Identifying State Dependence in Non-Stationary Processes

Timothy J. Halliday*

University of Hawaii at Mānoa

Department of Economics and John A. Burns School of Medicine

January 27, 2006

Abstract

We consider the identification of state dependence in a non-stationary process of binary outcomes within the context of the dynamic logit model with time-variant transition probabilities and an arbitrary distribution for the unobserved heterogeneity. We derive a simple identification result that allows us to calculate a test for state dependence in this model. We also consider alternative tests for state dependence that will have desirable properties only in stationary processes and derive their asymptotic properties when the true underlying process is non-stationary. Finally, we provide Monte Carlo evidence that shows a range of non-stationarity in which the effects of mis-specifying the binary process as stationary are not too large.

*Address: 2424 Maile Way; 533 Saunders Hall, Honolulu, HI 96822. Phone: (808) 956 - 8615. E-mail: halliday@hawaii.edu.

Key Words: Dynamic Panel Data Models, State Dependence, Non-Stationary Processes

1 Introduction

Economic data often display serial correlation. This is true in the case of labor force participation, crime, accident occurrence and numerous measures of health. However, the source of this persistence is often unclear. One possible source is an unobserved time-invariant propensity to experience a given economic outcome or unobserved heterogeneity. Another potential source is that experiencing a particular event today may alter a person's preferences or opportunities and, thus, impact the probability that the same event will occur in the future. Heckman (1981) refers to this second source of the persistence as "true state dependence." Identification of true state dependence is of particular interest to social scientists because its presence implies that policies that impact an economic outcome today will have dynamic consequences.

Because of this, econometricians have devoted much time and effort towards the identification of state dependence. Much of this work has used random effects estimators in which the researcher specifies a distribution for the heterogeneity and then maximizes a parametric likelihood function. However, this approach is limited as it imposes *ad hoc* distributional assumptions on the data. More recently, Honoré and Kyriazidou (2000) have relaxed the assumptions of the random effects approach and developed a fixed effects estimator for a discrete choice model with lagged dependent variables and unobserved heterogeneity. Their approach builds upon the conditional logit model of Chamberlain (1985) and, thus, imposes no assumptions on the distribution of the heterogeneity.

The Honoré and Kyriazidou estimator requires conditioning on subsets of the data for which the exogenous regressors are equal in at least two separate time periods. While this is certainly a weakness of the estimator, Hahn (2001) and Honoré and Tamer (2004) have speculated that this conditioning procedure is unavoidable and that point estimation of the model's parameters is impossible without it. Nevertheless, this procedure has an undesirable property in that it precludes the use of many explanatory variables such as age, cohort and/or time effects. Potentially, this is a major drawback since many economic outcomes including labor force participation and health vary with age and are, thus, non-stationary processes.

In this paper, we investigate the impact that non-stationarity in the underlying data generating process has on the identification of state dependence in the dynamic conditional logit model. To do this, first, we derive a simple result that allows us to identify state dependence in the presence of time-varying transition probabilities. If the model also includes an unbounded regressor, the logistic assumption on the unobserved period-specific shocks is not only sufficient, but also necessary for the identification result to hold. We also show how the result can easily be used to derive a test for the presence of state dependence. In addition, we consider the properties of tests for state dependence which erroneously specify the data generating process as stationary when the true underlying process is non-stationary. Finally, we conduct Monte Carlo experiments which suggest that there is an "acceptable" range of non-stationarity in which mis-specifying the data generating process as stationary does not matter too much.

The balance of this paper is organized as follows. Section 2 discusses our main identification result. Section 3 uses this result to derive a test statistic for the presence of state dependence in a non-stationary process. Section 4 discusses the asymptotic properties of some mis-specified tests. Section 5 concludes.

2 A Very Simple Identification Result

In this section, we establish a simple result that allows for the identification of state dependence in non-stationary processes. We let $\{y_{i,t}\}_{t=0}^{T}$ denote a sequence of binary outcomes such that $y_{i,t} \in \{0,1\}$. We assume that the data are generated by the following binary choice model:

$$y_{i,t} = 1(\alpha_i + y_{i,t-1}\gamma + f_t(x_{i,t}) + \varepsilon_{i,t} \geq 0) \tag{1}$$

for $i = 1, ..., N$ and $t = 1, ..., T$. In equation (1), α_i is an unobserved individual-specific effect, γ is the state-dependence coefficient, $x_{i,t}$ is a vector of strictly exogenous regressors and $\varepsilon_{i,t}$ is an unobserved error term. If $\gamma > 0$ ($\gamma < 0$), then the process $\{y_{i,t}\}_{t=0}^{T}$ exhibits positive (negative) state dependence. $f_t(x_{i,t})$ is a time-varying function of the strictly exogenous regressors. If $x_{i,t}$ is a constant then $f_t(x_{i,t})$ simply becomes a time dummy (*i.e* $f_t(x_{i,t}) = \delta_t$). We assume that $\varepsilon_{i,t}$ is *i.i.d.* across time, is independent of the vector $(\alpha_i, x_{i,1}, ..., x_{i,T}, y_{i,0})$ and follows the logistic distribution which we denote by $\Lambda(h) \equiv P(\varepsilon_{i,t} \leq h)$. Finally, we assume that we observe *i.i.d.* draws of $(y_{i,T}, ..., y_{i,0}, x_{i,1}, ..., x_{i,T})$ from some underlying population.

To derive our identification result, we define the events:

$$A_1 = \{y_{i,2} = 0, y_{i,1} = 1, y_{i,0} = 1\}; A_2 = \{y_{i,2} = 1, y_{i,1} = 0, y_{i,0} = 1\} \tag{2}$$

$$B_1 = \{y_{i,2} = 0, y_{i,1} = 1, y_{i,0} = 0\}; B_2 = \{y_{i,2} = 1, y_{i,1} = 0, y_{i,0} = 0\} \tag{3}$$

For $\gamma \geq 0$, the assumptions of the model are sufficient to show that

$$\begin{aligned}
\frac{P(A_1|x_i,\alpha_i)}{P(A_2|x_i,\alpha_i)} &= \frac{(1-\Lambda(\alpha_i+\gamma+f_2(x_{i,2})))\Lambda(\alpha_i+\gamma+f_1(x_{i,1}))}{\Lambda(\alpha_i+f_2(x_{i,2}))(1-\Lambda(\alpha_i+\gamma+f_1(x_{i,1})))} \\
&\geq \frac{(1-\Lambda(\alpha_i+\gamma+f_2(x_{i,2})))\Lambda(\alpha_i+\gamma+f_1(x_{i,1}))}{\Lambda(\alpha_i+\gamma+f_2(x_{i,2}))(1-\Lambda(\alpha_i+\gamma+f_1(x_{i,1})))} \\
&= \exp(f_1(x_{i,1})-f_2(x_{i,2})).
\end{aligned} \tag{4}$$

where $x_i \equiv (x_{i,1}, x_{i,2})$. Similarly, for $\gamma \geq 0$, we will also have that

$$\begin{aligned}
\frac{P(B_1|x_i,\alpha_i)}{P(B_2|x_i,\alpha_i)} &= \frac{(1-\Lambda(\alpha_i+\gamma+f_2(x_{i,2})))\Lambda(\alpha_i+f_1(x_{i,1}))}{\Lambda(\alpha_i+f_2(x_{i,2}))(1-\Lambda(\alpha_i+f_1(x_{i,1})))} \\
&\leq \frac{(1-\Lambda(\alpha_i+f_2(x_{i,2})))\Lambda(\alpha_i+f_1(x_{i,1}))}{\Lambda(\alpha_i+f_2(x_{i,2}))(1-\Lambda(\alpha_i+f_1(x_{i,1})))} \\
&= \exp(f_1(x_{i,1})-f_2(x_{i,2})).
\end{aligned} \tag{5}$$

In the presence of negative state dependence, the inequalities (4) and (5) are reversed. Accordingly, denoting $\Pi(x_i) \equiv [1+\exp(f_2(x_{i,2})-f_1(x_{i,1}))]^{-1}$, we will have that

$$P(A_1|A_1 \cup A_2, x_i, \alpha_i) \;\geq\; \Pi(x_i) \geq P(B_1|B_1 \cup B_2, x_i, \alpha_i) \text{ for } \gamma \geq 0 \text{ and all } \alpha_i \tag{6}$$

$$P(A_1|A_1 \cup A_2, x_i, \alpha_i) \;\leq\; \Pi(x_i) \leq P(B_1|B_1 \cup B_2, x_i, \alpha_i) \text{ for } \gamma \leq 0 \text{ and all } \alpha_i. \tag{7}$$

Clearly, when there is no state dependence, the model becomes the static conditional logit model and (4), (5), (6) and (7) will hold with strict equality.

