Non-Causality in Bivariate Binary Panel Data

Rocco Mosconi* Raffaello Seri†

Abstract

In this paper we develop a suitable dynamic discrete time bivariate probit model, in which the conditions for Granger non-causality may be represented and tested. The conditions for simultaneous independence are also worked out. The model is extended in order to allow for covariates, representing individual as well as time heterogeneity. The proposed model may be estimated by Maximum Likelihood; Granger non-causality, as well as simultaneous independence may be tested by Likelihood Ratio tests. A specialized version of the model, aimed at testing Granger non-causality with bivariate survival data is also discussed. The proposed tests are illustrated using data concerning the relation between marriage and fertility timing in a sample of 266 American women and the adoption of two interrelated technological innovations by 552 Italian metalworking plants.

Acknowledgments This paper has benefited at different stages by discussions with Per Kragh Andersen, Clelia Di Serio, Clive Granger, Philip Hougaard, Søren Johansen and Niels Keiding. The authors are particularly indebted to Fabio Sartori, who significantly contributed to the development of the main ideas presented in the paper. The usual disclaimer applies.

1 Introduction

The epistemological status of the statistical-probabilistic notion of causality based on predictability is still a matter of profound controversy among philosophers and methodologists (see Geweke, 1984). This notion fits, in a probabilistic sense, two key aspects of causation: the systematic conjunction of cause and effect, and the time precedence of the cause with respect to the effect. Nonetheless, it fails to account for what probably is the deepest, though empirically less helpful, aspect, *i.e.* the idea that the cause "forces" or "produces" the effect. Despite these limitations, the notion of causality based on predictability proved to be a valuable tool for applied research thanks to its operational usefulness

*Dipartimento di Economia e Produzione, Politecnico di Milano, P.za Leonardo da Vinci 32, 20133 MILANO, ITALY, email: rocco.mosconi@polimi.it

†CREST-LFA, Timbre J320, 15 bd Gabriel Péri, 92245 MALAKOFF CEDEX, FRANCE, email: seri@ensae.fr

in the construction, estimation, interpretation and application of econometric models.

In a general setting (see Florens and Fougère, 1996), a mathematically rigorous definition of non-causality based on predictability requires the specification of the stochastic process to be predicted, the available information set, and the reduced information set. Although several generalizations exist, we will briefly review here the concept of discrete time one step ahead strong non-causality (the terminology is drawn from Florens and Fougère,1996). Here *one step ahead* (as opposed to *global*) is referred to the prediction horizon, whereas *strong* (as opposed to *weak*) means that the focus is on predicting the whole distribution, rather than just the mean. Notice that Granger's (1969) original definition is stated in terms of the mean. Chamberlain (1982) and Florens and Mouchart (1982) propose the definition involving the whole distribution (see also Granger, 1988).

Let $\left\{Y_t = \left(Y_t^1, Y_t^2\right), t \in I \subseteq \mathbb{N} = \{0, 1, \ldots\}\right\}$, or $\{Y_t\}$ for short,[1] be a discrete time vector stochastic process. This means that, for any positive integer $t \in I$, Y_t is a vector random variable on a probability space $(\Omega, \mathcal{A}, P)$. P is an element of a family of probability measures, and the statistical problem of non-causality is to test whether P satisfies non-causality conditions.

The available information is described by $\mathcal{F}_t$, which is a sub-σ-field of $\mathcal{A}$. It is assumed that the family $\{\mathcal{F}_t, t \in I\}$, briefly $\{\mathcal{F}_t\}$ is a *filtration*, *i.e.* $\mathcal{F}_t \subset \mathcal{F}_{t'}$ for $t \leq t'$. For simplicity, we will assume here that $\mathcal{F}_t$ is the *canonical filtration* associated to the multivariate stochastic process $\{(Y_t, X_t)\} = \left\{\left(Y_t^1, Y_t^2, X_t\right)\right\}$,[2] where each of $\left\{Y_t^1\right\}$, $\left\{Y_t^2\right\}$ and $\{X_t\}$ may either be scalar or vector processes. This of course implies that $\{Y_t\}$ is *adapted* to $\{\mathcal{F}_t\}$, *i.e.* Y_t is $\mathcal{F}_t$-measurable for any $t \in I$.

The reduced information set is represented by the filtrations $\left\{\mathcal{G}_t^1\right\}$ and $\left\{\mathcal{G}_t^2\right\}$. We will assume that $\left\{\mathcal{G}_t^1\right\}$ is the canonical filtration of $\left\{\left(Y_t^1, X_t\right)\right\}$, and $\left\{\mathcal{G}_t^2\right\}$ is the canonical filtration of $\left\{\left(Y_t^2, X_t\right)\right\}$, which implies that Y_t^1 is $\mathcal{G}_t^1$-measurable and Y_t^2 is $\mathcal{G}_t^2$-measurable for any $t \in I$. Let then $\left\{\mathcal{Y}_t^1\right\}$, $\left\{\mathcal{Y}_t^2\right\}$ and $\{\mathcal{Y}_t\}$ be the canonical filtration associated to the processes $\left\{Y_t^1\right\}$, $\left\{Y_t^2\right\}$ and $\{Y_t\}$ respectively. Notice that $\mathcal{Y}_t^1 \subseteq \mathcal{G}_t^1 \subseteq \mathcal{F}_t, \forall t \in I$, and similarly $\mathcal{Y}_t^2 \subseteq \mathcal{G}_t^2 \subseteq \mathcal{F}_t, \forall t \in I$.

In the paper, we will adopt the following definitions, stated in terms of conditional independence of sub-σ-fields of $\mathcal{A}$ (see Florens and Mouchart, 1982, Appendix, for the relevant results about conditional independence):

Definition 1 - *Strong one step ahead non-causality (Granger non-causality):* $\mathcal{Y}_{t-1}^2$ *does not strongly cause* Y_t^1 *one step ahead, given* $\mathcal{G}_{t-1}^1$, *briefly* $Y^1 \nleftarrow Y^2$, *if*

$$\mathcal{Y}_t^1 \perp \mathcal{Y}_{t-1}^2 \mid \mathcal{G}_{t-1}^1 \qquad \forall t \in I \tag{1}$$

[1] The following notation is used through the paper: $\{Z_t\}$ denotes a stochastic process, Z_t being the value of the process at time t; $\{z_t\}$ and z_t represent the corresponding realizations. Moreover, $\Pr\{z_t \mid w_t\}$ is adopted as a short notation for $\Pr\{Z_t = z_t \mid W_t = w_t\}$.

[2] The *canonical* (or *self exciting*) filtration associated to the process $\{Z_t\}$ defined on $(\Omega, \mathcal{A}, P)$ is a family $\{\mathcal{F}_t\}$ of sub-σ-fields of $\mathcal{A}$, whose element $\mathcal{F}_t$ is the σ-field generated by the family of Z_s, $0 \leq s \leq t$. Intuitively, $\mathcal{F}_t$ embodies the knowledge of the history of $\{Z_t\}$ up to time t.

Similarly, $\mathcal{Y}_{t-1}^1$ does not strongly cause Y_t^2 one step ahead, given $\mathcal{G}_{t-1}^2$, briefly $Y^1 \nrightarrow Y^2$, if

$$\mathcal{Y}_t^2 \perp \mathcal{Y}_{t-1}^1 \mid \mathcal{G}_{t-1}^2 \qquad \forall t \in I \tag{2}$$

Definition 2 - *Strong simultaneous independence: Y_t^1 and Y_t^2 are strongly simultaneously independent given $\{\mathcal{F}_t\}$, briefly $Y^1 \nleftrightarrow Y^2$, if*

$$\mathcal{Y}_t^1 \perp \mathcal{Y}_t^2 \mid \mathcal{F}_{t-1} \qquad \forall t \in I \tag{3}$$

Notice that the term *simultaneous* in the latter definition has exactly the same meaning as *instantaneous* in Geweke (1984) and Granger (1988). A different term is suggested here since, rigorously, Florens and Fougère (1996) observe that one step ahead non-causality in discrete time has an analogue in continuous time when the time distance between "cause" and "effect" goes to zero, a circumstance that they define as *instantaneous causality*. Therefore, in discrete time, they use instantaneous as a synonym for *one step ahead*, while they do not give any definition similar to 3. Moreover, for the simultaneous condition (3), the term dependence is proposed instead of causality (as in Granger, 1988) or feedback (as in Geweke, 1984), since the notion is completely a-directional in nature (not even bi-directional).

In Economics, the notion of non-causality has been mainly used in modelling macroeconomic variables, and hence $\{Y_t^1\}$ and $\{Y_t^2\}$ are usually assumed to be continuous processes, the exogenous processes $\{X_t\}$ are often not included in the information set (so that $\mathcal{G}_t^1$ and $\mathcal{G}_t^2$ do coincide with $\mathcal{Y}_t^1$ and $\mathcal{Y}_t^2$ respectively), and one single realization of $\{Y_t\}$ is observed. In this framework, non-causality is usually tested assuming that $\{Y_t\}$ belongs to the class of Vector ARIMA processes. In microeconometric applications, where the variables are often qualitative and longitudinal data are usually available, the VARIMA framework is not appropriate, and including covariates to account for individual heterogeneity becomes an essential aspect of modelling. Therefore, a set of ad hoc tools has to be developed in order to make the above definitions of non-causality operational.

In this paper the case where $\{Y_t\}$ is a bivariate discrete time binary process will be addressed.[3] This case is shortly discussed in Chamberlain (1982). We will assume that N individual realizations ($i = 1, \ldots, N$) of the process are observed, with $t = 1, \ldots, T$. As we will see, depending on the dynamic structure of the model, N might have to be large with respect to T, but if very simple dynamic structures are assumed, a small N, or even $N = 1$, might be enough if T is large. Notice that, in this case, $\{X_t\}$ is needed to model individual heterogeneity, and might well include some time fixed variables.

In our framework, at any time $t \in \{1, \ldots, T\}$, the state space of $Y_t = \left(Y_t^1, Y_t^2\right)$ is given by the following states: $\{(0,0)\,;(1,0)\,;(0,1)\,;(1,1)\}$.

Essentially the model may be represented by the diagram in Figure 1, where each box represents one of the four states where the process could belong at time $(t-1)$, and the arrows represent the transitions which might occur at time t.

[3] The case where $\{Y_t^1\}$ and $\{Y_t^2\}$ are continuous time counting processes ($t \in I \subset \mathbb{R}^+$) is addressed in the literature (Florens and Fougère, 1996; Schweder, 1970; Aalen, 1987).

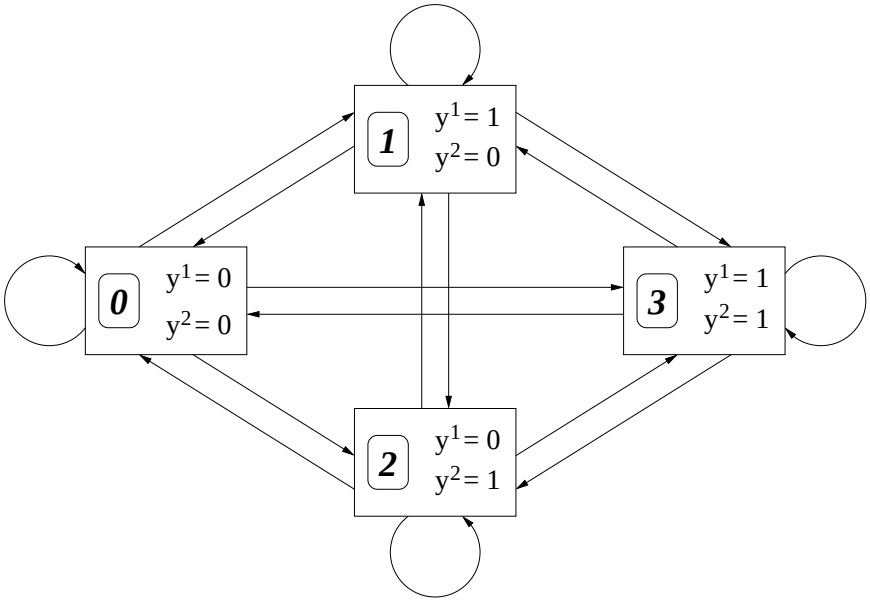


Figure 1: State-Transition Diagram for a Binary Bivariate Markov Model

Let us illustrate how definitions (2)-(3) may be made operational by applying them to a precise stochastic process and information set. To make the simplest possible example, let us restrict the information set to the canonical filtration associated to $\{Y_t\}$, and furthermore make the assumption that $\{Y_t\}$ is a *Markov process* (or *Markov chain*), so that

$$\Pr\{y_t \mid y_{t-1}, \ldots, y_0\} = \Pr\{y_t \mid y_{t-1}\}$$

The most restrictive definition of Markov process requires that the transition probabilities do not vary over time. More specifically, under this assumption the process is defined a Markov chain with *stationary transition probabilities*. Notice that the assumption of stationary transition probabilities alone does exclude any impact of covariates on the transition probabilities. In this simplified framework, the definitions given above specialize as follows:

Definition 3 - *Strong one step ahead non-causality for a Markov chain with stationary transition probabilities: Y^2_{t-1} does not strongly cause Y^1_t one step ahead, given Y^1_{t-1}, if* [4]

$$\Pr\left\{y^1_t \mid y_{t-1}\right\} = \Pr\left\{y^1_t \mid y^1_{t-1}\right\} \qquad \forall t \in \{1, \ldots, T\} \tag{4}$$

[4]The equivalence between (2) and (4) in this framework comes immediately by noticing that, under the Markov assumption and the assumption that the information set $\mathcal{F}_{t-1}$ coincides with $\mathcal{Y}_{t-1}$, the conditional independence statement (2) implies:

$$\Pr\left\{y^1_t, y^2_{t-1} \mid y^1_{t-1}\right\} = \Pr\left\{y^1_t \mid y^1_{t-1}\right\} \Pr\left\{y^2_{t-1} \mid y^1_{t-1}\right\} \qquad \forall t \in \{1, \ldots, T\}$$

Similarly, Y^1_{t-1} does not strongly cause Y^2_t one step ahead, given Y^2_{t-1} if

$$\Pr\left\{y^2_t \mid y_{t-1}\right\} = \Pr\left\{y^2_t \mid y^2_{t-1}\right\} \qquad \forall t \in \{1,\ldots,T\} \tag{5}$$

Definition 4 - *Strong simultaneous independence for a Markov chain with stationary transition probabilities: Y^1_t and Y^2_t are strongly simultaneously independent given Y_{t-1} if*

$$\Pr\left\{y_t \mid y_{t-1}\right\} = \Pr\left\{y^1_t \mid y_{t-1}\right\} \Pr\left\{y^2_t \mid y_{t-1}\right\} \qquad \forall t \in I \tag{6}$$

or equivalently

$$\Pr\left\{y^1_t \mid y^2_t, y_{t-1}\right\} = \Pr\left\{y^1_t \mid y_{t-1}\right\} \qquad \forall t \in I$$

or equivalently

$$\Pr\left\{y^2_t \mid y^1_t, y_{t-1}\right\} = \Pr\left\{y^2_t \mid y_{t-1}\right\} \qquad \forall t \in I$$

The appropriate statistical model where these conditions may be tested is the joint distribution of Y_t given Y_{t-1}. Granger non-causality conditions involve only the marginal distributions of Y^1_t and Y^2_t (conditional on Y_{t-1}), whereas testing for simultaneous independence requires the joint distribution to be fully specified, and compared to the product of the marginal distributions. Notice that, since Y_{t-1}, as well as Y_t, may belong to a finite set of four states, the most general model representing $\Pr\{y_t \mid y_{t-1}\}$ involves 16 parameters, corresponding to the transition probabilities from each of the states in $(t-1)$ to each of the states in t (or some one to one transformation of the transition probabilities). More precisely, since the sum of the transition probabilities for transitions starting from each of the states is equal to 1, just 12 parameters are enough to describe the conditional distribution completely.

