DEPARTMENT OF ECONOMICS

Working Paper

Duration Dependence and Time-Varying Variables
in Discrete Time Duration Models

Anna Christina D'Addio and Bo E. Honoré

Working Paper No. 2002-13

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

INSTITUT FOR ØKONOMI

AFDELING FOR NATIONALØKONOMI - AARHUS UNIVERSITET - BYGNING 350
8000 AARHUS C - ☎ 89 42 11 33 - TELEFAX 86 13 63 34

WORKING PAPER

Duration Dependence and Time-Varying Variables
in Discrete Time Duration Models

Anna Christina D'Addio and Bo E. Honoré

Working Paper No. 2002-13

DEPARTMENT OF ECONOMICS

SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Duration dependence and timevarying variables in discrete time duration models*

Anna Cristina D'Addio
CIM-Aarhus School of Business
Department of Economics
University of Aarhus

Bo E. Honoré
Department of Economics
Princeton University

September 2002

Abstract

This paper considers estimation of a dynamic discrete choice model with second order state dependence in the presence of strictly exogenous time–varying explanatory variables. We propose a new method for estimating such models, and a small Monte Carlo study suggests that the method performs well in practice. The method is used to test for duration dependence in labour market spells for French youth. The novelty in the application is that we are able to control for time–varying explanatory variables.

In a discrete time duration model, duration dependent will result in second order state dependence, and the paper therefore also adds to the literature on estimation of duration models with unobserved heterogeneity.

*Mailing addresses: University of Aarhus, Department of Economics, Universitetsparken, building 326, 8000 Aarhus C. Denmark (D'Addio) and Princeton University, Department of Economics, Princeton NJ 08544-1021 (Honoré). The authors gratefully acknowledge financial support from the Marie-Curie fellowships scheme of the European Commission, the National Science Foundation, The Gregory C. Chow Econometric Research Program at Princeton University, and the Danish National Research Foundation (through CAM at The University of Copenhagen). We also thank Marina Sallustro, Aureo de Paula and participants at the Tenth International Conference on Panel Data and at the 2002 European Meetings of the Econometric Society for comments.

1 Introduction

This paper is concerned with estimation and testing of dynamic discrete choice panel data models with second order state dependence. These models are closely related to two–state, discrete time duration models with duration dependence.

An individual who has experienced an event in the past, is frequently more likely to experience the same event in the future than an individual who has not experienced the event. Examples where one might expect this include unemployment, union participation, accident occurrence, purchase decisions, etc. Heckman (1981a, 1981b, 1981c) discusses two explanations for this serial correlation in the context of standard discrete choice threshold crossing models. The first explanation is the presence of *true state dependence*, in which case lagged choices/decisions enter the model in a structural way as explanatory variables. For example, second order state dependence, which is the topic of this paper, refers to the case where the choice probability is allowed to depend on whether the event happened in the two most recent periods. The second source of persistence is the presence of serial correlation in the unobserved error. Heckman calls this source of serial correlation *spurious state dependence.* The serial correlation in the unobserved error is frequently modeled by assuming that the error is composed of a time–invariant component (unobserved heterogeneity) and a time–specific, serially independent component.

Distinguishing between the two sources of persistence is important if one wants to evaluate the effect of economic policies that temporarily change the outcome of the dependent variable. If the serial correlation is due to unobserved heterogeneity, then such a policy will not change future choice probabilities, whereas these will change if the dependence is due to true state dependence.

In this paper, we consider estimation of a discrete choice model that accommodates both second order state dependence and unobserved heterogeneity. Specifically, we considered the model:

$$y_{it} = 1\left\{x_{it}\beta_{y_{i,t-1}} + \delta_{i1}y_{it-1} + \delta_{2,y_{i,t-1}}y_{it-2} + \alpha_i + \varepsilon_{it} > 0\right\} \tag{1}$$

where x_{it} is a vector of strictly exogenous variables for individual i in time–period t, ε_{it} is an unobservable error term, α_i and δ_{i1} are unobservable individual-specific effects, and β_0, β_1, $\delta_{2,0}$ and $\delta_{2,1}$ are the parameters of interest to be estimated or tested. In this model, we allow the effect x_{it} and y_{it} to depend on the lagged value of y_{it}. This is natural in situations where y_{it} is an indicator for whether an individual is in one of two states at time t. In that case, it is natural to allow the transition probabilities (the hazards rates), $P\left(y_{it}=1 \middle| y_{it-1}=0, x_{it}, y_{it-2}\right)$ and $P\left(y_{it}=0 \middle| y_{it-1}=1, x_{it}, y_{it-2}\right)$ to depend differently on x_{it} and y_{it-2}. (1) allows for that. In other situations, for example if one thinks of (1) as the result of some structural model in the spirit of Chintagunta, Kyriazidou, and Perktold (2001), it is natural to restrict β_0 and $\delta_{2,0}$ to equal β_1 and $\delta_{2,1}$, respectively. Therefore, in some of our discussion in this paper, we emphasize this case. Moreover, we will focus on a logit specification in which ε_{it} is i.i.d and logistically distributed, but the approach can also be used to construct estimators and tests for more semiparametric versions of the model.

The hypotheses $\beta_0 = \beta_1 = 0$ and $\delta_{2,0} = \delta_{2,1} = 0$ are of particular interest, and we will discuss how to test these in the model given in (1). On one hand, when both β_0 and β_1 equal 0, it is known that a conditional likelihood approach (see e.g., Chamberlain (1985)) can be used to estimate $\delta_{2,0}$ and $\delta_{2,1}$ and to test hypotheses regarding them. The resulting estimator and tests will have all the usual asymptotic properties such as consistency and root-n asymptotic normality (where n denotes the number of cross sectional units and the number of time periods is assumed to be fixed). On the other hand, when $\delta_{2,0} = \delta_{2,1} = 0$, the model becomes consistent with a two–state duration model with no duration dependence.[1] The reason is that in that case, the probability that an individual is in a given state at time t ($y_{it} = 1$) depends on whether the individual was in that state in the previous period ($y_{it-1} = 1$), but not whether she/he was in the state in periods before to that. Hence, that probability does not depend on the duration of time spent in the state.

The contribution of this paper is twofold. First, in section 2, we propose econometric methods

[1]On the other hand, one might estimate $\delta_{2,0}$ and $\delta_{2,1}$ to be nonzero even if there is no true duration dependence, but the unobserved errors are serially correlated in a way that is more complicated than the one assumed here.

to estimate the model in (1). As mentioned, when $\beta = 0$ a conditional likelihood approach can be used to estimate δ_2. See e.g. Magnac (2000) for a discussion of this. When $\delta_{2,0} = \delta_{2,1} = 0$ and δ_{i1} is constant, the model is similar to the model with first order state dependence discussed in Honoré and Kyriazidou (2000b).

The approach proposed in our paper is based on generalizations of the suggestions in Honoré and Kyriazidou (2000b). Like the estimator proposed by them, our estimator will depend on a bandwidth, which must shrink to zero (as $n \to \infty$) for the estimator to be consistent. As in Honoré and Kyriazidou (2000b), this will prevent our estimator from being root–n consistent. This is in contrast to the conditional likelihood method that one would use to estimate $\delta_{2,0}$ and $\delta_{2,1}$ in (1) if one knows that $\beta = 0$. That approach does not depend on a bandwidth, and it generally leads to a root–n consistent estimator. It is therefore interesting to note that the Wald–test of the hypothesis that $\beta = 0$ in (1) will have its usual χ^2–distribution (under the null), even if one considers asymptotics that holds the bandwidth fixed. A small scale Monte Carlo study presented in section 4 of this paper suggests that the estimator performs well in practice.

The second contribution of this paper is to reconsider the issue of second order state dependence in youth unemployment by estimating (1) using the difference in the (monthly) number of unemployed between t and $t-1$, as the time–varying explanatory variable. We interpret this variable as a proxy for business cycle effects, and finding that it has a significant effect in (1), should be considered as evidence that *some* time varying variable plays a role in (1). A recent paper by Magnac (2000) estimates a model like (1), as well as more complicated multi–state models, to study the dynamics of youth labour market behavior. Unfortunately, existing methods did not allow that paper to control for business cycle effects by including time–varying (macroeconomic) variables. It is exactly this problem that motivates the econometric developments in the next section. We therefore estimate model (1) using the same data as Magnac (2000), but also including the first difference of the number of unemployed in a month as explanatory variable and distinguishing according to the gender and the age. When we assume that the explanatory variables have the same effects on the probability of being unemployed next period for employed and non-employed, the estimate of on the effect of the macroeconomic variable tends to not be significant. When this assumption is removed, we find that the macroeconomic variable is statistically significant (and with the right sign) in some of the sub-samples considered, and that its effects vary according to the state occupied on the labour market. We also find that second order state dependence is important for the overall sample when β and δ do not depend on $y_{i,t-1}$. When those parameters depend on the prior state, the results are somehow different. In most cases duration dependence does not seem to be statistically significant. This is in agreement with most of the literature on youth labour market behaviour and more specifically for French young people (see for instance d'Addio (2000)). These results are discussed in Section 3.

2 Estimation

As explained in Chamberlain (1985), one can test for duration dependence in a duration model with point sampling by considering the hypothesis that $\delta_2 = 0$ in the model

$$P\left(y_{i,t} = 1 | \, y_{i,t-1}, y_{i,t-2}, \alpha_i\right) = \frac{\exp\left(\alpha_i + \delta_{1i} y_{i,t-1} + \delta_2 y_{i,t-2}\right)}{1 + \exp\left(\alpha_i + \delta_{1i} y_{i,t-1} + \delta_2 y_{i,t-2}\right)} \tag{2}$$

This model has been used, for example, by Chay, Hoynes, and Hyslop (2001) to estimate a model for welfare participation, and a multinomial version of it has been used by Magnac (2000) to estimate a model for labour market transitions. In the latter, the coefficient measuring the of duration dependence, δ_2, is allowed to depend on which state the individual is in. In other words, δ_2 in (2) is replaced by $\delta_{2,y_{it-1}}$. This is the model given in (1), but without the explanatory variables, x_{it}.

When $\delta_2 = 0$, (2) is a Markov chain for a given individual, i. Both α_i and δ_{1i} are allowed to differ across individuals in an arbitrary way, which implies that the logit assumption places no restrictions on the transition probabilities. (2) is therefore a nonparametric Markov chain when $\delta_2 = 0$. However, the logit assumption *does* impose restrictions when δ_2 does not equal 0, and the model given by (2) is therefore semiparametric. A parametric version of (2) can be obtained by making distributional assumptions on α_i and δ_{1i}.

The absence of time-varying effects in (2) is a limitation, which is sometimes undesirable in empirical applications. In order to allow for such time–varying effects, it is natural to generalize (2) by allowing the probability to also depend on a set of strictly exogenous time–varying explanatory variables, x_{it}. One way to do this, is to introduce x_{it} as additional explanatory variables in (2)

$$P\left(y_{i,t}=1 \mid y_{i,t-1}, y_{i,t-2}, x_i, \alpha_i\right) = \frac{\exp\left(\alpha_i + x_{i,t}\beta + \delta_{1i} y_{i,t-1} + \delta_2 y_{i,t-2}\right)}{1+\exp\left(\alpha_i + x_{i,t}\beta + \delta_{1i} y_{i,t-1} + \delta_2 y_{i,t-2}\right)} \tag{3}$$

where x_i denotes the set of all the explanatory variables in all time periods for individual i. While (3) is a natural extension of (2), it differs from (2) in that it imposes parametric assumptions even when $\delta_2 = 0$. The model can therefore no longer be considered nonparametric in this case.

