science Publishers chapter 22, pp. 1247–1318.

Flaig, G., and Stadler, M. (1994), "Success breeds success. The dynamics of the innovation process," *Empirical Economics,* 19, 55–68.

Gouriéroux, Christian, Monfort, Alain, Renault, Eric, and Trognon, Alain (1987), "Generalised residuals," *Journal of Econometrics,* 34, 5–32.

Heckman, J. J. (1981), "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process," In *Structural Analysis of Discrete Data with Econometric Applications,* ed. C. F. Manski and D. McFadden, Cambridge, MA: MIT Press pp. 179–195.

OECD (2007) *OECD Science, Technology and Industry Scoreboard 2007: Innovation and Performance in the Global Economy,* Paris: OECD; 2007 edition.

Orme, Chris D (2001), "Two-step inference in dynamic non-linear panel data models," University of Manchester, Mimeo.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2005), "Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects," *Journal of Econometrics,* 128, 301–323.

Raymond, W. (2007), "The dynamics of innovation and firm performance: An econometric panel data analysis," PhD Dissertation, Maastricht University.

Raymond, W., Mohnen, P., Palm, F., and Schim van der Loeff, S. (2009), "Innovative sales, R&D and total innovation expenditures: Panel evidence on their dynamics," CESifo Working Paper No. 2716.

Schmookler, J. (1966) *Invention and Economic Growth,* Cambridge: Harvard University Press.

Schumpeter, J. A. (1934) *The Theory of Economic Development,* Cambridge, MA: Harvard University Press.

—— (1942) *Capitalism, Socialism and Democracy,* New York: Harper and Brothers.

Wooldridge, J. M. (2002) *Econometric Analysis of Cross-Section and Panel Data,* Cambridge, MA: MIT Press.

—— (2005a), "Simple solutions to the initial conditions problem in dynamic nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics,* 20, 39–54.

—— (2005b), "Unobserved heterogeneity and estimation of average partial effects," In *Indentification and Inference for Econometric Models,* ed. D. W. K. Andrews and J. H. Stock, Cambridge: Cambridge University Press pp. 27–55.