The paper is organized as follows. Under the maintained assumption that $\{Y_t\}$ is a Markov chain with stationary transition probabilities, and that the information set is restricted to $\mathcal{Y}_{t-1}$, Section 2 shows how $\Pr\{y_t \mid y_{t-1}\}$ may be represented with no loss of generality by a dynamic bivariate probit model. Within this framework, the restrictions on the parameters implying (4), (5) or (6) are illustrated. Section 3 extends the simple dynamic probit model illustrated in Section 2 in two directions. First, the assumption of stationary transition probabilities is dropped, allowing the transition probabilities to depend on covariates. Then, the Markov assumption is also relaxed, allowing for more complex dynamic structures. Section 4 shows under which conditions the Maximum Likelihood estimates of the parameters of the proposed models, as well as the Likelihood Ratio tests for hypotheses on such parameters, display

---

which in turn implies (4), being

$$\Pr\left\{y^1_t \mid y^1_{t-1}, y^2_{t-1}\right\} = \frac{\Pr\left\{y^1_t, y^2_{t-1} \mid y^1_{t-1}\right\}}{\Pr\left\{y^2_{t-1} \mid y^1_{t-1}\right\}}$$

The same argument holds for the other definitions.

the usual asymptotic properties. Possible problems in finite samples are also illustrated. Section 5 shows how the proposed analysis does specialize when one is interested in a specific four states Markov chain, corresponding to discrete time bivariate survival data. Sections 6 and 7 illustrate the proposed methodology, using respectively data about marriage and fertility timing in a sample of 266 American women and about the adoption of two interrelated technologies by 552 Italian metalworking plants. Section 8 concludes.

2 A Markov Dynamic Bivariate Probit Model for Homogeneous Population

Essentially, the type of data set where we want to check for non-causality consists in observations on the choices of N individuals facing two interacting binary choices in discrete time. To do this, it seems then natural to use, as a statistical model, a dynamic version of a bivariate discrete choice model. The static univariate and multivariate probit model, viewed in a latent regression perspective, is briefly reviewed in Appendix A. In this Section a simple dynamic version of the bivariate probit model is presented,[5] derived under the following assumptions:

- the population is homogeneous (no covariates are introduced);
- the process is Markov (all the information from the history of the process which is relevant for the transition probabilities in t is embodied in the state of the process in $(t-1)$).

Both these simplifying assumptions will be relaxed in Section 3. Notice that the derivation of the dynamic model requires the events to take place only in discrete time. In fact, qualitative variable models as logit and probit are sensitive to the length of grouping: when analyzing events taking place in continuous time subject to grouping, the estimates obtained using discrete choice models can be subject to a severe bias. The phenomenon is much more prominent when analyzing multi-state models, in which interactions between events are under study.

In order to use the bivariate probit setting for representing the distribution of $Y_{i,t} = \left(Y^1_{i,t}, Y^2_{i,t}\right)'$ conditional on the state of the system in $(t-1)$, let us introduce the following notation:

$$s_{i,t-1} = \left(1, y^1_{i,t-1}, y^2_{i,t-1}, y^1_{i,t-1} y^2_{i,t-1}\right)'$$
$$D_{y_{i,t}} = 2\text{diag}\left(y_{i,t}\right) - I_2$$

It is worth to point out that s_{t-1} is an invertible linear transformation of

$$s^*_{t-1} = \left[\left(1-y^1_{t-1}\right)\left(1-y^2_{t-1}\right), y^1_{t-1}\left(1-y^2_{t-1}\right), \left(1-y^1_{t-1}\right) y^2_{t-1}, y^1_{t-1} y^2_{t-1}\right]$$

[5] Dynamic versions of the univariate probit model are discussed, for example, in Heckman (1981).

which involves four mutually exclusive dummies representing the four states of the process in $(t-1)$; in fact $s_{t-1} = Q s^*_{t-1}$, with

$$Q = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The reason for using s_{t-1} instead of s^*_{t-1} to describe the state of the system in $(t-1)$ is that, doing so, the non-causality restrictions are more easily written and interpreted.

As shown in Appendix A, the joint distribution of $Y_{i,t}$ conditional on the state of the system in $(t-1)$ can be written as follows:

$$\Pr\{y_{i,t} \mid y_{i,t-1}\} = \Phi_2\left(D_{y_{i,t}}\beta' s_{i,t-1}; 0, D_{y_{i,t}}\begin{bmatrix} 1 & \rho_{i,t} \\ \rho_{i,t} & 1 \end{bmatrix} D'_{y_{i,t}}\right) \tag{7}$$

where $\rho_{i,t}$ is given by

$$\rho_{i,t} = \frac{2\exp(\gamma' s_{i,t-1})}{1+\exp(\gamma' s_{i,t-1})} - 1 \tag{8}$$

$\beta = [\beta_1, \beta_2]$ and γ are parameter matrices of dimension 4×2 and 4×1 respectively, while $\Phi_2(\cdot; \mu, R)$ denotes the integrated bivariate normal with mean μ and correlation matrix R. The logit-type functional form in (8) is chosen so as to bound the correlation coefficient between -1 and 1: other choices are possible. As a whole, the distribution depends on 12 parameters freely varying in $\mathbb{R}^{12}$, and it is easily shown that the transition probabilities are a bijective transformation of β and γ. Notice that the marginal distribution of $Y^1_{i,t}$ and $Y^2_{i,t}$ (given $Y_{i,t-1}$) is given by

$$\Pr\left\{y^1_{i,t} \mid y_{i,t-1}\right\} = \Phi_1\left(\left(2y^1_{i,t}-1\right)\beta_1' s_{i,t-1}; 0, 1\right) \tag{9}$$

$$\Pr\left\{y^2_{i,t} \mid y_{i,t-1}\right\} = \Phi_1\left(\left(2y^2_{i,t}-1\right)\beta_2' s_{i,t-1}; 0, 1\right) \tag{10}$$

It may be useful to give an interpretation of the model in terms of latent regression model. Each individual i has to make two binary choices at time t, *i.e.* to choose the value of the binary bivariate vector $Y_{i,t}$. The latent regression approach assumes that the individual will choose $Y^1_{i,t} = 1$ when a latent continuous random variable $Y^{1*}_{i,t}$ crosses a threshold, which, in the current framework, is assumed to depend on the choice made in $(t-1)$. The same holds for $Y^2_{i,t}$. The latent regression here is:

$$y^{1*}_{i,t} = \beta_1 s_{i,t-1} + \varepsilon^1_{i,t}$$
$$y^{2*}_{i,t} = \beta_2 s_{i,t-1} + \varepsilon^2_{i,t}$$

where:

$$\varepsilon_{i,t} = \begin{pmatrix} \varepsilon^1_{i,t} \\ \varepsilon^2_{i,t} \end{pmatrix} \sim Nid\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{it} \\ \rho_{it} & 1 \end{bmatrix}\right)$$

Notice that the assumption that ε_t is independently distributed matches perfectly the Markov assumption, since failure of this condition means that there is some information left in the history of the process after conditioning on $Y_{i,t-1}$.

The conditions for strong one step ahead non-causality and strong simultaneous independence are easily stated as restriction on the parameter space of (7):

$$H_{1\nleftarrow 2} \ (Y^1 \nleftarrow Y^2): \qquad \beta_1 = H_1\varphi_1 \tag{11}$$

$$H_{1\nrightarrow 2} \ (Y^1 \nrightarrow Y^2): \qquad \beta_2 = H_2\varphi_2 \tag{12}$$

$$H_{1\not\Leftrightarrow 2} \ (Y^1 \not\Leftrightarrow Y^2): \qquad \gamma = 0 \tag{13}$$

where

$$H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad H_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{14}$$

Under $H_{1\nleftarrow 2}$, y^2_{t-1} and $y^1_{t-1}y^2_{t-1}$ are excluded from (9), so that $\Pr\left\{y^1_{i,t} \mid y_{i,t-1}\right\} = \Pr\left\{y^1_{i,t} \mid y^1_{i,t-1}\right\}$. Similarly, under $H_{1\nrightarrow 2}$, y^1_{t-1} and $y^1_{t-1}y^2_{t-1}$ are excluded from (10), so that $\Pr\left\{y^2_{i,t} \mid y_{i,t-1}\right\} = \Pr\left\{y^2_{i,t} \mid y^2_{i,t-1}\right\}$. Finally, under $H_{1\not\Leftrightarrow 2}$, $\rho_{i,t}$ is equal to zero, and hence the joint distribution (7) factors out in the product of the marginal distributions (9) and (10).

3 Introducing Covariates and Relaxing the Markov Hypothesis

In this Section, the model presented in Section 2 will be extended in two directions. First we will relax the assumption of stationary transition probabilities by introducing covariates, in order to account for individual and/or time heterogeneity. This will be done under the Markov assumption. Then we will drop the Markov assumption to allow for more complex dynamics. This will be done in the absence of covariates. Relaxing both hypotheses is straightforward and left to the reader.

Extending the information set. The information set available to predict Y_t is now enlarged to $\mathcal{F}_{t-1} = \mathcal{Y}_{t-1} \vee \mathcal{X}_{t-1}$.[6] Notice that replacing $\mathcal{X}_{t-1}$ by $\mathcal{X}_t$ is completely irrelevant for the following discussion. Let us first maintain the Markov assumption, and for notational simplicity let us also assume, without loss of generality, that all the information in $\mathcal{X}_{t-1}$ which is relevant for the transition probabilities in t is given by X_{t-1}.

Extending model (7) so that the transition probabilities depend on x_{t-1} may be easily done by replacing $s_{i,t-1}$ by

$$z^*_{i,t-1} = \left[s'_{i,t-1}, x^{*\prime}_{i,t-1}\right]' \tag{15}$$

[6] Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be σ-fields. $\mathcal{M}_1 \vee \mathcal{M}_2$ denotes the σ-field generated by $\mathcal{M}_1 \cup \mathcal{M}_2$. Hence $\{\mathcal{F}_t\} = \{\mathcal{Y}_t \vee \mathcal{X}_t\}$ corresponds to the canonical filtration associated to $\{(Y_t, X_t)\}$.

where $x^*_{i,t}$ is the part of $x_{i,t}$ which is linearly independent on $s_{i,t}$ (typically, if the constant is in both $x_{i,t}$ and $s_{i,t}$ it has to be dropped from $x_{i,t}$ to avoid perfect collinearity). If we denote by k the dimension of $x_{i,t}$, and by k^* the dimension of $x^*_{i,t}$, then B and γ will be now of dimension $(4+k^*) \times 2$ and $(4+k^*) \times 1$. It is important to point out that this way to include the covariates amounts to assuming that the impact on the transition probabilities is the same irrespective of $s_{i,t-1}$, so that the effect of the covariates is the same whatever state the individual belongs to in $(t-1)$. A more general model, allowing for interaction among the covariates and the state of the process in $(t-1)$, *i.e.* $s_{i,t-1}$, ensues from using in (7), instead of $s_{i,t-1}$,

$$z_{i,t-1} = s_{i,t-1} \otimes x_{i,t-1}. \tag{16}$$

Notice that, in this case, B and γ will be of dimension $4k \times 2$ and $4k \times 1$, so that many more parameters have to be estimated. We will refer to the model deriving from (16) as *saturated* model, while the model deriving from (15) will be referred to as *unsaturated*. Notice that the unsaturated model is nested in the saturated one, and therefore the decision about which one is convenient for describing the data may be empirically based on testing. A simple example may help understanding the difference between the two models. Assume that each individual i belongs, at any time t to either one or the other two mutually exclusive and exhaustive classes C_1 ad C_2. Define

$$D^1_{i,t} = 1_{\{\text{individual } i \in C_1 \text{ at time } t\}}$$
$$D^2_{i,t} = 1_{\{\text{individual } i \in C_2 \text{ at time } t\}}$$

so that $D^1_{i,t} + D^2_{i,t} = 1$. Let $X_{i,t} = \left(D^1_{i,t}, D^2_{i,t}\right)'$. In this setting, one may take $x^*_{i,t} = d^1_{i,t}$, and therefore

$$z^*_{i,t-1} = \left[s'_{i,t-1}, d^1_{i,t-1}\right]'$$
$$z_{i,t-1} = s_{i,t-1} \otimes x_{i,t-1} = \left[d^1_{i,t-1} s'_{i,t-1}, d^2_{i,t-1} s'_{i,t-1}\right]'$$

The most striking difference between the saturated and unsaturated model in this case is that, in the unsaturated model, $X_{i,t}$ (*i.e.* belonging to class C_1 or C_2) has the same impact (positive or negative or none) on the probability of $Y^1_{i,t}$ and $Y^2_{i,t}$ irrespective of the state in $(t-1)$. Conversely, in the saturated model, $X_{i,t}$ might have, say, a positive effect on the probability of $Y^1_{i,t}$ if $Y_{i,t-1} = (0,0)'$, and no effect on the probability of $Y^1_{i,t}$ if $Y_{i,t-1} = (1,0)'$.

