On IV estimation of the dynamic binary panel data model with fixed effects

Andrew Adrian Yu Pua*

March 30, 2015

Abstract

A big part of applied research still uses IV to estimate a dynamic linear probability model (LPM) when analyzing a panel of discrete choices. Setting aside the possibility that the average marginal effect may not be point-identified, these IV estimators are still inconsistent for the true average marginal effect. Consistent estimation for the true average marginal effect is possible when the state dependence parameter is equal to zero. Unfortunately, inferences will still depend on the distribution of the fixed effects.

1 Introduction

Consider a panel binary choice model with a strictly exogenous binary regressor. Assume that this regressor is a treatment where all individuals are not treated in the first period but all of them are treated in the second period. This means that the model is given by

$$\Pr\left(y_{it}=1|x_{i1},x_{i2}\alpha_i\right)=\Pr\left(y_{it}=1|x_{it},\alpha_i\right)=H\left(\alpha_i+\beta x_{it}\right),\ i=1,\ldots,n,\ t=1,2, \tag{1}$$

where $x_{i1}=0, x_{i2}=1$ for all i. Suppose you chose to estimate β by starting with a fixed-effects LPM, i.e.

$$y_{it}=\alpha_i+\beta x_{it}+\varepsilon_{it}.$$

The within estimator for the above model is given by

$$\widehat{\beta}=\frac{1}{n}\sum_{i=1}^{n}\left(y_{i2}-y_{i1}\right)1\left(y_{i1}+y_{i2}=1\right)=\frac{1}{n}\left(n_{01}-n_{10}\right).$$

*I have heard Martin Carree and Frank Windmeijer pose questions about the practice of using LPMs in panel data situations. I thank them for raising them in the workshops that I have attended. Comments from seminar participants at the University of Amsterdam, especially those of Frank Kleibergen and Maurice Bun, and Monash University are gratefully acknowledged. All errors remain mine.

It can be shown that as $n \to \infty$, we have

$$\widehat{\beta} \xrightarrow{p} \int \left[H(\alpha+\beta) - H(\alpha)\right] g(\alpha)\, d\alpha.$$

In this situation, the avergae marginal effect (or average treatment effect) Δ is given by

$$\Delta = E\left[E(y_t|x_t = 1, \alpha) - E(y_t|x_t = 0, \alpha)\right] = E\left[H(\alpha+\beta) - H(\alpha)\right].$$

Despite the inability of the within estimator to consistently estimate β[1], the within estimator does coincide with Δ even if the true model is nonlinear. This arises because of a lucky coincidence of factors – the independence of α_i and (x_{i1}, x_{i2}), the strict exogeneity of x_{it} (which follows from the first equality in (1)), and a time homogeneity assumption (which follows from H not depending on time).

Hahn's (2001) discussion of Angrist (2001) has already pointed out these special conditions under which the within estimator is able to estimate an average treatment effect. In addition, he emphasizes that the simple strategies suggested by Angrist (2001) require knowledge of the "structure of treatment assignment *and* careful reexpression of the new target parameter". Chernozhukov et al. (2013) also make the same point and further show that the within estimator converges to some weighted average of difference of means for a specific subset of the data.

Despite all the above concerns, applied researchers still insist on estimating LPMs with fixed effects. One may argue that the example above does not really arise in empirical applications but the example already gives an indication that complicated binary choice models estimated though a LPM are unlikely to produce intended results. Several applications of the dynamic LPM with fixed effects can be found in the literature, such as Hyslop (1999), Carrasco (2001), Bernard and Jensen (2004), Alessie, Hochguertel, and Soest (2004), Janvry et al. (2006), Acemoglu et al. (2009), Bolhaar, Lindeboom, and van der Klaauw (2012), Genakos and Pagliero (2012), to name a few.

