Taylor & Francis
Taylor & Francis Group

TESTING FOR STATE DEPENDENCE WITH TIME-VARIANT TRANSITION PROBABILITIES

Timothy J. Halliday □ *Department of Economics and John A. Burns
School of Medicine, University of Hawaii at Mānoa, Honolulu, Hawaii, USA*

□   *We derive a simple result that allows us to test for the presence of state dependence in a
dynamic Logit model with time-variant transition probabilities and an arbitrary distribution of
the unobserved heterogeneity. Monte Carlo evidence suggests that this test has desirable properties
even when there are some violations of the model's assumptions. We also consider alternative
tests that will have desirable properties only when the transition probabilities do not depend on
time and provide evidence that there is an "acceptable" range in which ignoring time-dependence
does not matter too much. We conclude with an application to the Barker Hypothesis.*

1. INTRODUCTION

Economic data often display serial correlation. This is true in the case
of labor force participation, accident occurrence, and numerous measures
of health. However, the source of this persistence is often unclear. One
possibility 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 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 (2006) 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.

In this paper, we present a simple result that allows us to detect state dependence in the presence of time-varying transition probabilities. We then go on to show how this result can be used to derive a consistent test for the presence of state dependence in Section 3. In Section 4, we conduct Monte Carlo experiments which suggest that there is an "acceptable" range of time dependence in which ignoring age or time effects does not matter too much. In this section, we also show that the test statistic still performs well even when some of the underlying assumptions of the model fail, although we also show that an egregious violation of the model's assumption of exogenous initial conditions will result in poor performance. In Section 5, we employ our test statistic to better understand the persistence of health. Section 6 concludes.

2. A VERY SIMPLE RESULT

In this section, we establish a simple result that allows us to identify the sign of the state dependence parameter in a dynamic Logit model when

the underlying process has time-variant transition probabilities.[1] 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 $i = 1, \ldots, N$ and $t = 1, \ldots, T$. In Eq. (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. We allow $f_t(x_{i,t})$ to be 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 $(\alpha_i, x_{i,1}, \ldots, x_{i,T}, y_{i,0})$ 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 $(y_{i,T}, \ldots, y_{i,0}, x_{i,T}, \ldots, x_{i,1})$ from some underlying population.

To derive our result, we define the events

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

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

As it turns out, there is a "difference-in-difference" intuition behind choosing these four events. First, we will show that we will be able to bound an estimable function of the events A_1 and A_2 by a number which does not depend on α_i. This is the first "difference." A similar argument holds for the events B_1 and B_2. However, while this bound will be purged of the heterogeneity, it will still depend on $f_t(x_{i,t})$ which is an unknown function. Thus, the second "difference" will subtract the function of A_1 and A_2 from B_1 and B_2 to eliminate the dependence on $f_t(x_{i,t})$. We now make these arguments formal.

We begin the derivation by noting that, for $\gamma \geq 0$, the assumptions of the model are sufficient to show that

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

---

[1]We must emphasize that this result does *not* allow us to *point*-identify the state dependence parameter.

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

$$\frac{P(B_1 \mid x_i, \alpha_i)}{P(B_2 \mid x_i, \alpha_i)} = \frac{(1 - \Lambda(\alpha_i + \gamma + f_2(x_{i,2})))\Lambda(\alpha_i + f_1(x_{i,1}))}{\Lambda(\alpha_i + f_2(x_{i,2}))(1 - \Lambda(\alpha_i + f_1(x_{i,1})))}$$

$$\leq \frac{(1 - \Lambda(\alpha_i + f_2(x_{i,2})))\Lambda(\alpha_i + f_1(x_{i,1}))}{\Lambda(\alpha_i + f_2(x_{i,2}))(1 - \Lambda(\alpha_i + f_1(x_{i,1})))}$$

$$= \exp(f_1(x_{i,1}) - f_2(x_{i,2})). \tag{5}$$

In the presence of negative state dependence, the inequalities in (4) and (5) are reversed. These inequalities provide bounds which are independent of the heterogeneity and, thus, constitute the first "difference." Next, denoting $\Pi(x_i) \equiv [1 + \exp(f_2(x_{i,2}) - f_1(x_{i,1}))]^{-1}$, we will have that

$$P(A_1 \mid A_1 \cup A_2, x_i, \alpha_i) \geq \Pi(x_i) \geq P(B_1 \mid B_1 \cup B_2, x_i, \alpha_i)$$

$$\text{for } \gamma \geq 0 \quad \text{and} \quad \text{all } \alpha_i \tag{6}$$

$$P(A_1 \mid A_1 \cup A_2, x_i, \alpha_i) \leq \Pi(x_i) \leq P(B_1 \mid B_1 \cup B_2, x_i, \alpha_i)$$

$$\text{for } \gamma \leq 0 \quad \text{and} \quad \text{all } \alpha_i. \tag{7}$$

Clearly, when there is no state dependence, the model becomes the static conditional Logit model and (4)–(7) will hold with strict equality. The system of inequalities in Eqs. (6) and (7) shows how we can construct bounds that do not depend on $f_t(x_{i,t})$ and, thus, provides us with the second "difference."