Note that all of our statements, thus far, have been conditional on the unobserved heterogeneity. This poses problems because, while the probabilities $P(A_1|A_1 \cup A_2, x_i)$ and $P(B_1|B_1 \cup B_2, x_i)$ can easily be estimated non-parametrically, the probabilities $P(A_1|A_1 \cup A_2, x_i, \alpha_i)$ and

$P(B_1|B_1 \cup B_2, x_i, \alpha_i)$ cannot be estimated as they depend on α_i. However, the fact that $\Pi(x_i)$ provides both an upper and a lower bound on the unobserved probabilities for all values of α_i and only varies across individuals through observables (*i.e via* x_i) allows us to make a statement that is no longer conditional on the heterogeneity.

To see this, let $G(\alpha_i|A_1 \cup A_2, x_i)$ denote the distribution of the heterogeneity conditional on $(A_1 \cup A_2, x_i)$ and let $F(\alpha_i|B_1 \cup B_2, x_i)$ denote the distribution of the heterogeneity conditional on $(B_1 \cup B_2, x_i)$. We impose no assumptions on either distribution. The inequalities in (6) and (7) imply that

$$P(A_1|A_1 \cup A_2, x_i) = \int P(A_1|A_1 \cup A_2, x_i, \alpha_i) dG(\alpha_i|A_1 \cup A_2, x_i) \geq \Pi(x_i) \text{ for } \gamma \geq 0 \tag{8}$$

and

$$P(B_1|B_1 \cup B_2, x_i) = \int P(B_1|B_1 \cup B_2, x_i, \alpha_i) dF(\alpha_i|B_1 \cup B_2, x_i) \leq \Pi(x_i) \text{ for } \gamma \geq 0. \tag{9}$$

These inequalities will be strict inequalities when $\gamma > 0$, but will hold with equality when $\gamma = 0$. When $\gamma < 0$, the inequalities will be reversed. This gives us Proposition 1 which is our key identification result.

Proposition 1 *Assume that the data generating process for $\{y_{i,t}\}_{t=2}^{3}$ is given by equation (1) and that $\varepsilon_{i,t}$ is logistically distributed and independent of $(\alpha_i, y_{i,0}, x_i)$. Then we will have that*

$$P(A_1|A_1 \cup A_2, x_i) \gtreqless P(B_1|B_1 \cup B_2, x_i) \; for \gamma \gtreqless 0.$$

One remaining question is whether or not the logistic assumption is necessary, in addition to sufficient for our results to obtain. Recent work by Magnac (2004) and older work by Chamberlain (1992) sheds light on this issue. In these papers, it is shown that, in a static binary choice model with unbounded exogenous covariates, the only distribution function such that $\varepsilon_{i,t}$ is independent across time and such that the sum of the binary variables is sufficient for α_i is the logistic distribution. This, in turn, implies that the logistic assumption would also be necessary for our results to hold provided that some element of $x_{i,t}$ has unbounded support. The reason is that our results depend crucially on the existence of a sufficient statistic for the heterogeneity when no state dependence is present since the sufficient statistics allow us to separate the probabilities $P(A_1|A_1 \cup A_2, x_i, \alpha_i)$ and $P(B_1|B_1 \cup B_2, x_i, \alpha_i)$ with $\Pi(x_i)$ which does not depend on α_i. Accordingly, without the logistic assumption, it is not be possible to separate these probabilities with a constant that does not depend on the unobserved heterogeneity.

3 Testing for State Dependence without Stationarity

It is a straightforward exercise to use the results of the previous section to derive a test statistic for the presence of state dependence in a non-stationary binary process. While it is fairly obvious from Proposition 1 how this can be done simply by constructing a test of a difference in means, we still provide the details for the sake of completeness. For the sake of simplicity, throughout the remainder of the paper, we only consider the case where the only element of $x_{i,t}$ is a time dummy so that $f_1(x_{i,1}) = \delta_1$ and $f_2(x_{i,2}) = \delta_2$. In this section, we provide the main ideas behind the test. In the appendix, we provide a more detailed argument for this section's main proposition.

We start out by defining $1_i(A_1)$ and $1_i(A_1 \cup A_2)$ to be indicators which are turned on when the events A_1 and $A_1 \cup A_2$ occur for individual i. We can easily estimate the probability $P(A_1|A_1 \cup A_2) \equiv \pi_A(\gamma, \delta_2, \delta_1)$ via

$$\widehat{\pi}_A = \frac{\sum_{i=1}^{N} 1_i(A_1)}{\sum_{i=1}^{N} 1_i(A_1 \cup A_2)}. \tag{10}$$

We define $\widehat{\pi}_B$, the estimate of $P(B_1|B_1 \cup B_2) \equiv \pi_B(\gamma, \delta_2, \delta_1)$, in an analogous way. Clearly, equation (10) can easily accommodate discrete regressors simply by counting the number of times the events A_1 and $A_1 \cup A_2$ occur among the sub-population for whom $x_i = d$. Next, we let $\widehat{\pi}_{A_1}$ and $\widehat{\pi}_{A_{12}}$ denote estimates of $P(A_1)$ and $P(A_1 \cup A_2)$. We define $\widehat{\pi}_{B_1}$ and $\widehat{\pi}_{B_{12}}$ in a similar fashion. In the appendix, we show that the asymptotic variances of $\widehat{\pi}_A$ and $\widehat{\pi}_B$ are

$$\widehat{\sigma}_A^2 = \frac{\widehat{\pi}_{A_1}}{\widehat{\pi}_{A_{12}}^3} \left(\widehat{\pi}_{A_{12}} - \widehat{\pi}_{A_1}\right) \tag{11}$$

and

$$\widehat{\sigma}_B^2 = \frac{\widehat{\pi}_{B_1}}{\widehat{\pi}_{B_{12}}^3} \left(\widehat{\pi}_{B_{12}} - \widehat{\pi}_{B_1}\right). \tag{12}$$