It seems natural to let $x_{i,t}$ and $y_{i,t-2}$ appear as in (3) because that makes it a simple generalization of a logit model. However, this functional form is unnatural if one interprets the model as a discrete time, two–state duration model. In that case (3) imposes restrictions between the exit rate from state 1 to state 0 and the exit rate from state 0 to state 1. For example, the same explanatory variables affect the two exit rates, and the relative importance of different explanatory variables is the same in the two rates (because they both depend on $x_{i,t}\beta$). Similarly, the relative importance of duration dependence is the same for the two states. To overcome this, it is therefore interesting to also consider a generalization of (3) that allows β and δ_2 to depend on the $y_{i,t-1}$,

$$P\left(y_{i,t}=1 \mid y_{i,t-1}, y_{i,t-2}, x_i, \alpha_i\right) = \frac{\exp\left(\alpha_i + x_{i,t}\beta_{y_{i,t-1}} + \delta_{1i} y_{i,t-1} + \delta_{2,y_{i,t-1}} y_{i,t-2}\right)}{1+\exp\left(\alpha_i + x_{i,t}\beta_{y_{i,t-1}} + \delta_{1i} y_{i,t-1} + \delta_{2,y_{i,t-1}} y_{i,t-2}\right)} \tag{4}$$

The objects of interest in this paper are δ_2 and β in (3) (or $\delta_{2,0}$, $\delta_{2,1}$, β_0 and β_1 in (4)). The δ_2's capture the degree of duration dependence, and the β's accounts for time–varying explanatory variables. Since the point of departure is (2), it is of particular interest to test $\beta = 0$.

The paper by Honoré and Kyriazidou (2000b) considered a model with first order state dependence which is the same across individuals, i.e. δ_1 is not individual specific. Their paper emphasized the case where β does not depend on the $y_{i,t-1}$. However, it is implicit in their discussion of generalizations to multinomial models that the approach generalizes to models with first order state dependence in which β is allowed to depend on the $y_{i,t-1}$.

Non–Bayesian estimation of non-linear models like (3) and (4) is usually justified by asymptotic arguments. These asymptotic arguments can be based on letting either the number of individuals, n, or the number of time periods, T, (or both) increase to infinity. Since most relevant data sets have many more individuals than time periods, the asymptotic arguments used to justify the proposed estimators of δ_2 and β will be based on letting the number of individuals, n, increase for fixed number of time periods, T.

When specifying and estimating (3) and (4) one has the choice of whether to take a random effects approach in which one specifies a distribution for (α_i, δ_{1i}), or a fixed effects approach, which attempts to estimate the β's and the δ_2's without making any assumptions on the distribution of (α_i, δ_{1i}) and on the way distribution relates to x_{it}. There is a trade-off between these approaches. The main advantage of the random effect approach is that it delivers a completely specified model. This means that one can calculate all probabilities of interest under any "what–if" scenario, provided, of course, that the model remains true. One disadvantage of the approach is that it requires one to specify the distribution of (α_i, δ_{1i}) conditional on the time–varying explanatory variables in all the time–periods. If one assumes that the basic structure of the model is correct no matter how many time–periods one observes, this often leads to specifications that are inconsistent with the observed distribution of the time–varying explanatory variables, unless one assumes that (α_i, δ_{1i}) is independent of the time–varying explanatory variables. See Honoré (2002) for a discussion of this point. A second, and possibly more severe, disadvantage of the approach is the initial conditions problem. Indeed, a random effects approach also needs to specify either the distribution of (α_i, δ_{1i}) conditional on (y_{i1}, y_{i2}, x_i), or the distribution of (y_{i1}, y_{i2}) conditional on (α_i, x_i). If values of y_{it} before the start of the sample were also generated by (3) or by (4), then the relationship between (α_i, δ_{1i}) and (y_{i1}, y_{i2}) would depend on the time–varying variables before the start of the sample in a complicated way. Hence, if one models either the distribution of (α_i, δ_{1i}) conditional on (y_{i1}, y_{i2}, x_i), or the distribution of (y_{i1}, y_{i2}) conditional on $(\alpha_i, \delta_{1i}, x_i)$, one is implicitly modeling the behavior of the time–varying explanatory variables.

There are also severe drawbacks associated with a fixed effects approach like the one proposed in this paper. First, it will not always be possible to estimate a nonlinear model with fixed effects. For example, the approach used here places restrictions on the behavior of support of the time–varying explanatory variable that may not be satisfied, and it is not known how to estimate the model without these restrictions. Secondly, the semiparametric nature of fixed effects models may lead to estimates that are much less precise than the corresponding random effects estimates. Thirdly, and perhaps most seriously, the parameters estimated by the fixed effects approach often do not allow one to calculate objects such as the average effect of the explanatory variable on the probability that y_{it} equals 1 (because this will depend on the distribution of (α_i, δ_{1i})).

In this paper we pursue the fixed effects approach to estimate (3) and (4) , but it should be clear from the discussion above that this will only provide a partial answer to the question of how one should estimate such models.

The key idea behind the construction of estimators of fixed effects panel data models is to find some characteristic of the distribution of some random variable (which can be constructed from the data) that does not depend on the fixed effects. In a textbook linear panel data model with strictly exogenous explanatory variables, this is the conditional mean of the dependent variable minus the individual specific averages of the dependent variable. In the conditional likelihood approach it is the distribution of the dependent variable conditional on the sufficient statistic for the fixed effects.

Our proposed methods for analyzing (3) and (4) are based on the following expressions which are derived in the appendix. They are all probability statements which are satisfied at the true parameter values and which do not depend on the fixed effects.

Define Ξ_{its} to be the sequence of all the y's for individual i, except for y_{it} and y_{is}, $\Xi_{its} = \{y_{i,1}, ..., y_{i,T}\} \setminus \{y_{i,t}, y_{i,s}\}$, where $3 \leq t < s \leq T-2$.

For $t = 3, ..., T-3$, we have

$$P(y_{i,t} = 1 |\, \Xi_{it,t+1}, y_{i,t} \neq y_{i,t+1}, y_{i,t-1} = y_{i,t+2},$$

$$x_{i,t+1}\beta_1 = x_{i,t+2}\beta_1, x_{i,t+1}\beta_0 = x_{i,t+2}\beta_0, x_{i,t+2}\beta_{y_{i,t-1}} = x_{i,t+3}\beta_{y_{i,t-1}})$$

$$= \frac{\exp\left((x_{i,t} - x_{i,t+1})\, b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\,(y_{i,t-2} - y_{i,t+3})\right)}{1 + \exp\left((x_{i,t} - x_{i,t+1})\, b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\,(y_{i,t-2} - y_{i,t+3})\right)} \tag{5}$$

For $t = 3, ..., T-4$

$$P(y_{i,t} = 1 |\, \Xi_{it,t+2}, y_t \neq y_{i,t+2}, y_{i,t-1} = y_{i,t+1} = y_{i,t+3},$$

$$x_{i,t+1}\beta_1 = x_{i,t+3}\beta_1, x_{i,t+1}\beta_0 = x_{i,t+3}\beta_0, x_{i,t+2}\beta_{y_{i,t-1}} = x_{i,t+4}\beta_{y_{i,t-1}})$$

$$= \frac{\exp\left((x_{i,t} - x_{i,t+2})\, b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\,(y_{i,t-2} - y_{i,t+4})\right)}{1 + \exp\left((x_{i,t} - x_{i,t+2})\, b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\,(y_{i,t-2} - y_{i,t+4})\right)} \tag{6}$$

Finally for $t = 3, ..., T-5$ and $s = t+3, ..., T-2$

$$P(y_{i,t} = 1 |\, \Xi_{its}, y_{i,t} \neq y_{i,s}, y_{i,t-1} = y_{i,s-1}, y_{i,t+1} = y_{i,s+1},$$

$$x_{i,t+1}\beta_1 = x_{i,s+1}\beta_1, x_{i,t+1}\beta_0 = x_{i,s+1}\beta_0, x_{i,t+2}\beta_{y_{i,t+1}} = x_{i,t+3}\beta_{y_{i,t+1}})$$

$$= \frac{\exp\left((x_{i,t} - x_{i,s})\, b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\,(y_{i,t-2} - y_{i,s-2}) + d_{2,y_{i,t+1}}\,(y_{i,t+2} - y_{i,s+2})\right)}{1 + \exp\left((x_{i,t} - x_{i,s})\, b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\,(y_{i,t-2} - y_{i,s-2}) + d_{2,y_{i,t+1}}\,(y_{i,t+2} - y_{i,s+2})\right)} \tag{7}$$

Although equations (5), (6) and (7) are derived by brute force in the appendix, they are motivated by a conditional likelihood approach. Consider a version of (4) with no exogenous variables. Such a model could be estimated by the conditional likelihood approach (see Chamberlain (1985)). This would involve conditioning on the first two and the last two observations, the sum of all the observations, as well as $\sum y_{it}y_{it-1}$. Equations (5), (6) and (7) are derived by considering any subset of the data starting at time period $t-2$ and ending in period $s+2$. We then first condition on the first two and the last two observations as well as on the sum of all the observations and $\sum y_{i,t}y_{i,t-1}$, as one would in a model with no exogenous variables. Because of the exogenous variables, this leads to expressions that depend on (α_i, δ_{1i}). To eliminate the terms depending on (α_i, δ_{1i}) one must condition on events related to the exogenous variables. As will be seen shortly, this is costly in terms of the rate of convergence of the proposed estimator. It is therefore desirable to minimize this type of conditioning. The smallest amount of conditioning done in this way is obtained if one further conditions on all but two of the dependent variables (recall that we also condition on the sum of all of the dependent variables, so this is the most conditioning that one can do without making the distribution degenerate). This is why the probabilities in (5), (6) and (7) condition on all the values of y_{it}, except for two. This gives raise to conditioning on events of the type $\Xi_{its}, y_{i,t} \neq y_{i,s}$. When there are no time–varying explanatory variables, one would want to be sure to be implicitly conditioning on $\sum y_{i,t}y_{i,t-1}$ (because then one would have conditioned on the sufficient statistic for (α_i, δ_{1i})). When $s = t+1$, this is achieved by conditioning on $y_{i,t-1} = y_{i,t+2}$. When $s = t + 2$, one would need to condition on $y_{i,t-1} = y_{i,t+3}$. However, for the case where $y_{i,t-1} = y_{i,t+3} \neq y_{i,t+1}$ to deliver expressions that do not depend on (α_i, δ_{1i}), one needs to condition on $x_{i,t+1} = x_{i,t+2} = x_{i,t+3} = x_{i,t+4}$. The curse of dimensionality suggests that this is undesirable (relative to the conditioning in (6)) when x is continuously distributed, and we therefore ignore these terms. When $s > t + 2$, conditioning on $\sum y_{i,t}y_{i,t-1}$, implies that $y_{t-1}+y_{t+1} = y_{s-1}+y_{s+1}$. However, conditioning on either $\{y_{t-1} = 0,\ y_{t+1} = 1,\ y_{s-1} = 1,\ y_{s+1} = 0\}$ or $\{y_{t-1} = 0,\ y_{t+1} = 1,\ y_{s-1} = 1,\ y_{s+1} = 0\}$ leads to expressions that do not involve (α_i, δ_{1i}) (for all values of β_0, β_1, $\delta_{2,0}$ and $\delta_{2,1}$) only if $x_{t+1} = x_{t+2} = x_{s+1} = x_{s+2}$. Since these are based on three equalities rather than two (as in the other expressions), we will not use these expressions. The probabilities in (5), (6) and (7) are therefore the only cases in which α_i and/or δ_{1i} do not appear (for all values of β_0, β_1, $\delta_{2,0}$ and $\delta_{2,1}$), and which require conditioning on only .two events of the type $x_{it} = x_{is}$.

If one assumes that $\beta = 0$, then the $x\beta$ terms in (5), (6) and (7) are all equal by construction and one can use them to set up a "partial" conditional likelihood function in order to estimate δ_2. This would be inefficient relative to the conditional maximum likelihood approach because the additional conditioning eliminates variability which is informative about δ_2. If β does not equal zero, then one can mimic the line of argument in Honoré and Kyriazidou (2000b) and use nonparametric regression techniques to essentially construct a sample analog of the conditional likelihood function based on (5), (6) and (7). The problem with this is that the conditioning sets in these equations depend on β. This can be overcome by noting that the calculations in the appendix also imply the following probability statements.