We can further expand upon the intuition behind using the two sets of events (A_1, A_2) and (B_1, B_2). First, both sets of events contain the same initial condition which enables us to purge the ratios in (4) and (5) of the probability of $y_{i,0}$. Second, the sum, $y_{i,0} + y_{i,1} + y_{i,2}$ is the same for both events in the pairs (A_1, A_2) and (B_1, B_2). For the pair (A_1, A_2), it is two and for the pair (B_1, B_2), it is unity. However, the sum is never zero or three. Thus, if no state dependence is present, then the model simply becomes the static conditional Logit model and $y_{i,0} + y_{i,1} + y_{i,2}$ serves as a sufficient statistic for the unobserved heterogeneity.

Finally, note that all of our statements, thus far, have been conditional on the unobserved heterogeneity. This poses problems because, while the probabilities $P(A_1 \mid A_1 \cup A_2, x_i)$ and $P(B_1 \mid B_1 \cup B_2, x_i)$ can easily be estimated nonparametrically, the probabilities $P(A_1 \mid A_1 \cup A_2, x_i, \alpha_i)$ and $P(B_1 \mid B_1 \cup B_2, x_i, \alpha_i)$ cannot be estimated as they depend on α_i. However, it is important to note 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). Accordingly, we can make a

statement that is no longer conditional on the heterogeneity by integrating out α_i.[2] This gives us Proposition 1 which is our key result.[3]

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

$$P(A_1 \mid A_1 \cup A_2, x_i) \gtreqqless P(B_1 \mid B_1 \cup B_2, x_i) \qquad for \; \gamma \gtreqqless 0.$$

3. THE TEST STATISTIC

It is a straightforward exercise to use the result of the previous section to derive a consistent test for the presence of state dependence. 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. To keep matters simple, throughout the remainder of the paper, we consider the case where the only elements of $x_{i,t}$ are time dummies 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 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.

---

[2]To see this formally, let $G(\alpha_i \mid A_1 \cup A_2, x_i)$ denote the distribution of the heterogeneity conditional on $(A_1 \cup A_2, x_i)$ and let $F(\alpha_i \mid B_1 \cup B_2, x_i)$ denote the distribution of the heterogeneity conditional on $(B_1 \cup B_2, x_i)$. The inequalities in (6) and (7) imply that

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

and

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

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

[3]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 $P(A_1 \mid A_1 \cup A_2, x_i, \alpha_i)$ and $P(B_1 \mid B_1 \cup B_2, x_i, \alpha_i)$ with $\Pi(x_i)$ which does not depend on α_i. Accordingly, without the logistic assumption, it is not possible to separate these probabilities with a constant that does not depend on the unobserved heterogeneity.

We can estimate the probability $P(A_1 \mid A_1 \cup A_2) \equiv \pi_A(\gamma)$ via

$$\hat{\pi}_A = \frac{\sum_{i=1}^{N} 1_i(A_1)}{\sum_{i=1}^{N} 1_i(A_1 \cup A_2)} \tag{8}$$

We define $\hat{\pi}_B$, the estimate of $P(B_1 \mid B_1 \cup B_2) \equiv \pi_B(\gamma)$, in an analogous way. Clearly, Eq. (8) 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 subpopulation for whom x_i takes on some discrete value.[4] Next, we let $\hat{\pi}_{A_1}$ and $\hat{\pi}_{A_{12}}$ denote estimates of $P(A_1)$ and $P(A_1 \cup A_2)$. We define $\hat{\pi}_{B_1}$ and $\hat{\pi}_{B_{12}}$ in a similar fashion. In the Appendix, we show that the asymptotic variances of $\hat{\pi}_A$ and $\hat{\pi}_B$ can be estimated with

$$\hat{\sigma}_A^2 = \frac{\hat{\pi}_{A_1}}{\hat{\pi}_{A_{12}}^3}(\hat{\pi}_{A_{12}} - \hat{\pi}_{A_1}) \tag{9}$$

and

$$\hat{\sigma}_B^2 = \frac{\hat{\pi}_{B_1}}{\hat{\pi}_{B_{12}}^3}(\hat{\pi}_{B_{12}} - \hat{\pi}_{B_1}). \tag{10}$$

In the Appendix, we also show that the covariance between $\hat{\pi}_A$ and $\hat{\pi}_B$ is zero. We now define the statistic

$$sd(\gamma) = \frac{\hat{\pi}_A - \hat{\pi}_B}{\left(\frac{\hat{\sigma}_{AB}^2}{N}\right)^{1/2}} \tag{11}$$

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

$$sd(\gamma) = \sqrt{N}\left(\underbrace{\left(\frac{\hat{\pi}_A - \pi_A(\gamma)}{\hat{\sigma}_{AB}}\right)}_{X_N(\gamma)} - \underbrace{\left(\frac{\hat{\pi}_B - \pi_B(\gamma)}{\hat{\sigma}_{AB}}\right)}_{Y_N(\gamma)} + \underbrace{\left(\frac{\pi_A(\gamma) - \pi_B(\gamma)}{\hat{\sigma}_{AB}}\right)}_{Z_N(\gamma)}\right). \tag{12}$$

In the Appendix, we show that $\sqrt{N}(X_N(\gamma) - Y_N(\gamma))$ will converge in distribution to a $N(0,1)$ random variable regardless of the values of $(\gamma, \delta_2, \delta_1)$. However, Proposition 1 tells us that $Z_N(\gamma)$ will only be zero when $\gamma = 0$; otherwise, it will be positive when $\gamma > 0$, and negative when $\gamma < 0$. Consequently, $sd(\gamma)$ will converge to a Standard Normal random variable

---

[4]The probabilities $P(A_1 \mid A_1 \cup A_2, x_i)$ and $P(B_1 \mid B_1 \cup B_2, x_i)$ can also be estimated nonparametrically when $x_{i,t}$ is continuous via kernel or sieve estimation. However, because $x_i = (x_{i,1}, x_{i,2})$, the dimension of the estimated expectation will be twice the dimension of $x_{i,t}$ and, thus, the "curse of dimensionality" is likely to be a major problem with continuous regressors.