Finally, note that because the events $A_1 \cup A_2$ and $B_1 \cup B_2$ are mutually exclusive and because the sample is *i.i.d.*, the covariance between $\widehat{\pi}_A$ and $\widehat{\pi}_B$ is zero. We can now calculate the statistic

$$sd_1(\gamma, \delta_2, \delta_1) = \frac{\widehat{\pi}_A - \widehat{\pi}_B}{\left(\frac{\widehat{\sigma}_{AB}^2}{N}\right)^{1/2}} \tag{13}$$

where $\widehat{\sigma}^2_{AB} \equiv \widehat{\sigma}^2_A + \widehat{\sigma}^2_B$. Next, we note that

$$sd_1(\gamma, \delta_2, \delta_1) =$$

$$\sqrt{N}\left(\underbrace{\left(\frac{\widehat{\pi}_A - \pi_A(\gamma, \delta_2, \delta_1)}{\widehat{\sigma}_{AB}}\right)}_{X_N(\gamma,\delta_2,\delta_1)} - \underbrace{\left(\frac{\widehat{\pi}_B - \pi_B(\gamma, \delta_2, \delta_1)}{\widehat{\sigma}_{AB}}\right)}_{Y_N(\gamma,\delta_2,\delta_1)} + \underbrace{\left(\frac{\pi_A(\gamma, \delta_2, \delta_1) - \pi_B(\gamma, \delta_2, \delta_1)}{\widehat{\sigma}_{AB}}\right)}_{Z_N(\gamma,\delta_2,\delta_1)}\right). \quad (14)$$

In the appendix, we show that $\sqrt{N}(X_N(\gamma, \delta_2, \delta_1) - Y_N(\gamma, \delta_2, \delta_1))$ will converge to a $N(0, 1)$ random variable regardless of the values of $(\gamma, \delta_2, \delta_1)$. However, Proposition 1 tells us that $Z_N(\gamma, \delta_2, \delta_1)$ will only be zero when $\gamma = 0$; otherwise, it will be positive when $\gamma > 0$ and negative when $\gamma < 0$. Consequently, $sd_1(\gamma, \delta_2, \delta_1)$ will converge to a standard normal random variable when no state dependence is present, but will explode otherwise. This gives us Proposition 2.

Proposition 2 *Under the hypotheses of Proposition 1, we will have that*

$$sd_1(\gamma, \delta_2, \delta_1) \xrightarrow{d} N(0, 1) \text{ for } \gamma = 0.$$

and

$$sd_1(\gamma, \delta_2, \delta_1) \rightarrow \pm\infty \text{ for } \gamma \gtrless 0$$

for all δ_1 and δ_2.

Proposition 2 can easily be used to construct a one-sided test of size φ of $H_0 : \gamma = 0$ against $H_a : \gamma > 0$. Particularly, if we let $\Phi(.)$ denote the CDF of a $N(0, 1)$ random variable and $z_\varphi \equiv \Phi^{-1}(1 - \varphi)$, then a test of size φ can be constructed if we reject the null whenever $sd_1(\gamma, \delta_2, \delta_1) > z_\varphi$. Because $sd_1(\gamma, \delta_2, \delta_1)$ shoots off to positive infinity whenever $\gamma > 0$, the power of this test will approach unity as the sample size increases. Thus, we have the following corollary.

Corollary 3 *Let $\varphi \in (0,1)$ and $z_\varphi \equiv \Phi^{-1}(1-\varphi)$. Under the hypotheses of Proposition 1, we will have that*

$$\begin{aligned}
\lim_{N\to\infty} P(sd_1(\gamma,\delta_2,\delta_1) &> z_\varphi;\gamma,\delta_2,\delta_1) = \varphi \textit{ for } \gamma = 0 \\
\lim_{N\to\infty} P(sd_1(\gamma,\delta_2,\delta_1) &> z_\varphi;\gamma,\delta_2,\delta_1) = 1 \textit{ for } \gamma > 0
\end{aligned}$$

for all δ_1 and δ_2.

4 Properties of Some Mis-Specified Tests

In this section, we explore the properties of some tests for state dependence which erroneously specify the data generating process as stationary. The goal of this exercise is to better understand the consequences of mis-specifying a non-stationary process as stationary. To do this, we consider two tests:

$$sd_2(\gamma,\delta_2,\delta_1) = \frac{\widehat{\pi}_A - \frac{1}{2}}{\left(\frac{\widehat{\sigma}_A^2}{N}\right)^{1/2}} \tag{15}$$

and

$$sd_3(\gamma,\delta_2,\delta_1) = \frac{\widehat{\pi}_B - \frac{1}{2}}{\left(\frac{\widehat{\sigma}_B^2}{N}\right)^{1/2}}. \tag{16}$$

When the underlying data generating process is stationary, arguments similar to those above suggest that $sd_2(\gamma,\delta_2,\delta_1)$ and $sd_3(\gamma,\delta_2,\delta_1)$ will have desirable properties. Particularly, whenever

$\delta_1 = \delta_2$, the statistics will have the same asymptotic properties as $sd_1(\gamma, \delta_2, \delta_1)$ and, thus, they will converge to a $N(0,1)$ random variable when $\gamma = 0$, but will explode when $\gamma \neq 0$. However, when the underlying data generating process is non-stationary, these statistics will have less desirable properties as they are predicated upon a mis-specification of the data generating process. For the remainder of this section, we assume that $\delta_2 > \delta_1$ and investigate the properties of $sd_2(\gamma, \delta_2, \delta_1)$ and $sd_3(\gamma, \delta_2, \delta_1)$ under these conditions.

First, we consider the properties of $sd_2(\gamma, \delta_2, \delta_1)$. These properties will depend crucially upon the behavior of $\pi_A(\gamma, \delta_2, \delta_1)$ as the degrees of state dependence and non-stationarity vary. These properties are summarized in the next lemma. A proof can be found in the appendix.

Lemma 4 $\pi_A(\gamma, \delta_2, \delta_1)$ *has the following properties:*

$$\frac{\partial \pi_A(\gamma, \delta_2, \delta_1)}{\partial \gamma} > 0, \frac{\partial \pi_A(\gamma, \delta_2, \delta_1)}{\partial \delta_2} < 0, \frac{\partial \pi_A(\gamma, \delta_2, \delta_1)}{\partial \delta_1} > 0$$

Moreover, we will also have that

$$\pi_A(\gamma, \delta_2, \delta_1) = \Pi(\delta_2, \delta_1) < \frac{1}{2} \text{ for } \gamma = 0 \text{ and } \delta_2 > \delta_1$$

and that

$$\lim_{\gamma \to \infty} \pi_A(\gamma, \delta_2, \delta_1) \equiv l_A(\delta_2, \delta_1) < 1 \text{ for any } (\delta_2, \delta_1)$$

where $l_A(\delta_2, \delta_1) > \frac{1}{2}$ *for* $\delta_2 = \delta_1$.

These properties of $\pi_A(\gamma, \delta_2, \delta_1)$ can be seen in Figure 1.[1] The figure shows $\pi_A(\gamma, \delta_2, \delta_1)$ as

[1] For all the functions in the figure, we allow $\delta_1 = 0$ but we vary δ_2 and γ. We assume that $\alpha_i \in \{-0.75, 0, 0.75\}$ where each of the mass points occurs with equal probability.

a function of γ for varying degrees of non-stationarity (*i.e.* different values of (δ_2, δ_1)). The top function corresponds to $(\delta_2, \delta_1) = (0.1, 0)$, the middle function corresponds to $(\delta_2, \delta_1) = (0.5, 0)$ and the bottom function corresponds to $(\delta_2, \delta_1) = (0.75, 0)$. We see that $\pi_A(\gamma, \delta_2, \delta_1)$ increases with γ over the interval $[0, \infty)$ and that $\pi_A(0, \delta_2, \delta_1) < \frac{1}{2}$ and $\pi_A(\infty, \delta_2, \delta_1) < 1$. Moreover, $\pi_A(\gamma, \delta_2, \delta_1)$ decreases as the degree of non-stationarity rises. For example, for the case where $(\delta_2, \delta_1) = (0.5, 0)$ and $(\delta_2, \delta_1) = (0.75, 0)$, $\pi_A(\gamma, \delta_2, \delta_1)$ is furthest away from $\frac{1}{2}$ when $\gamma = 0$ but becomes gradually closer to $\frac{1}{2}$ as the degree of state dependence increases. This is the exact opposite of we would like if $sd_2(\gamma, \delta_2, \delta_1)$ were to be used to test a null of no state dependence against an alternative hypothesis of positive state dependence. In fact, we will see that tests for state dependence that use $sd_2(\gamma, \delta_2, \delta_1)$ will have extremely low power in the presence of high degrees of non-stationarity.