In this paper, I show that estimating a dynamic LPM with fixed effects by IV is inappropriate. In particular, I show the large-n limit of the Anderson-Hsiao (1981; 1982) estimator (henceforth AH) is an average marginal effect but subject to incorrect weighting. As a result, estimating a dynamic LPM using GMM will inherit the problems of the AH estimator. Finally, I give examples to show that there are certain parameter configurations and fixed effect distributions for which the large-n limits of the AH estimators are outside the nonparametric bounds derived by Chernozhukov et al. (2013).

[1]Incidentally, Chamberlain (2010) shows that β is not even point identified in this example unless H is logistic. The result of Manski (1987) does not apply here. He shows that β is identified up to scale when one of the strictly exogenous regressors has unbounded support.

2 Main results

Consider the following specification of a dynamic discrete choice model with fixed effects

$$\Pr\left(y_{it}=1|y_i^t,\alpha_i\right)=\Pr\left(y_{it}=1|y_{i,t-1},\alpha_i,y_{i0}\right)=H\left(\alpha_i+\rho y_{i,t-1}\right),\ i=1,\ldots,n,\ t=1,2,3, \tag{2}$$

where $\{(y_{i0},y_{i1},y_{i2},y_{i3},\alpha_i):i=1,\ldots,n\}$ are independently drawn from their joint distribution. Here α_i is the individual-specific fixed effect, y_{i0} is an observable initial condition, and H is some cdf. I assume nothing about the joint distribution of (y_{i0},α_i) for every i. The data generating process I have specified satisfies Assumptions 1, 3, 5, and 6 of Chernozhukov et al. (2013).

If H happens to be the logistic function, then ρ can be estimated consistently using conditional logit (Hsiao, 2003). Furthermore, inference follows from usual likelihood theory. If H happens to be the standard normal cdf, then ρ is not even point-identified (Honoré and Tamer, 2006).[2] We also cannot point-identify the average marginal effect Δ, i.e.,

$$\Delta=\int\left[\Pr\left(y_{it}=1|y_{i,t-1}=1,\alpha,y_0\right)-\Pr\left(y_{it}=1|y_{i,t-1}=0,\alpha,y_0\right)\right]dF\left(\alpha,y_0\right)$$

even if we know H but leave the distribution of (y_{i0},α_i) unrestricted.

Despite these results, applied researchers insist on using a dynamic LPM on the grounds that linearity still provides a good approximation even if the true H is nonlinear.[3] I use this as a starting point and determine the large-n limit of IV estimators for the dynamic LPM. The model can be expressed as:

$$y_{it}=\alpha_i+\rho y_{i,t-1}+\epsilon_{it},\ i=1,\ldots,n,\ t=1,2,3,$$

where $\epsilon_{it}=y_{it}-E\left(y_{it}|y_i^{t-1},\alpha_i\right)$. As a consequence, $E\left(\epsilon_{it}|y_i^{t-1},\alpha_i\right)=0$ and $Var\left(\epsilon_{it}|y_i^{t-1},\alpha_i\right)=\left(\alpha_i+\rho y_{i,t-1}\right)\left(1-\alpha_i-\rho y_{i,t-1}\right)$. Take first-differences to eliminate α_i, i.e.,

$$\Delta y_{it}=\rho\Delta y_{i,t-1}+\Delta\epsilon_{it},\ i=1,\ldots,n,\ t=2,3.$$

Using lagged differences as instruments, the AH estimator can be written as

$$\widehat{\rho}_{AHd}=\frac{\sum_{i=1}^n\Delta y_{i1}\Delta y_{i3}}{\sum_{i=1}^n\Delta y_{i1}\Delta y_{i2}}.$$

Because of the binary nature of the sequences $\{(y_{i0},y_{i1},y_{i2},y_{i3}):i=1,\ldots,n\}$, it is certainly possible for the first differences to be equal to zero. Therefore, there are only certain types

[2] Honoré and Tamer (2006) actually show that the sign of ρ is identified for any strictly increasing cdf H and unrestricted distribution of (y_{i0},α_i).