For $t = 3, ..., T-3,$ we have

$$P\left(y_{i,t}=1 \middle| \Xi_{it,t+1}, y_{i,t} \neq y_{i,t+1}, y_{i,t-1}=y_{i,t+2}, x_{i,t+1}=x_{i,t+2}=x_{i,t+3}\right)$$

$$= \frac{\exp\left(\left(x_{i,t}-x_{i,t+1}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2}-y_{i,t+3}\right)\right)}{1+\exp\left(\left(x_{i,t}-x_{i,t+1}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2}-y_{i,t+3}\right)\right)} \tag{8}$$

For $t = 3, ..., T-4$

$$P\left(y_{i,t}=1 \middle| \Xi_{it,t+2}, y_t \neq y_{i,t+2}, y_{i,t-1}=y_{i,t+1}=y_{i,t+3}, x_{i,t+1}=x_{i,t+3}, x_{i,t+2}=x_{i,t+4}\right)$$

$$= \frac{\exp\left(\left(x_{i,t}-x_{i,t+2}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2}-y_{i,t+4}\right)\right)}{1+\exp\left(\left(x_{i,t}-x_{i,t+2}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2}-y_{i,t+4}\right)\right)} \tag{9}$$

Finally for $t = 3, ..., T-5$ and $s = t+3, ..., T-2$

$$P\left(y_{i,t}=1 \middle| \Xi_{its}, y_{i,t} \neq y_{i,s}, y_{i,t-1}=y_{i,s-1}, y_{i,t+1}=y_{i,s+1}, x_{i,t+1}=x_{i,s+1}, x_{i,t+2}=x_{i,t+3}\right)$$

$$= \frac{\exp\left(\left(x_{i,t}-x_{i,s}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2}-y_{i,s-2}\right) + d_{2,y_{i,t+1}}\left(y_{i,t+2}-y_{i,s+2}\right)\right)}{1+\exp\left(\left(x_{i,t}-x_{i,s}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2}-y_{i,s-2}\right) + d_{2,y_{i,t+1}}\left(y_{i,t+2}-y_{i,s+2}\right)\right)} \tag{10}$$

When x_{it} is continuously distributed, we therefore propose to estimate (β, δ_2) by maximizing[2]

$$\sum_i q_i\left(b, d_2\right) \tag{11}$$

[2]When elements of x_{it} are discrete, it is not necessary to use a kernel for those components of x_{it}.

where

$$q_i = \sum_{t=3}^{T-3} 1\{y_t \neq y_{t+1}\} 1\{y_{i,t-1} = y_{t+2}\} K\left(\frac{x_{i,t+1} - x_{i,t+2}}{h}\right) K\left(\frac{x_{i,t+2} - x_{i,t+3}}{h}\right) \tag{12}$$

$$\log\left(\frac{\exp\left(y_{i,t}\left((x_{i,t} - x_{i,t+1}) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}(y_{i,t-2} - y_{i,t+3})\right)\right)}{1 + \exp\left((x_{i,t} - x_{i,t+1}) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}(y_{i,t-2} - y_{i,t+3})\right)}\right)$$

$$+ \sum_{t=3}^{T-4} 1\{y_t \neq y_{t+2}\} 1\{y_{i,t-1} = y_{t+1} = y_{t+3}\} K\left(\frac{x_{i,t+1} - x_{i,t+3}}{h}\right) K\left(\frac{x_{i,t+2} - x_{i,t+4}}{h}\right)$$

$$\log\left(\frac{\exp\left(y_{i,t}\left((x_{i,t} - x_{i,t+2}) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}(y_{i,t-2} - y_{i,t+4})\right)\right)}{1 + \exp\left((x_{i,t} - x_{i,t+2}) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}(y_{i,t-2} - y_{i,t+4})\right)}\right)$$

$$+ \sum_{t=3}^{T-5} \sum_{s=t+3}^{T-2} 1\{y_t \neq y_s\} 1\{y_{t-1} = y_{s-1}\} 1\{y_{t+1} = y_{s+1}\} K\left(\frac{x_{i,t+1} - x_{i,s+1}}{h}\right) K\left(\frac{x_{i,t+2} - x_{i,s+2}}{h}\right)$$

$$\log\left(\frac{\exp\left(y_{i,t}\left((x_{i,t} - x_{i,s}) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}(y_{i,t-2} - y_{i,s-2}) + d_{2,y_{i,t+1}}(y_{i,t+2} - y_{i,s+2})\right)\right)}{1 + \exp\left((x_{i,t} - x_{i,s}) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}(y_{i,t-2} - y_{i,s-2}) + d_{2,y_{i,t+1}}(y_{i,t+2} - y_{i,s+2})\right)}\right),$$

$K(\cdot)$ is a kernel and h is a bandwidth that approaches 0 as the number of observations increase to ∞. In the empirical application in the next section we choose K to be an Epanichnikov kernel[3] and we experiment with the bandwidth h.

It is interesting to note that the first two sums in q_i separate into a sum of two terms, one that depends on $(b_0, d_{2,0})$ and one that depends on $(b_1, d_{2,1})$. This means that one can estimate the parameters of the two transition probabilities (from state 0 to state 1, and from state 1 to state 0) separately by considering only the first two sums in q_i. On the other hand, if one wants to use all

[3]This kernel is efficient (in a particular sense) in other settings, here we use it primarily because of its simplicity and because it is continuous and has finite support (which means that the many of the terms in the objective function are 0).

the terms in (12) that are informative about $(\delta_{2,0}, \beta_0)$ (or$(\delta_{2,1}, \beta_1)$), then one must simultaneously estimate $\delta_{2,1}$ (or $\delta_{2,0}$)

By arguments similar to those in Honoré and Kyriazidou (2000b),

$$\sqrt{nh^{2k}} \left(\widehat{\theta} - \theta\right) \widetilde{\longrightarrow} N\left(0, \Gamma^{-1} V \Gamma^{-1}\right) \tag{13}$$

under suitable regularity conditions, where $\widehat{\theta}$ and θ denote $\left(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\delta}_{2,0}, \widehat{\delta}_{2,1}\right)$ and $(\beta_0, \beta_1, \delta_{2,0}, \delta_{2,1})$, respectively, and k is the dimensionality of x_{it}. $avar\left(\widehat{\theta}\right) = \frac{1}{nh^{2k}} \Gamma^{-1} V \Gamma^{-1}$ can be estimated by

$$\left(\sum_i q_i'' \left(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\delta}_{2,0}, \widehat{\delta}_{2,1}\right)\right)^{-1} \tag{14}$$

$$\left(\sum_i q_i' \left(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\delta}_{2,0}, \widehat{\delta}_{2,1}\right)^T q_i' \left(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\delta}_{2,0}, \widehat{\delta}_{2,1}\right)\right) \left(\sum_i q_i'' \left(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\delta}_{2,0}, \widehat{\delta}_{2,1}\right)\right)^{-1}.$$

Since the contribution of this paper is to allow for time–varying explanatory variables in models like (2), it is useful to consider a test of $\beta_0 = \beta_1 = 0$. A natural test would be the Wald test based on (13) and (14). Such a test will be justified in the sense that it will have the usual χ^2 –distribution (under the null) even if the bandwidth is a fixed constant (i.e., does not decrease to 0 as the sample size increases). The reason for this is that if the true β_0 and β_1 equal 0, then (for fixed h), $E[q_i]$ would be maximized by making the terms inside the log's equal to the probability that $y_{it} = 1$ (conditional on the events in the indicator function for each of the terms in the sums in q_i). This happens at $b_1 = b_0 = 0$, $d_{2,0} = \delta_{2,0}$ and $d_{2,1} = \delta_{2,1}$. It therefore follows from standard asymptotics for M–estimators that if β equals 0, then the proposed estimator will be consistent and asymptotically normal as $n \to \infty$, but with h (and T) fixed. This in turn implies that the Wald test of the hypothesis $\beta = 0$ that uses $\widehat{avar}\left(\widehat{\theta}\right)$ in (14) will have the usual asymptotic χ^2 –distribution under the null (for h fixed). In other words, under the null, the Wald test has the same asymptotic distribution as it would in a parametric model. Unfortunately, it does not follow that the Wald test is consistent for h fixed. This is because the estimator of (β_0, β_1) is not guaranteed to be consistent under the alternative.

A random effects approach to estimating (3) or (4) has to deal with the fact that the model does not specify the distribution of the first two observations conditional on $(x_{i1}, ..., x_{iT}, \alpha_i, \delta_{1i})$. If the first two observations are also generated from (3) or (4), then their distribution (given $(x_{i1}, ..., x_{iT}, \alpha_i, \delta_{1i})$) will depend on the distribution of the explanatory variables in time periods prior to the sample (given $(x_{i1}, ..., x_{iT}, \alpha_i, \delta_{1i})$) which is typically unspecified. This is the initial conditions problem. It is therefore important to note that the probability statements in (5), (6) and (7) all condition on the first two observations for an individual. This is because Ξ_{its} always contains y_{i1} and y_{i2}. This means that the estimators based on (5), (6) and (7) do not suffer from the initial conditions problem.

The difference between basing an estimator on (8), (9) and (10) rather than on (5), (6) and (7) is bigger than it might appear. The reason is that in the latter case, the dimensionality of the nonparametric problem is $2k$, whereas it would be either 2 or 3 in the former (depending on whether $k = 1$ or $k > 1$). The curse of dimensionality implies that this can be very important. One way to exploit (5), (6) and (7) is to note that they can be used to construct moment conditions that are based on conditional probabilities given values of $x_{it}\beta$. However, it is not clear whether those moment conditions identify β. One should therefore combine them with the moment conditions based on conditional probabilities given values of x_{it}. If the former satisfy a local identification condition, then this would lead to an estimator whose rate of convergence is driven by the moment conditions based on the conditional probabilities given values of $x_{it}\beta$ (by an argument similar to that in Honoré and Hu (2000)). Since the application discussed in the next section has only one time–varying explanatory variable, we do not pursue this here.

3 The Empirical Application

In this section we use the methods outlined above to estimate equation (3) and (4) for a dataset composed of French youth. This is the same data set used by Magnac (2000). Since one of the specification used by him is similar to (2), it is especially interesting to test whether $\beta = 0$ in (3) and $\beta_0 = \beta_1 = 0$ in (4).

The data are extracted from the 1990-1992 waves of the French Labour Force Survey and an additional survey held in 1992 (Module Jeunes) focusing on individuals and their family background since they were 16 years old up to the survey date.

The French Labour Force Survey is a rotating panel on three years concerning approximately 60,000 households. One third of the sample is renewed each year implying that 20,000 households are present in the survey at three successive dates. The sub–sample used here consists of 5,824 young individuals aged between 18 and 29 in 1992. More specifically, it contains information about their histories on the labour market for the period going from January 1989 to March 1992. Surveys took place at three dates, January 1990, March 1991 and March 1992. At each survey date, the interviewer attempted to rebuild the individuals' labour market history through questions about their activities in each month of the previous year. The interviewer also asked about the current labour market activity of the individual. As a result of this, the information about the month of February 90 is missing.

The survey sampling scheme makes spells in various states left–censored at the beginning of the observation period and right–censored at the end. This complications matters if one models the durations using standard continuous time duration models. See d'Addio and Rosholm (2002). However, as discussed in the previous section, the initial conditions problem (which is similar to the problem of left censoring in duration models) plays no role for the approach proposed here. Right–censoring is also not a problem, provided that the censoring time is exogenous. This is clearly the case here, since censoring time is the final survey date.

For the empirical application, the dependent variable is defined to be 1 if the individual is unemployed or out of the labour force, and 0 otherwise. This differs from Magnac (2000) who estimated a multinomial model with a more disaggregated definition of the labour market states. If an individual is in school at the start of the survey, we ignore the data until the moment she/he enters the labour market. Later periods of schooling are treated as employment.

Table 1 presents summary statistics regarding the dynamic behavior of the dependent variable

(non-employment)

Table 1: Transition probabilities

$\Pr(y_t = 1 \mid y_{t-1} = 0)$	2.38
$\Pr(y_t = 1 \mid y_{t-1} = 1)$	92.68
$\Pr(y_t = 1 \mid y_{t-1} = 0, y_{t-2} = 0)$	2.26
$\Pr(y_t = 1 \mid y_{t-1} = 0, y_{t-2} = 1)$	7.61
$\Pr(y_t = 1 \mid y_{t-1} = 1, y_{t-2} = 0)$	84.90
$\Pr(y_t = 1 \mid y_{t-1} = 1, y_{t-2} = 1)$	93.10

Although Table 1 does not control for individual–specific heterogeneity, it appears that state dependence of order 1 (see the first two lines of Table 1) and order 2 (the last four lines of Table 1) is important in the non-employment behaviour of young individuals. Without first order state dependence, $\Pr(y_t = 1|y_{t-1} = 0)$ would equal $\Pr(y_t = 1|y_{t-1} = 1)$, but, as one would expect, the probability of nonemployment is much higher for an individual who was not employed in the previous month. Without second order state dependence, $\Pr(y_t = 1|y_{t-1} = 0, y_{t-2} = 0)$ would equal $\Pr(y_t = 1|y_{t-1} = 0, y_{t-2} = 1)$, and $\Pr(y_t = 1|y_{t-1} = 1, y_{t-2} = 0)$ would equal $\Pr(y_t = 1|y_{t-1} = 1, y_{t-2} = 1)$. However, the data clearly suggest that the employment status in time period $t-2$ plays a role for the employment history in period t, and this role is consistent with a decreasing hazard in both employment and unemployment.