Lemma 4 allows us to discuss the asymptotic behavior of $sd_2(\gamma, \delta_2, \delta_1)$ when $\delta_2 > \delta_1$. To facilitate the discussion, we begin by writing

$$sd_2(\gamma, \delta_2, \delta_1) = \sqrt{N}\left(\frac{\widehat{\pi}_A - \pi_A(\gamma, \delta_2, \delta_1)}{\widehat{\sigma}_A}\right) + \sqrt{N}\left(\frac{\pi_A(\gamma, \delta_2, \delta_1) - \frac{1}{2}}{\widehat{\sigma}_A}\right) \tag{17}$$

First, we consider the case where $l_A(\delta_2, \delta_1) > \frac{1}{2}$.[2] In this case, given the arguments above, there must exist some $\gamma^* > 0$ such that $\pi_A(\gamma^*, \delta_2, \delta_1) = \frac{1}{2}$ since $\pi_A(\gamma, \delta_2, \delta_1)$ increases continuously from $\Pi(\delta_2, \delta_1) < \frac{1}{2}$ to $l_A(\delta_2, \delta_1) > \frac{1}{2}$ with γ. Consequently, the second term in equation (17) will be zero when $\gamma = \gamma^*$ and $sd_2(\gamma^*, \delta_2, \delta_1)$ will converge to $N(0, 1)$. If $\gamma > \gamma^*$ ($\gamma < \gamma^*$), then

[2] Formally, a sufficient (but not necessary) condition for $l_A > \frac{1}{2}$ is

$$\frac{\exp(\alpha_i + 2\delta_2 - \delta_1)}{1 + \exp(\alpha_i + \delta_2)} < 1 \text{ for all } \alpha_i$$

which will be true provided that $\delta_2 - \delta_1$ is not too large. This can be seen in equation (37) in the appendix.

$sd_2(\gamma^*, \delta_2, \delta_1)$ will go to positive (negative) infinity. Next, in the case where $l_A(\delta_2, \delta_1) < \frac{1}{2}$, we will have that $\pi_A(\gamma, \delta_2, \delta_1) < \frac{1}{2}$ for all γ. Accordingly, if $l_A(\delta_2, \delta_1) < \frac{1}{2}$, $sd_2(\gamma, \delta_2, \delta_1)$ will always explode to minus infinity.[3] These arguments are summarized in Proposition 5.

Proposition 5 *Let $\delta_2 > \delta_1$. Under the hypotheses of Proposition 1 and for $l_A(\delta_2, \delta_1) > \frac{1}{2}$, we will have that*

$$sd_2(\gamma, \delta_2, \delta_1) \xrightarrow{d} N(0,1) \quad \textit{for } \gamma = \gamma^*$$

and

$$sd_2(\gamma, \delta_2, \delta_1) \to \pm\infty \textit{ for } \gamma \gtreqless \gamma^* \; .$$

where γ^* is implicitly defined by $\pi_A(\gamma^*, \delta_2, \delta_1) = \frac{1}{2}$. For $l_A(\delta_2, \delta_1) < \frac{1}{2}$, we will have that

$$sd_2(\gamma, \delta_2, \delta_1) \to -\infty \textit{ for any } \gamma \; .$$

To better understand the ramifications that Proposition 4 has for the detection of state dependence, once again, we consider a one-sided test of $H_0 : \gamma = 0$ against $H_a : \gamma > 0$ where we reject the null whenever $sd_2(\gamma, \delta_2, \delta_1) > z_\varphi$ where $z_\varphi \equiv \Phi^{-1}(1 - \varphi)$. Clearly, φ is the size of this test when $\delta_1 = \delta_2$.[4] φ is what the size of the test would be if the data generating process were correctly specified. A direct implication of Proposition 5 is Corollary 6 which summarizes the properties of this test's power function.

Corollary 6 *Let $\varphi \in (0,1)$, $z_\varphi \equiv \Phi^{-1}(1-\varphi)$ and $\delta_2 > \delta_1$. Under the hypotheses of Proposition*

[3] There is a degenerate case in which there exists a pair (δ_2, δ_1) such that $l_A(\delta_2, \delta_1) = \frac{1}{2}$. However, if $l_A(\delta_2, \delta_1) = \frac{1}{2}$ then this means that $\pi_A(\gamma, \delta_2, \delta_1)$ is only $\frac{1}{2}$ in the limit i.e. $\pi_A(\infty, \delta_2, \delta_1) = \frac{1}{2}$. In this event, we will have that $sd_2(\infty, \delta_2, \delta_1) \xrightarrow{d} N(0,1)$. While this is a theoretical possibility, it really is not of practical concern to us.

[4] If $\delta_1 = \delta_2 = \delta$, then $sd_2(0, \delta, \delta) \xrightarrow{d} N(0,1)$. Consequently, with stationary transition probabilities, the size of this test will be given by $\varphi = 1 - \Phi(z_\varphi)$.

1 and for $l_A(\delta_2,\delta_1) > \frac{1}{2}$, we will have that

$$\begin{aligned}
\lim_{N\to\infty} P(sd_2(\gamma,\delta_2,\delta_1) &> z_\varphi;\gamma,\delta_2,\delta_1) = 0 \text{ for } \gamma < \gamma^* \\
\lim_{N\to\infty} P(sd_2(\gamma,\delta_2,\delta_1) &> z_\varphi;\gamma,\delta_2,\delta_1) = \varphi \text{ for } \gamma = \gamma^* \\
\lim_{N\to\infty} P(sd_2(\gamma,\delta_2,\delta_1) &> z_\varphi;\gamma,\delta_2,\delta_1) = 1 \text{ for } \gamma > \gamma^*
\end{aligned}$$

where γ^* is implicitly defined by $\pi_A(\gamma^*,\delta_2,\delta_1) = \frac{1}{2}$. For $l_A(\delta_2,\delta_1) < \frac{1}{2}$, we will always have

$$\lim_{N\to\infty} P(sd_2 > z_\varphi;\gamma,\delta_2,\delta_1) = 0 \text{ for any } \gamma.$$

The above corollary tells us that asymptotically, for $l_A(\delta_2,\delta_1) > \frac{1}{2}$ (which should be true when the degree of non-stationarity is not too large), tests based on $sd_2(\gamma,\delta_2,\delta_1)$ will fail to detect any state dependence for $\gamma \in (0,\gamma^*)$. If γ^* is small, this should not be problematic particularly in smaller samples when the second term in (17) will not be that large. However, because γ^* gets larger as the process becomes more non-stationary, large degrees of non-stationarity will have more pernicious ramifications. In the case where $l_A(\delta_2,\delta_1) < \frac{1}{2}$, the test will always fail to detect state dependence even when γ is large. In summary, when the process is non-stationary, tests based on $sd_2(\gamma,\delta_2,\delta_1)$ will not detect state dependence often enough and will, thus, have low power.

Finally, we turn the discussion to the behavior of $sd_3(\gamma,\delta_2,\delta_1)$. The behavior of this statistic will depend critically on the properties of $\pi_B(\gamma,\delta_2,\delta_1)$ which we state in the next lemma. The proof is trivial. Nevertheless, it still can be found in the appendix.