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*

$$sd(\gamma) \overset{d}{\to} N(0,1) \qquad \text{for } \gamma = 0.$$

and

$$sd(\gamma) \to \pm\infty \qquad \text{for } \gamma \gtrless 0.$$

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 $\Phi(\cdot)$ denote the CDF of a $N(0,1)$ random variable and define $z_\varphi = \Phi^{-1}(1 - \varphi)$, then a test of size φ can be constructed if we reject the null whenever $sd(\gamma) > z_\varphi$. Because $sd(\gamma)$ shoots off to positive infinity whenever $\gamma > 0$, the power of this test will approach unity as the sample size increases. Thus, we have a consistent test.[5]

4. MONTE CARLO EVIDENCE

In this section, we investigate the performance of three tests for state dependence using Monte Carlo experiments. The first is based on the statistic $sd(\gamma)$ from the previous section. The second is based on the statistic

$$sd_A(\gamma) \equiv \frac{\hat{\pi}_A - \frac{1}{2}}{\left(\frac{\hat{\sigma}_A^2}{N}\right)^{1/2}} \tag{13}$$

and the third is based on the statistic

$$sd_B(\gamma) \equiv \frac{\hat{\pi}_B - \frac{1}{2}}{\left(\frac{\hat{\sigma}_B^2}{N}\right)^{1/2}}. \tag{14}$$

Note that the last two statistics will provide us with a consistent test for state dependence only if the transition probabilities do not depend on time or equivalently if $\delta_1 = \delta_2$. In such a scenario, we will have

---

[5]In the presence of a vector of strictly exogenous regressors, we can use the statistic in this section as an ad hoc specification test. The reason is that the inequality in Proposition 1 must hold for all x_i. Accordingly, if we find that the statistic $sd(\gamma)$ is positive and significantly different from zero for some values of x_i, but negative and significantly different from zero for other values of x_i, then it would suggest that the model in Eq. (1) may not be appropriate for all the possible values of x_i. Rather it may suggest that a more complicated model in which the state dependence parameter depends on the strictly exogenous regressors is a more appropriate model for the data. I am grateful to an anonymous referee for pointing this out.

that $\pi_A(0) = \frac{1}{2} = \pi_B(0)$, but $\pi_A(\gamma) > \frac{1}{2} > \pi_B(\gamma)$ for $\gamma > 0$. However, if the transition probabilities do depend on time, then these tests will have undesirable properties which we will see later on in this section.

The goals of the Monte Carlo exercises in this section are threefold. First, we want to better understand how time-variant transition probabilities will affect our ability to detect state dependence in a finite sample. Second, we want to explore how a mis-specification of the distribution for $\varepsilon_{i,t}$ will affect the properties of tests based on $sd(\gamma)$. Third, we want to investigate the performance of our statistic as the sample size becomes small.

To accomplish these ends, we generate data from the model

$$y_{i,1} = 1(\alpha_i + \varepsilon_{i,0} \geq 0) \tag{15}$$

and

$$y_{i,t} = 1(\alpha_i + y_{i,t-1}\gamma + \rho * t/10 + \varepsilon_{i,t} \geq 0) \qquad \text{for } t = 1, 2. \tag{16}$$

The residual, $\varepsilon_{i,t}$, is i.i.d. across time and, in most of the exercises, is independent of α_i. We experiment with different distributional assumptions on $\varepsilon_{i,t}$ such as the Logistic as well as Normal distributions both with and without exogenous initial conditions. 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 all but the last three experiments. For each simulation, we consider a test of $H_0 : \gamma = 0$ against $H_a : \gamma > 0$. We calculate the power functions for each of the three statistics for different values of ρ. Each figure plots the percentage of times that the test statistic surpasses the 95% critical value as a function of the parameters γ and ρ. Figures 1 through 12 each correspond to a separate experiment. Tests 1–3 refer to the statistics $sd(\gamma)$, $sd_A(\gamma)$, and $sd_B(\gamma)$, respectively.

First, we consider Figures 1 through 3. In Figure 1, we plot the power function for a test that uses $sd(\gamma)$, the test statistic from the previous section. We allow ρ to vary between 0.0 and 1.0 in increments of 0.25 and we assume Logistic disturbances. As can be seen in the figure, the test has desirable properties. However, this is not too surprising since everything is correctly specified. In Figures 2 and 3, we plot the respective power functions for $sd_A(\gamma)$ and $sd_B(\gamma)$ using the same values of ρ that were used in Figure 1 and, once again, assuming Logistic disturbances. Both figures show that these alternative tests have desirable properties when $\rho = 0$, but when ρ is large, the properties of the tests become highly undesirable.