As mentioned earlier, it is interesting to control for macro–economic business cycle effects when studying unemployment. Here, we use the difference between t and $t-1$ in the number of French unemployed as the time–varying explanatory variable. Strictly speaking, this variable will not satisfy the regularity conditions needed for the consistency of the estimator discussed in the previous section because it does not vary by individual. However, the Wald tests will have the correct asymptotic distribution (under the null).

Figure 1 shows the time–path for the difference between t and $t-1$ in the number of French unemployed over the relevant time–period.

[insert Fig. 1 here]

We think of this variable as a proxy for broad business cycle effects and more particularly for labour market conditions.

Several arguments suggest that macroeconomic effects (frequently summarized by the overall unemployment rate) may have an important impact on the unemployment of young individuals. For instance, it has been argued that when labour market conditions allow for it, young people are more likely than adults to "shop around" before finding a stable job and to quit voluntarily (see e.g. Blanchflower (1996)). The reason may be that their opportunity cost of changing jobs is lower than the adults'. This phenomenon has been found to be especially important for very young workers and to become less important with age. Other reasons have been advocated to explain the higher turnover of young individuals. For instance, within human capital theory (Becker (1993)), turnover acquires a particular meaning when costs (linked to human capital) for both the firm and the workers appear. One of those costs is related to training. For young people this cost (and the related investment) is lower compared to that supported for the adults, therefore firms are more likely to separate from the former. In addition, firms can more easily fire young workers since the protection offered by the employment legislation to people hired under an apprenticeship or a limited term contract is often very weak. Besides this, the flow of young people into unemployment is often considered as a mechanism of adjustment: during a recession period firms cut at first jobs held by young people to protect those of adults (more experienced). This would confirm the importance of the LIFO (Last In First Out) criterion in the firing decisions of young individuals as suggested by Layard, Nickell, and Jackman (1991) and would imply that, in recession periods, newly employed young individuals would suffer more than the adults. Moreover, especially for young individuals firms face a great uncertainty about their skills. Signaling (Spence (1973)) ranking and filtering (Arrow (1973)) mechanisms can intervene in the hiring process as suggested for example in Blanchard and Diamond (1994). On one hand, young individuals entering the labour market have generally no previous professional experience. Those with better labour market histories will be likely to be more successful; being older (among the young individuals) and having experienced unemployment in the past may affect the labour market prospects, especially in recession periods. On the other hand, within the same cohort, better educated (or experienced) individuals would provide potential employers with the best signals. These mechanisms are latent in dual market theories that account also for various forms of information asymmetries;

Because of this we will consider subsets of the sample based on age. Since the behaviour of young individuals has been shown to differ between men and women, we will also consider subsets of sample based on gender.

All the data used in this paper were collected by the French National Institute of Statistics (INSEE).

3.1 Results

In this section we present the results from estimating (3) and (4)using the data described above. As previously mentioned, it is likely that duration dependence and the effect of the business cycle differ across age groups. One might also expect these to differ according to gender. In order to account for this, we estimate the models using the total sample as well as four different sub–samples. Specifically, we consider the samples

1. The overall sample;
2. Men in the sample aged less than 25;[4]
3. Women in the sample aged less than 25;
4. Men in the sample aged 25 or more;
5. Women in the sample aged 25 or more;

Note that the individuals who turn 25 during the period will be used in both the under 25 and the over 25 samples, but only for the periods for which they were under 25 or over 25, respectively.

[4]This corresponds to the ILO definition of young people.

This makes some of the samples unbalanced, but the number of observations for an individual is exogenous.

The contribution of a single individual to the objective function is a sum of $\binom{T_i-4}{2} = \frac{(T_i-4)(T_i-5)}{2}$ terms (although many of them will be 0), where T_i is the number of observations for individual i. Heuristically, this seems to give too much weight to individuals with a large value of T_i. Following Honoré and Kyriazidou (2000a), we therefore use a weighted version of (3) with weights given by $\frac{1}{T_i-4}$, with the standard errors adjusted appropriately. This does not affect the consistency of the estimator, but while we do not claim the weights to be optimal, we suspect that they will result in an improvement in efficiency over the unweighted estimator. Our motivation for using these weights is that by conditioning on the first two and last two observations, we essentially have $T_i - 4$ terms for each individual. The contribution to the objective function is then defined by all pairwise comparisons of observations taken from those. Hence, we have $\binom{T_i-4}{2}$ terms. The deviations from mean estimator of the standard linear panel data model can also be written as the minimizer of a weighted sum (across individual) of terms that are defined by all pairwise comparisons of observations for that individual. The weight given to an individual in that case is the inverse of the number of observations for that individual minus 1. This is the reason for using $\frac{1}{T_i-4}$ as the weight.

The estimator defined by minimizing (11) depends on a bandwidth and a kernel to be chosen by the researcher. The choice of kernel is usually less critical than the choice of bandwidth in applications of semi– and nonparametric methods. We therefore use only one kernel, which is the Epanichnikov kernel given by

$$K(u) = \max\left\{0, 1 - u^2\right\} \tag{15}$$

The fact that $K(\cdot)$ has bounded support implies that many of the terms in the objective function are 0. This make it computationally much more tractable than, say, a normal kernel. Since we expect that the choice of bandwidth is more important than the choice of kernel, we use three different values of the bandwidth h.

The results from estimating (3) and (4) using the method described in Section 2 are presented in Table 2 and in Table 3 respectively. Tables 2 also presents results from a linear probability model with individual specific intercepts, and from a logit model that treats α_i as parameters to be estimated,[5]

$$\left(\widehat{\delta}_1, \widehat{\delta}_2, \widehat{\beta}\right) = \underset{d_1,d_2,b}{\operatorname{argmin}} \sum_{i=1}^{n} \max_{a_i} \sum_t \big\{ y_{it} \log\left(\Lambda\left(x_{it}b + d_1 y_{i,t-1} + d_2 y_{i,t-2} + a_i\right)\right) \tag{16}$$

$$+ \left(1 - y_{it}\right) \log\left(1 - \Lambda\left(x_{it}b + d_1 y_{i,t-1} + d_2 y_{i,t-2} + a_i\right)\right)\big\}$$

Table 3 presents results from a linear probability model with individual specific intercepts and coefficients on $y_{i.t-1}$ and with y_{it-2} and x_{it} interacted with y_{it-1}, and from a logit model that treats α_i and δ_{i1} as parameters to be estimated and allows the coefficients on y_{it-2} and x_{it} to depends on y_{it-1},

$$\left(\widehat{\delta}_{20}, \widehat{\delta}_{21}, \widehat{\beta}_0, \widehat{\beta}_1\right) = \underset{d_{20},d_{21},b_0,b_1}{\operatorname{argmin}} \sum_{i=1}^{n} \max_{a_i,\delta_{1i}} \sum_t \big\{ y_{it} \log\left(\Lambda\left(x_{it}b_{y_{i,t-1}} + \delta_{1i} y_{i,t-1} + \delta_{2,y_{i,t-1}} y_{i,t-2} + a_i\right)\right) \tag{17}$$

$$+ \left(1 - y_{it}\right) \log\left(1 - \Lambda\left(x_{it}b_{y_{i,t-1}} + \delta_{1i} y_{i,t-1} + \delta_{2,y_{i,t-1}} y_{i,t-2} + a_i\right)\right)\big\}$$

As mentioned in the previous section, y_{it} is missing for the month February 1990. Since this is exogenous, the sums in (12) can be replaced by the same sums excluding terms that involve February of 1990 without affecting the asymptotic properties of the estimator (except for the loss of efficiency). For the linear probability model and the maximum likelihood estimator defined in (16) and (17), we also ignore terms that involve February of 1990. This also will not affect the interpretation of our results, although it does mean that (16) and (17) are not the maximum likelihood estimators (but rather a quasi maximum likelihood estimators with the same properties).

Both the linear probability model and the model estimated by (16) assume that δ_1 is constant across individuals, and the estimators would be consistent as $T \to \infty$ for fixed n. One should

[5] We calculate the asymptotic standard errors of the estimator by treating it as an m—estimator of the form $\widehat{\beta} = \operatorname{argmin}_b \sum_{i=1}^{n} q_i$, where q_i is the term inside $\{\cdot\}$ in 16.

therefore not necessarily expect similar results across the three models. One should also expect the scales of the estimates obtained from (16) to differ from those estimated by the linear probability model because the two models make different scale normalizations.

[Table 2 to be inserted here]

[Table 3 to be inserted here]

The results in Table 2 are as one would expect. An increase in the aggregate number of unemployed increases the probability that an individual is unemployed. This is true for all the estimators and all the subsamples, except the estimates from (3) for men over 25 and for women over 25 for one of the bandwidths (and the estimates are insignificant in those cases). Takes as a whole the results in Table 2 also suggest negative[6] duration dependence (with the exceptions being insignificant). Specifically, the probability of being unemployed (or employed) is lower if the person was unemployed (employed) two weeks earlier. The results in Table 3 that allow for the coefficients on the change in the number of unemployed and on y_{it-2} to depend on y_{it-1} are somewhat more interesting. First, it is clear that coefficients differ depending on y_{it-1}. This is not at all surprising, since there is no reason why the hazard for employment and unemployment spells should be related in the way that is enforced by (3). Secondly, it appears that the estimates are unstable across the different subsamples. Presumably, the reason for this is that the effective samples size (the number of nonzero terms in the objective function) is small for some of the subsamples. We therefore focus on the results for the full sample. Here, the effect of the macroeconomic variable is significant with the expected sign for the people who are currently employed and insignificant for the people who are unemployed. The estimates for the linear probability model and the estimates based on (16) and (17) imply *positive* duration dependence whereas the estimates based on (4) suggest no or positive duration dependence. The Monte Carlo results reported in the next section suggest that the incidental parameters problem which is associated with the linear probability model and

[6] Note that a *positive* coefficient on $y_{i,t-2}$ is associated with *negative* duration dependence.

the estimates based on (16) and (17), results in severe bias in the direction of positive duration dependence. This is consistent with the finding in Table 3.

As a general statement we would expect that higher values of h will lead to more biased but less variable estimators. It is impossible to comment on the bias based on Table 2 and Table 3 (because the true data–generating process is unknown), but the estimated variances are consistent with the standard errors of the estimator being bigger when h is small.

4 Monte Carlo Investigation

In this section we report the results of a small Monte Carlo study designed to investigate the behavior of the estimator defined in equation (11) as well as the maximum likelihood estimator defined in (16). We focus on the models (and estimators) that restrict the coefficients on x_{it} and y_{it-2} to not depend on y_{it-1}. The reason for this is that it makes the model more similar to a standard logit model and the Monte Carlo study presented here more comparable to the one in Honoré and Kyriazidou (2000b).

We consider four versions of the model

$$P\left(y_{i,t}=1 \middle| y_{i,t-1},y_{i,t-2},x_i,\alpha_i\right)=\frac{\exp\left(\alpha_i+x_{i,t}\beta+\delta_{1i}y_{i,t-1}+\delta_2 y_{i,t-2}\right)}{1+\exp\left(\alpha_i+x_{i,t}\beta+\delta_{1i}y_{i,t-1}+\delta_2 y_{i,t-2}\right)} \tag{18}$$

All designs have x_{it} distributed as independent normal random variables with mean zero and variance 2. This choice is based on convenience. The other common features of (18) are that $\beta=1$ and $\delta_2=1$ in all the designs. All the data sets have $n=1000$, and data is generated from (18) for time periods 1 to $T+10$, where $y_{i,1}$ and $y_{i,2}$ are generated from (18) with $y_{i,0}=0$ and $y_{i,-1}=0$. Only the last T time periods are used for the estimation. This essentially means that the initial observations are drawn for the stationary distribution of (y_{it},y_{it+1}). The differences in the designs are summarized in Table 4.