Lemma 7 *$\pi_B(\gamma, \delta_2, \delta_1)$ has the following properties:*

$$\frac{\partial \pi_B(\gamma, \delta_2, \delta_1)}{\partial \gamma} < 0, \frac{\partial \pi_B(\gamma, \delta_2, \delta_1)}{\partial \delta_2} < 0, \frac{\partial \pi_B(\gamma, \delta_2, \delta_1)}{\partial \delta_1} > 0$$

Moreover, we will also have that

$$\pi_B(\gamma, \delta_2, \delta_1) = \Pi(\delta_2, \delta_1) < \frac{1}{2} \textit{ for } \gamma = 0 \textit{ and any } (\delta_2, \delta_1)$$

and that

$$\lim_{\gamma \to \infty} \pi_B(\gamma, \delta_2, \delta_1) = 0 \textit{ for any } (\delta_2, \delta_1).$$

These properties of $\pi_B(\gamma, \delta_2, \delta_1)$ are illustrated in Figure 2. As was the case for Figure 1, this figure shows $\pi_B(\gamma, \delta_2, \delta_1)$ as a function of γ for varying degrees of non-stationarity. Once again, the top function corresponds to $(\delta_2, \delta_1) = (0.1, 0)$, the middle function corresponds to $(\delta_2, \delta_1) = (0.5, 0)$ and the bottom function corresponds to $(\delta_2, \delta_1) = (0.75, 0)$. We see that $\pi_B(0, \delta_2, \delta_1) = \Pi(\delta_2, \delta_1) < \frac{1}{2}$ and that $\pi_B(\gamma, \delta_2, \delta_1)$ decreases with γ thereafter over the interval $[0, \infty)$ until it asymptotes to 0. As was the case with $\pi_A(\gamma, \delta_2, \delta_1)$, $\pi_B(\gamma, \delta_2, \delta_1)$ is decreasing in the degree of non-stationarity. Consequently, $\pi_B(\gamma, \delta_2, \delta_1) < \frac{1}{2}$ for all $\gamma \geq 0$ whenever $\delta_2 > \delta_1$ and, thus, a calculation similar to the one in equation (17) suggests that $sd_3(\gamma, \delta_2, \delta_1)$ will go to negative infinity as the sample size increases. Proposition 8 summarizes this result.

Proposition 8 *Let $\delta_2 > \delta_1$. Under the hypotheses of Proposition 1, we will have that*

$$sd_3(\gamma, \delta_2, \delta_1) \to -\infty \textit{ for all } \gamma \geq 0.$$

The above proposition suggests that the power of a one-sided test of $H_0 : \gamma = 0$ against $H_a : \gamma > 0$ will have power that approaches unity as the samples size increases. To see this, once again, we let $\varphi \in (0, 1)$ and consider a test where we reject H_0 whenever $sd_3(\gamma, \delta_2, \delta_1) < -z_\varphi$ where $z_\varphi \equiv \Phi^{-1}(1 - \varphi)$. As was the case for $sd_2(\gamma, \delta_2, \delta_1)$, when the process is stationary, this test will have size φ under the null. However, Proposition 8 tells us that if $\delta_2 > \delta_1$ then $sd_3(\gamma, \delta_2, \delta_1)$ will explode to minus infinity as the sample grows and, thus, asymptotically, this test will always reject H_0. This gives us the following corollary.

Corollary 9 *Let $\varphi \in (0, 1)$, $z_\varphi \equiv \Phi^{-1}(1 - \alpha)$ and $\delta_2 > \delta_1$. Under the hypotheses of Proposition 1, we will have that*

$$\lim_{N\to\infty} P(sd_3(\gamma, \delta_2, \delta_1) < -z_\varphi; \gamma, \delta_2, \delta_1) = 1 \textit{ for all } \gamma \geq 0$$

5 Monte Carlo Evidence

In this section, we conduct some Monte Carlo experiments to investigate the performance of the statistics $sd_1(\gamma, \delta_2, \delta_1)$, $sd_2(\gamma, \delta_2, \delta_1)$ and $sd_3(\gamma, \delta_2, \delta_1)$. The goal of this exercise is to better understand how non-stationarity will affect the ability to detect state dependence in a finite sample. To do this, we generate data from the model

$$y_{i,1} = 1(\alpha_i + \varepsilon_{i,0} \geq 0) \tag{18}$$

and

$$y_{i,t} = 1(\alpha_i + y_{i,t-1}\gamma + \rho * t/10 + \varepsilon_{i,t} \geq 0) \text{ for } t = 1, 2. \tag{19}$$

$\varepsilon_{i,t}$ has a Logistic distribution, is *i.i.d.* across time and is independent of α_i. We allow α_i to take on values in $\{-0.75, 0, 0.75\}$ with equal probability. We simulate the model 1000 times and use a sample size of $N = 1500$. For each simulation, we consider a test of $H_0 : \gamma = 0$ against $H_a : \gamma > 0$. In Figures 3 through 5, we calculate the power functions for each of the three statistics when ρ varies between 0.0 and 1.0.

Figure 3 plots the percentage of times that $sd_1(\gamma, \rho)$ exceeds $1.645 = \Phi^{-1}(0.95)$ as a function of (γ, ρ). Accordingly, the figure shows $P(sd_1(\gamma, \rho) > 1.645; \gamma, \rho)$ for $N = 1500$. Not surprisingly, the figure shows that the power function for this statistic is well behaved. Under the null, we see that the probability of rejection is 5% regardless of the size of ρ. For $\gamma > 0.7$, the probability of rejecting the null is essentially unity for all values of ρ.

Figure 4 displays $P(sd_2(\gamma, \rho) > 1.645; \gamma, \rho)$. The top function shows the power function when $\rho = 0$. Because the data generating process is correctly specified when $\rho = 0$, we see that the size of this test is 5%. However, as ρ increases, the power and size of the test are greatly diminished which is exactly what we would expect given Proposition 5 and its corollary. In fact, for $\rho = 1.0$, we observe that the probability of detecting state dependence is less that 5% for $\gamma < 0.25$.

Figure 5 displays $P(sd_3(\gamma, \rho) < -1.645; \gamma, \rho)$. It is noteworthy that even when the process is stationary this test has significantly more power than the previous test. The reason for this is that, according to Lemma 4, for a sufficiently large γ, $sd_2(\gamma, \rho)$ will explode at rate $\sqrt{N}\left(l_A(\rho) - \frac{1}{2}\right)$ where $l_A(\rho) < 1$ for any ρ whereas Lemma 7 tells us that $sd_3(\gamma, \rho)$ will explode at rate $\sqrt{N}\left(\frac{1}{2}\right)$. Consequently, test statistics based on $sd_3(\gamma, \rho)$ will explode at a faster rate than tests based on $sd_2(\gamma, \rho)$ and, thus, in a finite sample, they will will tend to have more power which is exactly what Figures 4 and 5 depict. In fact, for $\rho = 1.0$, the figure shows that the probability of rejecting the null of no state dependence when $\gamma = 0$ is around 25%.

While Figures 4 through 5 show us that large degrees of non-stationarity can have harmful effects on the ability of $sd_2(\gamma, \rho)$ and $sd_3(\gamma, \rho)$ to detect state dependence, this is not at all surprising given that both tests are predicated upon an erroneous assumption. A slightly more interesting exercise is to investigate how smaller degrees of non-stationarity will affect the ability of these two statistics to detect state dependence. Such an exercise may shed some light on whether or not there is an acceptable range of non-stationarity in which erroneously assuming stationarity does not have too large of an impact on the detection of state dependence.