However, this is not too surprising given that we are allowing ρ to take on rather large values. Perhaps a more interesting exercise would be to investigate the properties of $sd_A(\gamma)$ and $sd_B(\gamma)$ when we confine ρ to values between 0.0 and 0.20 so that the transition probabilities only have
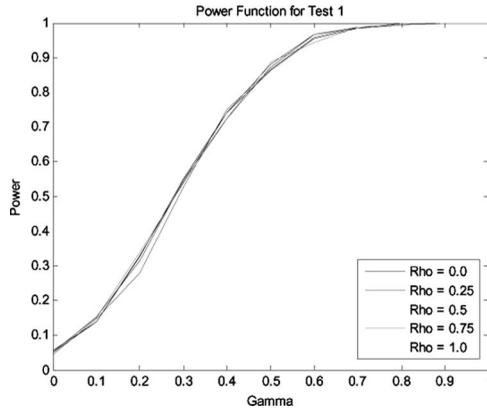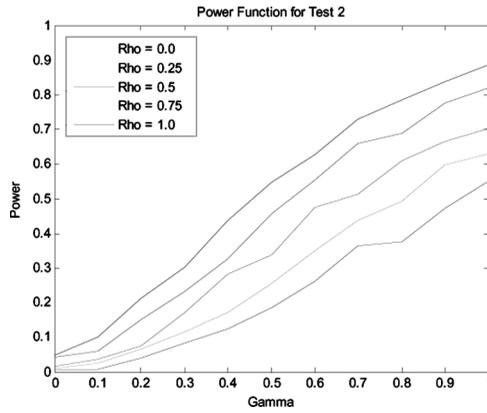
FIGURE 1 $sd(\gamma)$ with Logistic Residuals.

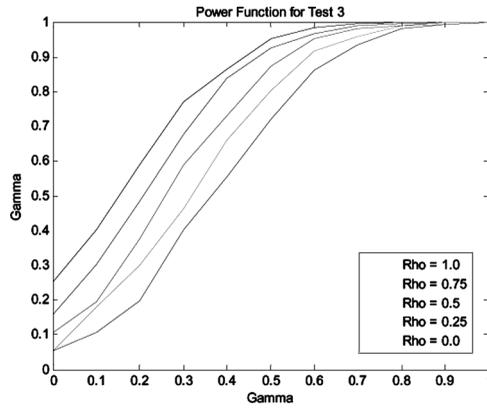**FIGURE 2** $sd_A(\gamma)$ with Logistic Residuals.

FIGURE 3 $sd_B(\gamma)$ with Logistic Residuals.
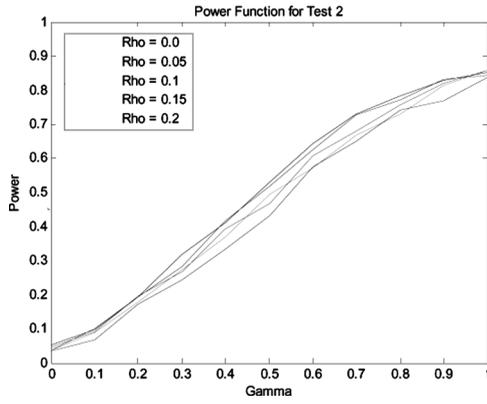
T. J. Halliday

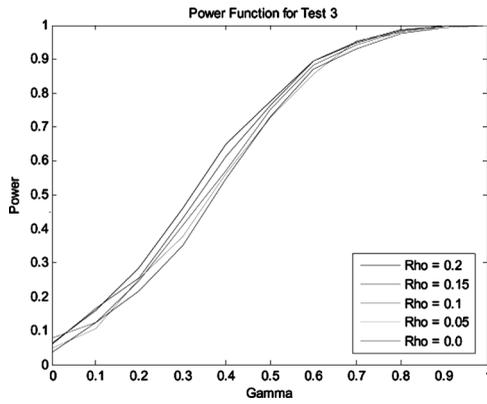**FIGURE 4** $sd_A(\gamma)$ with Logistic Residuals.

FIGURE 5 $sd_B(\gamma)$ with Logistic Residuals.
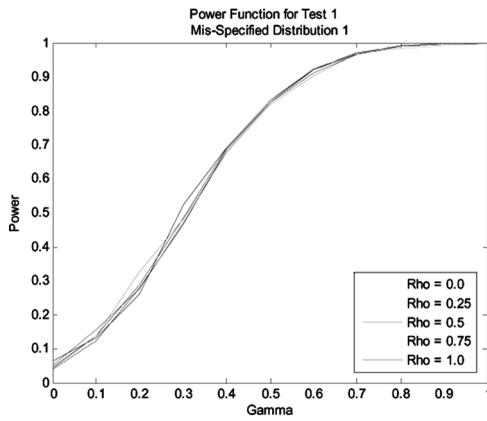

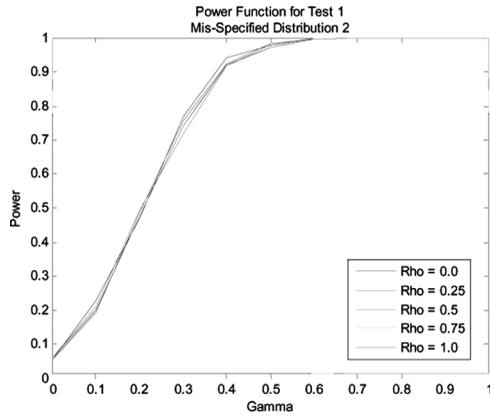**FIGURE 6** $sd(\gamma)$ with Normal Residuals, $\text{Var} = \frac{\pi^2}{3}$.