[3] The dynamic LPM is really a special case of (2), where H is the identity function.

of sequences that enter into the expression above. If we enumerate all these 16 possible sequences, we can rewrite the estimator as

$$\widehat{\rho}_{AHd}=\frac{n_{0110}+n_{1001}-n_{1010}-n_{0101}}{n_{0100}+n_{1010}+n_{0101}+n_{1011}},$$

where $n_{abcd}=\sum_{i=1}^{n}\mathbf{1}\left(y_{i0}=a,y_{i1}=b,y_{i2}=c,y_{i3}=d\right)$ denotes the number of observations in the data for which we observe the sequence $abcd$. Note that we cannot just drop those sequences for which $y_{i1}+y_{i2}=1$, like in conditional logit. If we do this, the resulting AH estimator becomes

$$\widetilde{\rho}_{AHd}=\frac{-n_{1010}-n_{0101}}{n_{0100}+n_{1010}+n_{0101}+n_{1011}},$$

which is always negative regardless of the sign of ρ or Δ. It is quite apparent that identification arguments based on the conditional logit do not necessarily translate to other link functions, including that of the identity link. Furthermore, the IV estimator is using other binary sequences beyond that considered in the conditional logit.

It can be shown[4] that the large-n limit of $\widehat{\rho}_{AHd}$ is

$$\begin{aligned}\widehat{\rho}_{AHd}\quad&\overset{p}{\rightarrow}\quad\frac{\displaystyle\int H\left(\alpha\right)\left(1-H\left(\alpha+\rho\right)\right)\left(H\left(\alpha+\rho\right)-H\left(\alpha\right)\right)g\left(\alpha\right)d\alpha}{\displaystyle\int H\left(\alpha\right)\left(1-H\left(\alpha+\rho\right)\right)g\left(\alpha\right)d\alpha} \\ &= \int w_d\left(\alpha,\rho\right)\left(H\left(\alpha+\rho\right)-H\left(\alpha\right)\right)d\alpha \\ &= \int w_d\left(\alpha,\rho\right)\left[\Pr\left(y_{it}=1|y_{i,t-1}=1,\alpha,y_0\right)-\Pr\left(y_{it}=1|y_{i,t-1}=0,\alpha,y_0\right)\right]d\alpha\end{aligned}\tag{3}$$

where

$$w_d\left(\alpha,\rho\right)=\frac{H\left(\alpha\right)\left(1-H\left(\alpha+\rho\right)\right)g\left(\alpha\right)}{\displaystyle\int H\left(\alpha\right)\left(1-H\left(\alpha+\rho\right)\right)g\left(\alpha\right)d\alpha}.$$

Note that the weighting function $w_d\left(\alpha,\rho\right)$ depends on the true value of ρ and the marginal distribution of the fixed effects $g\left(\alpha\right)$ (rather than the joint distribution of $\left(y_0,\alpha\right)$). Therefore, $\widehat{\rho}_{AHd}$ is inconsistent for Δ because of the incorrect weighting of the individual marginal dynamic effect $H\left(\alpha+\rho\right)-H\left(\alpha\right)$.

The analysis above can be extended to the AH estimator which uses levels as the instrument set. It can be shown that this AH estimator has the following form:

$$\begin{aligned}\widehat{\rho}_{AHl}\quad&=\quad\frac{\sum_{i=1}^{n}\sum_{t=2}^{3}y_{i,t-2}\Delta y_{it}}{\sum_{i=1}^{n}\sum_{t=2}^{3}y_{i,t-2}\Delta y_{i,t-1}} \\ &=\quad\frac{n_{0110}-n_{0101}+n_{1110}-n_{1010}+n_{1100}-n_{1011}}{n_{1010}+n_{1000}+n_{1001}+n_{1011}+n_{0100}+n_{1100}+n_{0101}+n_{1101}}.\end{aligned}$$

[4]A part of the derivation can be found in section after the concluding remarks.