Table 4: Monte Carlo Designs

Design 1	$\delta_{1i} = 1$	$\alpha_i = 0$
Design 2	$\delta_{1i} \sim N(1,1)$	$\alpha_i = 0$
Design 3	$\delta_{1i} = 1$	$\alpha_i \sim N(1,1)$
Design 4	$\delta_{1i} \sim N(1,1)$	$\alpha_i \sim N(1,1)$

For each of the designs we generate 500 datasets for both $T = 10$ and for $T = 20$. For each dataset, we estimate the β and δ_2 using the estimator defined in maximizing (11) and (16). For the latter, we also obtain estimates of δ_1. Since one might suspect that the estimator (11) is sensitive to the choice of bandwidth, we calculate the estimator using three values of h, with $h = 0.5$, $h = 1.0$, $h = 1.5$. We use the Epanichnikov kernel in (15).

Tables 4 and 5 report the mean bias and root mean square error of the three estimators of δ_2 and β for each of the eight designs. Those tables also report the median bias and median absolute errors of the estimators. The results for $T = 10$ are in Table 4, and the results for $T = 20$ are in Table 5.

[Tables 5 and 6 to be inserted here]

A number of interesting patterns emerge from Tables 5 and 6,

- The estimator proposed in this paper generally performs much better than the fixed effects maximum likelihood estimator that treats the individual–specific effects as parameters to be estimated.
- While all the estimators improve when T is increased from 10 to 20, the improvement is particularly pronounced for the bias of the fixed effects estimator.
- The performance of the estimators does not vary dramatically with the sample design. This is presumably because the individual specific effects and the state dependence parameter are

relatively small compared to the exogenous explanatory variables and to the (implicit) error in the logit model.

- The estimator defined in (16) is severely biased in favor of positive duration dependence. We conjecture that this is due to the incidental parameters problem. This would explain why the estimates in the first two columns of Tables 2 and 3 suggest duration dependence parameters that are much lower than the ones based on the estimator proposed here.

5 Conclusion

Existing methods do not allow one to estimate and test for second order state dependence in dynamic discrete choice models with unrestricted individual–specific effects. Building on Honoré and Kyriazidou (2000b), this paper proposes methods for doing this in the context of a logit model. We discuss the large sample properties of the estimator and a small Monte Carlo study illustrates its performance in finite samples. An extension to the semiparametric case following the logic of Honoré and Kyriazidou (2000b) is relatively straightforward, and the resulting maximum score estimator based on Manski (1987) would be consistent.

The paper also applies the method to estimate a simple dynamic discrete choice model of youth unemployment which allows for a time–varying macroeconomic explanatory variable. The results suggest that such variables are indeed important in practice.

6 Appendix: Derivation of objective function

Recall that the model is

$$P\left(y_{i,t}=1 \middle| y_{i,t-1},y_{i,t-2},x_i,\alpha_i\right)=\frac{\exp\left(\alpha_i+x_{i,t}\beta_{y_{t-1}}+\delta_{1i}y_{t-1}+\delta_{2,y_{t-1}}y_{t-2}\right)}{1+\exp\left(\alpha_i+x_{i,t}\beta_{y_{t-1}}+\delta_{1i}y_{t-1}+\delta_{2,y_{t-1}}y_{t-2}\right)}$$

an that the estimation of $(\beta_0,\beta_1,\delta_{2,0},\delta_{2,1})$ is based on minimizing

$$\sum_i q_i\left(b,d_2\right)$$

where

$$q_i = \sum_{t=3}^{T-3} 1\left\{y_t \neq y_{t+1}\right\} 1\left\{y_{i,t-1} = y_{t+2}\right\} K\left(\frac{x_{i,t+1} - x_{i,t+2}}{h}\right) K\left(\frac{x_{i,t+2} - x_{i,t+3}}{h}\right)$$

$$\log\left(\frac{\exp\left(y_{i,t}\left(\left(x_{i,t} - x_{i,t+1}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2} - y_{i,t+3}\right)\right)\right)}{1 + \exp\left(\left(x_{i,t} - x_{i,t+1}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2} - y_{i,t+3}\right)\right)}\right)$$

$$+ \sum_{t=3}^{T-4} 1\left\{y_t \neq y_{t+2}\right\} 1\left\{y_{i,t-1} = y_{t+1} = y_{t+3}\right\} K\left(\frac{x_{i,t+1} - x_{i,t+3}}{h}\right) K\left(\frac{x_{i,t+2} - x_{i,t+4}}{h}\right)$$

$$\log\left(\frac{\exp\left(y_{i,t}\left(\left(x_{i,t} - x_{i,t+2}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2} - y_{i,t+4}\right)\right)\right)}{1 + \exp\left(\left(x_{i,t} - x_{i,t+2}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2} - y_{i,t+4}\right)\right)}\right)$$

$$+ \sum_{t=3}^{T-5} \sum_{s=t+3}^{T-2} 1\left\{y_t \neq y_s\right\} 1\left\{y_{t-1} = y_{s-1}\right\} 1\left\{y_{t+1} = y_{s+1}\right\} K\left(\frac{x_{i,t+1} - x_{i,s+1}}{h}\right) K\left(\frac{x_{i,t+2} - x_{i,s+2}}{h}\right)$$

$$\log\left(\frac{\exp\left(y_{i,t}\left(\left(x_{i,t} - x_{i,s}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2} - y_{i,s-2}\right) + d_{2,y_{i,t+1}}\left(y_{i,t+2} - y_{i,s+2}\right)\right)\right)}{1 + \exp\left(\left(x_{i,t} - x_{i,s}\right) b_{y_{i,t-1}} + d_{2,y_{i,t-1}}\left(y_{i,t-2} - y_{i,s-2}\right) + d_{2,y_{i,t+1}}\left(y_{i,t+2} - y_{i,s+2}\right)\right)}\right),$$

$K(\cdot)$ is a Kernel and h is a bandwidth which will in principle depend on the sample size.

The objective function is defined by considering two sequences, A and B, each sequence is of length $T \geq 6$. The two sequences differ only in the t^{th} and s^{th} coordinate, where $2 < t < s < T-1$.

We will now justify the objective function above by considering three cases based on whether t and s differ by one, two or more than two. In each case we will compare $P\left(A \mid x_{i,1}, ..., x_{i,T}, \alpha_i, \delta_{1i}\right)$ to $P\left(B \mid x_{i,1}, ..., x_{i,T}, \alpha_i, \delta_{1i}\right)$. For notational convenience, we will denote these by $P(A)$ and $P(B)$, and we will drop the subscript i on x, y, α and δ_1.

6.1 Case 1. $s = t + 1$

Without loss of generality assume that A has $y_t = 1$, $y_{t+1} = 0$ (otherwise switch A and B)

$$\frac{P(A)}{P(B)} = \frac{F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}} y_{t-2}\right)}{1 - F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}} y_{t-2}\right)} \tag{19}$$

$$\times \frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1} y_{t-1})}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0} y_{t-1})}$$

$$\times \frac{F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})^{y_{t+2}} (1 - F(\alpha + x_{t+2}\beta_0 + \delta_{2,0}))^{1-y_{t+2}}}{F(\alpha + x_{t+2}\beta_1 + \delta_1)^{y_{t+2}} (1 - F(\alpha + x_{t+2}\beta_1 + \delta_1))^{1-y_{t+2}}}$$

$$\times \frac{F\left(\alpha + x_{t+3}\beta_{y_{t+2}} + \delta_1 y_{t+2}\right)^{y_{t+3}} \left(1 - F\left(\alpha + x_{t+3}\beta_{y_{t+2}} + \delta_1 y_{t+2}\right)\right)^{1-y_{t+3}}}{F\left(\alpha + x_{t+3}\beta_{y_{t+2}} + \delta_1 y_{t+2} + \delta_{2,y_{t+2}}\right)^{y_{t+3}} \left(1 - F\left(\alpha + x_{t+3}\beta_{y_{t+2}} + \delta_1 y_{t+2} + \delta_{2,y_{t+2}}\right)\right)^{1-y_{t+3}}}$$

If $y_{t-1} = y_{t+2} = 1$, $x_{t+1}\beta_0 = x_{t+2}\beta_0$ and $x_{t+1}\beta_1 = x_{t+2}\beta_1 = x_{t+3}\beta_1$ then (19) becomes

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1} y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1} y_{t-2})}$$

$$\times \frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})}$$

$$\times \frac{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})}{F(\alpha + x_{t+1}\beta_1 + \delta_1)}$$

$$\times \frac{F(\alpha + x_{t+1}\beta_1 + \delta_1)^{y_{t+3}} (1 - F(\alpha + x_{t+1}\beta_1 + \delta_1))^{1-y_{t+3}}}{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})^{y_{t+3}} (1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1}))^{1-y_{t+3}}}$$

or

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}$$

$$\times \frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})^{y_{t+3}} (1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1}))^{1-y_{t+3}}}$$

$$\times \frac{F(\alpha + x_{t+1}\beta_1 + \delta_1)^{y_{t+3}} (1 - F(\alpha + x_{t+1}\beta_1 + \delta_1))^{1-y_{t+3}}}{F(\alpha + x_{t+1}\beta_1 + \delta_1)}$$

There are then two cases. If $y_{t+3} = 1$ then

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})}$$

If $y_{t+3} = 0$ then

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1)}{F(\alpha + x_{t+1}\beta_1 + \delta_1)}$$

Either way

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1}y_{t+3})}{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1}y_{t+3})} \tag{20}$$

On the other hand, if $y_{t-1} = y_{t+2} = 0$ and $x_{t+1}\beta_0 = x_{t+2}\beta_0 = x_{t+3}\beta_0$ and $x_{t+1}\beta_1 = x_{t+2}\beta_1$then (19) becomes

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_2 y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_2 y_{t-2})}$$

$$\times \frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1)}{F(\alpha + x_{t+1}\beta_0)}$$

$$\times \frac{1 - F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})}{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1)}$$

$$\times \frac{F(\alpha + x_{t+1}\beta_0)^{y_{t+3}} (1 - F(\alpha + x_{t+1}\beta_0))^{1-y_{t+3}}}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})^{y_{t+3}} (1 - F(\alpha + x_{t+1}\beta_0 + \delta_{2,0}))^{1-y_{t+3}}}$$

or

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}$$

$$\times \frac{1 - F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})}{F(\alpha + x_{t+1}\beta_0)}$$

$$\times \frac{F(\alpha + x_{t+1}\beta_0)^{y_{t+3}} (1 - F(\alpha + x_{t+1}\beta_0))^{1-y_{t+3}}}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})^{y_{t+3}} (1 - F(\alpha + x_{t+1}\beta_0 + \delta_{2,0}))^{1-y_{t+3}}}$$

There are again two cases. If $y_{t+3} = 0$

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+1}\beta_0)}{F(\alpha + x_{t+1}\beta_0)}$$

If $y_{t+3} = 1$

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})}$$

Either way

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+1}\beta_0 + \delta_{2,0}y_{t+3})}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0}y_{t+3})} \tag{21}$$

Combining (20) and (21) we get that if $y_{t-1} = y_{t+2}$, $x_{t+1}\beta_{1-y_{t-1}} = x_{t+2}\beta_{1-y_{t-1}}$ and $x_{t+1}\beta_{y_{t-1}} = x_{t+2}\beta_{y_{t-1}} = x_{t+3}\beta_{y_{t-1}}$ then

$$\frac{P(A)}{P(B)} \tag{22}$$

$$= \frac{F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t-2}\right)}{1 - F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t-2}\right)} \times \frac{1 - F\left(\alpha + x_{t+1}\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t+3}\right)}{F\left(\alpha + x_{t+1}\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t+3}\right)}$$

The logit assumption, $F(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$, implies $\frac{F(\eta)}{1-F(\eta)} = \exp(\eta)$, and therefore (22) becomes

$$\frac{P(A)}{P(B)} = \frac{\exp\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t-2}\right)}{\exp\left(\alpha + x_{t+1}\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t+3}\right)} = \exp\left((x_t - x_{t+1})\beta_{y_{t-1}} + \delta_{2,y_{t-1}}(y_{t-2} - y_{t+3})\right)$$

In other words,

$$P(A|A\cup B) = \frac{P(A)}{P(A)+P(B)} = \frac{P(A)/P(B)}{P(A)/P(B)+1}$$

$$= \frac{\exp\Big(\left(x_t - x_{t+1}\right)\beta_{y_{t-1}} + \delta_{2,y_{t-1}}\left(y_{t-2} - y_{t+3}\right)\Big)}{1+\exp\Big(\left(x_t - x_{t+1}\right)\beta_{y_{t-1}} + \delta_{2,y_{t-1}}\left(y_{t-2} - y_{t+3}\right)\Big)}$$

It is easy to see that $y_{t-1} = y_{t+2}$, $x_{t+1}\beta_{1-y_{t-1}} = x_{t+2}\beta_{1-y_{t-1}}$ and $x_{t+1}\beta_{y_{t-1}} = x_{t+2}\beta_{y_{t-1}} = x_{t+3}\beta_{y_{t-1}}$ is the only case in which α and $\delta_1$cancel. In particular without $y_{t-1} = y_{t+2}$, the sum $\sum y_t y_{t-1}$ would differ for the sequences A and B, so in that case conditioning on $A\cup B$ will not condition on what would be the sufficient statistics for α and δ_1 in a model without time–varying explanatory variables.

