Specification Tests for Dynamic Binary Response Models with State Dependence

Alexandru M. Lefter[1], Brian P. McCall[2]
[1]Mount Royal University, Calgary, AB
[2]University of Michigan, Ann Arbor, MI

Abstract

This paper develops a specification test for a class of dynamic binary response models with state dependence in which true state dependence is distinguished from spurious state dependence. The test uses the property that if the models are correctly specified, the residuals follow a martingale difference sequence. One application of this test is to assess the validity of the over-identifying restrictions used to achieve identification.

Key Words: binary response; martingale; specification test

1. Introduction

This paper develops a specification test for dynamic binary response models with state dependence and unobserved heterogeneity. Dynamic binary response models with state dependence have been used in economics to investigate topics such as married women's labor supply (Hyslop, 1999), welfare dependency (Card and Hyslop, 2005), illicit drug use (Deza, 2015), and repeat use of unemployment insurance (Lefter and McCall, 2016). These models aim to separate true state dependence from spurious state dependence by making assumptions about the nature of the unobserved heterogeneity and the manner in which the exogenous variables affect the outcome variable. However, if these assumptions are invalid, estimates of the extent of true state dependence are likely to be biased. For this reason, it is important to assess the validity of such assumptions.

The test developed in this paper relies on the fact that under the null hypothesis that the model is correctly specified, a set of residuals can be constructed that form a martingale difference sequence. This test can be used to assess whether the over-identifying or exclusion restrictions that are typically imposed in empirical applications in order to identify true state dependence from spurious state dependence are valid. Similar tests based on martingales have been applied to hazard models (see, for example, Therneau, Grambsch, and Fleming, 1990; McCall, 1994; Therneau and Grambsch, 2000).

2. Dynamic Binary Response Models with State Dependence and Unobserved Heterogeneity

Let y_{it} be a binary random variable, with $i = 1, \dots, N$ indexing cross-sectional units (individuals, firms, schools, etc.), and $t = 1, \dots, T_i$ indexing time periods, where we assume that the number of periods of observation is individual-specific subject to the restrictions explained below. We assume that there exists a latent variable y_{it}^* with $y_{it}^* = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1} + \epsilon_{it}$, where $\mathbf{x}_{it}$ is a $J \times 1$ vector of contemporaneous exogenous explanatory variables, $\mathbf{y}_i^{t-1} = (y_{i1}, y_{i2}, \dots, y_{i,t-1})'$ is a $(t-1) \times 1$ vector of previous realizations of y_{it}, α_i is an unobserved random variable that is independent of $\mathbf{x}_{it}$ (i.e., a random effect) and has cumulative distribution function *G*, and ϵ_{it} are independent and identically distributed random variables that are symmetric around zero and have cumulative distribution function *F*. Implicit in this assumption is that y_{it}^* does not depend on past values of the exogenous explanatory variables. We further assume that $y_{it} = I(y_{it}^* > 0)$, where *I* is an indicator function equal to 1 whenever the statement in brackets is true, and 0 otherwise. Thus, conditional on $\mathbf{x}_{it}$, $\mathbf{y}_i^{t-1}$, and α_i, we have:

$$\begin{aligned} p_{it\alpha} &\equiv \Pr(y_{it} = 1 \mid \mathbf{x}_{it}, \mathbf{y}_i^{t-1}, \alpha_i) \\ &= \Pr(y_{it}^* > 0 \mid \mathbf{x}_{it}, \mathbf{y}_i^{t-1}, \alpha_i) \\ &= \Pr(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1} + \epsilon_{it} > 0 \mid \mathbf{x}_{it}, \mathbf{y}_i^{t-1}, \alpha_i) \\ &= \Pr(\epsilon_{it} > -(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}) \mid \mathbf{x}_{it}, \mathbf{y}_i^{t-1}, \alpha_i) \\ &= 1 - F\left(-(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1})\right) \\ &= F(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}). \end{aligned} \tag{1}$$