Calculations similar to (3) will allow us to derive the large-n limit of $\widehat{\rho}_{AHl}$, i.e.

$$\begin{aligned}
\widehat{\rho}_{AHl} &\overset{p}{\to} \frac{\int (1-H(\alpha+\rho))(1+H(\alpha+\rho))(H(\alpha+\rho)-H(\alpha))f(\alpha,1)\,d\alpha+}{\int [(1-H(\alpha+\rho))(1+H(\alpha+\rho))f(\alpha,1)+(1-H(\alpha+\rho))H(\alpha)f(\alpha,0)]\,d\alpha} \\
&\quad + \frac{\int (1-H(\alpha+\rho))H(\alpha)(H(\alpha+\rho)-H(\alpha))f(\alpha,0)\,d\alpha}{\int [(1-H(\alpha+\rho))(1+H(\alpha+\rho))f(\alpha,1)+(1-H(\alpha+\rho))H(\alpha)f(\alpha,0)]\,d\alpha} \\
&= \int w_l(\alpha,\rho,y_0)(H(\alpha+\rho)-H(\alpha))\,dy_0 d\alpha \\
&= \int w_l(\alpha,\rho,y_0)\left[\Pr\left(y_{it}=1|y_{i,t-1}=1,\alpha,y_0\right)-\Pr\left(y_{it}=1|y_{i,t-1}=0,\alpha,y_0\right)\right]dy_0 d\alpha,
\end{aligned}$$

where

$$\begin{aligned}
w_l(\alpha,\rho,0) &= \frac{(1-H(\alpha+\rho))H(\alpha)f(\alpha,0)}{\int [(1-H(\alpha+\rho))(1+H(\alpha+\rho))f(\alpha,1)+(1-H(\alpha+\rho))H(\alpha)f(\alpha,0)]\,d\alpha}, \\
w_l(\alpha,\rho,1) &= \frac{(1-H(\alpha+\rho))(1+H(\alpha+\rho))f(\alpha,1)}{\int [(1-H(\alpha+\rho))(1+H(\alpha+\rho))f(\alpha,1)+(1-H(\alpha+\rho))H(\alpha)f(\alpha,0)]\,d\alpha}.
\end{aligned}$$

I denote $f(\alpha,0)=\Pr(y_0=0|\alpha)\,g(\alpha)$ and $f(\alpha,1)=\Pr(y_0=1|\alpha)\,g(\alpha)$. Like the large-$n$ limit of $\widehat{\rho}_{AHd}$, the weighting function $w_l(\alpha,\rho,y_0)$ also depends on the true value of ρ. But this weighting function also depends on the joint distribution of (y_0,α). Therefore, we have a similar inconsistency result for $\widehat{\rho}_{AHl}$.[5]

3 Practical implications

Based on the results of the previous section, we should not be using IV estimators for the dynamic LPM. It is difficult to give a general indication of whether we overestimate or underestimate Δ, because the results depend on the joint distribution of (y_0,α). If it happens that $\rho=0$ (so that $\Delta=0$), then the AH estimators are consistent for Δ. Using this observation to construct a test of the hypothesis that $\Delta=0$ is not so straightforward. The appropriate standard errors for the AH estimators depend on the unknown joint distribution of (y_0,α). It is unclear whether there exists a consistent estimator for this unknown joint distribution. Although of practical interest, testing the hypothesis $\Delta=0$ may still be infeasible.

To further persuade applied researchers not to use IV for the dynamic LPM, I adopt

[5]For the case where we have one less time period, i.e. we observe sequences of the form $\{(y_{i0},y_{i1},y_{i2}):i=1,\ldots,n\}$, the large-$n$ limit of $\widehat{\rho}_{AHl}$ depends only on $f(\alpha,1)$.

the example in Chernozhukov et al. (2013) to show that, even in the simplest of cases, we cannot ignore the distortion of the weighting function. Chernozhukov et al. (2013) consider a data generating process where H is the standard normal cdf, $y_{i0} \perp \alpha_i$, and $\Pr(y_{i0} = 1) = 0.5$. I use the following fixed effects distributions:

	Distribution			
Property	$N(0,1)$	$0.5N(0,1) + 0.5N(2,0.5)$	$Beta(4,2)$	$0.5N(-2,0.1) + 0.5N(-1,1)$
Mean	0	1	0.667	-1.5
Variance	1	1.625	0.032	0.755
Skewness	0	-0.543	-0.468	1.132
Kurtosis	3	2.402	2.625	4.070
Multimodal?	Unimodal	Bimodal	Unimodal	Bimodal

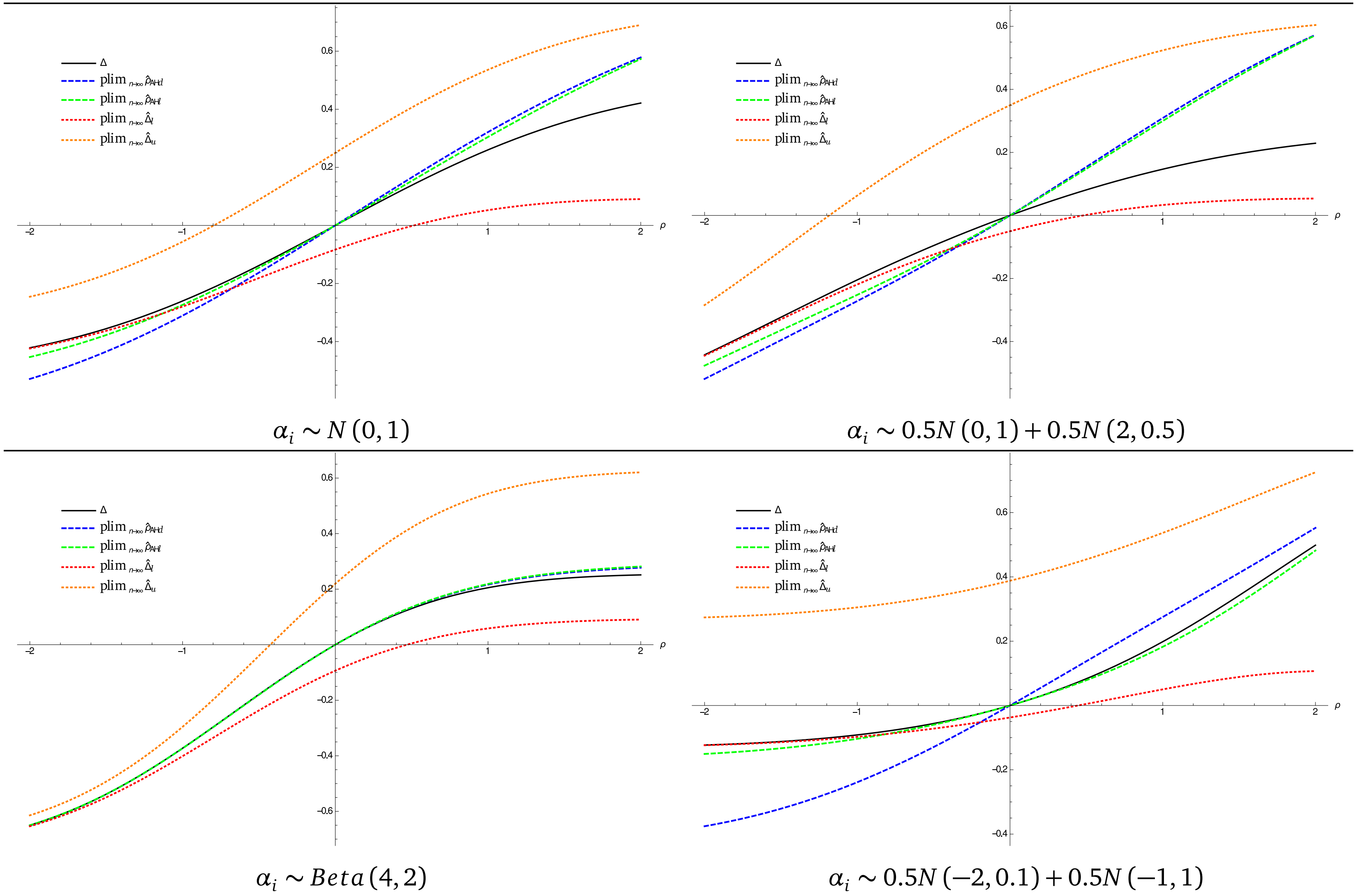
In Figure 1, I calculate[6] the large-n limits of the AH estimators (in blue and green) and the nonparametric bounds proposed by Chernozhukov et al. (2013) (in red and orange) for $\rho \in [-2, 2]$. I also calculate the true Δ (in black) where knowledge of the distribution of (y_0, α) was exploited.