6.2 Case 2 $s = t + 2$

Without loss of generality assume that A has $y_t = 1$, $y_{t+2} = 0$ (otherwise switch A and B)

$$\frac{P(A)}{P(B)} = \frac{F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}} y_{t-2}\right)}{1 - F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2.y_{t-1}} y_{t-2}\right)} \tag{23}$$

$$\times \left(\frac{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1}y_{t-1})}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0}y_{t-1})}\right)^{y_{t+1}} \left(\frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1}y_{t-1})}{1 - F(\alpha + x_{t+1}\beta_0 + \delta_{2,0}y_{t-1})}\right)^{1-y_{t+1}}$$

$$\times \frac{1 - F\left(\alpha + x_{t+2}\beta_{y_{t+1}} + \delta_1 y_{t+1} + \delta_{2,y_{t+1}}\right)}{F\left(\alpha + x_{t+2}\beta_{y_{t+1}} + \delta_1 y_{t+1}\right)}$$

$$\times \left(\frac{F(\alpha + x_{t+3}\beta_0 + \delta_{2,0}y_{t+1})}{F(\alpha + x_{t+3}\beta_1 + \delta_1 + \delta_{2,1}y_{t+1})}\right)^{y_{t+3}} \left(\frac{1 - F(\alpha + x_{t+3}\beta_0 + \delta_{2,0}y_{t+1})}{1 - F(\alpha + x_{t+3}\beta_1 + \delta_1 + \delta_{2,1}y_{t+1})}\right)^{1-y_{t+3}}$$

$$\times \left(\frac{F\left(\alpha + x_{t+4}\beta_{y_{t+3}} + \delta_1 y_{t+3}\right)}{F\left(\alpha + x_{t+4}\beta_{y_{t+3}} + \delta_1 y_{t+3} + \delta_{2,y_{t+3}}\right)}\right)^{y_{t+4}} \left(\frac{1 - F\left(\alpha + x_{t+4}\beta_{y_{t+3}} + \delta_1 y_{t+3}\right)}{1 - F\left(\alpha + x_{t+4}\beta_{y_{t+3}} + \delta_1 y_{t+3} + \delta_{2,y_{t+3}}\right)}\right)^{1-y_{t+4}}$$

Throughout this section, we will condition on the event $x_{t+1}\beta_1 = x_{t+3}\beta_1$, $x_{t+1}\beta_0 = x_{t+3}\beta_0$ and $x_{t+2}\beta_{y_{t-1}} = x_{t+4}\beta_{y_{t-1}}$, and we will consider two cases.

Consider first the case where $y_{t-1} = y_{t+1} = y_{t+3} = 0$. In this case

If $y_{t+4} = 1$ (23) simplifies to

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}{F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}$$

If $y_{t+4} = 0$ (23) simplifies to

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+2}\beta_0)}{F(\alpha + x_{t+2}\beta_0)}$$

Either way

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})} \times \frac{1 - F(\alpha + x_{t+2}\beta_0 + \delta_{2,0}y_{t+4})}{F(\alpha + x_{t+2}\beta_0 + \delta_{2,0}y_{t+4})}$$

Consider next the case where $y_{t-1} = y_{t+1} = y_{t+3} = 1$. In this case

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0})}\right)$$

$$\times \frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+2}\beta_1 + \delta_1)}$$

$$\times \left(\frac{F(\alpha + x_{t+1}\beta_0 + \delta_2)}{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})}\right)$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1)}{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{y_{t+4}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{1-y_{t+4}}$$

which simplifies to

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})} \frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1}y_{t+4})}{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1}y_{t+4})}$$

Thus if $y_{t-1} = y_{t+1} = y_{t+3}$

$$\frac{P(A)}{P(B)} = \frac{F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t-2}\right)}{1 - F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t-2}\right)} \frac{1 - F\left(\alpha + x_{t+2}\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t+4}\right)}{F\left(\alpha + x_{t+2}\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t+4}\right)}$$

In the case of a the logit model, this becomes

$$\frac{P(A)}{P(B)} = \frac{\exp\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t-2}\right)}{\exp\left(\alpha + x_{t+2}\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}}y_{t+4}\right)} = \exp\left((x_t - x_{t+2})\,\beta_{y_{t-1}} + \delta_{2,y_{t-1}}(y_{t-2} - y_{t+4})\right).$$

The individual specific effect α and δ_1 will only cancel in the expression for $\frac{P(A)}{P(B)}$ if $y_{t-1} = y_{t+3}$. The reason for this is that this is the condition for A and B to have the same value of $\sum y_t y_{t-1}$. The reason why we do not use the terms corresponding to $y_{t-1} = y_{t+3} \neq y_{t+1}$ is that in order to use those, we would also need to condition on $x_{t+1} = x_{t+2} = x_{t+3} = x_{t+4}$.

6.3 Case 3. $s > t + 2$

Without loss of generality assume that A has $y_t = 1$, $y_s = 0$ (otherwise switch A and B)

$$\frac{P(A)}{P(B)} = \frac{F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}} y_{t-2}\right)}{1 - F\left(\alpha + x_t\beta_{y_{t-1}} + \delta_1 y_{t-1} + \delta_{2,y_{t-1}} y_{t-2}\right)}$$

$$\times \left(\frac{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1} y_{t-1})}{F(\alpha + x_{t+1}\beta_0 + \delta_{2,0} y_{t-1})}\right)^{y_{t+1}} \left(\frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1} y_{t-1})}{1 - F(\alpha + x_{t+1}\beta_0 + \delta_{2,0} y_{t-1})}\right)^{1-y_{t+1}}$$

$$\times \left(\frac{F\left(\alpha + x_{t+2}\beta_{y_{t+1}} + \delta_1 y_{t+1} + \delta_{2,y_{t+1}}\right)}{F\left(\alpha + x_{t+2}\beta_{y_{t+1}} + \delta_1 y_{t+1}\right)}\right)^{y_{t+2}} \left(\frac{1 - F\left(\alpha + x_{t+2}\beta_{y_{t+1}} + \delta_1 y_{t+1} + \delta_{2,y_{t+1}}\right)}{1 - F\left(\alpha + x_{t+2}\beta_{y_{t+1}} + \delta_1 y_{t+1}\right)}\right)^{1-y_{t+2}}$$

$$\times \frac{1 - F\left(\alpha + x_s\beta_{y_{s-1}} + \delta_1 y_{s-1} + \delta_{2,y_{s-1}} y_{s-2}\right)}{F\left(\alpha + x_s\beta_{y_{s-1}} + \delta_1 y_{s-1} + \delta_{2,y_{s-1}} y_{s-2}\right)}$$

$$\times \left(\frac{F(\alpha + x_{s+1}\beta_0 + \delta_{2,0} y_{s-1})}{F(\alpha + x_{s+1}\beta_1 + \delta_1 + \delta_{2,1} y_{s-1})}\right)^{y_{s+1}} \left(\frac{1 - F(\alpha + x_{s+1}\beta_0 + \delta_{2,0} y_{s-1})}{1 - F(\alpha + x_{s+1}\beta_1 + \delta_1 + \delta_{2,1} y_{s-1})}\right)^{1-y_{s+1}}$$

$$\times \left(\frac{F\left(\alpha + x_{s+2}\beta_{y_{s+1}} + \delta_1 y_{s+1}\right)}{F\left(\alpha + x_{s+2}\beta_{y_{s+1}} + \delta_1 y_{s+1} + \delta_{2,y_{s+1}}\right)}\right)^{y_{s+2}} \left(\frac{1 - F\left(\alpha + x_{s+2}\beta_{y_{s+1}} + \delta_1 y_{s+1}\right)}{1 - F\left(\alpha + x_{s+2}\beta_{y_{s+1}} + \delta_1 y_{s+1} + \delta_{2,y_{s+1}}\right)}\right)^{1-y_{s+2}}$$

Expressions that do not depend on α can only be obtained when $y_{t-1} + y_{t+1} = y_{s-1} + y_{s+1}$. This can be seen by inspection, or by noting that without time–varying explanatory variables, one would have to condition of $\sum y_t y_{t-1}$, which implies that the sequences A and B must satisfy $y_{t-1} + y_{t+1} = y_{s-1} + y_{s+1}$. There are six such cases.

6.3.1 $y_{t-1}=0,\, y_{t+1}=0,\, y_{s-1}=0,\, y_{s+1}=0$

In this case

$$\frac{P(A)}{P(B)}=\frac{F(\alpha+x_t\beta_0+\delta_{2,0}y_{t-2})}{1-F(\alpha+x_t\beta+\delta_2 y_{t-2})}$$

$$\times\left(\frac{1-F(\alpha+x_{t+1}\beta_1+\delta_1)}{1-F(\alpha+x_{t+1}\beta_0)}\right)$$

$$\times\left(\frac{F(\alpha+x_{t+2}\beta_0+\delta_{2,0})}{F(\alpha+x_{t+2}\beta_0)}\right)^{y_{t+2}}\left(\frac{1-F(\alpha+x_{t+2}\beta_0+\delta_{2,0})}{1-F(\alpha+x_{t+2}\beta_0)}\right)^{1-y_{t+2}}$$

$$\times\frac{1-F(\alpha+x_s\beta_0+\delta_{2,0}y_{s-2})}{F(\alpha+x_s\beta_0+\delta_{2,0}y_{s-2})}$$

$$\times\left(\frac{1-F(\alpha+x_{s+1}\beta_0)}{1-F(\alpha+x_{s+1}\beta_1+\delta_1)}\right)$$

$$\times\left(\frac{F(\alpha+x_{s+2}\beta_0)}{F(\alpha+x_{s+2}\beta_0+\delta_{2,0})}\right)^{y_{s+2}}\left(\frac{1-F(\alpha+x_{s+2}\beta_0)}{1-F(\alpha+x_{s+2}\beta_0+\delta_{2,0})}\right)^{1-y_{s+2}}$$

This simplifies only in the event $x_{t+1}\beta_1 = x_{s+1}\beta_1$, $x_{t+1}\beta_0 = x_{s+1}\beta_0$ and $x_{t+2}\beta_0 = x_{s+2}\beta_0$, in which case

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}{F(\alpha + x_{t+2}\beta_0)}\right)^{y_{t+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}{1 - F(\alpha + x_{t+2}\beta_0)}\right)^{1-y_{t+2}}$$

$$\times \frac{1 - F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}{F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_0)}{F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}\right)^{y_{s+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_0)}{1 - F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}\right)^{1-y_{s+2}}$$

$$= \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})} \times \frac{1 - F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}{F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}$$

$$\times \frac{F(\alpha + x_{t+2}\beta_0 + 1\{y_{t+2} > y_{s+2}\} \cdot \delta_{2,0})}{1 - F(\alpha + x_{t+2}\beta_0 + 1\{y_{t+2} > y_{s+2}\} \cdot \delta_{2,0})} \times \frac{1 - F(\alpha + x_{t+2}\beta_0 + 1\{y_{t+2} < y_{s+2}\} \cdot \delta_{2,0})}{F(\alpha + x_{t+2}\beta_0 + 1\{y_{t+2} < y_{s+2}\} \cdot \delta_{2,0})}$$