Let $T = \max_i\{T_i\}$, and let c_{it} be a dichotomous censoring variable equal to 1 if $t \le T_i$, and 0 otherwise, $t = 1, \dots, T$. We assume that c_{it} is I_t-measurable with respect to the sigma algebra generated by $\mathbf{x}_{i1}, \dots, \mathbf{x}_{it}$ and $\mathbf{y}_i^{t-1}$, $i = 1, \dots, N$, where I_t is the information accumulated at "just before time t" (and is assumed to include $\mathbf{x}_{it}$). Finally, we assume that *F* is a known function (e.g., standard normal), and that *G* can be parameterized by a finite-dimensional, $H \times 1$, vector $\boldsymbol{\eta}$. Estimation proceeds by maximum likelihood estimation with the log likelihood function given by:

$$\begin{aligned} \log L &= \sum_{i=1}^{N} \log L_i \\ &= \sum_{i=1}^{N} \log \int \left\{ \prod_{t=1}^{T} [{p_{it\alpha}}^{y_{it}} (1 - p_{it\alpha})^{1-y_{it}}]^{c_{it}} \right\} dG(\alpha) \\ &= \sum_{i=1}^{N} \log \int \left(\prod_{t=1}^{T} \ell_{it\alpha} \right) dG(\alpha), \end{aligned}$$

where

$$\ell_{it\alpha} = [{p_{it\alpha}}^{y_{it}} (1 - p_{it\alpha})^{1-y_{it}}]^{c_{it}} \tag{2}$$

and represents individual i's contribution to the likelihood at time t, conditional on α. We will let $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\gamma}}', \widehat{\boldsymbol{\eta}}'\right)'$ represent the $K \times 1$ vector of maximum likelihood estimates, where $K = J + (T-1) + H$, and $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\gamma}_0', \boldsymbol{\eta}_0')'$ the $K \times 1$ vector of true population values of the parameters. Furthermore, we will denote the estimated $K \times K$ variance-covariance matrix by $\widehat{\boldsymbol{\Lambda}}$.

3. Specification Test Based on Martingale Residuals

To develop a specification test of the model, let $\mathbf{d}_{it}(I_t, \boldsymbol{\theta}_0) = \mathbf{z}_{it} c_{it} [y_{it} - p_{it}(I_t, \boldsymbol{\theta}_0)]$, where $\mathbf{z}_{it}$ is any I_t-measurable $R \times 1$ vector and $p_{it}(I_t, \boldsymbol{\theta}_0) = \Pr(y_{it} = 1 | I_t, \boldsymbol{\theta}_0)$, and let $\mathbf{d}_t(I_t, \boldsymbol{\theta}_0) = \sum_{i=1}^{N} \boldsymbol{d}_{it}(I_t, \boldsymbol{\theta}_0)$, $t = 1, \dots, T$. If the model is correctly specified, $\mathbf{d}_t(I_t, \boldsymbol{\theta}_0)$ forms a vector of martingale difference sequences, that is, $\mathrm{E}(\mathbf{d}_t(I_t, \boldsymbol{\theta}_0)) = 0$ for all t. As a result, the I_t-measurable $R \times R$ variance-covariance process, which we denote by $\mathbf{V}_t(\boldsymbol{\theta}_0)$, satisfies:

$$\begin{aligned} \mathbf{V}_t(\boldsymbol{\theta}_0) &= \mathrm{Var}\big(\mathbf{d}_t(I_t, \boldsymbol{\theta}_0)\big) \\ &= \mathrm{E}(\mathbf{d}_t(I_t, \boldsymbol{\theta}_0)\mathbf{d}_t(I_t, \boldsymbol{\theta}_0)') - \mathrm{E}\big(\mathbf{d}_t(I_t, \boldsymbol{\theta}_0)\big) E(\mathbf{d}_t(I_t, \boldsymbol{\theta}_0))' \\ &= \mathrm{E}(\mathbf{d}_t(I_t, \boldsymbol{\theta}_0)\mathbf{d}_t(I_t, \boldsymbol{\theta}_0)') \end{aligned}$$

for all t. So, by the central limit theorem, $N^{-1/2}\mathbf{d}_t(\boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{Normal}(\mathbf{0}, \mathbf{V}_t(\boldsymbol{\theta}_0))$ for all t.

