Should we use IV to estimate dynamic linear probability models with fixed effects?*

Andrew Adrian Yu Pua†

July 31, 2017

Abstract

Researchers have applied linear dynamic panel data methods to analyze a panel of binary choices while allowing for individual-specific unobserved heterogeneity and dynamics. This leads to IV/GMM estimation of a dynamic linear probability model (LPM) with fixed effects. In this paper, I give a set of pros and cons of this procedure and conclude that this procedure should be treated with caution, especially in fixed-T settings. Even if we ignore the possibility that average marginal effects may not be point-identified, directly applying IV/GMM estimators to this dynamic LPM identifies incorrectly-weighted average marginal effects, which may differ from the true average marginal effect, under large-n, fixed-T or large-n, large-T asymptotics. I also show that there exist certain DGPs that can push the large-n, fixed-T limits of these IV estimators outside the identified set for the true average marginal effect. The only good news is that nonparametrically testing the point null of zero first-order state dependence is possible with default routines. Unfortunately, this nonparametric test can have low power. In relation to this, I demonstrate through an empirical example that the resulting IV/GMM estimates of the average treatment effect of fertility on female labor force participation are outside the nonparametric bounds under monotonicity.

Keywords: Linear probability model, Average marginal effects, Anderson-Hsiao estimator, Arellano-Bond estimator, Nonparametric bounds, State dependence, Predetermined binary treatment
JEL Classification: C23, C25, C26

*I thank Martin Carree and Frank Windmeijer for posing questions during two different Amsterdam Panel Data Workshops regarding the practice of estimating LPMs in panel data situations. Comments from Maurice Bun, Geert Dhaene, Artūras Juodis, Iván Fernández-Val, Shakeeb Khan, Frank Kleibergen, Ryo Okui, Frank Windmeijer, and Jeffrey Wooldridge are greatly appreciated. I also thank participants of the 2017 Singapore Economic Review Conference, 2017 Asian Meeting of the Econometric Society and 21st Panel Data Conference, and seminar participants at the University of Amsterdam, Monash University, the University of Passau, and Xiamen University. A previous version of this paper was circulated as "On IV estimation of a dynamic linear probability model with fixed effects".

†The Wang Yanan Institute for Studies in Economics, Xiamen University, andrewypua@gmail.com

1 Introduction

Angrist and Pischke (2009) make their case for applying simple linear methods to recover marginal effects of interest without resorting to complicated nonlinear methods. Since many researchers find their arguments convincing, it is not surprising that these simple linear methods continue to be applied in situations outside their domain of validity. One such situation, that has almost risen to "folk-theorem" status, is the use of linear panel data methods to analyze a panel of binary choices. This situation is really an extension of the "success" of the linear probability model (henceforth, LPM) in cross-sectional settings to panel data settings that allow for unobserved heterogeneity, feedback, and dynamics. Several applications of this panel-based LPM can be found in papers published in top journals[1].

A more suitable approach is to use limited dependent variable models when analyzing these binary choices. Unfortunately, the inclusion of fixed effects creates an incidental parameter problem that complicates the estimation of average marginal effects, especially when the time dimension is small (see the survey by Arellano and Bonhomme (2011)). Resorting to a random effects or correlated random effects approach may require specifying the full distribution of the fixed effects and initial conditions – something that researchers may be unwilling to do because of the lack of specific subject matter knowledge to construct such a distribution. Another suitable approach would be to use the nonparametric bounds derived by Chernozhukov et al. (2013). These nonparametric bounds may be calculated in software with some programming but the resulting bounds may be viewed as being too wide. Linear dynamic panel data methods naturally become simple and attractive procedures that allow for fixed effects, dynamics, predetermined regressors, fewer functional form restrictions, and even allow for heteroscedasticity.