Even in the benchmark case where $\alpha_i \sim N(0,1)$, both the large-n limits of the AH estimators are larger than Δ when $\rho > 0$. Further note that when $\rho < -0.5$, both these large-n limits are outside the identified set. For $\alpha_i \sim 0.5N(0,1) + 0.5N(2,0.5)$, both the large-$n$ limits of the AH estimators coincide and are much larger than Δ even for less persistent state dependence. For $\alpha_i \sim Beta(4,2)$, the large-n limits of the AH estimators are practically the same as Δ and can be found in the identified set. The key seems to be that the restricted support for the fixed effect, which is $(0,1)$. Finally, the large-n limit of the AH estimator using levels as the instrument set is smaller than Δ for $\alpha_i \sim 0.5N(-2,0.1) + 0.5N(-1,1)$.

A direct application of GMM for the dynamic LPM need not produce a consistent estimator for Δ (even if the latter is point-identified). Furthermore, computation of standard errors is more difficult because of the absence of a suitable plug-in for the unknown marginal distribution of the fixed effects and for the optimal weighting matrix. All these results, including those discussed in this paper, suggest that we should avoid using IV for the dynamic LPM. It is also more appropriate to use the nonparametric bounds proposed by Chernozhukov et al. (2013), especially if one is unwilling to say something about the distribution of (y_0, α). Furthermore, their bounds allows for an arbitrary H as long as this H satisfies their assumptions.

[6]A Mathematica notebook containing the calculations are available upon request.

Figure 1: Large-n limits of the AH estimators under different distributions for the fixed effects



Some calculations for (3)

We calculate $E\left[\mathbf{1}\left(y_{i0}=0, y_{i1}=1, y_{i2}=1, y_{i3}=0\right)\right]$ in detail since the other expressions follow similarly. This expression is equal to

$$\begin{aligned}
& \Pr\left(y_{i0}=0, y_{i1}=1, y_{i2}=1, y_{i3}=0\right) \\
= & \int \Pr\left(y_{i0}=0, y_{i1}=1, y_{i2}=1, y_{i3}=0 \mid \alpha\right) g(\alpha)\, d\alpha \\
= & \int \Pr\left(y_{i3}=0 \mid y_{i0}=0, y_{i1}=1, y_{i2}=1, \alpha\right) \Pr\left(y_{i2}=1 \mid y_{i0}=0, y_{i1}=1, \alpha\right) \times \\
& \Pr\left(y_{i1}=1 \mid y_{i0}=0, \alpha\right) \Pr\left(y_{i0}=0 \mid \alpha\right) g(\alpha)\, d\alpha \\
= & \int \Pr\left(y_{i3}=0 \mid y_{i2}=1, \alpha\right) \Pr\left(y_{i2}=1 \mid y_{i1}=1, \alpha\right) \Pr\left(y_{i1}=1 \mid y_{i0}=0, \alpha\right) f(\alpha, 1)\, d\alpha \\
= & \int (1-H(\alpha+\rho)) H(\alpha+\rho) H(\alpha) f(0, \alpha)\, d\alpha,
\end{aligned}$$