The conditions for Granger non-causality in the presence of covariates are formally identical to (11) and (12), but the restriction matrices are defined as follows for the unsaturated model:

$$H^*_1 = \begin{bmatrix} H_1 & 0 \\ 0 & I_{k^*} \end{bmatrix} \quad , \quad H^*_2 = \begin{bmatrix} H_2 & 0 \\ 0 & I_{k^*} \end{bmatrix}$$

while, for the saturated model the matrices are

$$H^*_1 = I_k \otimes H_1 \quad , \quad H^*_2 = I_k \otimes H_2$$

It is easily checked that these restrictions matrices exclude all the regressors involving y^2_{t-1} from $\Pr\left\{y^1_{i,t} \mid y_{i,t-1}, x_{i,t-1}\right\}$, and all the regressors involving y^1_{t-1} from $\Pr\left\{y^2_{i,t} \mid y_{i,t-1}, x_{i,t-1}\right\}$. As for the simultaneous independence condition (13), it remains unchanged, since $\rho_{i,t}$ must be identically equal to zero for all (i,t) in order to factor out the joint distribution (7) into the product of the marginal (9) and (10), which requires that $\rho_{i,t}$ does not depend on covariates.

Relaxing the Markov assumption. Let us now relax the Markov assumption. For the sake of simplicity, we go back to the assumption that the information set available in t is $\mathcal{Y}_{t-1}$. Consider first the case where the relevant information for the transition probabilities is given by the last two states visited by an individual, rather than the last one only. There are therefore 16 possible paths followed in $(t-2)$, $(t-1)$, at the end of which the individual may choose among 4 states. Hence, the most general model one may use to describe $\Pr\{y_{i,t} \mid y_{i,t-1}, y_{i,t-2}\}$ requires $16 \times (4-1) = 48$ transition probabilities.[7]

This model may be written in the form (7) by replacing s_{t-1} by

$$s^2_{i,t-1} = s_{i,t-1} \otimes s_{i,t-2}$$

i.e. using the saturated model with $x_{i,t-1} = s_{i,t-2}$. To generalize to the case in which the last ℓ states visited are relevant for the transition probabilities, then

$$s^\ell_{i,t-1} = s_{i,t-1} \otimes s_{i,t-2} \otimes \ldots \otimes s_{i,t-\ell}$$

has to be used in (7) instead of $s_{i,t-1}$. It seems natural to refer to this model as bivariate Vector AutoRegressive Probit model of order ℓ, or *VAP(ℓ)*. We will call *VAPX(ℓ)* the model where exogenous covariates are also included. Notice that the number of parameters does increase very rapidly, since B and γ will be of dimension $4^\ell \times 2$ and $4^\ell \times 1$. The dynamic structure of the process may be simplified by using the unsaturated model rather than the saturated one, which would dramatically reduce the number of parameters to $3\,(3\ell+1)$, although the interpretation of the ensuing model is unclear. A further simplification could be based on the following underlying latent regression:

$$y^*_{i,t} = \mu + \sum_{j=i}^{\ell} A_j y_{i,t-j} + \varepsilon_{it}$$

where $y^*_{i,t} = \left(y^{1*}_{i,t}, y^{2*}_{i,t}\right)$, A_j $(j = 1, \ldots \ell)$ are 2×2 parameter matrices, μ is a 2×1 parameter vector, and:

$$\varepsilon_{i,t} = \begin{pmatrix} \varepsilon^1_{i,t} \\ \varepsilon^2_{i,t} \end{pmatrix} \sim Niid\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

Notice that this model implies that

$$\mathbb{E}\{y_{i,t} \mid y_{i,t-1}, \ldots, y_{i,t-\ell}\} = \mu + \sum_{j=i}^{\ell} A_j y_{i,t-j}$$

[7] It is easily shown that this non-Markov 4 states model may be rewritten as a Markov 16 states model, where 192 out of the 16^2 transition probabilities are set to zero, while 16 transition probabilities may be written as linear functions of the remaining 48.

or equivalently

$$y_{i,t} = \mu + \sum_{j=i}^{\ell} A_j y_{i,t-j} + \eta_{i,t} \tag{17}$$

which is formally identical to the standard VAR model, with the only difference that the distribution of the error term, which is usually assumed to be Gaussian, takes here into account the binary nature of $Y_{i,t}$.[8] The total number of parameters is further reduced to $4\ell + 3$. It is important to explore the precise meaning of these restrictions on the dynamics of the process, and find other, more interpretable ways to set up a priori restrictions of the parameter space of the *VAP* models. Notice however that the problems related to using overparameterized models are here mitigated, with respect to the standard Vector AutoRegressive literature, by the availability of individual data, since the usable data points are here $N \times T$, rather than T (see also Section 4).

For the unrestricted *VAP(ℓ)* model, the Granger non-causality conditions are formally identical to (11) and (12), but the restriction matrices are defined as follows:

$$H_1^* = \underbrace{H_1 \otimes H_1 \otimes \ldots \otimes H_1}_{\ell \text{ times}} \quad , \quad H_2^* = \underbrace{H_2 \otimes H_2 \otimes \ldots \otimes H_2}_{\ell \text{ times}}$$

In this case, the restrictions matrices exclude all the regressors involving $y^2_{t-1}, \ldots, y^2_{t-\ell}$ from $\Pr\left\{y^1_{i,t} \mid y_{i,t-1}, \ldots, y_{i,t-\ell}\right\}$, and all the regressors involving $y^1_{t-1}, \ldots, y^1_{t-\ell}$ from $\Pr\left\{y^2_{i,t} \mid y_{i,t-1}, \ldots, y_{i,t-\ell}\right\}$. Again, the simultaneous independence condition (13) remains unchanged. The restriction matrices for the restricted versions of the *VAP*, as well as those needed for the *VAPX* may be obtained accordingly.

4 Estimation and Testing

The purpose of this Section is to discuss the properties of the parameters estimates in model (7), as well as the properties of the tests for the hypotheses (11)-(13). Some hints will also be given about the generalizations illustrated in Section 3. We will discuss the asymptotic properties of Maximum Likelihood estimates and LR tests, although several other standard procedures for estimating and testing may be used. Some finite sample results will be also illustrated.

We assume that each individual i, $i = 1, \ldots N$, is observed at each time t during a period of known length $t = 1, \ldots, T$; the extension to the case in which every individual i is observed for a length T_i is straightforward if we suppose

[8] Notice also that, in this model, dropping the constant would impose the following restriction, probably uninteresting in most applications:

$$\Pr\left\{Y^1_{i,t} = 1 \mid Y_{i,t-1} = \ldots = Y_{i,t-\ell} = (0,0)'\right\} = 0.5$$
$$\Pr\left\{Y^2_{i,t} = 1 \mid Y_{i,t-1} = \ldots = Y_{i,t-\ell} = (0,0)'\right\} = 0.5$$

that T_i is a Markov time with respect to the filtration $\mathcal{F}_{T_i-1}$. The likelihood of the sample can be written in compact form as:

$$\mathsf{L}(\theta; y) = \prod_{i=1}^{N}\prod_{t=1}^{T} \Pr\{y_{i,t}|y_{i,t-1}\} =$$
$$= \prod_{i=1}^{N}\prod_{t=1}^{T} \Phi\left(D_{y_{i,t}} B s_{i,t-1}; 0, D_{y_{i,t}} \begin{bmatrix} 1 & \rho_{i,t} \\ \rho_{i,t} & 1 \end{bmatrix} D'_{y_{i,t}}\right),$$

where $\rho_{i,t}$ is given by (8), and is therefore a function of $s_{i,t-1}$ and the parameters γ.

Under some conditions, developed in the following, maximization of the log-likelihood leads to consistent, asymptotically normal and asymptotically efficient estimates of the parameters of the model, which ensures that LR tests are asymptotically χ^2 distributed under the null hypothesis.[9] Let us discuss the asymptotics involved, by considering three cases:

- $T \to \infty$, N finite
- T finite, $N \to \infty$
- $T \to \infty$, $T \to \infty$

To keep the notation simple, notice that the Markov model for homogeneous population can be written in a compact notation, since it is equivalent to a Markov model for the process $\{S_t\}$ which, at any time t, takes on values in a finite state-space $\mathcal{S} = \{0, 1, 2, 3\}$,[10] with a stationary (or *time-homogeneous*) transition probability matrix $P = \|P_{hk}\|$, $(h, k) \in \mathcal{S} \times \mathcal{S}$; clearly P is a stochastic matrix, that is, $P_{hk} \geq 0\ \forall h, k$, and $\sum_{k \in \mathcal{S}} P_{hk} = 1$. Moreover we define

$$P^n = \underbrace{P \cdot ... \cdot P}_{n \text{ times}} = \|P^n_{hk}\|$$

where $P^n_{hk} = \Pr\{S_{t+n} = k | S_t = h\}$.

Essentially, what we need to have the usual asymptotic properties for the ML estimates and LR tests is that *all* the transitions whose probabilities have to be estimated (*i.e.* are not known) *can* be observed infinitely many times as T and/or N go to infinity. It is intuitive that this ensures consistent and asymptotically normal estimates of the transition probabilities. This requires the following:

[9] McFadden (1984) proves consistency and asymptotic normality for a general multinomial model under more general conditions, but the extension in this case seems to be difficult, because he postulates that the explanatory variables are independent identically distributed for each observation. This does not seem to hold generally for multiperiod models; nevertheless, McFadden's result can be applied to our homogeneous model.

[10] The elements of $\mathcal{S}$ correspond element-wise with the elements of $\mathcal{Y}$, so defined:

$$\mathcal{Y} = \{(0,0), (0,1), (1,0), (1,1)\}.$$

representing the state space of the process $\{Y_t\}$ at any time t.

1. (necessary condition) Each state with at least one unknown exiting transition probability must be visited infinitely often with probability 1 as either T or N or both go to infinity.

2. (sufficient condition) Infinitely many of the individuals who have reached each state with at least one unknown exiting transition probability must be observed for at least one time period in that state.

In the following we will enunciate some results about the conditions on P under which both the necessary and sufficient conditions are fulfilled. Let us first state the condition when $T \to \infty$ with N fixed.

Proposition 5 *Assume that P is such that each state with at least one unknown exiting transition probability is persistent. Then $T \to \infty$ ensures the fulfillment of condition 1 for any $N \geq 1$.*[11] *Condition 2 is obviously fulfilled for $T \to \infty$.*

The proof of the first part is in Billingsley (1986), Theorem 8.2. Notice that each state of an irreducible Markov chain with finite state space is persistent (Billingsley, 1986, Example 8.7). Therefore in our case assuming that the chain is irreducible ensures that it is persistent, which in turn ensures that ML estimates show the usual asymptotic properties when $T \to \infty$ for any $N \geq 1$, finite or infinite.

A similar result can be obtained when $N \to \infty$, but it depends on the initial conditions of the process. These are defined by a vector of *initial probabilities* $p = \|p_h\|$, representing, $\forall h \in \mathcal{S}$, the probability of being in state h in $T = 0$. Of course, $p_h \geq 0 \ \forall h \in \mathcal{S}$, and $\sum_{h \in \mathcal{S}} p_h = 1$.

Proposition 6 *Assume that P and p are such that $p_{h_k} P_{h_k k}^{n_k} > 0$ for each state $k \in \mathcal{S}$ with at least one unknown exiting transition probability, for at least one $h_k \in \mathcal{S}$ and for some finite integer n_k. Then, $N \to \infty$ ensures the fulfillment of condition 1 for any $T \geq \max\{\overline{n}_k, k \in \mathcal{S}\}$, where $\overline{n}_k$ is, for each k, the minimum n_k such that $p_{h_k} P_{h_k k}^{n_k} > 0$. Moreover, condition 2 is fulfilled if $T \geq \max\{\overline{n}_k, k \in \mathcal{S}\} + 1 = T^*$.*

The condition on p and P ensures that there is at least one initial state and one route which allows each state with at least one unknown exiting transition probability to be reached in a finite number of steps. Then, if T is large enough, and $N \to \infty$, each state with at least one unknown exiting transition probability will be reached infinitely many times. One additional time period is needed to estimate the transition probabilities. Notice that if $p_h \neq 0 \ \forall h \in \mathcal{S}$, and the chain is irreducible, then the conditions on p and P are met, and moreover $T^* = 2$, so that the ML estimates show the usual asymptotic properties when $N \to \infty$ for $T \geq 2$, finite or infinite.

[11] For definitions of irreducibility and persistence, see Billingsley (1986), Section 8.

As a whole, the conditions

$$p_h \neq 0, \quad \forall h \in \mathcal{S}$$
$$P \quad \text{irreducible}$$

are sufficient, although non necessary, for consistency and asymptotic normality of the Maximum Likelihood estimator with either $T \to \infty$ or $N \to \infty$, where in the latter case the T must be at least equal to 2. Of course, since these conditions are only sufficient, the less stringent assumptions stated in Propositions 5 and 6 lead to the same result.

Extension to non-Markov chains such as the *VAP(ℓ)* model are straightforward, since it is always possible to rewrite a bivariate *VAP(ℓ)* model as a finite state-space Markov chain with 4^ℓ states, and hence the conditions stated in Propositions 5 and 6 apply to the transition probabilities matrix and initial probabilities vector corresponding to the new Markov chain.

When covariates are introduced, it is an obvious requirement that they are weakly exogenous with respect to the transition probabilities (see Engle, Hendry and Richard (1983)) in order to achieve asymptotic efficiency of the ML estimates. Under this additional assumption, the same conditions stated in Propositions 5 and 6 ensure consistency and asymptotic normality.

Let us briefly address some problems possibly arising in finite samples. It may be easily shown that, in the homogeneous population model, the transition probabilities are estimated essentially as the ratio of the number of cases where the transition occurred (say, N_{hk}) over the number of cases where it could have occurred (say, N_h). In any finite sample, the distribution of N_{hk} given N_h will be Binomial, and will hence converge in distribution to the Normal with $N_h \to \infty$. On the other hand, the distribution of N_h depends on N , T, the transition probabilities matrix P and the initial probabilities vector p; in the light of this, propositions 5 and 6 state the conditions on P which ensures divergence of such distribution with either N (Proposition 5) or T (Proposition 6). This divergence leads to normality of the distribution of the ratio $\frac{N_{hk}}{N_h}$. Notice however that, in finite samples, this distribution will not be normal. In particular, for states which have been visited few times (N_h small), the distribution of N_{hk} given N_h may be very far from normality, especially when the true transition probability P_{hk} is close to zero or one, which will give highly skewed distributions. Clearly, the distribution of Wald type tests for hypotheses involving parameters whose estimates are affected by this problem will be very far from the asymptotic χ^2 distribution, while the asymptotic approximation might work quite well when the parameters involved in the restriction are estimated based on large N_h's. Mimicking similar results in other models, LR type tests might perform better in the first case, and worst in the second, but a careful analysis of finite sample properties is needed to make precise statements. In any case, in order to get an idea of how reliable the asymptotic distribution can be, it is convenient to check the number of data points *in each* state with at least one unknown exiting transition probability. In fact, even if $N \times T$ is large, the information about some of the transitions might be quite scarce.