In Figures 6 and 7, we plot the power functions for $sd_2(\gamma, \rho)$ and $sd_3(\gamma, \rho)$ when ρ varies between 0.0 and 0.20. In Figure 6, we see that the size of tests based on $sd_2(\gamma, \rho)$ is not greatly affected as ρ varies in this range. Moreover, unlike Figure 4 where higher values of ρ systematically reduced the power of the test, in Figure 6, we do not see this type of a systematic relationship. In Figure 7, we see that higher values of ρ have a larger effect on the properties of $sd_3(\gamma, \rho)$. For $\rho = 0.20$, we see that the size of the test is almost 10% which is double what it should be. However, for values of ρ between 0 and 0.15, the properties of the test improve somewhat and we see that the size of the test is closer to 5%. Overall, Figures 6 and 7 suggest that values of ρ between 0.0 and 0.15 do not adversely impact the properties of the mis-specified tests too much.

6 Conclusions

In this paper, we explored the identification of state dependence in the presence of non-stationary transition probabilities. We presented a very simple result that allows us to identify state dependence in the dynamic conditional logit model with fixed effects. We then showed how it is a straight-forward exercise to use this result to derive a test for the presence of state dependence in a non-stationary process. Finally, we concluded the paper with an investigation of the impact of non-stationarity on tests for state dependence which erroneously specify the underlying data generating process as stationary in both large and finite samples.

One future research avenue that is suggested by this note is an investigation into whether or not it is possible to point-estimate the parameters of a dynamic binary choice model when the underlying process is non-stationary and without imposing assumptions on the heterogeneity. Recent work by Honoré and Tamer (2004) sheds some light on this issue. In their paper, they construct bounds on the parameters of the model in equation (1). While the identified regions in their exercise are often small, they are never singletons. This suggests that the matching strategy of Honoré and Kyriazidou (2000) is essential for point-identification and, thus, that it is not possible to point-estimate the parameters of such a model.

7 Appendix - Proofs

7.1 Proposition 2

Proof. We begin by defining $\pi_{A_1} \equiv P(A_1)$, $\pi_{A_{12}} \equiv P(A_1 \cup A_2)$, $\pi_{B_1} \equiv P(B_1)$ and $\pi_{B_{12}} \equiv P(B_1 \cup B_2)$. Next, using the notation from the body of the paper, we will have that

$$\sqrt{N}\begin{pmatrix} \widehat{\pi}_{A_1} - \pi_{A_1} \\ \widehat{\pi}_{A_{12}} - \pi_{A_{12}} \\ \widehat{\pi}_{B_1} - \pi_{B_1} \\ \widehat{\pi}_{B_{12}} - \pi_{B_{12}} \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \mathbf{\Sigma}_A & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{\Sigma}_B \end{pmatrix}\right). \tag{20}$$

where

$$\mathbf{\Sigma}_A \equiv \begin{pmatrix} \pi_{A_1}(1-\pi_{A_1}) & \pi_{A_1}(1-\pi_{A_{12}}) \\ \pi_{A_1}(1-\pi_{A_{12}}) & \pi_{A_{12}}(1-\pi_{A_{12}}) \end{pmatrix}, \tag{21}$$

$$\mathbf{\Sigma}_B \equiv \begin{pmatrix} \pi_{B_1}(1-\pi_{B_1}) & \pi_{B_1}(1-\pi_{B_{12}}) \\ \pi_{B_1}(1-\pi_{B_{12}}) & \pi_{B_{12}}(1-\pi_{B_{12}}) \end{pmatrix} \tag{22}$$

and $\mathbf{0}_4$ is a 4 by 4 matrix of zeros. The asymptotic covariance is block diagonal since the events A_1 and A_{12} and the events B_1 and B_{12} are mutually exclusive and because the sample is random. If we define the mapping $f(x_1, y_1, x_2, y_2) = (\frac{x_1}{y_1}, \frac{x_2}{y_2})$ and apply the δ-method to (20), we obtain

$$\sqrt{N}\begin{pmatrix} \widehat{\pi}_A - \pi_A(\gamma, \delta_2, \delta_1) \\ \widehat{\pi}_B - \pi_B(\gamma, \delta_2, \delta_1) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix}\right). \tag{23}$$

where

$$\sigma_A^2 = \frac{\pi_{A_1}}{\pi_{A_{12}}^3}\left(\pi_{A_{12}} - \pi_{A_1}\right) \tag{24}$$

and

$$\sigma_B^2 = \frac{\pi_{B_1}}{\pi_{B_{12}}^3} \left(\pi_{B_{12}} - \pi_{B_1}\right) \tag{25}$$

Equations (11) and (12) are simply the sample analogues of the asymptotic variances above. Next, we write

$$\sqrt{N}\left(\widehat{\pi}_A - \widehat{\pi}_B\right) = \tag{26}$$

$$\sqrt{N}\left(\underbrace{(\widehat{\pi}_A - \pi_A(\gamma,\delta_2,\delta_1))}_{A_N(\gamma,\delta_2,\delta_1)} - \underbrace{(\widehat{\pi}_B - \pi_B(\gamma,\delta_2,\delta_1))}_{B_N(\gamma,\delta_2,\delta_1)} + \underbrace{(\pi_A(\gamma,\delta_2,\delta_1) - \pi_B(\gamma,\delta_2,\delta_1))}_{C_N(\gamma,\delta_2,\delta_1)}\right) \tag{27}$$

Now, because Proposition 1 tells us that $\pi_A(\gamma,\delta_2,\delta_1) = \pi_B(\gamma,\delta_2,\delta_1)$ when $\gamma = 0$ and because the asymptotic covariance between $\widehat{\pi}_A$ and $\widehat{\pi}_B$ is zero, in the absence of state dependence, we will have that

$$\sqrt{N}\left(\widehat{\pi}_A - \widehat{\pi}_B\right) \xrightarrow{d} N(0,\sigma_{AB}^2) \text{ for } \gamma = 0 \tag{28}$$

where

$$\sigma_{AB}^2 \equiv \frac{\pi_{A_1}}{\pi_{A_{12}}^3}\left(\pi_{A_{12}} - \pi_{A_1}\right) + \frac{\pi_{B_1}}{\pi_{B_{12}}^3}\left(\pi_{B_{12}} - \pi_{B_1}\right). \tag{29}$$

The Slutsky Theorem then gives us that

$$sd_1(0,\delta_2,\delta_1) \xrightarrow{d} N(0,1). \tag{30}$$

Next, in the case where γ is not zero, $\sqrt{N}(A_N(\gamma,\delta_2,\delta_1) - B_N(\gamma,\delta_2,\delta_1))$ will converge to $N(0,\sigma_{AB}^2)$ random variable, whereas $\sqrt{N}C_N(\gamma,\delta_2,\delta_1)$ will explode since $\pi_A(\gamma,\delta_2,\delta_1) \neq \pi_B(\gamma,\delta_2,\delta_1)$ in the presence of state dependence. In particular, if $\gamma > 0$, then Proposition 1 tells us that

$\pi_A(\gamma, \delta_2, \delta_1) > \pi_B(\gamma, \delta_2, \delta_1)$ and, thus, $\sqrt{N}C_N(\gamma, \delta_2, \delta_1)$ will go to positive infinity. If $\gamma < 0$, then the reverse is true. Consequently, we will have that

$$sd_1(\gamma, \delta_2, \delta_1) \to \pm\infty \text{ for } \gamma \gtrless 0. \tag{31}$$