Now, if $\boldsymbol{\theta}_0$ were known, a consistent estimate of $\mathbf{V}_t(\boldsymbol{\theta}_0)$ would be:

$$\frac{1}{N}\sum_{i=1}^{N} c_{it}\mathbf{z}_{it}\mathbf{z}_{it}' \, p_{it}(I_t, \boldsymbol{\theta}_0)[1 - p_{it}(I_t, \boldsymbol{\theta}_0)].$$

This follows from the fact that:

$$\begin{aligned} \mathrm{E}(\mathbf{d}_{it}(I_t, \boldsymbol{\theta}_0)\mathbf{d}_{it}(I_t, \boldsymbol{\theta}_0)') &= \mathrm{E}\big\{\big[c_{it}\mathbf{z}_{it}\big(y_{it} - p_{it}(I_t, \boldsymbol{\theta}_0)\big)\big]\big[c_{it}\mathbf{z}_{it}\big(y_{it} - p_{it}(I_t, \boldsymbol{\theta}_0)\big)\big]'\big\} \\ &= c_{it}\mathbf{z}_{it}\mathbf{z}_{it}'\mathrm{E}\{[y_{it} - p_{it}(I_t, \boldsymbol{\theta}_0)]^2\} \\ &= c_{it}\mathbf{z}_{it}\mathbf{z}_{it}'\mathrm{E}\big[y_{it}^2 - 2p_{it}(I_t, \boldsymbol{\theta}_0)y_{it} + p_{it}(I_t, \boldsymbol{\theta}_0)^2\big] \\ &= c_{it}\mathbf{z}_{it}\mathbf{z}_{it}' p_{it}(I_t, \boldsymbol{\theta}_0)[1 - p_{it}(I_t, \boldsymbol{\theta}_0)]. \end{aligned} \quad (3)$$

However, we only have an estimate of $\boldsymbol{\theta}_0$. Applying the delta method (van der Vaart, 1998) to $\mathbf{d}_{it}\big(I_t, \widehat{\boldsymbol{\theta}}\big)$, we get $\mathbf{d}_{it}\big(I_t, \widehat{\boldsymbol{\theta}}\big) = \mathbf{d}_{it}(I_t, \boldsymbol{\theta}_0) - \mathbf{D}_{it}(I_t, \boldsymbol{\theta}_0)\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_P\left(N^{-\frac{1}{2}}\right)$, which gives us:

$$\begin{aligned} &\mathbf{d}_{it}\big(I_t, \widehat{\boldsymbol{\theta}}\big)\mathbf{d}_{it}\big(I_t, \widehat{\boldsymbol{\theta}}\big)' = \\ &\quad = \left[\mathbf{d}_{it}(I_t, \boldsymbol{\theta}_0) - \mathbf{D}_{it}(I_t, \boldsymbol{\theta}_0)\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_P\left(N^{-1/2}\right)\right] \times \\ &\quad\quad \times \left[\mathbf{d}_{it}(I_t, \boldsymbol{\theta}_0) - \mathbf{D}_{it}(I_t, \boldsymbol{\theta}_0)\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_P\left(N^{-1/2}\right)\right]' \\ &\quad = \left[\mathbf{d}_{it}(I_t, \boldsymbol{\theta}_0)\mathbf{d}_{it}(I_t, \boldsymbol{\theta}_0)' - \mathbf{D}_{it}(I_t, \boldsymbol{\theta}_0)\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)'\mathbf{D}_{it}(I_t, \boldsymbol{\theta}_0)' + o_P(N^{-1})\right]', \end{aligned}$$

where $\mathbf{D}_{it}(I_t, \boldsymbol{\theta}_0) = \partial \mathbf{d}_{it}\big(I_t, \widehat{\boldsymbol{\theta}}\big)/\partial \boldsymbol{\theta}'$.