5 Non-Causality with Survival Data

Special cases of the model discussed in Sections 2 and 3 can be obtained when some of the transition probabilities are set at 0. In the following paragraph we will deal with the case of survival models,[12] *i.e.* models in which the states with $Y_t^j = 0$ are not accessible from the states with $Y_{t-1}^j = 1$, $j = \{1, 2\}$; this implies that every decision with respect to a variable Y_t^j is in a certain sense irreversible. This imposes on the model the following constraints:

$$\begin{aligned}
&\Pr\left\{Y_{i,t}^1 = 0 \mid Y_{i,t-1}^1 = 1\right\} = 0, \\
&\Pr\left\{Y_{i,t}^2 = 0 \mid Y_{i,t-1}^2 = 1\right\} = 0,
\end{aligned}$$

or, in a different form:

$$\begin{aligned}
&\Pr\left\{Y_{i,t} = (0,0)' \mid Y_{i,t-1} = (1,0)'\right\} = 0, \\
&\Pr\left\{Y_{i,t} = (0,0)' \mid Y_{i,t-1} = (0,1)'\right\} = 0, \\
&\Pr\left\{Y_{i,t} = (0,0)' \mid Y_{i,t-1} = (1,1)'\right\} = 0, \\
&\Pr\left\{Y_{i,t} = (1,0)' \mid Y_{i,t-1} = (0,1)'\right\} = 0, \\
&\Pr\left\{Y_{i,t} = (1,0)' \mid Y_{i,t-1} = (1,1)'\right\} = 0, \\
&\Pr\left\{Y_{i,t} = (0,1)' \mid Y_{i,t-1} = (1,0)'\right\} = 0, \\
&\Pr\left\{Y_{i,t} = (0,1)' \mid Y_{i,t-1} = (1,1)'\right\} = 0.
\end{aligned}$$

Then the state-transition diagram of this kind of model can be represented as in Figure 2.

Probably the simpler way to think about this model is to consider a bivariate probit model starting from state 0, that is from $Y_t = (0,0)'$ with transition probabilities:

$$\Pr\left\{y_{i,t} \mid Y_{i,t-1} = (0,0)'\right\} = \Phi_2\left(D_{y_{i,t}} \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix}; 0, D_{y_{i,t}} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} D'_{y_{i,t}}\right),$$

where we have eliminated the subscript from ρ since no other correlation coefficient is present in the model.

Transitions from states 1 and 2 have respectively probabilities:

$$\begin{aligned}
&\Pr\left\{y_{i,t}^1 \mid Y_{i,t-1} = (0,1)'\right\} = \Phi_1\left(\left(2y_{i,t}^1 - 1\right)\left(\beta_{10} + \beta_{12}\right); 0, 1\right), \\
&\Pr\left\{y_{i,t}^2 \mid Y_{i,t-1} = (1,0)'\right\} = \Phi_1\left(\left(2y_{i,t}^2 - 1\right)\left(\beta_{20} + \beta_{21}\right); 0, 1\right).
\end{aligned}$$

[12] A standard reference for survival models is Kalbfleisch and Prentice (1980). A counting processes perspective on these models is in Andersen *et al.* (1993). A review of multivariate survival models is Hougaard (1987). The model presented in this Section is discussed in greater detail in Mosconi, Sartori and Seri (1998), where in particular the effect of time aggregation on causality tests is discussed.

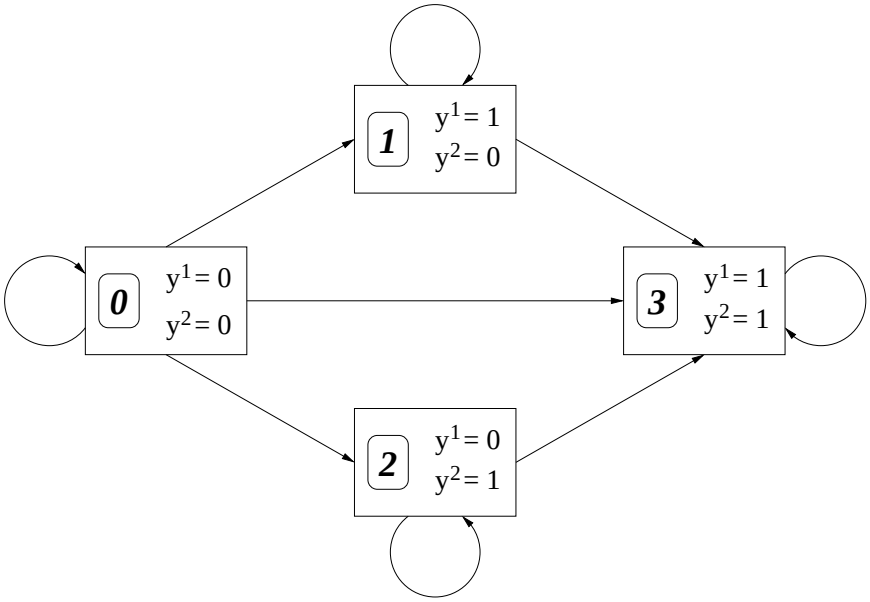


Figure 2: State-Transition Diagram for a Bivariate Survival Model

Conditions for simultaneous independence and Granger non-causality can be easily adapted from those introduced in Section 2, and take the particularly simple form:

$$\begin{aligned} H_{1\not\leftarrow 2} \ (Y^1 \not\leftarrow Y^2) : & \qquad \beta_{12} = 0 \\ H_{1\not\rightarrow 2} \ (Y^1 \not\rightarrow Y^2) : & \qquad \beta_{21} = 0 \\ H_{1\not\Leftrightarrow 2} \ (Y^1 \not\Leftrightarrow Y^2) : & \qquad \rho = 0 \end{aligned}$$

Also the introduction of exogenous variables can be accounted for by simply paralleling the solutions presented in Section 3. The same holds true as regards the introduction of the lagged dependent variables $y_{i,t-j}$; saturated and unsaturated models are defined as in Section 3, but a comment is in order: in this case conditioning on the lagged endogenous variables is meaningful only when the individual is in state 1 or 2, since when he is in state 0, conditioning on this information induces no change in the transition probabilities, as state 0 cannot be accessed from any other state of the model.

As for the asymptotic properties of the estimates and tests, notice that for this model, even in the case of homogeneous populations, Proposition 5 does not hold, since this model can be represented as a Markov chain with an absorbing state corresponding to state 3. This means that $N \to \infty$ is needed in order to ensure consistency and asymptotic normality of the estimates. Conversely, $T \to \infty$ does not lead to the usual asymptotic properties. The intuition is that, for any finite N, when $T \to \infty$ all individuals will eventually end up in the

absorbing state 3. When this happens, no additional information is obtained by increasing T further. Notice however that, up to this point, the estimates become more efficient and, loosely speaking, closer to normality, as T increases.

6 An Illustrative Example for Panel Data

Notwithstanding the interest in timing fertility and in working choices, the joint study of marital duration and fertility has not been a deeply studied topic in Sociometrics: Waite and Lillard (1991), Lillard and Waite (1993) and Lillard (1993) are almost the only contributions to its study.

However, the picture they draw is quite clear. Married couples with children appear to be less likely to end their marriages than childless couples. This seems to suggest that children affect the chances that they parents divorce; however, the process may not be so simple, since the chances that the marriage will last may affect couples' willingness to have children. The economic model used by the authors is quite sophisticated and able to capture a wide variety of situations: the problem with Lillard's and Waite's approach is that the method of estimation they use is not able to yield consistent estimators and therefore the inferences they draw from the model are fallacious.[13]

Our analysis will be based on data from the PSID database and therefore a brief description appears to be necessary. The Panel Study of Income Dynamics (PSID), begun in 1968, is a longitudinal study of a representative sample of U.S. individuals (men, women, and children) and the family units in which they reside. It emphasizes the dynamic aspects of economic and demographic behavior, but its content is broad, including sociological and psychological measures. Starting with a national sample of 5,000 U.S. households in 1968, the PSID has reinterviewed individuals from those households every year since that time, whether or not they are living in the same dwelling or with the same people. Adults have been followed as they have grown older, and children have been observed as they have advanced through childhood and into adulthood, forming family units of their own. The study is conducted at the Survey Research Center, Institute for Social Research, University of Michigan. Information about the original 1968 sample individuals and their current co-residents (spouses, co-habitors, children, and anyone else living with them) is collected each year. This has allowed us to select 266 women appearing in the database from 1968 to 1993: subjects with missing covariates have been eliminated and the variables has been elaborated in order to yield a certain uniformity over time.

More information about the variables can be found in Section B: each column represents a variable and for each year and variable the code is reported; the

[13]The Lillard and Waite approach suffers from a problem of endogeneity since they use a model with unobserved heterogeneity that is valid only under the hypothesis of strong exogeneity of the regressors. However, since they use as explicative variables some lagged dependent variables, this condition seems to be violated.

The methods of Arellano and Carrasco (1996) and Honoré and Kiriazidou (1997) could be of interest here.

selected range is reported in parentheses. Therefore the following variables have been selected:

- `Y1`: the variable is set to 1 in t if the individual has had (at least) a child during the year t, 0 otherwise;
- `Y2`: this variable is set to 1 in t if the individual was married during the year t, 0 otherwise;
- `AGE`: this is equivalent to the PSID variable `AGE OF THE INDIVIDUAL`;
- `INCOME`: it is equivalent to `MONEY INCOME` (in the period 1968-1974) or `TAXABLE INCOME` plus `TRANSFER INCOME` (in the period 1975-1990) or the sum of `TRANSFER INCOME`, `LABOR INCOME` and `ASSET INCOME` (in the period 1991-1993); this is due to the increased precision of the variables in the PSID database;
- `HOURS`: it is the PSID `HOURS WORKED` variable;
- `EDU`: it is essentially the PSID `YEARS OF SCHOOL COMPLETED` variable: however it has been checked for incongruences with `HIGHEST GRADE COMPLETED` and `COMPLETED EDUCATION`.

To provide an application of causality testing for binary bivariate processes, six models belonging to the class presented in Section 3 have been estimated on PSID data.[14] We are interested in performing a series of tests on these models: in particular, we would like to choose a model that fits parsimoniously the data and to test for non-causality between the two processes. In order to select the model that seems to fit the data better we have to consider separately the case in which one of the two models is nested in the other one, and the non-nested case. In the nested case, Likelihood Ratio tests can be used; on the other side, to test among non-nested models, we consider two well-known information criteria, the Akaike Information Criterion and Schwarz' Bayesian Information Criterion. Then we study through Wald tests the non-causality relations between the two processes $\{Y_t^1\}$ and $\{Y_t^2\}$.

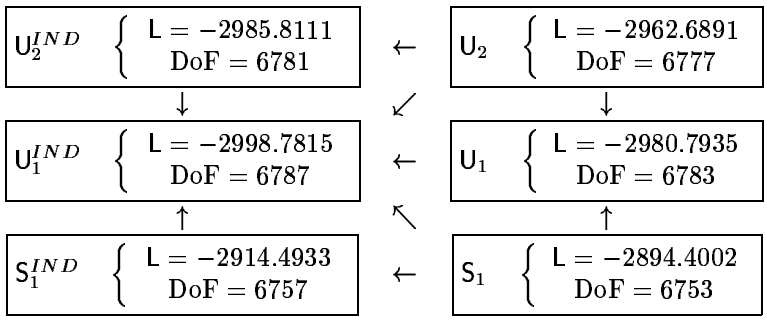
In the following we will use the mnemonics for the estimated models reported in Table 1. The full estimates of the proposed models are reported in Appendix C; for our purposes, it is enough to consider the log-likelihoods of the estimated

[14] Estimation has been performed using GAUSS-386i. The covariance matrix of the estimates have been calculated through the cross-product of first derivatives.

Mnemonics	Contents
U_1^{IND}	One Lag Unsaturated Model with No Simultaneous Dependence
U_2^{IND}	Two Lag Unsaturated Model with No Simultaneous Dependence
S_1^{IND}	One Lag Saturated Model with No Simultaneous Dependence
U_1	One Lag Unsaturated Model
U_2	Two Lag Unsaturated Model
S_1	One Lag Saturated Model

Table 1: Mnemonics for the estimated models

models, together with the degrees of freedom as reported in the following scheme:



The possible restrictions are indicated by an arrow connecting the two boxes: in particular, if a box connects model M^U (U for *unrestricted*) to model M^R (R for *restricted*), we say that model M^R is *nested* in model M^U. Therefore, we can test all the restrictions through a Likelihood Ratio test;[15] in particular, if we indicate by L_U the log-likelihood of the unrestricted model and by L_R the log-likelihood of the model obtained after constraining some of the parameters, the Likelihood Ratio test consists in calculating the statistic

$$2 \cdot (\mathsf{L}_R - \mathsf{L}_U) .$$

This statistic is asymptotically distributed as a χ^2_p where p is the number of parameters constrained in the restricted model. The results are reported in Table 3 and show that all the restrictions are refused at any conventional significance level (90%, 95%, 99%). Clearly this procedure is not able to discriminate between two models that are not directly linked by a restriction, and therefore, we cannot say whether U_2 fits the data better than S_1, since they are *non-nested*. Therefore, in order to select a model, we calculate two widely used criteria for the comparison of non-nested models: the AIC (Akaike Information Criterion) and the BIC (Bayesian Information Criterion, due to Schwarz, 1978); in both cases the model maximizing the criterion is selected. The formulas are the

[15] See Gouriéroux, Monfort and Renault (1987) for a similar testing strategy.