I show that usual linear dynamic panel data methods are inappropriate for estimating average marginal effects even if the goal is just to approximate an average marginal

[1]Examples include assessing the magnitude of state dependence in female labor force participation (Hyslop 1999), examining the factors that affect exporting decisions (Bernard and Jensen 2004), determining the effect of income on transitions in and out of democracy (Acemoglu et al. 2009), and determining how overnight rates affect a bank's decision to provide loans (Jiménez et al. 2014).

effect. In particular, I show the large-n limit of the Anderson-Hsiao (1981; 1982) IV estimator (henceforth AH) is a marginal effect subject to incorrect weighting. Given that the AH estimator is a special case of GMM, estimators in the spirit of Arellano and Bond (1991) will be subject to the same problem. I also show that the effect of this incorrect weighting does not disappear even when T is large. We are unable to use a similar analysis as Loken, Mogstad, and Wiswall (2012), who compute the weights only from observables, to study the incorrect weighting function because the incorrect weighting function depends on unobservables. Furthermore, I give examples to show that there exist certain parameter configurations and fixed effect distributions (not all of which are esoteric) for which the large-n limit of the AH estimator is outside the nonparametric bounds derived by Chernozhukov et al. (2013).

Despite these issues, the usual linear dynamic panel data methods can be used to nonparametrically test the null hypothesis of zero first-order state dependence or the null hypothesis of zero effect for a predetermined binary treatment. For the case of state dependence, the test is nonparametric in the sense that the test does not require the functional form for the inverse link function and the joint distribution of unobserved heterogeneity and the initial conditions. For the case of the effect of a predetermined binary treatment, the test does not require knowledge of the feedback effects of past choices on current treatment. Unfortunately, it is unclear whether this good news will have wide applicability because this nonparametric test may have low power.

Much research has been done on whether using the LPM is suitable, especially for the cross-sectional case. A particularly eye-catching example was provided by Lewbel, Dong, and Yang (2012). They show, in a toy example, that OLS applied to the LPM cannot even get the correct sign of the treatment effect even in the situation where there is just a binary exogenous regressor and a high signal-to-noise ratio. On the other hand, Wooldridge (2010) argues that "the case for the LPM is even stronger if most the regressors are discrete and take on only a few values"[2].

[2]Problem 15.1 of his book asks the reader to show that we need not worry about success probabilities being outside $[0, 1]$ in a saturated model, a result also demonstrated in Section 3.4 of Angrist and Pischke (2009).

Some results are available for the panel data case. If we specialize the results in Wooldridge (2005) and Murtazashvili and Wooldridge (2008) to the LPM, then fixed-effects (henceforth, FE) estimation applied to the LPM with strictly exogenous regressors or to the LPM with continuously endogenous regressors can be used to consistently estimate average marginal effects under a specific correlated random coefficients condition. In contrast, I consider the situation where we have a lagged binary dependent variable or a predetermined binary treatment.

Hahn's (2001) discussion of Angrist (2001) has already pointed out the lucky coincidence of factors under which the within estimator is able to estimate an average treatment effect in a two-period static panel data model. 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 individual difference of means for a specific subset of the data. They also show that this weighted average is not the average marginal effect of interest because of incorrect weighting. Their results are obtained under a strict exogeneity condition. In contrast, I consider the case where one has predetermined binary regressors.

The previously cited papers and this paper can also be situated in the studies involving misspecification in panel data models. Galvao and Kato (2014) and Okui (2015) represent some attempts to understand exactly what linear dynamic panel data methods recover when the panel data model is incorrectly specified. Although the message of this paper is primarily for large-n, fixed-T settings, I use their results to obtain the behavior of the FE and GMM estimators in large-n, large-T settings.

I organize the rest of the paper as follows. In Section 2, I derive analytically the consequences of using linear dynamic panel data methods when interest centers on the average marginal effect of state dependence for the cases of $T = 3$ and $T \to \infty$. These analytical results help in developing a list of pros and cons of using linear dynamic panel data methods to study dynamic discrete choice. Next, I revisit one of the empirical applications in Chernozhukov et al. (2013) on female labor force participation and fertility in Section 3. The last section contains concluding remarks followed by an appendix containing some details of the derivations in this paper.