The implication of this result is that when estimates are obtained by maximum likelihood estimation, we need to add the following adjustment term:

$$\frac{1}{N}\sum_{i=1}^{N} c_{it}\mathbf{z}_{it}\frac{\partial p_{it}(I_t,\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\widehat{\boldsymbol{\Lambda}}\left(\frac{\partial p_{it}(I_t,\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right)'\mathbf{z}_{it}'.$$

Therefore, the estimate of the variance-covariance matrix is:

$$\widehat{\mathbf{V}}_t = \frac{1}{N}\left\{\sum_{i=1}^{N} c_{it}\mathbf{z}_{it}\mathbf{z}_{it}'\,p_{it}\big(I_t,\widehat{\boldsymbol{\theta}}\big)\big[1-p_{it}\big(I_t,\widehat{\boldsymbol{\theta}}\big)\big] + \sum_{i=1}^{N} c_{it}\mathbf{z}_{it}\frac{\partial p_{it}(I_t,\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\widehat{\boldsymbol{\Lambda}}\left(\frac{\partial p_{it}(I_t,\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right)'\mathbf{z}_{it}'\right\}.$$

We can use these results to construct a specification test for the binary random effects model with state dependence. Under the null hypothesis that the econometric model is correctly specified, we have that $\boldsymbol{d}_t\widehat{\mathbf{V}}_t\boldsymbol{d}_t'$ is distributed as a χ^2 with *R* degrees of freedom. Moreover, if we sum these test statistics across all *T* time periods, then, under the null hypothesis, $\sum_{t=1}^{T}\boldsymbol{d}_t\widehat{\mathbf{V}}_t\boldsymbol{d}_t'$ is distributed as a χ^2 with *TR* degrees of freedom.

To compute this test statistic, we need to calculate the terms $p_{it}\big(I_t,\widehat{\boldsymbol{\theta}}\big)$ and $\partial p_{it}(I_t,\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}$. With the specification of $p_{it\alpha}$ given in (1), and the assumption that α is independent of $\mathbf{x}_{it}$ and has cumulative distribution function *G*, we have:

$$\begin{aligned} p_{it}\big(I_t,\widehat{\boldsymbol{\theta}}\big) &= \int \hat{p}_{it\alpha}\,dG(\alpha|\mathbf{X}_i^t,\mathbf{y}_i^{t-1},\boldsymbol{c}_i^t,\widehat{\boldsymbol{\eta}}) \\ &= \int F\big(\alpha+\widehat{\boldsymbol{\beta}}'\mathbf{x}_{it}+\widehat{\boldsymbol{\gamma}}'\mathbf{y}_i^{t-1}\big)dG(\alpha|\mathbf{X}_i^t,\mathbf{y}_t^{t-1},\boldsymbol{c}_i^t,\widehat{\boldsymbol{\eta}}), \end{aligned}$$

where $\mathbf{X}_i^t = (\mathbf{x}_{i1},\dots,\mathbf{x}_{it})$ and $\boldsymbol{c}_i^t = (c_{i1},\dots,c_{it})$. To calculate $dG(\alpha|\mathbf{X}_i^t,\mathbf{y}_i^{t-1},\boldsymbol{c}_i^t,\widehat{\boldsymbol{\eta}})$, we can apply Bayes' theorem to get:

$$dG\big(\alpha|\mathbf{X}_i^t,\mathbf{y}_i^{t-1},\boldsymbol{c}_i^t\big) = \frac{\Pr\big(\mathbf{X}_i^t,\mathbf{y}_i^{t-1},\boldsymbol{c}_i^t|\alpha\big)\,dG(\alpha)}{\int\Pr\big(\mathbf{X}_i^t,\mathbf{y}_i^{t-1},\boldsymbol{c}_i^t|\alpha\big)\,dG(\alpha)}.$$