Models	L	p	AIC	BIC
U_1^{IND}	-2998.7815	18	-2980.7815	-3078.21022
U_2^{IND}	-2985.8111	24	-2961.8111	-3091.71605
S_1^{IND}	-2914.4933	48	-2866.4933	-3126.30321
U_1	-2980.7935	22	-2958.7935	-3077.87304
U_2	-2962.6891	31	-2931.6891	-3099.48300
S_1	-2894.4002	52	-2842.4002	-3123.86094

Table 2: Information criteria for the estimated models

Restrictions	χ^2 Test	DoF	Signif.
$\mathsf{U}_1 \rightarrow \mathsf{U}_1^{IND}$	35.9760	4	2.92678E-07
$\mathsf{U}_2^{IND} \rightarrow \mathsf{U}_1^{IND}$	25.9408	6	2.28367E-04
$\mathsf{S}_1^{IND} \rightarrow \mathsf{U}_1^{IND}$	168.5764	30	3.10543E-21
$\mathsf{S}_1 \rightarrow \mathsf{U}_1^{IND}$	208.7626	34	5.20616E-27
$\mathsf{U}_2 \rightarrow \mathsf{U}_1^{IND}$	72.1848	10	1.67555E-11
$\mathsf{U}_2 \rightarrow \mathsf{U}_2^{IND}$	46.2440	4	2.19107E-09
$\mathsf{S}_1 \rightarrow \mathsf{S}_1^{IND}$	40.1862	4	3.96112E-08
$\mathsf{S}_1 \rightarrow \mathsf{U}_1$	172.7866	30	5.31910E-22
$\mathsf{U}_2 \rightarrow \mathsf{U}_1$	36.2088	6	2.51064E-06

Table 3: Restrictions on the estimated models

following ones:

$$\mathsf{AIC} = \mathsf{L}_M - p,$$
$$\mathsf{BIC} = \mathsf{L}_M - \frac{p}{2} \cdot \ln M,$$

where M is a shortcut for the number of observation, be it N, T or mixed (in our case, $M = N \cdot T = 6,805$). It is important to remark that Akaike's criterion is not in general *consistent*, that is the probability of choosing the correct model does not go to 1 as long as the number of observations go to infinity: moreover, AIC has the unpleasant tendency to select overparameterized models; on the other side, the BIC is consistent. From Table 2, we see that AIC selects model S_1, while BIC selects U_1.Even more interesting and more directly linked to the topic of this Section would be to test for the presence of causality relations between marriage decisions and fertility timing. To do so, we have performed a Wald test as described in the previous Sections, whose results are displayed in Table 4. The results of all of these tests are univocal:

- the hypothesis $\mathsf{H}_{1 \nleftarrow 2}$, concerning the non-causality of Y^2 towards Y^1 is strongly rejected at any conventional significance level: therefore, we cannot accept the hypothesis that Y^2 does not cause Y^1. Therefore, a marital relation seems to increase significantly the probability of having a child, as common sense suggests.

Models	$\mathsf{H}_{1\not\leftarrow 2}$:	$Y^1 \not\leftarrow Y^2$	$\mathsf{H}_{1\not\rightarrow 2}$:	$Y^1 \not\rightarrow Y^2$	$\mathsf{H}_{1\not\Leftrightarrow 2}$:	$Y^1 \not\Leftrightarrow Y^2$
	T-Stat (DoF)	Signif.	T-Stat (DoF)	Signif.	T-Stat (DoF)	Signif.
U_1^{IND}	71.204432 (2)	3.4526626e-16	2.4087217 (2)	0.29988360	-	-
U_2^{IND}	83.657807 (4)	2.9219652e-17	3.1373153 (4)	0.53511467	-	-
S_1^{IND}	111.08184 (12)	3.6553090e-18	5.8918288 (12)	0.92143632	-	-
U_1	76.619847 (2)	2.3025637e-17	2.1034362 (2)	0.34933704	27.508227 (4)	1.5688357e-05
U_2	88.789666 (4)	2.3799911e-18	2.5789416 (4)	0.63055802	30.390133 (7)	8.0519688e-05
S_1	111.93028 (12)	2.4825481e-18	5.9715420 (12)	0.91750980	30.140922 (4)	4.5815270e-06

Table 4: Causality testing

- the hypothesis $\mathsf{H}_{1\not\rightarrow 2}$, concerning the non-causality of Y^1 towards Y^2 is accepted at any conventional significance level: therefore, we can accept the hypothesis that Y^1 does not cause Y^2. Hence, fertility timing does not seem to have any impact on the marriage and divorce decisions of American women.

- the hypothesis $\mathsf{H}_{1\not\Leftrightarrow 2}$, concerning the simultaneous independence between Y^2 and Y^1 is rejected at any conventional significance level: the same result had been obtained when testing for restrictions in the estimated models. Indeed, testing hypothesis $\mathsf{H}_{1\not\Leftrightarrow 2}$ is equivalent to testing for the restriction of a model with simultaneous dependence to a model without simultaneous dependence: clearly, the numerical results are hardly comparable, but we expect the two testing procedures to be asymptotically equivalent.

7 An Illustrative Example for Survival Data

To show a potential application of the model developed in Section 5, we investigate the causal relationship between the adoption of two technologies introduced in the 70's in the Italian metalworking industry. The dataset involves survival data, namely the spell of non adoption for both technologies in a sample of Italian plants, and therefore the analysis described in Section 5 will be performed. The two technologies considered are Computer Aided Design or Manufacturing (CADCAM), which will be labelled by 1, and Flexible Manufacturing Systems (FMS) which will be labelled by 2. Both technologies are originated from the Flexible Automation (FA) paradigm and therefore they are expected to display significative interactions.

Data on the diffusion of FA within the Italian metalworking industry are provided by the FLAUTO database, developed at Politecnico di Milano. FLAUTO monitors adoption of FA technologies by a sample composed of 782 Italian metalworking plants with 10 or more employees. The sample is stratified by size class, industry, and geographical area, so as to faithfully represent the universe of Italian metalworking plants with 10 or more employees. The dataset

originates from a retrospective survey carried on in June 1989. While all the plants included in FLAUTO are involved in production activities (*i.e.*, manufacturing and/or assembly), the same does not apply to design/engineering function. According to the information available in the database, 230 plants had no full time employee in the design/engineering function. Such plants were quite unlikely to belong to the population of realistic prospective adopters of CADCAM; therefore, they were excluded from the joint analysis of the diffusion of both technologies, reducing the sample size to 552. CADCAM and FMS have been introduced in Italy around 1970, hence the observation window is assumed to begin in this year and calendar time t is set to 0 in 1969. Notice that, since the survey is made in 1989, t never exceeds $T = 20$.

For each plant $i = 1, \ldots, 552$, we observe the year of adoption of both technologies, say t_i^1 and t_i^2. To fit these survival type data into our framework, it is convenient to transform them as follows:

$$\begin{array}{ll} y_{i,t}^1 = 1_{\{\text{plant } i \text{ adopts CADCAM at time } t\}} & t = t_i^E, \ldots, \min\left[t_i^1, T\right] \\ y_{i,t}^1 = 1_{\{\text{plant } i \text{ adopts FMS at time } t\}} & t = t_i^E, \ldots, \min\left[t_i^2, T\right] \end{array}$$

The different time span of the individual data is explained as follows. First, about 30% of the plants in the sample entered the metalworking industry after 1970: therefore, firm i contributes to the likelihood function only from the year of entrance in the sector, denoted as t_i^E. Moreover, in terms of the Markov chain depicted in Figure 2, when the plant adopts both technologies it enters an absorbing state, and therefore does not give any additional contribution to the likelihood (the probability of exiting the state is zero). Notice also that, as illustrated in Section 5, when $Y_{i,t-1} = (1,0)'$ the bivariate distribution $\Pr\left\{y_{i,t} \mid Y_{i,t-1} = (1,0)'\right\}$ collapses into its marginal $\Pr\left\{y_{i,t}^2 \mid Y_{i,t-1} = (1,0)'\right\}$, since $\Pr\left\{y_{i,t}^1 = 1 \mid Y_{i,t-1} = (1,0)'\right\} = 1$. Therefore, as obvious, no additional information is involved in recording $y_{i,t}^1$ for $t > t_i^1$. The same argument holds for Y_t^2. Finally, the data are right-censored because the firms are not observed after 1989, so that t never exceeds $T = 20$. Notice that this kind of censoring involves no bias, since time of censoring is a Markov time w.r.t. the filtration generated by the process.

In this framework, the use of a discrete time model can be justified supposing that the adoption of these technologies is decided during the budgeting phase, *i.e.* it is a discrete time event. This assumption seems to be correct in this case since these technologies are very expensive and requires often a reorganization of the structure of the firm (revision of the information system, skills of the employees, etc.): this renders plausible that the decision is taken in a formal moment as the discussion of budget.

Let us now introduce the covariates, *i.e.* the vector $x_{i,t}$ in our notation. The economic theory related to the diffusion of technological innovation recognizes four different groups of factors affecting the delay in the adoption of an innovation (see for example Colombo and Mosconi, 1995). *Rank effects* explain the delay in adoption as a consequence of firm heterogeneity: characteristics of the firm affect adoption probability, independently of the behavior of other

firms. *Epidemic effects* relate the dynamic pattern of diffusion with the spread of information about the features of the technology and, therefore, with the cumulative number of previous adopters. *Stock* and *order effects* originate from game-theoretic models: early adopters have a first-mover advantage over the others because of the possibility of preemption on resources critical for the use of the technology.

The model presented in the following is simplified and involves a reduced number of regressors: it should be considered only as an example of the methods introduced in the previous sections. The only time-invariant covariate considered is the size of the plant expressed in thousands of employees at June 1989: previous studies show the evidence of a positive and highly significant effect of size on the probability of adoption. This could be due to the fact that the profitability of a technological innovation requires a critical mass to become effective; moreover, great plants enjoy advantages connected with the reduction of the risk, because of preferential accesses to the capital markets and to managerial and organizational resources.

In addition we consider different time scales. We conform to Colombo and Mosconi (1995) in using both the calendar time t, and the duration of non adoption $\tau_{i,t} = t - \max\left(0, t_i^E\right)$. For plants that entered the sector before 1970, the two time scales coincide. We expect calendar time to reflect phenomena which do not depend on the existence of the firm, notably changes in the prices and performance improvements of the technologies over time, epidemic effects, and other time varying factors. Instead, the duration of non adoption captures effects related to the existence of the firm: insofar as is negatively related to the age of plant i, we may expect a negative duration dependence as younger firms and plants exhibit on average higher growth opportunities than older ones and thus more frequently face investment decisions.

As a whole, the covariates introduced in the model are the following:

$$x_{i,t} = (size_i, t, \tau_{i,t})\,;$$

in our application, for expositional purposes, ρ is assumed not to depend of $x_{i,t}$.

We have estimated two different classes of models:[16] the first group is formed by the saturated ones; the second is composed of the unsaturated ones. In each class, the baseline model, which will be referred to as H_U, is estimated under no restriction on the parameters. Within this model it is possible to test for non-causality through Wald or likelihood ratio tests.

Table 5 reports the estimates of the unrestricted unsaturated model, which allows for a more parsimonious representation of the economic aspects.[17] These estimates might suffer some bias due to the omission of several variables that have proved significant in previous studies: however, all of our hypotheses have been confirmed. The size of the firm is highly significant for both CADCAM

[16] Estimation has been performed as in the previous Section.

[17] The saturated models give no significant improvement of fit in this case, as can be seen comparing through a Likelihood Ratio test the value of the loglikelihoods reported in Table 6.

			coeff.	std.dev.	t-test	p-value
$\Pr\{y^1_{i,t} \mid y_{i,t-1}, x_{i,t}\}$ (CADCAM)	β_{10}	Constant	-3.798	0.129	-29.51	0.00
	β_{12}	$y^2_{i,t-1}$	0.469	0.125	3.76	0.00
	β_{15}	$size_i$	0.285	0.043	6.64	0.00
	β_{16}	t	0.139	0.011	12.48	0.00
	β_{17}	$\tau_{i,t}$	-0.003	0.009	-0.30	0.38
$\Pr\{y^2_{i,t} \mid y_{i,t-1}, x_{i,t}\}$ (FMS)	β_{20}	Constant	-3.562	0.192	-18.56	0.00
	β_{21}	y^1_{t-1}	0.198	0.109	1.81	0.03
	β_{25}	$size_i$	0.252	0.057	4.41	0.00
	β_{26}	t	0.082	0.018	4.61	0.00
	β_{27}	$\tau_{i,t}$	-0.006	0.014	-0.41	0.34
	ρ	Correlation	0.246	0.081	3.05	0.00
	Observations					9853
	Loglikelihood					-1475.488

Table 5: Estimates of the Unsaturated Model

and FMS: it has a positive sign, confirming the presence of an effect of size on the returns from adoption. Calendar time is also highly significant, because of its positive correlation with the performance of both CADCAM and FMS, and its negative correlation with prices. Duration of non adoption results insignificant, but this is probably due to a problem of misspecification, since a more complete study on the same dataset evidences a negative and moderately significant impact of this variable on the probability of adoption of FMS, while no effect is detected on CADCAM.

Based on Table 5, Wald type non-causality tests may be done by simply analyzing the $t-test$ for the parameters β_{12}, β_{21} and ρ. β_{12} is positive and significant (the hypothesis of Granger non-causality is rejected), which suggest a positive effect of adoption of FMS on the following adoption of CADCAM: this has a simple economic interpretation and confirms the presence of an interaction of the two technologies. On the other hand, β_{21} is positive but non significant, which means that the Wald test accepts the hypothesis that CADCAM does not Granger cause FMS $\left(Y^1 \nrightarrow Y^2\right)$. The correlation between the error terms of the latent regressions relative to CADCAM and FMS (ρ) is positive and highly significant: this implies a positive interaction in the simultaneous adoption of the two technologies. It is worth noticing that the results of non-causality tests depend on the information set, and therefore one might think that the evidence presented here depends on the very limited information supplied to the model. Actually, the results of the non-causality analysis are substantially unchanged even when the variables conditioned upon are all those included in Colombo and Mosconi (1995).