■

7.2 Lemma 4

Proof. We begin by noting that

$$P(A_1|A_1 \cup A_2, \alpha_i) = \frac{1 + \exp(\alpha_i + \delta_2)}{1 + \exp(\alpha_i + \delta_2) + \exp(\alpha_i + 2\delta_2 - \delta_1) + \exp(-\gamma + \delta_2 - \delta_1)} \equiv \pi_A(\gamma, \delta_2, \delta_1, \alpha_i). \tag{32}$$

If we integrate over $G(\alpha_i|A_1 \cup A_2)$ we obtain that

$$\pi_A(\gamma, \delta_2, \delta_1) = \int \pi_A(\gamma, \delta_2, \delta_1, \alpha_i) dG(\alpha_i|A_1 \cup A_2). \tag{33}$$

Differentiating $\pi_A(\gamma, \delta_2, \delta_1, \alpha_i)$ with respect to δ_2, δ_1 and γ, we obtain

$$\frac{\partial \pi_A(\gamma, \delta_2, \delta_1, \alpha_i)}{\partial \delta_2} = \frac{-\exp(\alpha_i + 2\delta_2 - \delta_1)(\exp(\alpha_i + \delta_2) + 2) - \exp(-\gamma + \delta_2 - \delta_1)}{(1 + \exp(\alpha_i + \delta_2) + \exp(\alpha_i + 2\delta_2 - \delta_1) + \exp(-\gamma + \delta_2 - \delta_1))^2} < 0, \tag{34}$$

$$\frac{\partial \pi_A(\gamma, \delta_2, \delta_1, \alpha_i)}{\partial \delta_1} = \frac{\exp(\alpha_i + 2\delta_2 - \delta_1)(1 + \exp(\alpha_i + \delta_2) + \exp(-\gamma + \delta_2 - \delta_1))}{(1 + \exp(\alpha_i + \delta_2) + \exp(\alpha_i + 2\delta_2 - \delta_1) + \exp(-\gamma + \delta_2 - \delta_1))^2} > 0 \tag{35}$$

and

$$\frac{\partial \pi_A(\gamma, \delta_2, \delta_1, \alpha_i)}{\partial \gamma} = \frac{(1+\exp(\alpha_i+\delta_2))\exp(-\gamma+\delta_2-\delta_1)}{(1+\exp(\alpha_i+\delta_2)+\exp(\alpha_i+2\delta_2-\delta_1)+\exp(-\gamma+\delta_2-\delta_1))^2} > 0. \tag{36}$$

This verifies the first part of the lemma. To verify the second part of the proposition, it suffices to note that Proposition 1 together with the fact that $\delta_2 > \delta_1$ imply that

$$\pi_A(\gamma, \delta_2, \delta_1) = \Pi(\delta_2, \delta_1) < \frac{1}{2} \text{ for } \gamma = 0$$

since $\Pi(\delta_2, \delta_1) = [1+\exp(\delta_2-\delta_1)]^{-1}$. The final part of the theorem can be proven by letting γ approach infinity in equation (32). Doing this, we see that

$$\begin{aligned} \lim_{\gamma \to \infty} \pi_A(\gamma, \delta_2, \delta_1) &= l_A(\delta_2, \delta_1) \\ &= \int \frac{1+\exp(\alpha_i+\delta_2)}{1+\exp(\alpha_i+\delta_2)+\exp(\alpha_i+2\delta_2-\delta_1)} dG(\alpha_i|A_1 \cup A_2) \\ &\equiv \int l_A(\delta_2, \delta_1, \alpha_i) dG(\alpha_i|A_1 \cup A_2) \end{aligned} \tag{37}$$

Clearly, we will have that $l_A(\delta_2, \delta_1, \alpha_i) < 1$ for all α_i and (δ_2, δ_1) and, thus, we will have that $l_A(\delta_2, \delta_1) < 1$ for all (δ_2, δ_1), as well. In addition, for $\delta_1 = \delta_2 = \delta$, we can write

$$l_A(\delta, \delta, \alpha_i) = \left(1 + \frac{\exp(\alpha_i+\delta)}{1+\exp(\alpha_i+\delta)}\right)^{-1} > \frac{1}{2} \text{ for all } \alpha_i \tag{38}$$

which, in turn, gives us that $l_A(\delta, \delta) > \frac{1}{2}$ which proves the final part of the lemma. ■

7.3 Lemma 7

Proof. The proof of this lemma is almost the same as Lemma 4. We start out by writing

$$P(B_1|B_1 \cup B_2, \alpha_i) = \frac{1+\exp(\alpha_i+\delta_2)}{1+\exp(\alpha_i+\delta_2)+\exp(\alpha_i+2\delta_2-\delta_1)+\exp(\gamma+\delta_2-\delta_1)} \equiv \pi_B(\gamma,\delta_2,\delta_1,\alpha_i) \tag{39}$$

and

$$\pi_B(\gamma,\delta_2,\delta_1) = \int \pi_B(\gamma,\delta_2,\delta_1,\alpha_i)dF(\alpha_i|B_1 \cup B_2). \tag{40}$$

Simple calculations reveal that

$$\frac{\partial \pi_B(\gamma,\delta_2,\delta_1,\alpha_i)}{\partial \delta_2} = \frac{-\exp(\alpha_i+2\delta_2-\delta_1)(\exp(\alpha_i+\delta_2)+2)-\exp(\gamma+\delta_2-\delta_1)}{(1+\exp(\alpha_i+\delta_2)+\exp(\alpha_i+2\delta_2-\delta_1)+\exp(\gamma+\delta_2-\delta_1))^2} < 0 \tag{41}$$

$$\frac{\partial \pi_B(\gamma,\delta_2,\delta_1,\alpha_i)}{\partial \delta_1} = \frac{\exp(\alpha_i+2\delta_2-\delta_1)(1+\exp(\alpha_i+\delta_2)+\exp(\gamma+\delta_2-\delta_1))}{(1+\exp(\alpha_i+\delta_2)+\exp(\alpha_i+2\delta_2-\delta_1)+\exp(\gamma+\delta_2-\delta_1))^2} > 0 \tag{42}$$

and

$$\frac{\partial \pi_B(\gamma,\delta_2,\delta_1,\alpha_i)}{\partial \gamma} = \frac{-(1+\exp(\alpha_i+\delta_2))\exp(\gamma+\delta_2-\delta_1)}{(1+\exp(\alpha_i+\delta_2)+\exp(\alpha_i+2\delta_2-\delta_1)+\exp(\gamma+\delta_2-\delta_1))^2} < 0 \tag{43}$$

which verifies the first part of the lemma. The second part of the lemma follows from Proposition 1 and the observation that $\Pi(\delta_2,\delta_1) = [1+\exp(\delta_2-\delta_1)]^{-1} < \frac{1}{2}$ whenever $\delta_2 > \delta_1$. Finally, the third part of the lemma follows from allowing γ to go to infinity in (39) ■

References

[1] Chamberlain, G. (1985): "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, edited by J.J. Heckman and B. Singer. Cambridge: Cambridge University Press.

[2] Chamberlain, G. (1992): "Binary Response Models for Panel Data: Identification and Information," unpublished mimeo.

[3] Hahn, J. (2001): "The Information Bound of a Dynamic Logit Model with Fixed Effects," *Econometric Theory*, 17, 913-932.

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

[5] Honoré, B. and Kyriazidou, K. (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839 - 874.

[6] Honoré, B. and Tamer, E. (2004): "Bounds on Parameters in Dynamic Discrete Choice Models," unpublished mineo.

[7] Magnac, T. (2004): "Panel Binary Variables and Sufficiency: Generalizing Conditional Logit," *Econometrica*, 72, 1859-1876.