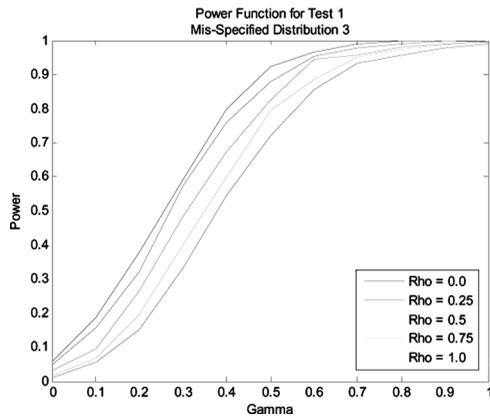**FIGURE 7** $sd(\gamma)$ with Normal Residuals, Var $= 1$.

FIGURE 8 $sd(\gamma)$ with Normal Residuals, Var $= 1^{y_{i,0}} 5^{1-y_{i,0}}$.
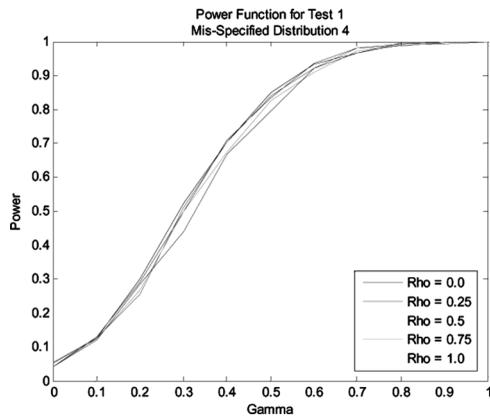

FIGURE 9 $sd(\gamma)$ with Normal Residuals, Var $= 2.5^{y_{i,0}} 3.5^{1-y_{i,0}}$.
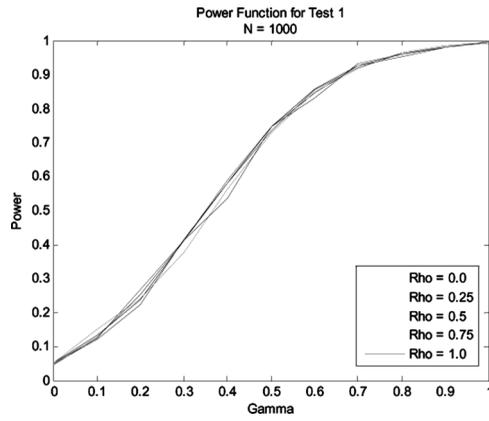
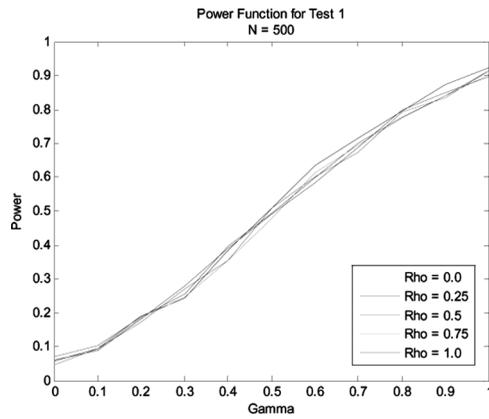**FIGURE 10** $sd(\gamma)$ with Logistic Residuals.

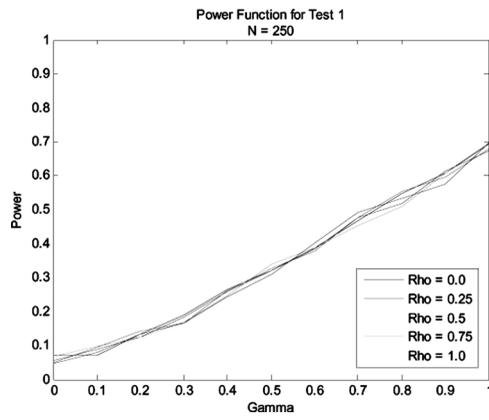**FIGURE 11** $sd(\gamma)$ with Logistic Residuals.

FIGURE 12 $sd(\gamma)$ with Logistic Residuals.

a small degree of time dependence. We report the results of this exercise in Figures 4 and 5 for $sd_A(\gamma)$ and $sd_B(\gamma)$, respectively. Once again, we have assumed Logistic disturbances. What we now see is that these alternative tests have far better properties than before. This suggests that ignoring time-dependent transition probabilities may not matter too much provided that the degree of time dependence is small.[6]

We now investigate the properties of $sd(\gamma)$ when the distribution of $\varepsilon_{i,t}$ is mis-specified. We consider four different mean zero Normal distributions with variances equal to $\pi^2/3, 1, 1^{y_{i,0}}5^{1-y_{i,0}}$ and $2.5^{y_{i,0}}3.5^{1-y_{i,0}}$. The results of these exercises are reported in Figures 6 through 9 for each of the respective distributions.[7] Note that the initial condition is no longer exogenous in the last two distributions. In Figure 6, we see that our test still has desirable properties. However, this is not too surprising since this distribution is very close to a Logistic distribution. In Figure 7, we assume a Standard Normal distribution for the residuals. We still see that the test has desirable properties which, perhaps, is slightly more surprising. In Figures 8 and 9, we allow for heteroskedasticity. In Figure 8, where the variance of the residual is $1^{y_{i,0}}5^{1-y_{i,0}}$, we now see that the desirable properties of the test are starting to break down. In Figure 9, we allow for a more muted degree of heteroskedasticity and allow the variance of the residual to be $2.5^{y_{i,0}}3.5^{1-y_{i,0}}$. We observe, once again, that the properties of the test are desirable.