Let us now illustrate a more complete non-causality analysis, including some interesting joint hypotheses. Testing is based on likelihood ratio tests; Wald tests have been also computed, getting essentially identical results. For both the saturated and the unsaturated models, four restrictions on H_U are considered:

	Saturated		Unsaturated	
Model	Loglikelihood	Par.	Loglikelihood	Par.
H_U	−1472.925	17	−1475.488	11
$H_{1\nrightarrow 2}$	−1481.004	13	−1482.533	10
$H_{1\nleftarrow 2}$	−1476.176	13	−1477.141	10
$H_{1\nleftrightarrow 2}$	−1484.106	9	−1484.106	9
$H_{1\perp 2}$	−1488.510	8	−1488.510	8

Table 6: Likelihood of the Estimated Models

Hypothesis	LR test	dgf	p-value
$H_{1\nrightarrow 2}\, vs.\, H_U$	14.090	1	0.00
$H_{1\nleftarrow 2}\, vs.\, H_U$	3.307	1	0.07
$H_{1\nleftrightarrow 2}\, vs.\, H_U$	17.236	2	0.00
$H_{1\perp 2}\, vs.\, H_U$	26.043	3	0.00

Table 7: LR Non-Causality Tests for the Unsaturated Model

$$
\begin{aligned}
&H_{1\nrightarrow 2} : Y^1 \nrightarrow Y^2,\\
&H_{1\nleftarrow 2} : Y^2 \nrightarrow Y^1,\\
&H_{1\nleftrightarrow 2} : Y^1 \nrightarrow Y^2, Y^2 \nrightarrow Y^1,\\
&H_{1\perp 2} : Y^1 \nrightarrow Y^2, Y^2 \nrightarrow Y^1, Y^1 \nleftrightarrow Y^2.
\end{aligned}
$$

The maximized loglikelihood, together with the corresponding number of parameters, is reported in Table 6 for each of the estimated models.

It is straightforward to note that likelihood ratio tests fail to detect, at usual significance level, any difference between a saturated model and the corresponding unsaturated one. It is therefore not surprising that also the non-causality tests yield about the same results. Table 7 reports the tests for the unsaturated model. The results support the economic consideration that the previous adoption of an FMS rises the probability of adoption of a CADCAM, while in the opposite direction the evidence is much weaker, and statistically non significant. In fact, the economic intuition suggests that CADCAM is a powerful design tool even without an FMS. The strong simultaneous dependence suggest that the decision of joint simultaneous adoption occurs more often than what would be expected if the two decisions were taken independently.

8 Conclusions

In this paper we make a step towards rendering an important tool of applied macroeconometric analysis, such as Granger non-causality, available and operational for those situations in which the processes involved in the analysis are

binary, as often happens in microeconometric analysis. The paper is grounded on a rigorous mathematical definition of non-causality, which is shown to be easily fitted into a dynamic version of the bivariate probit model. Particular attention is placed in including covariates into the analysis, and in specializing the definitions for longitudinal data sets. Panel type data for heterogeneous individuals are in fact typical in microeconometric applications.

The paper implicitly suggests so many extensions to fill up a research agenda. To make some examples: a more parsimonious representation of the dynamics; allowing for unobserved heterogeneity; generalizing to a multivariate setting; mixing binary and continuous variables; analyzing the impact of time aggregation.

References

[1] Aalen, O.O., 1987, Dynamic Modelling and Causality, *Scandinavian Actuarial Journal*, 177-190

[2] Amemiya, T., 1981, Qualitative Response Models: A Survey, *Journal of Economic Literature*, **XIX**, 1483-1536

[3] Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N., 1993, *Statistical Models Based on Counting Processes*, Springer Verlag, Berlin

[4] Arellano, M., and Carrasco, R., 1996, Binary Choice Panel Data Models with Predetermined Variables, CEMFI, Working paper No. 9618

[5] Ashford, J.R., Sowden, R.R., 1970, Multi-Variate Probit Analysis, *Biometrics*, 535-546

[6] Billingsley, P., 1986, *Probability and Measure, 2nd Ed.*, Wiley, New York

[7] Chamberlain, G., 1982, The General Equivalence of Granger and Sims Causality, *Econometrica*, **50**, 3, 569-581

[8] Colombo, M., Mosconi, R., 1995, Complementarity and Cumulative Learning Effects in the Early Diffusion of Multiple Technologies, *Journal of Industrial Economics*, **XLIII**, 13-48

[9] Cox, D.R., 1972, Regression Models and Life Tables, *Journal of the Royal Statistical Society, Series B*, **34**, 187-220

[10] Engle, R.F., Hendry, D.F., Richard, J.F., 1983, Exogeneity, *Econometrica*, **51**, 277-304

[11] Florens, J-P., Fougère, D., 1996, Noncausality in Continuous Time, *Econometrica*, **64**, 5, 1195-1212

[12] Florens, J-P., Mouchart, M., 1982, A Note on Noncausality, *Econometrica*, **50**, 3, 583-591

[13] Geweke, J., 1984, Inference and Causality in Economic Time Series Models, in Griliches, Z. and Intriligator, M.D., *Handbook of Econometrics*, Vol. 2, Ch. 19, Horth-Holland, Amsterdam

[14] Gouriéroux, C., Monfort, A., Renault, E., 1987, Kullback causality measures, *Ann. Économ. Statist.*, **6-7**, 369–410.

[15] Granger, C.W.J., 1969, Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, **37**, 424-438

[16] Granger, C.W.J., 1988, Recent Developments in a Concept of Causality, *Journal of Econometrics*, **39**, 199-211

[17] Heckman, J.J., 1981, Statistical Models for Discrete Panel Data, in Manski, C.F., McFadden, D., *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge (MA)

[18] Honoré, B.E. and Kiriazidou, E., 1997, Panel Data Discrete Choice Models with Lagged Dependent Variables, Working Paper, forthcoming in *Econometrica*

[19] Hougaard, P., 1987, Modelling Multivariate Survival, *Scandinavian Journal of Statistics*, **14**, 291-304

[20] Kalbfleisch, J.D., Prentice, R.L., 1980, *The Statistical Analysis of Failure Time Data*, Wiley, New York

[21] Lillard, L.A., 1993, Simultaneous Equations for Hazards. Marriage Duration and Fertility Timing, *Journal of Econometrics*, **56**, 189-217

[22] Lillard, L.A. and Waite, L.J., 1993, A Joint Model of Marital Childbearing and Marital Disruption, *Demography*, **30**, 653-681

[23] McFadden, D.L., 1984, Econometric Analysis of Qualitative Response Models, in Griliches, Z. and Intriligator, M.D., *Handbook of Econometrics*, Vol. 2, Ch. 24, 1395-1457

[24] Mosconi, R., Sartori, F., Seri, R., 1998, The Effect of Time Aggregation in Analyzing Causality with Survival Data, *Working Paper*, Politecnico di Milano

[25] Schwarz, G., 1978, Estimating the Dimension of a Model, *The Annals of Statistics*, **6**, 461-464

[26] Schweder, T., 1970, Composable Markov Processes, *Journal of Applied Probability*, **7**, 400-410

[27] Styan, G.P.H., 1973, Hadamard Products and Multivariate Statistical Analysis, *Linear Algebra and Its Applications*, **6**, 217-240

[28] Waite, L.J. and Lillard, L.A., 1991, Children and Marital Disruption, *American Journal of Sociology*, **96**, 930-953

A Multivariate Discrete Choice Models

The binary case. Choice among discrete alternatives is a deeply studied topic in modern econometrics. Consider an individual facing a binary choice. This situation may be modelled by assuming the existence of a latent continuous variable, supposed to be an index of the propensity to undertake this decision. A *threshold model* is a model in which the choice of the individual is supposed to be caused by this index crossing a deterministic value.[18] In the following we will consider a threshold of 0, since no restriction is imposed in this setting.

If we indicate with Y_i^* the latent continuous variable, we have:

$$Y_i = \begin{cases} 1, \text{ if } Y_i^* > 0 \\ 0, \text{ otherwise} \end{cases}$$

or

$$Y_i = 1_{\{Y_i^* > 0\}}$$

where $1_{\{\cdot\}}$ is the indicator function. If Y_i^* has a cumulative distribution function $F_{Y^*}(\cdot)$, we can express the probability of Y_i taking a value 1 as:

$$\Pr(Y_i = 1) = \mathbb{E}(Y_i) = \mathbb{E}\left(1_{\{Y_i^* > 0\}}\right) = \Pr(Y_i^* > 0) = 1 - F_{Y^*}(0)$$

Similarly the probability of having a value of 0 is:

$$\Pr(Y_i = 0) = 1 - \Pr(Y_i = 1) = \Pr(Y_i^* \leq 0) = F_{Y^*}(0)$$

A particular case of this general framework is the *latent regression model*, that can be obtained when Y_i^* is a (usually linear) function of observed and unobserved characteristics of the individual and of the environment:

$$\begin{aligned} \mathbb{E}(Y_i^* \mid \boldsymbol{x}_i) &= \beta' \boldsymbol{x}_i \\ Y_i^* &= \beta' \boldsymbol{x}_i + \varepsilon_i \end{aligned}$$

Here $\boldsymbol{x}_i$ is a vector of observed covariates, while ε_i represents the unobserved ones. Assuming that ε_i has a c.d.f. $F_\varepsilon(\cdot)$, the probability of the discrete binary variable Y_i is:

$$\begin{aligned} \Pr(Y_i = 0 \mid \boldsymbol{x}_i) &= \Pr(Y_i^* \leq 0) = \Pr(\beta' \boldsymbol{x}_i + \varepsilon_i \leq 0) \\ &= \Pr(\varepsilon_i \leq -\beta' \boldsymbol{x}_i) = F_\varepsilon(-\beta' \boldsymbol{x}_i) \\ \Pr(Y_i = 1 \mid \boldsymbol{x}_i) &= 1 - \Pr(Y_i = 0) = 1 - F_\varepsilon(-\beta' \boldsymbol{x}_i) \end{aligned}$$

If the density function $f_\varepsilon(\cdot)$ is symmetric around 0, then $F(-y) = 1 - F(y)$ and:

$$\Pr(Y_i = 1 \mid \boldsymbol{x}_i) = 1 - F_\varepsilon(-\beta' \boldsymbol{x}_i) = F_\varepsilon(\beta' \boldsymbol{x}_i)$$

[18] Stochastic threshold are commonly advocated in biometrical response study: here an individual is treated with a deterministic quantity of a reagent; the reaction (death or survival) is caused by the crossing of a stochastic threshold, function of observed (age, etc.) and unobserved (frailty or proneness, etc.) characteristics of the individual. In economics, it seems more natural to suppose the threshold deterministic.

The assumption of symmetry is particularly well suited when there is no natural order in the responses, *i.e.* when 1 and 0 are logically interchangeable.[19] In particular, when ε_i is a normal r.v.

$$\varepsilon_i \sim N\left(0, \sigma^2\right)$$

we recover the standard probit model:

$$\begin{aligned}\Pr\left(Y_i = 0 \mid x_i\right) &= \Phi\left(-\beta' x_i; 0, \sigma^2\right) \\ \Pr\left(Y_i = 1 \mid x_i\right) &= \Phi\left(\beta' x_i; 0, \sigma^2\right)\end{aligned}$$

or, compactly, for $y \in \{0, 1\}$:

$$\Pr\left(Y_i = y \mid x_i\right) = \Phi\left((2 \cdot y - 1) \cdot \beta' x_i; 0, \sigma^2\right)$$

However, in this model β and σ^2 are not separately identified and this can be easily shown considering a new latent continuous random variable $Y_i^\circ = c \cdot Y_i^* = c \cdot (\beta x_i + \varepsilon_i) = \gamma x_i + \eta_i$, with $c > 0$. In fact, in this case:

$$\begin{aligned}\Pr\left(Y_i = 1 \mid x_i\right) &= \Pr\left(Y_i^\circ > 0\right) = \Phi\left(\gamma x_i; 0, (c \cdot \sigma)^2\right) \\ &= \int_{-\infty}^{\gamma x_i} \frac{1}{\sqrt{2\pi (c \cdot \sigma)^2}} \exp\left(-\frac{y^2}{2(c \cdot \sigma)^2}\right) dy \\ &= \int_{-\infty}^{\gamma x_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y/c)^2}{2\sigma^2}\right) d\left(\frac{y}{c}\right) = \int_{-\infty}^{\beta x_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz \\ &= \Pr\left(Y_i^* > 0\right)\end{aligned}$$

To avoid this problem σ^2 is customarily set at the arbitrarily value of 1, and β is evaluated under this condition.

A multivariate extension. The latent regression approach can be generalized to two or more binary decisions.[20] Dropping the i subscript for notational ease, we can suppose that every decision j can be expressed as a function of a latent continuous regression:

$$\begin{aligned}Y_j^* &= \beta_j' x + \varepsilon_j = v_j + \varepsilon_j, \qquad\qquad j = 1, ..., n \\ Y_j &= 1_{\{Y_j^* > 0\}}\end{aligned}$$

[19] This happens in most cases. However, suppose a group of individuals in a controlled trial is observed on a window $[0, T]$: assuming the DGP to be a Cox (1972) Proportional Hazards Model, survival probability in T is the c.d.f. of an Extreme Value Type I or Gumbel distribution. Clearly, it is a skewed distribution.

[20] The first attempt in this direction is due to Ashford and Sowden (1970); several papers have followed this one, expanding the subject and proposing estimation metods different from maximum likelihood. We refer to Amemiya (1981) for a survey of multivariate binary regression models.

or, in compact vectorial form:

$$\begin{aligned} Y^* &= Bx + \varepsilon = v + \varepsilon \\ Y &= \left[1_{\{Y_j^* > 0\}}\right] \end{aligned}$$

where Y^*, Y, v and ε are $(n \times 1)$ vectors of elements Y_j^*, Y_j, $\beta_j x$ and ε_j, whereas $B = [\beta_1, \cdots, \beta_n]'$. Notice that no loss of generality is involved in assuming that the same vector x enters all equations, since some of the coefficients in B may be zero.

The probability of Y_j can be easily calculated by adapting the results of the previous section:

$$\Pr(Y_j = y_j \mid x) = \Phi\left((2 \cdot y_j - 1) \cdot \beta_j x; 0, \sigma^2\right)$$

i.e., the marginal distribution of Y_j is a simple univariate probit model; this result can be very useful in applications, since it shows that the distribution of Y_j depends only on the parameters of the corresponding latent regression Y_j^*.