8 Tables and Figures

Figure 1

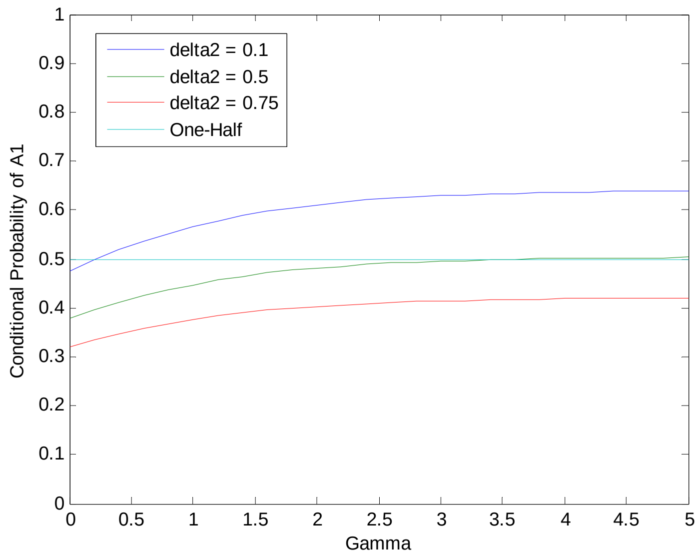
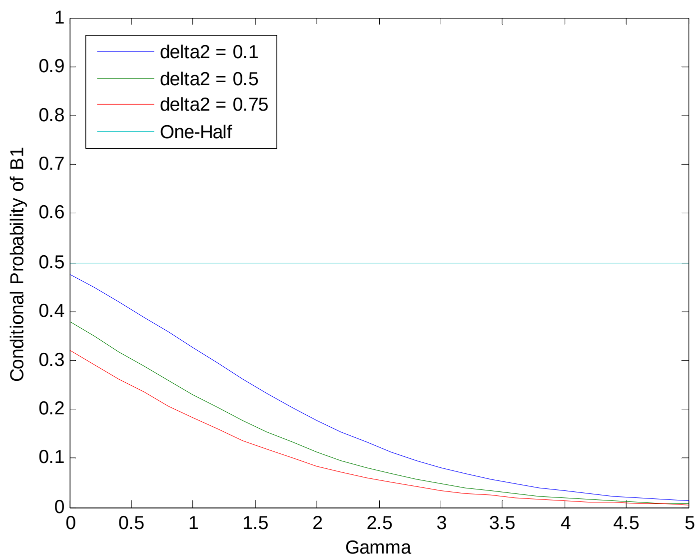


Figure 2



*In both figures, $\delta_1 = 0$.

Figure 3

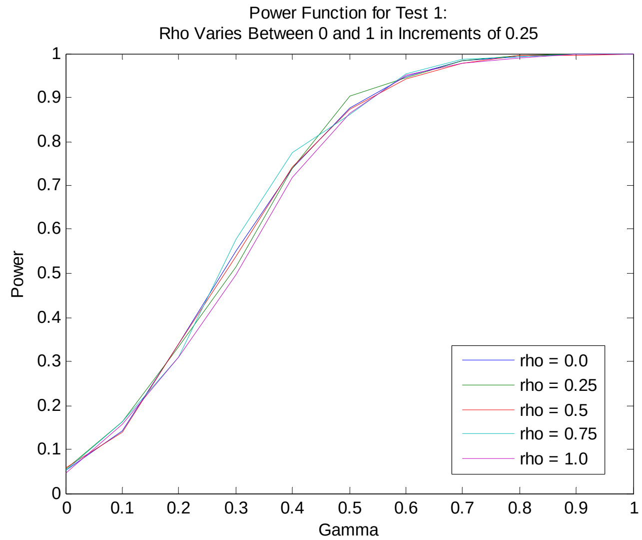


Figure 4

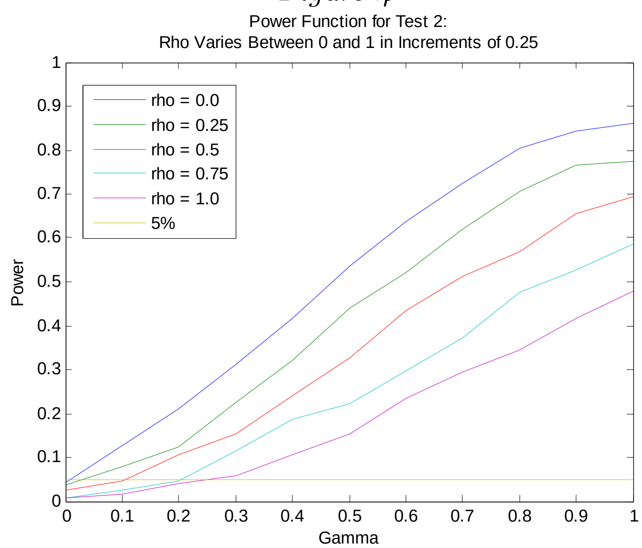


Figure 5

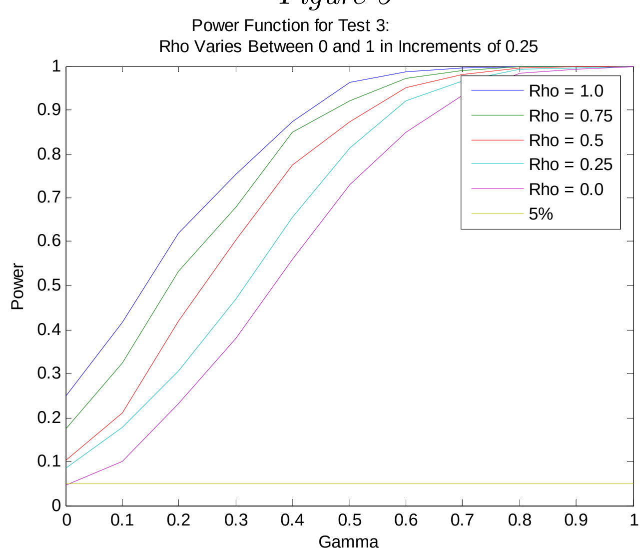


Figure 6

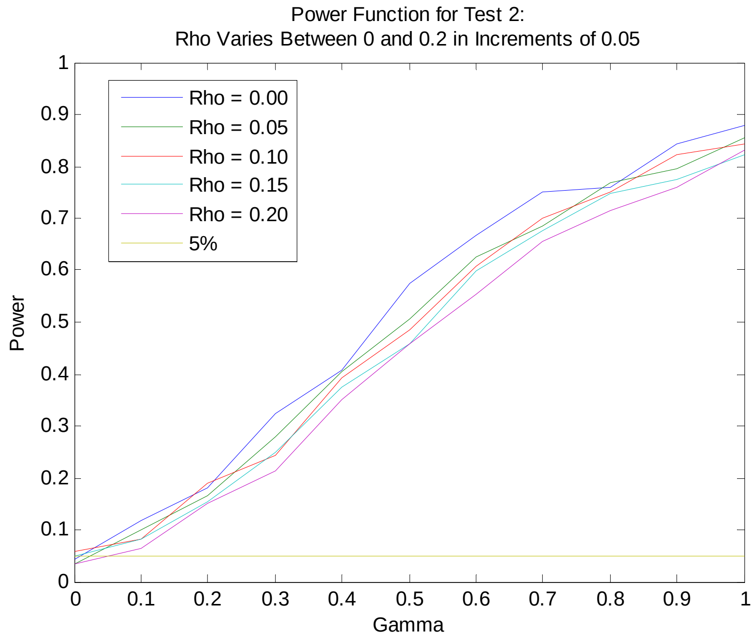


Figure 7