2 Pros and cons

2.1 Model

Consider the following specification of a dynamic discrete choice model with fixed effects and no additional regressors:

$$\Pr\left(y_{it}=1|x_i^t,\alpha_i\right)=H\left(\alpha_i+\rho x_{it}\right),\ i=1,\ldots,n,\ t=1,\ldots,T, \tag{1}$$

where $x_i^t=(x_{i1},x_{i2},\ldots,x_{it})$ is the past and present observation of x, α_i is an individual-specific fixed effect, and $H:\mathbb{R}\to[0,1]$ is some unspecified inverse link function. Thus, the model covers the case of a lagged binary dependent variable or a predetermined binary regressor. All the analytical results will focus on the case where $x_{it}=y_{i,t-1}$ since the main ideas do not change once you allow for a predetermined binary regressor. The only change in the analytical results will be additional terms involving of an unspecified model describing the feedback of past values of y on the current value of x, i.e., $\Pr\left(x_{it}=1|x_i^{t-1},y_i^{t-1},\alpha_i\right)$.

Assume that $(y_{i0},y_{i1},y_{i2},y_{i3},\alpha_i)$ are i.i.d. draws from their joint distribution. I leave the joint density of (α_i,y_{i0}), denoted by f, unspecified. This data generating process satisfies Assumptions 1, 3, 5, and 6 of Chernozhukov et al. (2013). We cannot point-identify the average marginal effect of state dependence, denoted by Δ:

$$\Delta=\sum_{y_0\in\{0,1\}}\int\left[\Pr\left(y_{it}=1|y_{i,t-1}=1,\alpha,y_0\right)-\Pr\left(y_{it}=1|y_{i,t-1}=0,\alpha,y_0\right)\right]f\left(\alpha,y_0\right)\,d\alpha \tag{2}$$

even if we know H but leave the density of (y_{i0},α_i) unspecified. This average marginal effect is of practical interest because it measures the effect of first-order state dependence in the presence of individual-specific unobserved heterogeneity.

Despite this negative result, researchers still insist on using a dynamic LPM on the grounds that linearity still provides a good approximation even if the true H is nonlinear. Presumably, the linear model researchers have in mind can be expressed as

$$y_{it} = \eta_i + \gamma y_{i,t-1} + \nu_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T.$$

Taking first-differences to eliminate η_i, we have

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta \nu_{it},\ i = 1, \ldots, n,\ t = 2, \ldots, T.$$

Because the differenced regressor $\Delta y_{i,t-1}$ is correlated with the differenced error $\Delta \nu_{it}$, it seems natural to use IV or GMM estimators to estimate γ and hope that we are able to recover Δ. The estimators I will study are the following:

1. AH estimator using lagged differences as instruments (for $T = 3$ and $T \to \infty$)
2. AH estimator using lagged levels as instruments (for $T = 3$ and $T \to \infty$)
3. First-difference estimator (for $T \to \infty$)
4. FE/Within-groups estimator (for $T \to \infty$)
5. Arellano-Bond estimator (for $T \to \infty$)

2.2 CON: IV estimators recover an incorrectly-weighted average marginal effect.

2.2.1 The case where $T = 3$

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

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

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 some of the first differences to be equal to zero. Therefore, there are only certain types of binary sequences that enter into the expression above. If we enumerate all these 16 possible sequences, we can simplify the estimator as

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

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 $abcd$.

It can be shown (as seen in the Appendix) that the large-n limit of $\widehat{\rho}_{AHd}$ is

$$\begin{aligned}
\widehat{\gamma}_{AHd} &\overset{p}{\to} \frac{\displaystyle\int H(\alpha)\left(1 - H(\alpha+\rho)\right)\left(H(\alpha+\rho) - H(\alpha)\right) g(\alpha)\ d\alpha}{\displaystyle\int H(\alpha)\left(1 - H(\alpha+\rho)\right) g(\alpha)\ d\alpha} \qquad (3) \\
&= \int w_d(\alpha,\rho)\left(H(\alpha+\rho) - H(\alpha)\right)\ d\alpha \\
&= \sum_{y_0\in\{0,1\}} \int w_d(\alpha,\rho)\left[\Pr\left(y_t = 1 | y_{t-1} = 1, \alpha, y_0\right) - \Pr\left(y_t = 1 | y_{t-1} = 0, \alpha, y_0\right)\right]\ d\alpha
\end{aligned}$$

where

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

Note that the weighting function $w_d(\alpha,\rho)$ depends on the true value of ρ and the marginal distribution of the fixed effects $g(\alpha)$. The correct weighting function should have been the joint density of (y_0, α) as in (2). Therefore, $\widehat{\gamma}_{AHd}$ is inconsistent for Δ because of the incorrect weighting of the individual marginal dynamic effect $H(\alpha+\rho) - H(\alpha)$. It is not possible 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 $\widehat{\gamma}_{AHd}$ is consistent for Δ.