where f is the joint distribution of (α, y_0). Similarly, we have the following:

$$\begin{aligned}
E\left[\mathbf{1}\left(y_{i0}=1, y_{i1}=0, y_{i2}=0, y_{i3}=1\right)\right] &= \int H(\alpha)(1-H(\alpha))(1-H(\alpha+\rho)) f(\alpha, 1)\, d\alpha \\
E\left[\mathbf{1}\left(y_{i0}=1, y_{i1}=0, y_{i2}=1, y_{i3}=0\right)\right] &= \int (1-H(\alpha+\rho)) H(\alpha)(1-H(\alpha+\rho)) f(\alpha, 1)\, d\alpha \\
E\left[\mathbf{1}\left(y_{i0}=0, y_{i1}=1, y_{i2}=0, y_{i3}=1\right)\right] &= \int H(\alpha)(1-H(\alpha+\rho)) H(\alpha) f(\alpha, 0)\, d\alpha \\
E\left[\mathbf{1}\left(y_{i0}=0, y_{i1}=1, y_{i2}=0, y_{i3}=0\right)\right] &= \int (1-H(\alpha))(1-H(\alpha+\rho)) H(\alpha) f(\alpha, 0)\, d\alpha \\
E\left[\mathbf{1}\left(y_{i0}=1, y_{i1}=0, y_{i2}=1, y_{i3}=1\right)\right] &= \int H(\alpha+\rho) H(\alpha)(1-H(\alpha+\rho)) f(\alpha, 1)\, d\alpha
\end{aligned}$$

Assembling these expressions together in the expression for the large-sample limit of $\widehat{\rho}_{AHd}$ gives (3).

References

Acemoglu, D, S Johnson, JA Robinson, and P Yared (2009). Reevaluating the modernization hypothesis. *Journal of Monetary Economics* **56.**(8), 1043–1058.

Alessie, R, S Hochguertel, and A van Soest (2004). Ownership of Stocks and Mutual Funds: A Panel Data Analysis. *The Review of Economics and Statistics* **86.**(3), 783–796.

Anderson, T and C Hsiao (1981). Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association* **76.**(375), 598–606.

Anderson, T and C Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* **18.**(1), 47–82.

Angrist, JD (2001). Estimations of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice (with discussion). *Journal of Business & Economic Statistics* **19**.(1), 2–16.

Bernard, AB and JB Jensen (2004). Why Some Firms Export. *The Review of Economics and Statistics* **86**.(2), 561–569.

Bolhaar, J, M Lindeboom, and B van der Klaauw (2012). A dynamic analysis of the demand for health insurance and health care. *European Economic Review* **56**.(4), 669–690.

Carrasco, R (2001). Binary Choice with Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Participation. *Journal of Business & Economic Statistics* **19**.(4), 385–94.

Chamberlain, G (2010). Binary Response Models for Panel Data: Identification and Information. *Econometrica* **78**.(1), 159–168.

Chernozhukov, V, I Fernández-Val, J Hahn, and W Newey (2013). Average and Quantile Effects in Nonseparable Panel Models. *Econometrica* **81**.(2), 535–580.

Genakos, C and M Pagliero (2012). Interim Rank, Risk Taking, and Performance in Dynamic Tournaments. *Journal of Political Economy* **120**.(4), 782–813.

Hahn, J (2001). Comment: Binary Regressors in Nonlinear Panel-Data Models with Fixed Effects. **19**.(1), 16–17.

Honoré, BE and E Tamer (2006). Bounds on Parameters in Panel Dynamic Discrete Choice Models. *Econometrica* **74**.(3), 611–629.

Hsiao, C (2003). *Analysis of Panel Data*. 2nd. Cambridge University Press.

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

Janvry, A de, F Finan, E Sadoulet, and R Vakis (2006). Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks? *Journal of Development Economics* **79**.(2), 349–373.

Manski, C (1987). Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data. *Econometrica* **55**.(2), 357–62.