But given our assumptions about $\mathbf{X}_i^t,\mathbf{y}_i^{t-1}$, and $\boldsymbol{c}_i^t$, and the specification of $\ell_{it\alpha}$ given in (2), we have that $\Pr\big(\mathbf{X}_i^t,\mathbf{y}_i^{t-1},\boldsymbol{c}_i^t|\alpha\big) = \prod_{s=1}^{t-1}\ell_{is\alpha}$. Therefore:

$$dG\big(\alpha|\mathbf{X}_i^t,\mathbf{y}_i^{t-1},\boldsymbol{c}_i^t\big) = \frac{\prod_{s=1}^{t-1}\ell_{is\alpha}dG(\alpha)}{\int\prod_{s=1}^{t-1}\ell_{is\alpha}dG(\alpha)},$$

and

$$p_{it}\big(I_t,\widehat{\boldsymbol{\theta}}\big) = \int F\big(\alpha+\widehat{\boldsymbol{\beta}}'\mathbf{x}_{it}+\widehat{\boldsymbol{\gamma}}'\mathbf{y}_i^{t-1}\big)\frac{\prod_{s=1}^{t-1}\ell_{is\alpha}dG(\alpha,\widehat{\boldsymbol{\eta}})}{\int\prod_{s=1}^{t-1}\ell_{is\alpha}dG(\alpha,\widehat{\boldsymbol{\eta}})}. \tag{4}$$

One useful application of the test would be to check the appropriateness of the exclusion restrictions that are generally used to identify state dependence. If we assume that only the current values of the exogenous explanatory variables affect the probability that y_{it} equals 1 at time *t*, then a vector $\mathbf{z}_{it}$ can be constructed from previous values of those exogenous explanatory variables that are time-varying. This vector can then be used to calculate $\mathbf{d}_t(\widehat{\boldsymbol{\theta}})$ and, ultimately, $\sum_{t=1}^{T} \boldsymbol{d}_t(I_t,\widehat{\boldsymbol{\theta}})\widehat{\mathbf{V}}_t(I_t,\widehat{\boldsymbol{\theta}})\boldsymbol{d}_t'(I_t,\widehat{\boldsymbol{\theta}})$.

As a particular example, assume that *G* has a discrete mass-point distribution with *J* mass points, $\alpha_1, \dots, \alpha_J$, where point α_j has mass q_j, *j* = 1, …, *J*, with $\sum_{j=1}^{J} q_j = 1$. Thus, we have $\boldsymbol{\eta} = (\boldsymbol{\alpha}', \boldsymbol{q}')'$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)'$, $\boldsymbol{q} = (q_2, \dots, q_J)'$, and q_1 is set to 1 for identification purposes. Then, from (4), we have that the posterior probability for person *i* at time *t*, q_{ij}^t, satisfies:

$$q_{ij}^t = \frac{\prod_{s=1}^{t-1} \ell_{isj} q_j}{\sum_{r=1}^{J} \prod_{s=1}^{t-1} \ell_{isr}\, q_r}.$$

This implies that:

$$\mathbf{d}_{it}(I_t,\widehat{\boldsymbol{\theta}}) = \mathbf{z}_{it} c_{it}\big[y_{it} - p_{it}\big(I_t,\widehat{\boldsymbol{\theta}}\big)\big] = \mathbf{z}_{it} c_{it}\big[y_{it} - \textstyle\sum_{j=1}^{J} q_{ij}^t p_{itj}\big(I_t,\widehat{\boldsymbol{\theta}}\big)\big].$$

In order to calculate $\widehat{\mathbf{V}}_t$, we need to compute:

$$\frac{\partial p_{it}(I_t,\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \sum_{j=1}^{J} q_{ij}^t p_{itj}\big(I_t,\widehat{\boldsymbol{\theta}}\big)}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{J} \left(q_{ij}^t \frac{\partial p_{itj}}{\partial \boldsymbol{\theta}} + \frac{\partial q_{ij}^t}{\partial \boldsymbol{\theta}} p_{itj}\big(I_t,\widehat{\boldsymbol{\theta}}\big) \right).$$

Now,

$$\frac{\partial p_{itj}\big(I_t,\widehat{\boldsymbol{\theta}}\big)}{\partial \boldsymbol{\beta}} = f\big(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big)\, \mathbf{x}_{it}\,,$$

and

$$\frac{\partial p_{itj}\big(I_t,\widehat{\boldsymbol{\theta}}\big)}{\partial \boldsymbol{\gamma}} = f\big(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big)\mathbf{y}_i^{t-1}.$$