In the logit case, this becomes

$$\frac{P(A)}{P(B)} = \exp((x_t - x_s)\beta_0 + \delta_{2,0}(y_{t-2} - y_{s-2} + y_{t+2} - y_{s+2}))$$

or

$$P(A|\,A \cup B) = \frac{\exp((x_t - x_s)\beta_0 + \delta_{2,0}(y_{t-2} - y_{s-2} + y_{t+2} - y_{s+2}))}{1 + \exp((x_t - x_s)\beta_0 + \delta_{2,0}(y_{t-2} - y_{s-2} + y_{t+2} - y_{s+2}))}$$

6.3.2 $y_{t-1} = 0,\, y_{t+1} = 1,\, y_{s-1} = 0,\, y_{s+1} = 1$

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+1}\beta_1 + \delta_1)}{F(\alpha + x_{t+1}\beta_0)}\right)$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{y_{t+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{1-y_{t+2}}$$

$$\times \frac{1 - F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}{F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}$$

$$\times \frac{F(\alpha + x_{s+1}\beta_0)}{F(\alpha + x_{s+1}\beta_1 + \delta_1)}$$

$$\times \left(\frac{F(\alpha + x_{s+2}\beta_1 + \delta_1)}{F(\alpha + x_{s+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{y_{s+2}} \left(\frac{1 - F(\alpha + x_{s+2}\beta_1 + \delta_1)}{1 - F(\alpha + x_{s+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{1-y_{s+2}}$$

This simplifies only if $x_{t+1}\beta_1 = x_{s+1}\beta_1$, $x_{t+1}\beta_0 = x_{s+1}\beta_0$ and $x_{t+2}\beta_1 = x_{s+2}\beta_1$, in which case

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}$$

$$\times \frac{1 - F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}{F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{y_{t+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{1-y_{t+2}}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1)}{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{y_{s+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{1-y_{s+2}}$$

$$= \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})} \times \frac{1 - F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}{F(\alpha + x_s\beta_0 + \delta_{2,0}y_{s-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1)}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)} \frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{1\{y_{s+2} > y_{t+2}\}}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})} \frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)}{F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{1\{y_{s+2} < y_{t+2}\}}$$

In the logit case this becomes

$$\frac{P(A)}{P(B)} = \exp\left((x_t - x_s)\beta_0 + \delta_{2,0}(y_{t-2} - y_{s-2}) + \delta_{2,1}(y_{t+2} - y_{s+2})\right)$$

or

$$P(A|\, A \cup B) = \frac{\exp\left((x_t - x_s)\beta_0 + \delta_{2,0}(y_{t-2} - y_{s-2}) + \delta_{2,1}(y_{t+2} - y_{s+2})\right)}{1 + \exp\left((x_t - x_s)\beta_0 + \delta_{2,0}(y_{t-2} - y_{s-2}) + \delta_{2,1}(y_{t+2} - y_{s+2})\right)}$$

6.3.3 $y_{t-1} = 0$, $y_{t+1} = 1$, $y_{s-1} = 1$, $y_{s+1} = 0$

In this case

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}{1 - F(\alpha + x_t\beta_0 + \delta_{2,0}y_{t-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+1}\beta_1 + \delta_1)}{F(\alpha + x_{t+1}\beta_1)}\right)$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{y_{t+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{1-y_{t+2}}$$

$$\times \frac{1 - F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}{F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}$$

$$\times \frac{1 - F(\alpha + x_{s+1}\beta_0 + \delta_{2,0})}{1 - F(\alpha + x_{s+1}\beta_1 + \delta_1 + \delta_{2,1})}$$

$$\times \left(\frac{F(\alpha + x_{s+2}\beta_0)}{F(\alpha + x_{s+2}\beta_0 + \delta_{2,0})}\right)^{y_{s+2}} \left(\frac{1 - F(\alpha + x_{s+2}\beta_0)}{1 - F(\alpha + x_{s+2}\beta_0 + \delta_2)}\right)^{1-y_{s+2}}$$

In this case, expressions that do not involve α (for all values of β) can only be obtained if equalities across the time periods $t+1$, $t+1$, $s+1$ and $s+2$ are satisfied. Since these are based on three equalities (across time) rather than two (as in the other expressions), we will not use these expressions.

6.3.4 $y_{t-1} = 1,\ y_{t+1} = 0,\ y_{s-1} = 1,\ y_{s+1} = 0$

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}$$

$$\times \frac{1 - F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})}{1 - F(\alpha + x_{t+1}\beta_0 + \delta_2)}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_0 + \delta_2)}{F(\alpha + x_{t+2}\beta_0)}\right)^{y_{t+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_0 + \delta_2)}{1 - F(\alpha + x_{t+2}\beta_0)}\right)^{1-y_{t+2}}$$

$$\times \frac{1 - F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}{F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}$$

$$\times \frac{1 - F(\alpha + x_{s+1}\beta_0 + \delta_{2,0})}{1 - F(\alpha + x_{s+1}\beta_1 + \delta_1 + \delta_{2,1})}$$

$$\times \left(\frac{F(\alpha + x_{s+2}\beta_0)}{F(\alpha + x_{s+2}\beta_0 + \delta_{2,0})}\right)^{y_{s+2}} \left(\frac{1 - F(\alpha + x_{s+2}\beta_0)}{1 - F(\alpha + x_{s+2}\beta_0 + \delta_{2,0})}\right)^{1-y_{s+2}}$$

This simplifies only if $x_{t+1}\beta_1 = x_{s+1}\beta_1$, $x_{t+1}\beta_0 = x_{s+1}\beta_0$ and $x_{t+2}\beta_0 = x_{s+2}\beta_0$, in which case in which case

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}{F(\alpha + x_{t+2}\beta_0)}\right)^{y_{t+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}{1 - F(\alpha + x_{t+2}\beta_0)}\right)^{1-y_{t+2}}$$

$$\times \frac{1 - F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}{F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_0)}{F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}\right)^{y_{s+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_0)}{1 - F(\alpha + x_{t+2}\beta_0 + \delta_{2,0})}\right)^{1-y_{s+2}}$$

In the case of a logit

$$\frac{P(A)}{P(B)} = \exp\left((x_t - x_s)\beta_1 + \delta_{2,1}(y_{t-2} - y_{s-2}) + \delta_{2,0}(y_{t+2} - y_{s+2})\right)$$

or

$$P(A|\, A \cup B) = \frac{\exp\left((x_t - x_s)\beta_1 + \delta_{2,1}(y_{t-2} - y_{s-2}) + \delta_{2,0}(y_{t+2} - y_{s+2})\right)}{1 + \exp\left((x_t - x_s)\beta_1 + \delta_{2,1}(y_{t-2} - y_{s-2}) + \delta_{2,0}(y_{t+2} - y_{s+2})\right)}$$

6.3.5 $y_{t-1} = 1$, $y_{t+1} = 0$, $y_{s-1} = 0$, $y_{s+1} = 1$

As the case $y_{t-1} = 0$, $y_{t+1} = 1$, $y_{s-1} = 1$, $y_{s+1} = 0$, expressions that do not involve α (for all values of β) can only be obtained if equalities across the time periods $t+1$, $t+1$, $s+1$ and $s+2$ are satisfied. Since these are based on three equalities (across time) rather than two (as in the other expressions), we will not use these expressions.

6.3.6 $y_{t-1} = 1$, $y_{t+1} = 1$, $y_{s-1} = 1$, $y_{s+1} = 1$

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+1}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+1}\beta_0 + \delta_2)} \right)$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+2}\beta_1 + \delta_1)} \right)^{y_{t+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)} \right)^{1-y_{t+2}}$$

$$\times \frac{1 - F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}{F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}$$

$$\times \left(\frac{F(\alpha + x_{s+1}\beta_0 + \delta_2)}{F(\alpha + x_{s+1}\beta_1 + \delta_1 + \delta_{2,1})} \right)$$

$$\times \left(\frac{F(\alpha + x_{s+2}\beta_1 + \delta_1)}{F(\alpha + x_{s+2}\beta_1 + \delta_1 + \delta_{2,1})} \right)^{y_{s+2}} \left(\frac{1 - F(\alpha + x_{s+2}\beta_1 + \delta_1)}{1 - F(\alpha + x_{s+2}\beta_1 + \delta_1 + \delta_{2,1})} \right)^{1-y_{s+2}}$$

This simplifies only if $x_{t+1}\beta_1 = x_{s+1}\beta_1$, $x_{t+1}\beta_0 = x_{s+1}\beta_0$ and $x_{t+2}\beta_1 = x_{s+2}\beta_1$, in which case in which case

$$\frac{P(A)}{P(B)} = \frac{F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}{1 - F(\alpha + x_t\beta_1 + \delta_1 + \delta_{2,1}y_{t-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{y_{t+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)}\right)^{1-y_{t+2}}$$

$$\times \frac{1 - F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}{F(\alpha + x_s\beta_1 + \delta_1 + \delta_{2,1}y_{s-2})}$$

$$\times \left(\frac{F(\alpha + x_{t+2}\beta_1 + \delta_1)}{F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{y_{s+2}} \left(\frac{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1)}{1 - F(\alpha + x_{t+2}\beta_1 + \delta_1 + \delta_{2,1})}\right)^{1-y_{s+2}}$$

In the case of a logit

$$\frac{P(A)}{P(B)} = \exp\left((x_t - x_s)\beta_1 + \delta_{2,1}(y_{t-2} - y_{s-2} + y_{t+2} - y_{s-2})\right)$$

or

$$P(A|\, A \cup B) = \frac{\exp\left((x_t - x_s)\beta_1 + \delta_{2,1}(y_{t-2} - y_{s-2} + y_{t+2} - y_{s+2})\right)}{1 + \exp\left((x_t - x_s)\beta_1 + \delta_{2,1}(y_{t-2} - y_{s-2} + y_{t+2} - y_{s+2})\right)}.$$

References

Arrow, K. (1973): "Higher Education as a Filter," *Journal of Public Economics*, 2, 193–216.

Becker, G. (1993): *Human Capital*. The University of Chicago Press.

Blanchard, O., and P. Diamond (1994): "Ranking, Unemployment Duration and Wages," *Review of Economic Studies*, 61, 417–435.

Blanchflower, D. G. (1996): "Youth Labour Markets in Twenty Three Countries: A Comparison using Micro Data," CEP Discussion Paper n. 284, LSE, London.

Chamberlain, G. (1985): "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. Singer, no. 10

in Econometric Society Monographs series,, pp. 3–38. Cambridge University Press, Cambridge, New York and Sydney.

Chay, K. Y., H. Hoynes, and D. Hyslop (2001): "A Non–Experimental Analysis of True State Dependence in Monthly Welfare Participation Sequences," University of California, Berkeley.

Chintagunta, P., E. Kyriazidou, and J. Perktold (2001): "Panel Data Analysis of Household Brand Choices," *Journal of Econometrics*, 103(1-2), 111–53.

d'Addio, A. C. (2000): "Mobility of Young People on the French Labour Market: Methodological Considerations and Empirical Analyses," Collection des Thèses n. 360, Université Catholique de Louvain.

d'Addio, A. C., and M. Rosholm (2002): "Left-Censoring in Duration Data: Theory and Applications," mimeo, University of Aarhus.

Heckman, J. J. (1981a): "Heterogeneity and State Dependence," *Studies in Labor Markets, S. Rosen (ed)*.

——— (1981b): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," *Structural Analysis of Discrete Panel Data with Econometric Applications, C. F. Manski and D. McFadden (eds)*, pp. 179–195.

——— (1981c): "Statistical Models for Discrete Panel Data," *Structural Analysis of Discrete Panel Data with Econometric Applications, C. F. Manski and D. McFadden (eds)*, pp. 114–178.

Honoré, B. E. (2002): "Nonlinear Models with Panel Data," Forthcoming, Portuguese Economic Journal.

Honoré, B. E., and L. Hu (2000): "Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity," Princeton University.

Honoré, B. E., and E. Kyriazidou (2000a): "Estimation of Tobit-Type Models with Individual Specific Effects," *Econometric Reviews*, 19, 341–66.

——— (2000b): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839–874.

Layard, R., S. Nickell, and R. Jackman (1991): *Unemployment, Macroeconomic Performance and the Labour Market*. Oxford University Press.

Magnac, T. (2000): "State Dependence and Unobserved Heterogegenity in Youth Employment Histories," *Economic Journal*, 110, 805–837.

Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–62.

Spence, M. (1973): "Job market Signaling," *Quarterly Journal of Economics*, 87 (3), 355–374.