However, in this case, it is more interesting to express the probability of a vector $y = [y_1, \ldots, y_n]'$ whose elements are only 0's or 1's; in order to do so, define the diagonal matrix D_y

$$D_y = 2 \cdot \operatorname{diag}(y) - I_n$$

Then the probability of y is

$$\begin{aligned} \Pr(Y = y \mid x) &= \Pr(D_y(Bx + \varepsilon) > 0) \\ &= \Pr(D_y B x + D_y \varepsilon > 0) \\ &= \Pr(-D_y \varepsilon < D_y B x) \end{aligned}$$

Now suppose that ε is a random Gaussian vector, so that

$$\varepsilon \sim N(0, \boldsymbol{\Sigma})$$

Then $-D_y \varepsilon$ is itself a random Gaussian vector

$$-D_y \varepsilon \sim N\left(0, D_y \boldsymbol{\Sigma} D_y'\right)$$

and the probability of observing y is:

$$\Pr(Y = y \mid x) = \Phi_n\left(D_y B x; 0, D_y \boldsymbol{\Sigma} D_y'\right)$$

This formula allows to express the probability of y in compact form, involving only an n-dimensional integral parameterized as a function of y. Here, $D_y \boldsymbol{\Sigma} D_y'$ is a positive definite matrix: in fact, consider d_y, a column vector equal to the main diagonal of D_y; then, by Equation (2.11) in Styan (1973)

$$D_y \boldsymbol{\Sigma} D_y' = \left(d_y d_y'\right) * \boldsymbol{\Sigma}$$

where $*$ is the Hadamard product. But $\mathbf{\Sigma}$ and $\left(d_y d_y'\right)$ are both positive definite, and so is their Hadamard product (see Theorem 3.1 in Styan (1973)).

Alternatively, we can write the probabilities $\Pr\left(Y = y \mid x\right)$ as

$$\begin{aligned}\Pr\left(Y = y \mid x\right) &= \Pr\left(D_y\left(Bx + \varepsilon\right) > 0\right) \\ &= \Pr\left(D_y Bx + D_y \varepsilon > 0\right)\end{aligned}$$

where $D_y Bx + D_y \varepsilon$ is a random Gaussian vector

$$D_y Bx + D_y \varepsilon \sim N\left(D_y Bx, D_y \mathbf{\Sigma} D_y'\right)$$

Then the probability of observing y is:

$$\begin{aligned}\Pr\left(Y = y \mid x\right) &= \Pr\left(D_y Bx + D_y \varepsilon > 0\right) \\ &= \int_{\varsigma=0}^{+\infty} \phi_n\left(D_y Bx, D_y \mathbf{\Sigma} D_y'\right) d\varsigma \\ &= \int_{\mathbb{R}^n} 1_{\{\varsigma \in \mathbb{R}_+^n\}} \cdot \phi_n\left(D_y Bx, D_y \mathbf{\Sigma} D_y'\right) d\varsigma \\ &= \mathbb{E}_\varsigma\left(1_{\{\varsigma \in \mathbb{R}_+^n\}}\right)\end{aligned}$$

where $\mathbb{R}_+^n$ is the non-negative orthant of the n-dimensional Euclidean space.

Similar to the univariate case, it is simple to show that this model suffers from an identification problem, insofar as B and $\mathbf{\Sigma}$ cannot be recovered unambiguously. As in the univariate case, the problem is solved by standardizing, which in the multivariate setting amounts at replacing the variance covariance matrix $\mathbf{\Sigma}$ with the correlation matrix R.

B Variables of the Application

This Appendix simply reports the variables used in the application of Section 6. For every year and every variable, we have reproduced the field of the PSID database corresponding to the couple: as an example `V30020` is the field corresponding to the variable `INTERVIEW` in 1969. In parentheses, the selection rule we applied is reported: since for `V30020` a recorded value of 0 corresponds to a missing value, we have selected only the individuals having a recorded value `V30020` greater than 0.

YEAR	INTERVIEW	PERSON NUMBER	RELATIONSHIP TO HEAD	AGE OF INDIVIDUAL	MARRIED PAIRS INDICATOR
1968	V30001	V30002	V30003 (>0,<9)	V30004 (>0,<99)	V30005
1969	V30020 (>0)		V30022 (>0)	V30023 (>0,<99)	V30024
1970	V30043 (>0)		V30045 (>0)	V30046 (>0,<99)	V30047
1971	V30067 (>0)		V30069 (>0)	V30070 (>0,<99)	V30071
1972	V30091 (>0)		V30093 (>0)	V30094 (>0,<99)	V30095
1973	V30117 (>0)		V30119 (>0)	V30120 (>0,<99)	V30121
1974	V30138 (>0)		V30140 (>0)	V30141 (>0,<99)	V30142
1975	V30160 (>0)		V30162 (>0)	V30163 (>0,<99)	V30164
1976	V30188 (>0)		V30190 (>0)	V30191 (>0,<99)	V30192
1977	V30217 (>0)		V30219 (>0)	V30220 (>0,<99)	V30221
1978	V30246 (>0)		V30248 (>0)	V30249 (>0,<99)	V30250
1979	V30283 (>0)		V30285 (>0)	V30286 (>0,<99)	V30287
1980	V30313 (>0)		V30315 (>0)	V30316 (>0,<99)	V30317
1981	V30343 (>0)		V30345 (>0)	V30346 (>0,<99)	V30347
1982	V30373 (>0)		V30375 (>0)	V30376 (>0,<99)	V30377
1983	V30399 (>0)		V30401 (>0)	V30402 (>0,<99)	V30405
1984	V30429 (>0)		V30341 (>0)	V30342 (>0,<99)	V30345
1985	V30463 (>0)		V30465 (>0)	V30466 (>0,<99)	V30469
1986	V30498 (>0)		V30500 (>0)	V30501 (>0,<99)	V30504
1987	V30535 (>0)		V30537 (>0)	V30538 (>0,<99)	V30541
1988	V30570 (>0)		V30572 (>0)	V30573 (>0,<99)	V30576
1989	V30606 (>0)		V30608 (>0)	V30609 (>0,<99)	V30612
1990	V30642 (>0)		V30644 (>0)	V30645 (>0,<99)	V30648
1991	V30689 (>0)		V30691 (>0)	V30692 (>0,<99)	V30695
1992	V30733 (>0)		V30735 (>0)	V30736 (>0,<99)	V30739
1993	V30806 (>0)		V30808 (>0)	V30809 (>0,<99)	V30812

YEAR	YEARS OF SCHOOL COMPLETED	HIGHEST GRADE COMPLETED	COMPLETED EDUCATION	HOURS WORKED	TYPE OF INCOME
1968	V30010 (<99)			V30013 (<9999)	V30011 (<9)
1969				V30034 (<9998)	V30032 (<9)
1970		V30052 (<99)		V30058 (<9999)	V30056 (<9)
1971		V30076 (<99)		V30082 (<9999)	V30080 (<9)
1972		V30100 (<99)	V30110 (<99)	V30107 (<9999)	V30105 (<9)
1973		V30126 (<99)		V0131 (<9999)	V30129 (<9)
1974		V30147 (<99)		V30153 (<7800)	V30150 (<9)
1975		V30169 (<99)	V30181 (<51)	V30177 (<6074)	
1976	V30197 (<99)			V30204 (<6800)	
1977	V30226 (<99)			V30233 (<6000)	
1978	V30255 (<99)			V30270 (<7280)	
1979	V30296 (<99)			V30300 (<9999)	
1980	V30326 (<99)			V30330	
1981	V30356 (<99)			V30360	
1982	V30384 (<99)			V30388	
1983	V30413 (<99)			V30417	
1984	V30443 (<99)			V30447	
1985	V30478 (<99)			V30482	
1986	V30513 (<99)			V30517	
1987	V30549 (<99)			V30553	
1988	V30584 (<99)			V30588	
1989	V30620 (<99)			V30624	
1990	V30657 (<99)			V30661	
1991	V30703 (<99)			V30709	
1992	V30748 (<99)			V30754	
1993	V30820 (<99)			V30823	

YEAR	TYPE OF INCOME	TYPE OF TAXABLE INCOME	MONEY INCOME	TAXABLE INCOME
1968	V30011 (<9)		V30012 (<9999)	
1969	V30032 (<9)		V30033 (<99999)	
1970	V30056 (<9)		V30057 (<99999)	
1971	V30080 (<9)		V30081 (<99999)	
1972	V30105 (<9)		V30106 (<99999)	
1973	V30129 (<9)		V30130 (<99999)	
1974	V30150 (<9)		V30152 (<99999)	
1975		V30171		V30173 (>-9999,<99999)
1976		V30201		V30202 (>-9999,<99999)
1977		V30230		V30231 (>-9999,<99999)
1978		V30267		V30268 (>-9999,<99999)
1979		V30297		V30298 (>-9999,<99999)
1980		V30327		V30328 (>-9999,<99999)
1981		V30357		V30358 (>-9999,<99999)
1982		V30385		V30386 (>-9999,<99999)
1983		V30414		V30415 (>-99999,<999999)
1984		V30444		V30445 (>-99999,<999999)
1985		V30479		V30480 (>-99999,<999999)
1986		V30514		V30515 (>-99999,<999999)
1987		V30550		V30551 (>-99999,<999999)
1988		V30585		V30586 (>-99999,<999999)
1989		V30621		V30622 (>-99999,<999999)
1990		V30658		V30659 (>-99999,<999999)
1991		V30704		
1992		V30749		
1993				

YEAR	TRANSFER INCOME	LABOR INCOME	ASSET INCOME
1968			
1969			
1970			
1971			
1972			
1973			
1974			
1975	V30175 (<99999)		
1976	V30209 (<99999)		
1977	V30238 (<99999)		
1978	V30275 (<99999)		
1979	V30305 (<99999)		
1980	V30335 (<99999)		
1981	V30365 (<99999)		
1982	V30391 (<99999)		
1983	V30420 (<99999)		
1984	V30455 (<99999)		
1985	V30490 (<99999)		
1986	V30525 (<99999)		
1987	V30561 (<99999)		
1988	V30596 (<99999)		
1989	V30632 (<99999)		
1990	V30669 (<99999)		
1991	V30717 (<99999)	V30705 (<999999)	V30707 (>-99999,<999999)
1992	V30762 (<99999)	V30750 (<999999)	V30752 (>-99999,<999999)
1993	V30825 (<999999)	V30821 (<999999)	V30822 (>-99999,<999999)

C Results of the Analysis

In this Appendix we report the maximum likelihood estimates of the models proposed in Section 6.

C.1 One Lag Unsaturated Model with No Simultaneous Dependence

Usable Observations		6805	Degrees of Freedom		6787
Function Value		-2998.7815			
	Variable	Coeff.	Std Error	T-Stat	Signif.
Children bearing equation					
1	CONST	-3.9287607	0.41655370	-9.4315827	4.0394717e-21
2	Y1_1	-0.32050908	0.21250474	-1.5082444	0.13149199
3	Y2_1	0.54406852	0.064495085	8.4358137	3.2890629e-17
4	Y1_1*Y2_1	-0.34575150	0.24271859	-1.4244954	0.15430313
5	AGE	0.23158938	0.032769702	7.0671799	1.5811396e-12
6	AGESQ	-0.53712013	0.059146188	-9.0812299	1.0735483e-19
7	INCOME	0.63728419	0.58817630	1.0834918	0.27859020
8	HOURS	-8.8763739	3.6303022	-2.4450786	0.014482060
9	EDU	0.015035067	0.0078050051	1.9263366	0.054062361
Marriage equation					
1	CONST	-2.4620387	0.42361628	-5.8119549	6.1747486e-09
2	Y1_1	-0.16410528	0.19331724	-0.84889108	0.39594190
3	Y2_1	3.3146366	0.072020815	46.023315	4.4501477e-308
4	Y1_1*Y2_1	-0.031148739	0.24185658	-0.12879013	0.89752372
5	AGE	0.037612061	0.032169351	1.1691893	0.24232737
6	AGESQ	-0.11340404	0.057506010	-1.9720380	0.048605268
7	INCOME	-0.046565355	0.63036349	-0.073870640	0.94111332
8	HOURS	1.2291636	3.8531471	0.31900252	0.74972460
9	EDU	0.081639921	0.0093039187	8.7747888	1.7122495e-18

C.2 Two Lag Unsaturated Model with No Simultaneous Dependence

Usable Observations		6805	Degrees of Freedom		6781
Function Value		-2985.8111			
Variable		Coeff.	Std Error	T-Stat	Signif.
Children bearing equation					
1	CONST	-3.9008416	0.42869897	-9.0992559	9.0952825e-20
2	Y1_1	-0.31333696	0.21574547	-1.4523455	0.14640557
3	Y2_1	0.79788606	0.10477661	7.6151159	2.6345506e-14
4	Y1_1*Y2_1	-0.34135754	0.24644282	-1.3851389	0.16600999
5	Y1_2	0.52122204	0.14298555	3.6452777	0.00026710321
6	Y2_2	-0.26678643	0.10419056	-2.5605624	0.010450290
7	Y1_2*Y2_2	-0.50563460	0.17084695	-2.9595764	0.0030806227
8	AGE	0.22376862	0.033771258	6.6260077	3.4488715e-11
9	AGESQ	-0.51696968	0.060666119	-8.5215552	1.5742500e-17
10	INCOME	0.62729488	0.58394945	1.0742280	0.28272049
11	HOURS	-8.0916553	3.7213930	-2.1743620	0.029677962
12	EDU	0.015170073	0.0078334484	1.9365766	0.052797124
Marriage equation					
1	CONST	-2.4851768	0.42749726	-5.8133162	6.1247208e-09
2	Y1_1	-0.16621892	0.19382431	-0.85757521	0.39112707
3	Y2_1	3.3040134	0.14620983	22.597752	4.5600171e-113
4	Y1_1*Y2_1	-0.038842085	0.24246218	-0.16019853	0.87272469
5	Y1_2	-0.071594182	0.19637191	-0.36458464	0.71542148
6	Y2_2	0.025019494	0.14862745	0.16833697	0.86631819
7	Y1_2*Y2_2	-0.043416742	0.24711266	-0.17569614	0.86053266
8	AGE	0.040037726	0.032427876	1.2346700	0.21695335
9	AGESQ	-0.11850395	0.057914644	-2.0461829	0.040738380
10	INCOME	-0.042200248	0.63209053	-0.066762981	0.94677039
11	HOURS	1.0763038	3.8575362	0.27901328	0.78023464
12	EDU	0.081756990	0.0093644551	8.7305656	2.5340212e-18