The previous analysis 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:

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

Calculations similar to (3) allow us to derive the large-n limit of $\widehat{\rho}_{AHl}$:

$$\widehat{\gamma}_{AHl} \xrightarrow{p} \sum_{y_0\in\{0,1\}} \int w_l(\alpha,\rho,y_0)\left[\Pr(y_{it}=1|y_{i,t-1}=1,\alpha,y_0)-\Pr(y_{it}=1|y_{i,t-1}=0,\alpha,y_0)\right] d\alpha,$$

where

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

I denote $f(\alpha,0)=\Pr(y_0=0|\alpha)g(\alpha)$ and $f(\alpha,1)=\Pr(y_0=1|\alpha)g(\alpha)$. Note that the weighting function $w_l(\alpha,\rho,y_0)$ depends on the true value of ρ and the joint distribution of (y_0,α). Once again, we have an incorrect weighting function $w_l(\alpha,\rho,y_0)$ instead of the joint distribution of (y_0,α). As a result, $\widehat{\gamma}_{AHl}$ is inconsistent for Δ.

2.2.2 The case where $T\to\infty$

A natural question to ask is whether the inconsistency results extend to the case where the number of time periods T is large. An intuitive response would be to say that as $T\to\infty$, the fixed effects α_i can be estimated consistently or that their removal via differencing should not pose an issue. Therefore, we should be able to estimate average marginal effects consistently. Unfortunately, this intuition may be mistaken.

To address this issue, I use sequential asymptotics where I let $T\to\infty$ and then $n\to\infty$ (see Phillips and Moon (1999)). Explicit calculations found in the appendix indicate that under this specific asymptotic scheme, we have

$$\widehat{\gamma}_{AHd} \overset{p}{\to} \int w_d(\alpha,\rho)\left(H(\alpha+\rho)-H(\alpha)\right)\, d\alpha, \tag{4}$$

$$\widehat{\gamma}_{AHl} \overset{p}{\to} \int w_l(\alpha,\rho)\left(H(\alpha+\rho)-H(\alpha)\right)\, d\alpha, \tag{5}$$

$$\widehat{\gamma}_{FD} \overset{p}{\to} \frac{1}{2}\left[1-\int w_l(\alpha,\rho)\left(H(\alpha+\rho)-H(\alpha)\right)\, d\alpha\right], \tag{6}$$

where the weighting functions are given by

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

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

A noteworthy aspect of the derivation is that $w_d(\alpha,\rho)$ is the same regardless of whether $T=3$ or $T\to\infty$. Thus, the inconsistency does not disappear and actually stays the same with respect to size even when $T\to\infty$. The result (6) is very troubling. When $\rho=0$ (so that the true average marginal effect is 0), $\widehat{\gamma}_{FD}$ converges to 0.5, grossly overstating the true Δ.

As for the behavior of the FE estimator in the large-T case, I rely on Proposition 3.1 of Galvao and Kato (2014). In the context I consider, the linear probability model is misspecified and the true model is the nonlinear model (1). As a result, the conditional mean $\mathbb{E}\left(y_{it}|y_{i,t-1},\alpha_i\right)$ is misspecified as additive and linear when in fact it is nonlinear. Under their assumptions A1 to A3, they show that the FE estimator converges to the following pseudo-true parameter:

$$\beta_0 = \frac{\mathbb{E}\left(\widetilde{y}_{it}\widetilde{y}_{i,t-1}\right)}{\mathbb{E}\left(\widetilde{y}_{i,t-1}^2\right)},$$

where $\widetilde{y}_{it} = y_{it}-\mathbb{E}\left(y_{it}|\alpha_i\right)$. Assumption A1 of their paper require that the marginal distribution of $(\alpha_i, y_{it}, y_{i,t-1})$ is invariant with respect to (i,t). As a result, the initial condition is drawn from the stationary distribution conditional on α_i. Although I did not impose this assumption in all previous derivations, this assumption is plausible because $T \to \infty$. For convenience, I now impose this assumption to derive the probability limit of the FE estimator. In the appendix, I show that this probability limit is the pseudo-true parameter is given by