These experiments suggest that heteroskedasticity of the form discussed above is a potentially large threat to the performance of the test statistic from Section 3. This is not too surprising since it violates the assumption that the disturbances must be independent of the initial condition. When the variance of the residual depends on the initial condition, the probabilities $\pi_A(\gamma)$ and $\pi_B(\gamma)$ will be generated by two distinct data generating processes. Thus, under the null of no state dependence, these two probabilities are apt to be quite different from each other and, consequently, we would expect the performance of the test statistic to be poor.

We conclude the Monte Carlo exercises by investigating the impact of smaller sample sizes on the performance of $sd(\gamma)$. In Figures 10 through 12, we allow the sample size to be 1000, 500, and 250, respectively. In all three figures, we assume Logistic disturbances. We see that the test still

---

[6]These two simulations suggest an ad hoc criterion to use when deciding whether or not the Honoré and Kyriazidou estimator is appropriate in which the dependent variable should be regressed on a time trend (divided by ten). If the parameter estimate is less than 0.2, then we hypothesize that the performance of the dynamic conditional Logit model may not be too bad. An interesting exercise would be to investigate this hypothesis using Monte Carlo experiments.

[7]The initial condition in for the simulations in each of these figures was generated using a normal distribution with mean zero and variance $\pi^2/3$.

performs well when the sample size is 1000. However, when the sample sizes are either 500 or 250, the power of the test is diminished substantially.

5. EMPIRICAL APPLICATION: THE BARKER HYPOTHESIS

In this section, we use our test for state dependence to better understand the determinants of adult health. Health provides us with an excellent application since it is persistent and its transition probabilities are highly dependent on age. Following Halliday (2005), we decompose the persistence of health into two components: state dependence and unobserved heterogeneity. If the former source of persistence is important then idiosyncratic shocks to health today will have large dynamic consequences. If the latter source is important then an adult's health is primarily determined by unobserved individual characteristics. This would lend credence to the Barker Hypothesis (Barker, 1997) which states that health in adulthood is to a large degree determined by events that have occurred in early childhood and, perhaps, in utero. Recent work by Almond (2006) and Case et al. (2002) has provided considerable evidence in favor of these early childhood factors. On the other hand, a relatively important role for state dependence would suggest that adult health is largely determined by idiosyncratic events that have occurred in adulthood.

The two different sources of persistence have different implications for how health policy should be conducted. If unobserved heterogeneity is relatively more important then policy should focus on pregnant mothers and young children. However, in such a scenario, interventions that target adults will not have large impacts on health. On the contrary, a large role for state dependence suggests that health policy can still target adults and have large effects on societal health.

The data that we employ come from the years 1984 to 1997 of the Panel Study of Income Dynamics (PSID). Our measure of health is Self-Reported Health Status (SRHS) which is a categorical variable that takes on integer values between one and five. One is the best category and five is the worst category. We map SRHS into a binary variable that equals one when SRHS is either four or five, but is otherwise zero. While these data are subjective, Halliday (2006) has shown that they are highly informative of mortality in the PSID. We use data on people between ages 30 and 32 and ages 50 and 52. We calculate the statistics by gender, by age and by race.

We report the results of calculating the statistics $sd(\gamma), sd_A(\gamma),$ and $sd_B(\gamma)$ in Table 1. We report the latter two statistics to see what happens if a researcher uses tests that only have desirable properties if the transition probabilities do not depend on time. In addition, we report N_A and N_B which are the number of times that the events $A_1 \cup A_2$ and $B_1 \cup B_2$ occur in

TABLE 1 Empirical application

	N	N_A^1	N_B^2	$sd(\gamma)$	$sd_A(\gamma)$	$sd_B(\gamma)$
Men ages 30 to 32						
White	925	12	35	−0.09	0.00	0.17
Black	475	13	35	1.36*	0.86	−1.21
All	2025	33	94	1.50*	1.25	−0.83
Women ages 30 to 32						
White	940	13	43	−0.67	−0.86	−0.15
Black	569	20	60	1.31*	0.00	−2.74**
All	2270	44	152	1.13	−0.30	−3.01**
Men ages 50 to 52						
White	305	14	15	−0.54	0.00	0.79
Black	105	5	15	0.00	−0.46	−0.79
All	691	29	39	−0.53	−0.56	0.16
Women ages 50 to 52						
White	346	20	33	0.54	0.00	−0.88
Black	180	18	18	−2.18	−0.47	2.83
All	836	50	73	−0.07	0.00	0.12

*Denotes significance at the 90% level.
**Denotes significance at the 99% level.
[1]N_A is the number of times that either A_1 or A_2 occurred in the data.
[2]N_B is the number of times that either B_1 or B_2 occurred in the data.

our sub-samples. We call these numbers the "effective" sample sizes. These numbers are important because they will have a large impact on the power of our test. The results of this exercise reveal several interesting findings.