C.3 One Lag Saturated Model with No Simultaneous Dependence

Usable Observations		6805	Degrees of Freedom		6757
Function Value		-2914.4933			
Variable		Coeff.	Std Err	T-Stat	Signif.
Children bearing equation					
1	CONST	-4.8735379	0.65039905	-7.4931504	6.7239775e-14
2	AGE	0.31072124	0.053455339	5.8127261	6.1463588e-09
3	AGESQ	-0.65054205	0.10169710	-6.3968592	1.5860540e-10
4	INCOME	0.24132310	0.57987937	0.41616085	0.67729230
5	HOURS	-14.763850	6.5982369	-2.2375447	0.025250759
6	EDU	-0.0018936892	0.0093267122	-0.20303931	0.83910431
7	Y1_1	-10.047147	49.970481	-0.20106165	0.84065037
8	AGE*Y1_1	1.3797660	5.6107123	0.24591636	0.80574698
9	AGESQ*Y1_1	-4.4478974	15.617096	-0.28480951	0.77579009
10	INCOME*Y1_1	8.0161405	47.043394	0.17039885	0.86469647
11	HOURS*Y1_1	-11.391349	122.62530	-0.092895586	0.92598651
12	EDU*Y1_1	-0.031549941	0.062175793	-0.50743126	0.61185226
13	Y2_1	4.2243331	0.97781866	4.3201601	1.5591607e-05
14	AGE*Y2_1	-0.31575156	0.074969275	-4.2117462	2.5340418e-05
15	AGESQ*Y2_1	0.49935018	0.13686845	3.6483951	0.00026388356
16	INCOME*Y2_1	1.6468916	2.0430136	0.80610898	0.42018002
17	HOURS*Y2_1	2.6111429	8.9176786	0.29280522	0.76967103
18	EDU*Y2_1	0.076113598	0.020455341	3.7209646	0.00019846326
19	Y1_1*Y2_1	11.009568	50.118805	0.21966940	0.82612864
20	AGE*Y1_1*Y2_1	-1.5035929	5.6192869	-0.26757717	0.78902480
21	AGESQ*Y1_1*Y2_1	4.7953336	15.629250	0.30681790	0.75898198
22	INCOME*Y1_1*Y2_1	-17.775702	51.021053	-0.34839936	0.72754028
23	HOURS*Y1_1*Y2_1	23.638582	130.62733	0.18096199	0.85639741
24	EDU*Y1_1*Y2_1	-0.030071537	0.10695451	-0.28116192	0.77858621

Marriage equation					
1	CONST	-4.3051076	0.55604436	-7.7423815	9.7571725e-15
2	AGE	0.20807817	0.043255330	4.8104633	1.5058088e-06
3	AGESQ	-0.50483522	0.079661763	-6.3372338	2.3392682e-10
4	INCOME	-0.57009967	0.74761720	-0.76255557	0.44572849
5	HOURS	17.061523	5.4778995	3.1146105	0.0018418798
6	EDU	0.089131232	0.010881296	8.1912332	2.5856309e-16
7	Y1_1	-3.7828244	12.403857	-0.30497163	0.76038776
8	AGE*Y1_1	0.35047798	1.1361518	0.30847814	0.75771853
9	AGESQ*Y1_1	-0.86860562	2.4639995	-0.35251859	0.72444938
10	INCOME*Y1_1	-0.74514310	41.524962	-0.017944462	0.98568316
11	HOURS*Y1_1	-18.635797	77.077420	-0.24178024	0.80895045
12	EDU*Y1_1	0.021238654	0.22577481	0.094070080	0.92505349
13	Y2_1	6.7878177	1.3462539	5.0420040	4.6068155e-07
14	AGE*Y2_1	-0.29346610	0.094372375	-3.1096610	0.0018730217
15	AGESQ*Y2_1	0.65689023	0.16220333	4.0497949	5.1262538e-05
16	INCOME*Y2_1	0.80757461	2.4524051	0.32929902	0.74192969
17	HOURS*Y2_1	-29.132226	9.3311771	-3.1220312	0.0017960793
18	EDU*Y2_1	-0.033579455	0.026885961	-1.2489587	0.21168018
19	Y1_1*Y2_1	-0.34148454	12.674678	-0.026942265	0.97850578
20	AGE*Y1_1*Y2_1	-0.081892745	1.1516904	-0.071106561	0.94331295
21	AGESQ*Y1_1*Y2_1	0.33553799	2.4862817	0.13495574	0.89264687
22	INCOME*Y1_1*Y2_1	4.2785058	44.739994	0.095630451	0.92381408
23	HOURS*Y1_1*Y2_1	8.0711418	87.344049	0.092406317	0.92637522
24	EDU*Y1_1*Y2_1	0.046845289	0.23730446	0.19740585	0.84350995

C.4 One Lag Unsaturated Model

Usable Observations		6805	Degrees of Freedom		6783
Function Value		-2980.7935			
Variable		Coeff.	Std Error	T-Stat	Signif.
Children bearing equation					
1	CONST	-3.8710789	0.41740376	-9.2741831	1.7898520e-20
2	Y1_1	-0.30850117	0.26301707	-1.1729321	0.24082303
3	Y2_1	0.56075765	0.064121485	8.7452379	2.2254615e-18
4	Y1_1*Y2_1	-0.35951336	0.28759497	-1.2500683	0.21127459
5	AGE	0.22642425	0.032884968	6.8853420	5.7648859e-12
6	AGESQ	-0.52605231	0.059449243	-8.8487638	8.8494075e-19
7	INCOME	0.68654642	0.58540116	1.1727794	0.24088428
8	HOURS	-9.4790776	3.6149914	-2.6221577	0.0087374987
9	EDU	0.014061426	0.0078607515	1.7888145	0.073644701
Marriage equation					
1	CONST	-2.4320292	0.42146893	-5.7703642	7.9100396e-09
2	Y1_1	-0.16917955	0.29296006	-0.57748333	0.56361299
3	Y2_1	3.3204214	0.071907156	46.176508	4.4501477e-308
4	Y1_1*Y2_1	-0.027000173	0.32669618	-0.082646125	0.93413292
5	AGE	0.034825314	0.032028056	1.0873377	0.27688758
6	AGESQ	-0.11019258	0.057367692	-1.9208125	0.054755351
7	INCOME	0.021975033	0.62818356	0.034981866	0.97209420
8	HOURS	1.4269067	3.8308028	0.37248243	0.70953369
9	EDU	0.082983945	0.0092095999	9.0105917	2.0494764e-19
Correlation					
1	CONST	0.64444888	0.13631158	4.7277634	2.2700646e-06
2	Y1_1	-2.1714184	433.67551	-0.0050070118	0.99600500
3	Y2_1	-0.38883914	0.23845645	-1.6306505	0.10296409
4	Y1_1*Y2_1	1.0259004	433.67734	0.0023655845	0.99811254

C.5 Two Lag Unsaturated Model

Usable Observations		6805	Degrees of Freedom		6783
Function Value		-2962.6891			
Variable		Coeff.	Std Error	T-Stat	Signif.
Children bearing equation					
1	CONST	-3.8604569	0.43182982	-8.9397647	3.8999996e-19
2	Y1_1	-0.30198455	0.27413569	-1.1015878	0.27064090
3	Y2_1	0.81396602	0.10537835	7.7242242	1.1253659e-14
4	Y1_1*Y2_1	-0.35560938	0.29868616	-1.1905787	0.23381901
5	Y1_2	0.52778634	0.14157346	3.7280033	0.00019300284
6	Y2_2	-0.26319774	0.10488934	-2.5092898	0.012097419
7	Y1_2*Y2_2	-0.52096131	0.16986227	-3.0669631	0.0021624558
8	AGE	0.21994753	0.034025508	6.4641954	1.0183923e-10
9	AGESQ	-0.50903854	0.061128191	-8.3273942	8.2641185e-17
10	INCOME	0.69075523	0.58219222	1.1864728	0.23543564
11	HOURS	-8.9174606	3.7135150	-2.4013531	0.016334567
12	EDU	0.014695627	0.0078375691	1.8750236	0.060789483
Marriage equation					
1	CONST	-2.4666318	0.42754514	-5.7692899	7.9606264e-09
2	Y1_1	-0.17238381	0.27786527	-0.62038629	0.53500349
3	Y2_1	3.3176873	0.15029129	22.075047	5.4906952e-108
4	Y1_1*Y2_1	-0.023736809	0.31445704	-0.075485062	0.93982878
5	Y1_2	-0.090833587	0.19300173	-0.47063614	0.63790060
6	Y2_2	0.011617392	0.15337995	0.075742575	0.93962390
7	Y1_2*Y2_2	-0.0091351286	0.24567103	-0.037184395	0.97033798
8	AGE	0.037996252	0.032418840	1.1720423	0.24118007
9	AGESQ	-0.11614342	0.057971474	-2.0034581	0.045128145
10	INCOME	0.040205691	0.63227099	0.063589333	0.94929723
11	HOURS	0.71454148	3.8742702	0.18443253	0.85367415
12	EDU	0.083478473	0.0092094117	9.0644740	1.2520688e-19
Correlation					
1	CONST	0.62253673	0.14633884	4.2540773	2.0991275e-05
2	Y1_1	-1.7417061	58.443183	-0.029801698	0.97622520
3	Y2_1	-1.4421480	0.58435671	-2.4679241	0.013589915
4	Y1_1*Y2_1	0.54487353	58.446087	0.0093226692	0.99256169
5	Y1_2	0.14190050	0.54791224	0.25898399	0.79564760
6	Y2_2	1.3170857	0.60339931	2.1827763	0.029052290
7	Y1_2*Y2_2	-0.58325856	0.78929053	-0.73896561	0.45992788

C.6 One Lag Saturated Model

Usable Observations		6805	Degrees of Freedom		6753
Function Value		-2894.4002			
Variable		Coeff.	Std Error	T-Stat	Signif
Children bearing equation					
1	CONST	-4.8995224	0.66206553	-7.4003587	1.3581707e-13
2	AGE	0.31243338	0.054773479	5.7040996	1.1695964e-08
3	AGESQ	-0.65308305	0.10501907	-6.2187089	5.0126219e-10
4	INCOME	0.25935497	0.57931629	0.44769149	0.65437587
5	HOURS	-14.755512	6.5331291	-2.2585673	0.023910311
6	EDU	-0.0019310214	0.0092950440	-0.20774742	0.83542619
7	Y1_1	-10.378821	58.787881	-0.17654694	0.85986427
8	AGE*Y1_1	1.4146545	6.6005603	0.21432340	0.83029487
9	AGESQ*Y1_1	-4.5376286	18.387923	-0.24677223	0.80508451
10	INCOME*Y1_1	7.8477387	55.510797	0.14137319	0.88757513
11	HOURS*Y1_1	-10.069800	142.09459	-0.070866881	0.94350370
12	EDU*Y1_1	-0.031486580	0.074604693	-0.42204557	0.67299176
13	Y2_1	4.2552770	0.98626230	4.3145490	1.5992905e-05
14	AGE*Y2_1	-0.31757682	0.075951888	-4.1812893	2.8986072e-05
15	AGESQ*Y2_1	0.50206420	0.13940484	3.6014833	0.00031640677
16	INCOME*Y2_1	1.6456195	2.0432482	0.80539383	0.42059245
17	HOURS*Y2_1	2.5383113	8.8713416	0.28612486	0.77478249
18	EDU*Y2_1	0.075900742	0.020445777	3.7122944	0.00020538886
19	Y1_1*Y2_1	11.402111	58.907510	0.19355955	0.84652078
20	AGE*Y1_1*Y2_1	-1.5472339	6.6074236	-0.23416599	0.81485612
21	AGESQ*Y1_1*Y2_1	4.8979175	18.397419	0.26622851	0.79006322
22	INCOME*Y1_1*Y2_1	-18.132995	58.771818	-0.30853214	0.75767745
23	HOURS*Y1_1*Y2_1	24.095482	148.80461	0.16192699	0.87136335
24	EDU*Y1_1*Y2_1	-0.024302196	0.11818617	-0.20562639	0.83708276

Marriage equation					
1	CONST	-4.3763693	0.55324962	-7.9102978	2.5677386e-15
2	AGE	0.21418401	0.043139929	4.9648671	6.8748145e-07
3	AGESQ	-0.51795580	0.079614934	-6.5057619	7.7300593e-11
4	INCOME	-0.51145351	0.74056524	-0.69062586	0.48980069
5	HOURS	17.289871	5.4319801	3.1829777	0.0014576886
6	EDU	0.089312581	0.010703995	8.3438546	7.1907376e-17
7	Y1_1	-3.7919139	14.189318	-0.26723722	0.78928652
8	AGE*Y1_1	0.35270333	1.3112221	0.26898823	0.78793874
9	AGESQ*Y1_1	-0.87365174	2.7940551	-0.31268236	0.75452200
10	INCOME*Y1_1	-0.43636696	42.413562	-0.010288383	0.99179120
11	HOURS*Y1_1	-18.759856	83.030988	-0.22593801	0.82124963
12	EDU*Y1_1	0.019435893	0.24358026	0.079792561	0.93640224
13	Y2_1	6.8288263	1.3384097	5.1021943	3.3573779e-07
14	AGE*Y2_1	-0.29792580	0.093888851	-3.1731755	0.0015078134
15	AGESQ*Y2_1	0.66746225	0.16147712	4.1334787	3.5731335e-05
16	INCOME*Y2_1	0.74751274	2.4496238	0.30515409	0.76024879
17	HOURS*Y2_1	-29.203653	9.3341623	-3.1286849	0.0017559047
18	EDU*Y2_1	-0.033526863	0.026827225	-1.2497328	0.21139718
19	Y1_1*Y2_1	-0.38054667	14.415970	-0.026397576	0.97894023
20	AGE*Y1_1*Y2_1	-0.077200587	1.3240386	-0.058306899	0.95350417
21	AGESQ*Y1_1*Y2_1	0.32627412	2.8128047	0.11599601	0.90765570
22	INCOME*Y1_1*Y2_1	4.3701346	45.419422	0.096217309	0.92334798
23	HOURS*Y1_1*Y2_1	7.7221368	92.321434	0.083644030	0.93333946
24	EDU*Y1_1*Y2_1	0.045841159	0.25406993	0.18042733	0.85681710
Correlation					
1	CONST	0.70090188	0.13873026	5.0522640	4.3660371e-07
2	Y1_1	-3.0617732	58905.697	-5.1977539e-05	0.99995853
3	Y2_1	-0.42309825	0.25439364	-1.6631636	0.096279688
4	Y1_1*Y2_1	1.8801661	58905.696	3.1918239e-05	0.99997453