Table 2: Controlling for the change in the number of unemployed

	LP w. FE	Logit	Based on (3)		
			$h = 0.3$	$h = 0.4$	$h = 0.5$
Everybody ($n = 3590$)					
δ_1	0.755*	4.397*	—	—	—
	0.006	0.074			
δ_2	0.009	0.256*	0.764*	0.672*	0.599*
	0.005	0.071	0.182	0.156	0.140
β	0.006*	0.139*	0.670*	0.646*	0.624*
	0.110	0.033	0.204	0.160	0.134
Young Men ($n = 1625$)					
δ_1	0.743*	4.296*	—	—	—
	0.009	0.106			
δ_2	0.022*	0.253*	0.511	0.395	0.319
	0.008	0.100	0.298	0.246	0.224
β	0.008*	0.120*	0.195	0.394	0.426*
	0.002	0.053	0.290	0.239	0.206
Young Women ($n = 1436$)					
δ_1	0.736*	4.457*	—	—	—
	0.010	0.138			
δ_2	−0.015	−0.213	−0.236	−0.043	0.070
	0.009	0.134	0.338	0.290	0.269
β	0.011*	0.346*	0.047	0.133	0.110
	0.002	0.064	0.335	0.276	0.234

Table 2 (continued)

	LP w. FE	Logit	Based on (3)		
			$h = 0.18$	$h = 0.24$	$h = 0.30$
Men over 25 ($n = 831$)					
δ_1	0.752^* 0.018	4.453^* 0.276	—	—	—
δ_2	0.006 0.016	0.109 0.265	2.075^* 0.828	2.342^* 0.703	2.291^* 0.571
β	0.001 0.002	0.019 0.130	−0.326 0.608	−0.495 0.539	−0.462 0.470
Women over 25 ($n = 894$)					
δ_1	0.744^* 0.015	4.420^* 0.282	—	—	—
δ_2	0.004 0.014	0.052 0.303	0.648 0.657	0.762 0.511	0.822^* 0.423
β	0.000 0.002	0.039 0.139	−0.086 0.549	0.110 0.455	0.213 0.399

Table 3: Controlling for the change in the number of unemployed

	LP w. FE	Logit	Based on (4) $h = 0.3$	$h = 0.4$	$h = 0.5$
Everybody ($n = 3590$)					
$\delta_{2,0}$	-0.062*	-0.890*	-0.259	-0.255	-0.190
	0.006	0.124	0.583	0.415	0.362
$\delta_{2,1}$	-0.067*	-0.594*	0.258	0.911	0.979*
	0.008	0.106	0.579	0.472	0.473
β_0	0.004*	0.297*	1.488	1.385*	1.345*
	0.001	0.058	0.602	0.422	0.355
β_1	0.007*	0.012	-0.597	-0.282	-0.203
	0.003	0.028	0.490	0.416	0.408
Young Men ($n = 1625$)					
$\delta_{2,0}$	-0.066*	-1.079*	-1.526	-0.342	0.006
	0.009	0.206	1.160	0.601	0.531
$\delta_{2,1}$	-0.039*	-0.322*	-1.103	-0.754	-0.641
	0.014	0.151	0.815	0.725	0.650
β_0	0.005*	0.257*	2.589	2.200*	2.062*
	0.002	0.090	1.697	0.875	0.680
β_1	0.009	0.031	-0.126	-0.224	-0.189
	0.006	0.120	0.819	0.666	0.595
Young Women ($n = 1436$)					
$\delta_{2,0}$	-0.080*	-1.402*	-1.040	0.320	0.613
	0.010	0.236	0.885	0.824	0.822
$\delta_{2,1}$	-0.103*	-1.251*	2.1222	1.698	1.805
	0.014	0.217	1.260	1.256	1.233
β_0	0.011*	0.628*	2.191*	1.827*	1.581*
	0.002	0.109	0.858	0.567	0.476
β_1	0.011*	0.194	-2.336*	-2.202*	-2.142*
	0.005	0.115	0.934	0.720	0.647

Table 3 (continued)

	LP w. FE	Logit	Based on (4) $h = 0.3$	$h = 0.4$	$h = 0.5$
Men over 25 ($n = 831$)					
$\delta_{2,0}$	-0.055*	-0.796*	-1.220	0.287	0.705
	0.018	0.350	1.191	0.880	0.870
$\delta_{2,1}$	-0.093*	-1.111*	51.330*	19.650*	19.874*
	0.025	0.440	0.000	0.010	0.008
β_0	0.000	0.053	2.173*	0.954	0.602
	0.001	0.184	1.131	1.092	0.971
β_1	0.003	0.010	-48.215*	-7.240*	-6.385*
	0.013	0.127	0.003	2.897	2.697
Women over 25 ($n = 894$)					
$\delta_{2,0}$	-0.071*	-0.936*	-0.012	-0.412	0.055
	0.016	0.335	1.441	1.281	0.946
$\delta_{2,1}$	-0.088*	-1.067*	1.304	1.453	1.111
	0.019	0.309	0.877	0.875	0.923
β_0	-0.001	0.013*	0.317	0.478	0.874
	0.002	0.033	0.965	0.818	0.743
β_1	-0.003	-0.149	0.689	0.196	0.222
	0.004	0.165	1.749	1.598	1.523

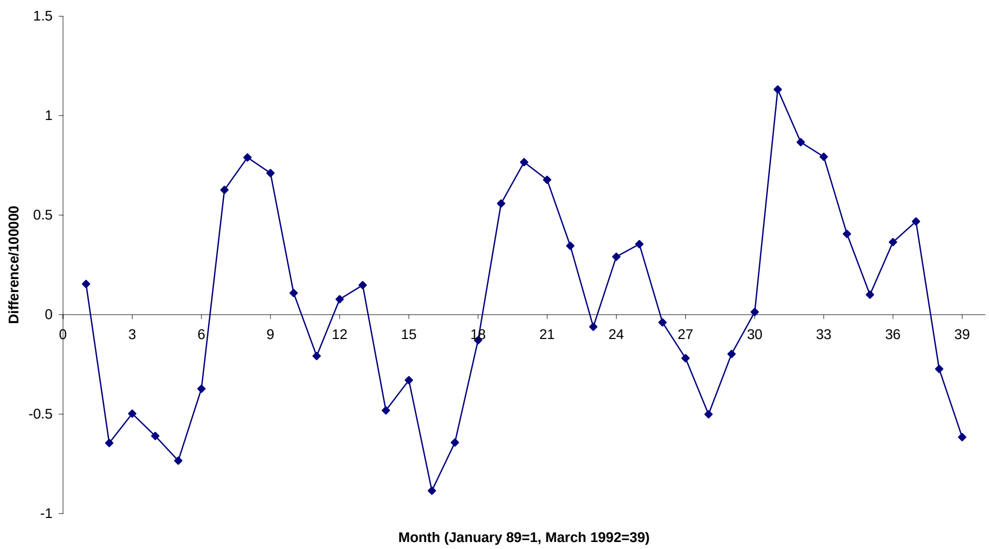
Table 5: Monte Carlo Results ($T = 10$)

	δ_2				β			
	Mean Bias	RMSE	Med. Bias	MAE	Mean Bias	RMSE	Med. Bias	MAE
Design 1								
MLE	−0.605	0.612	−0.610	0.610	0.235	0.239	0.236	0.236
$h = 0.5$	0.680	7.155	0.080	0.689	0.472	4.011	0.128	0.283
$h = 1.0$	0.045	0.510	0.014	0.302	0.056	0.229	0.024	0.131
$h = 1.5$	0.001	0.344	−0.000	0.220	0.034	0.147	0.019	0.089
Design 2								
MLE	−0.508	0.516	−0.510	0.510	0.199	0.202	0.198	0.198
$h = 0.5$	0.747	8.840	0.115	0.688	0.864	7.503	0.112	0.262
$h = 1.0$	0.008	0.480	−0.028	0.317	0.049	0.216	0.030	0.129
$h = 1.5$	−0.012	0.338	−0.018	0.208	0.032	0.146	0.006	0.082
Design 3								
MLE	−0.625	0.632	−0.622	0.622	0.240	0.243	0.238	0.238
$h = 0.5$	0.377	3.023	−0.059	0.728	0.652	4.824	0.115	0.261
$h = 1.0$	−0.001	0.501	−0.027	0.332	0.073	0.220	0.035	0.129
$h = 1.5$	−0.017	0.351	−0.014	0.230	0.048	0.151	0.030	0.091
Design 4								
MLE	−0.517	0.524	−0.522	0.522	0.199	0.203	0.197	0.197
$h = 0.5$	1.224	10.709	0.145	0.739	1.141	9.241	0.140	0.295
$h = 1.0$	0.041	0.513	−0.025	0.328	0.085	0.274	0.039	0.145
$h = 1.5$	−0.005	0.325	−0.026	0.223	0.041	0.156	0.030	0.097

Table 6: Monte Carlo Results ($T = 20$)

	δ_2				β			
	Mean Bias	RMSE	Med. Bias	MAE	Mean Bias	RMSE	Med. Bias	MAE
Design 1								
MLE	−0.248	0.253	−0.248	0.248	0.089	0.091	0.089	0.089
$h = 0.5$	0.009	0.300	−0.007	0.202	0.028	0.151	0.010	0.090
$h = 1.0$	−0.012	0.161	−0.006	0.108	0.005	0.073	0.001	0.049
$h = 1.5$	−0.023	0.122	−0.015	0.077	0.004	0.051	0.001	0.036
Design 2								
MLE	−0.160	0.169	−0.159	0.159	0.057	0.060	0.057	0.057
$h = 0.5$	0.044	0.309	0.026	0.186	0.032	0.142	0.020	0.089
$h = 1.0$	0.003	0.162	−0.010	0.113	0.009	0.074	0.005	0.049
$h = 1.5$	−0.011	0.118	−0.015	0.084	0.005	0.053	0.005	0.035
Design 3								
MLE	−0.256	0.260	−0.257	0.257	0.091	0.093	0.091	0.091
$h = 0.5$	0.024	0.322	0.003	0.217	0.030	0.146	0.018	0.093
$h = 1.0$	0.006	0.176	−0.004	0.121	0.007	0.075	0.002	0.051
$h = 1.5$	−0.013	0.126	−0.007	0.092	0.005	0.054	0.005	0.036
Design 4								
MLE	−0.161	0.169	−0.163	0.163	0.058	0.061	0.058	0.058
$h = 0.5$	0.036	0.330	0.038	0.216	0.034	0.145	0.017	0.088
$h = 1.0$	−0.002	0.165	0.002	0.115	0.010	0.075	0.002	0.052
$h = 1.5$	−0.013	0.122	−0.013	0.086	0.007	0.055	0.004	0.035

Figure 1: Difference in the number of unemployed between t and t-1 over the observation period (1989-1992)



Working Paper

2002-1:	Peter Jensen, Michael Rosholm and Mette Verner: A Comparison of Different Estimators for Panel Data Sample Selection Models.
2002-2:	Torben M. Andersen: International Integration and the Welfare State.
2002-3:	Bo Sandemann Rasmussen: Credibility, Cost of Reneging and the Choice of Fixed Exchange Rate Regime.
2002-4:	Bo William Hansen and Lars Mayland Nielsen: Can Nominal Wage and Price Rigidities Be Equivalent Propagation Mechanisms? The Case of Open Economies.
2002-5:	Anna Christina D'Addio and Michael Rosholm: Left-Censoring in Duration Data: Theory and Applications.
2002-6:	Morten Ørregaard Nielsen: Efficient Inference in Multivariate Fractionally Integration Models.
2002-7:	Morten Ørregaard Nielsen: Optimal Residual Based Tests for Fractional Cointegration and Exchange Rate Dynamics.
2002-8:	Morten Ørregaard Nielsen: Local Whittle Analysis of Stationary Fractional Cointegration.
2002-9:	Effrosyni Diamantoudi and Licun Xue: Coalitions, Agreements and Efficiency.
2002-10:	Effrosyni Diamantoudi and Eftichios S. Sartzetakis: International Environmental Agreements - The Role of Foresight.
2002-11:	N.E. Savin and Allan H. Würtz: Testing the Semiparametric Box–Cox Model with the Bootstrap.
2002-12:	Morten Ø. Nielsen, Spectral Analysis of Fractionally Cointegrated Systems
2002-13:	Anna Christina D'Addio and Bo E. Honoré, Duration Dependence and Timevarying Variables in Discrete Time Duration Models.