First and perhaps most importantly for applied researchers, is that, despite having more than 500 individuals in most of our sub-samples, the "effective" sample size is very low. For example, we see that for white men between ages 30 and 32, $A_1 \cup A_2$ is only observed 12 times and $B_1 \cup B_2$ is only observed 35 times. These small "effective" sample sizes suggest that the test statistic may have low power.

To better understand how low the power of the test might be, we conduct some simple calculations and refer back to Figures 11 and 12 from the Monte Carlo experiments. If we simulate the model in Eqs. (15) and (16) five times with Logistic disturbances and the same distribution for α_i that we assumed in Section 4, we obtain average values for N_A and N_B of 74.6 and 91.8 when $\gamma = 1.0$, $\rho = 0.6$, and $N = 500$. When $N = 250$ and for the same parameter values, the average values of N_A and N_B drop to 36.6 and 41.4. The values of N_A in Table 1 for the pooled samples (i.e., those that are not stratified by race) are broadly consistent with the simulations when $N = 250$, whereas the values of N_B for the pooled samples are closer to the simulations when $N = 500$. Accordingly, we would expect the power of $sd(\gamma)$ in this application to be somewhere between the power functions which are depicted in Figures 11 and 12 which is not terribly high.

Second, looking at the statistics that are calculated with the sub-sample that pools all races, we do not find strong evidence for state dependence. The only significant evidence of positive state dependence is for men between ages 30 and 32. All other statistics cannot reject the null of no state dependence. However, it is important to emphasize once again that the "effective" sample sizes are still on the low end. It is also important to mention that the calculation of $sd_B(\gamma)$ for women between ages 30 and 32 rejects the null of no state dependence, whereas the calculation of $sd(\gamma)$ does not which suggests that tests that are not robust to time-variant transition probabilities may yield incorrect conclusions.[8]

6. CONCLUSIONS

In this paper, we presented a very simple result that allows us to detect state dependence in a dynamic Logit model with fixed effects and time-variant transition probabilities. We then showed how it is a straight-forward exercise to use this result to derive a test for the presence of state dependence when age or time effects are present. We conducted Monte Carlo experiments that suggested that this test statistic still has nice properties when some of the assumptions of the model fail. We also provided evidence that ignoring time effects may not matter too much provided that the degree of time dependence is modest. We concluded by using our test statistic to better understand the persistence of health. The test did not uncover evidence of state dependence in health. However, we are careful to subject the reader to the caveat that the power of the test in this application is likely to be low.

One future research avenue that is suggested by this paper is an investigation into whether or not it is possible to point-estimate the parameters of a dynamic binary choice model when the underlying transition probabilities depend on time and without imposing assumptions on the heterogeneity. Recent work by Honoré and Tamer (2006) sheds some light on this issue. In their paper, they construct bounds on the parameters of the model in Eq. (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.

---

[8]The findings in Table 1 are consistent with results in Halliday (2005) who adopts a random effects approach to estimate a model of health dynamics similar to Eq. (1). While he did find that the estimates of the state dependence parameters were statistically different from zero, he also showed that the difference between the probabilities $\Lambda(\alpha_i + \gamma + f_i(x_{i,t})) - \Lambda(\alpha_i + f_i(x_{i,t}))$ was quite small for a large percentage of the population. In addition, he found that there was a lot of variation in the distribution of the heterogeneity suggesting an important role for unobservables.

7. APPENDIX: PROOF OF 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 $\pi_{B_{12}} \equiv P(B_1 \cup B_2)$. By the Central Limit Theorem and using the notation from the body of the paper, we will have that

$$\sqrt{N} \begin{pmatrix} \hat{\pi}_{A_1} - \pi_{A_1} \\ \hat{\pi}_{A_{12}} - \pi_{A_{12}} \\ \hat{\pi}_{B_1} - \pi_{B_1} \\ \hat{\pi}_{B_{12}} - \pi_{B_{12}} \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \Sigma_A & \Sigma_{AB} \\ \Sigma'_{AB} & \Sigma_B \end{pmatrix}\right) \tag{17}$$

where

$$\Sigma_A \equiv \begin{pmatrix} \pi_{A_1}(1 - \pi_{A_1}) & \pi_{A_1}(1 - \pi_{A_{12}}) \\ \pi_{A_1}(1 - \pi_{A_{12}}) & \pi_{A_{12}}(1 - \pi_{A_{12}}) \end{pmatrix}, \tag{18}$$

$$\Sigma_B \equiv \begin{pmatrix} \pi_{B_1}(1 - \pi_{B_1}) & \pi_{B_1}(1 - \pi_{B_{12}}) \\ \pi_{B_1}(1 - \pi_{B_{12}}) & \pi_{B_{12}}(1 - \pi_{B_{12}}) \end{pmatrix} \quad \text{and} \tag{19}$$

$$\Sigma_{AB} \equiv \begin{pmatrix} -\pi_{A_1}\pi_{B_1} & -\pi_{A_1}\pi_{B_{12}} \\ -\pi_{A_{12}}\pi_{B_1} & -\pi_{A_{12}}\pi_{B_{12}} \end{pmatrix}. \tag{20}$$