If we let $\mathbf{f}\big(\boldsymbol{\alpha} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big)$ be the $J \times 1$ vector given by $\big[f\big(\alpha_1 + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big), \dots$ $\dots, f\big(\alpha_J + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big)\big]'$, then:

$$\frac{\partial p_{itj}\big(I_t,\widehat{\boldsymbol{\theta}}\big)}{\partial \boldsymbol{\alpha}} = \mathbf{f}\big(\boldsymbol{\alpha} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big).$$

In a similar fashion, if we let $\mathbf{F}\big(\boldsymbol{\alpha}_{(1)} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big)$ be the $(J-1) \times 1$ vector given by $\big[F\big(\alpha_2 + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big), \dots, F\big(\alpha_J + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big)\big]'$, then:

$$\frac{\partial p_{itj}\big(I_t,\widehat{\boldsymbol{\theta}}\big)}{\partial \boldsymbol{q}} = \mathbf{F}\big(\boldsymbol{\alpha}_{(1)} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{y}_i^{t-1}\big).$$

Furthermore, it can be shown that:

$$\frac{\partial q_{ij}^t}{\partial \boldsymbol{\beta}} = q_{ij}^t \left[\sum_{s=1}^{t} (2y_{is} - 1) \frac{\partial p_{isj}}{\partial \boldsymbol{\beta}} \ell_{isj}^{-1} - \sum_{k=1}^{J} q_k \sum_{s=1}^{t} \sum_{s=1}^{t} (2y_{is} - 1) \frac{\partial p_{isj}}{\partial \boldsymbol{\beta}} \ell_{isj}^{-1} \right],$$

$$\frac{\partial q_{ij}^t}{\partial \alpha_j} = (1 - q_{ij}^t) q_{ij}^t \sum_{s=1}^{t} (2y_{is} - 1) \frac{\partial p_{isj}}{\partial \alpha_j} \ell_{isj}^{-1} \text{ for } m = j,$$

$$\frac{\partial q_{ij}^t}{\partial \alpha_m} = -q_{im}^t q_{ij}^t \sum_{s=1}^{t} (2y_{is} - 1) \frac{\partial p_{ism}}{\partial \alpha_m} \ell_{ism}^{-1} \text{ for } m \neq j,$$

$$\frac{\partial q_{ij}^t}{\partial q_j} = q_j^{-1} q_{ij}^t (1 - q_{ij}^t) \text{ for } m = j,$$

$$\frac{\partial q_{ij}^t}{\partial q_m} = -q_j^{-1} q_{ij}^t q_{im}^t \text{ for } m \neq j.$$

4. Conclusion

In this paper we developed a specification test for dynamic binary response models with state dependence and unobserved heterogeneity. This test can be applied to check the validity of the exclusion restrictions used to identify true state dependence from spurious state dependence or other aspects of the models. The test is based on the fact that the residuals can be constructed to follow a martingale difference sequence under the null hypothesis. If the null hypothesis is rejected, plots of these residuals may also help a researcher determine the source of the model rejection.

References

Card, D., and Hyslop, D. R. (2005), "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers," *Econometrica*, 73(6), 1723-1770.

Deza, M. (2015), "Is There a Stepping Stone Effect in Drug Use? Separating State Dependence from Unobserved Heterogeneity Within and Between Illicit Drugs," *Journal of Econometrics*, 184(1), 193-207.

Hyslop, D. R. (1999), "State Dependence, Serial Correlation, and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67(6), 1255-1294.

Lefter, A. M., and McCall, B. P. (2016), "Repeat Use of Unemployment Insurance in the United States: Evidence from the NLSY79," Unpublished Working Paper.

McCall, B. P. (1994), “Specification Diagnostics for Duration Models: A Martingale Approach,” *Journal of Econometrics*, 60(1-2), 293-312.

Therneau, T. M., and Grambsch, P. M. (2000), *Modeling Survival Data: Extending the Cox Model*, Springer, New York, NY, USA.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990), “Martingale-Based Residuals for Survival Models,” *Biometrika*, 77(1), l47-160.

van der Vaart, A .W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge, United Kingdom.