Next, 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 (17), we obtain

$$\sqrt{N} \begin{pmatrix} \hat{\pi}_A - \pi_A(\gamma) \\ \hat{\pi}_B - \pi_B(\gamma) \end{pmatrix} \xrightarrow{d} N(0, \mathbf{V}\Sigma\mathbf{V}') \tag{21}$$

where

$$\mathbf{V} \equiv \begin{pmatrix} \frac{1}{\pi_{A_{12}}} & -\frac{\pi_{A_1}}{\pi_{A_{12}}^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\pi_{B_{12}}} & -\frac{\pi_{B_1}}{\pi_{B_{12}}^2} \end{pmatrix} \equiv (\mathbf{V}_A \mathbf{V}_B). \tag{22}$$

Consequently, we will have that

$$\mathbf{V}\Sigma\mathbf{V}' = \mathbf{V}_A\Sigma_A\mathbf{V}'_A + \mathbf{V}_B\Sigma_B\mathbf{V}'_B + \mathbf{V}_A\Sigma_{AB}\mathbf{V}'_B + \mathbf{V}_B\Sigma'_{AB}\mathbf{V}'_A. \tag{23}$$

The sum of the first two terms can be shown to equal

$$\mathbf{V}_A\Sigma_A\mathbf{V}'_A + \mathbf{V}_B\Sigma_B\mathbf{V}'_B = \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix} \tag{24}$$

where

$$\sigma_A^2 = \frac{\pi_{A_1}}{\pi_{A_{12}}^3}(\pi_{A_{12}} - \pi_{A_1}) \tag{25}$$

and

$$\sigma_B^2 = \frac{\pi_{B_1}}{\pi_{B_{12}}^3}(\pi_{B_{12}} - \pi_{B_1}). \tag{26}$$

Because each of the last two terms can be shown to be a matrix of zeros, we will have that

$$\mathbf{V\Sigma V'} = \mathbf{V}_A \Sigma_A \mathbf{V}_A' + \mathbf{V}_B \Sigma_B \mathbf{V}_B'. \tag{27}$$

Equations (9) and (10) are simply the sample analogues of the asymptotic variances above. Next, we write

$$\sqrt{N}(\hat{\pi}_A - \hat{\pi}_B) \tag{28}$$

$$= \sqrt{N}\left(\underbrace{(\hat{\pi}_A - \pi_A(\gamma))}_{A_N(\gamma)} - \underbrace{(\hat{\pi}_B - \pi_B(\gamma))}_{B_N(\gamma)} + \underbrace{(\hat{\pi}_A - \pi_B(\gamma))}_{C_N(\gamma)}\right). \tag{29}$$

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

$$\sqrt{N}(\hat{\pi}_A - \hat{\pi}_B) \xrightarrow{d} N(0, \sigma_{AB}^2) \qquad \text{for } \gamma = 0 \tag{30}$$

where

$$\sigma_{AB}^2 \equiv \sigma_A^2 + \sigma_B^2. \tag{31}$$

The Slutsky Theorem then gives us that

$$sd(0) \xrightarrow{d} N(0,1). \tag{32}$$

Next, in the case where γ is not zero, $\sqrt{N}(A_N(\gamma) - B_N(\gamma))$ will still converge to a random variable with distribution $N(0, \sigma_{AB}^2)$, whereas $\sqrt{N}C_N(\gamma)$ will explode since $\pi_A(\gamma) \neq \pi_B(\gamma)$ in the presence of state dependence. In particular, if $\gamma > 0$, then Proposition 1 tells us that $\pi_A(\gamma) > \pi_B(\gamma)$ and, thus, $\sqrt{N}C_N(\gamma)$ will go to positive infinity. If $\gamma < 0$, then the reverse is true. Consequently, we will have that

$$sd(\gamma) \to \pm\infty \qquad \text{for } \gamma \gtrless 0. \tag{33}$$

ACKNOWLEDGMENT

REFERENCES

Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of *in utero* influenza exposure in the post-1940 US population. *Journal of Political Economy* 114:672–712.

Barker, D. J. P. (1997). Maternal nutrition, fetal nutrition and diseases in later life. *Nutrition* 13:807–813.

Case, A., Lubotsky, D., Paxson, C. (2002). Economic status and health in childhood: the origins of the gradient. *American Economic Review* 92:1308–1334.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In: Heckman, J. J., Singer, B., eds. *Longitudinal Analysis of Labor Market Data.* Cambridge: Cambridge University Press.

Chamberlain, G. (1992). Binary response models for panel data: Identification and information. Unpublished mimeo.

Hahn, J. (2001). The information bound of a dynamic logit model with fixed effects. *Econometric Theory* 17:913–932.

Halliday, T. (2005). Heterogeneity, state dependence and health. Unpublished mimeo.

Halliday, T. (2006). Income risk and health. Unpublished mimeo.

Heckman, J. J. (1981). Statistical models for discrete panel data. In: Manski, C. F., McFadden, D., eds. *Structural Analysis of Discrete Panel Data with Econometric Applications.* Cambridge, MA: MIT Press.

Honoré, B., Kyriazidou, K. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica* 68:839–874.

Honoré, B., Tamer, E. (2006). Bounds on parameters in dynamic discrete choice models. *Econometrica* 74:611–630.

Magnac, T. (2004). Panel binary variables and sufficiency: generalizing conditional logit. *Econometrica* 72:1859–1876.