$$\beta_0 = \frac{\mathbb{E}\left[\left(H\left(\alpha+\rho\right)-H\left(\alpha\right)\right)\Pr\left(y_{t-1}=1|\alpha\right)\left(1-\Pr\left(y_{t-1}=1|\alpha\right)\right)\right]}{\mathbb{E}\left[\Pr\left(y_{t-1}=1|\alpha\right)\left(1-\Pr\left(y_{t-1}=1|\alpha\right)\right)\right]}, \tag{7}$$

where the expectations are calculated with respect to the marginal distribution of α. Clearly, the FE estimator does not converge to the correct average marginal effect and the weighting function (which also depends on the stationary distribution of y as seen in (7)) is given by

$$w_{FE}\left(\alpha,\rho\right) = \frac{\dfrac{H\left(\alpha\right)\left(1-H\left(\alpha+\rho\right)\right)}{\left[1-H\left(\alpha+\rho\right)+H\left(\alpha\right)\right]^2}g\left(\alpha\right)}{\displaystyle\int \dfrac{H\left(\alpha\right)\left(1-H\left(\alpha+\rho\right)\right)}{\left[1-H\left(\alpha+\rho\right)+H\left(\alpha\right)\right]^2}g\left(\alpha\right)d\alpha}.$$

To derive the large-n, large-T limit of GMM estimators applied to the dynamic LPM, I use an existing result on heterogeneous dynamics by Okui (2015). He has shown that the Arellano and Bond (1991) estimator, the GMM estimator based on level moment conditions proposed by Arellano and Bover (1995), and the FE estimator all converge to the same probability limit under sequential asymptotics where first $n \to \infty$ followed by $T \to \infty$. This is different from the sequential asymptotics I have adopted in this paper and from the joint asymptotics adopted by Galvao and Kato (2014). I show in the Appendix that, applying his Theorem 5 to the case I consider (which satisfies his Assumption 2), the probability limit obtained is exactly the same as the one I obtain for the FE estimator earlier.

Thus, the two AH estimators, the GMM estimators, and the FE estimator are able to consistently estimate Δ when $\rho = 0$ (so that $\triangle = 0$). Unfortunately, for all other values of ρ, all these estimators still cannot consistently estimate the correct Δ because of incorrect weighting in (4), (5), and (7). The appropriate weighting function is now the marginal distribution of the fixed effects $g(\alpha)$, because the effect of the initial condition disappears as $T \to \infty$. Just as in the fixed-T case considered earlier, it is still not possible to determine the direction of inconsistency of all the discussed estimators.

2.2.3 Illustrations of the weighting function

The CON just discussed indicates that we have inconsistent estimation of the average marginal effect for ρ different from zero. To further persuade researchers not to use IV for the dynamic LPM in these situations, I adopt the example in Chernozhukov et al. (2013) to show that, even in the simplest of cases, we cannot ignore the distortion brought about by the incorrect weighting function.

Consider the data generating process used in Chernozhukov et al. (2013) where H is the standard normal cdf, y_{i0} is independent of α_i, $\Pr(y_{i0} = 1) = 0.5$, and $T = 3$ (later $T \to \infty$). I use four different distributions for the fixed effects, as described in Table 1. The first is the standard normal distribution, which is a usual choice in Monte Carlo simulations and in random-effects estimation. The second is a mixture of three normals that is symmetric but has three modes. Mixtures of normals are prominent in Bayesian models used in marketing research (e.g., Rossi, Allenby, and McCulloch (2012)) and have been used as flexible specifications for α_i in econometrics (e.g., Burda, Harding, and Hausman (2015)). The third is a distribution which favors the LPM because the support of α_i is a bounded interval $[0, 1]$. Finally, the fourth is a mixture of a standard normal and a normal distribution with mean 2 and variance 0.5^2. This mixture makes it more likely for cross-sectional units to have $y_{it} = 1$ across time. In comparison to the first two distributions, the last two distributions have nonzero mean.

Table 1: Distribution of fixed effects for computations

	$N(0,1)$	$1/3N(-1,1) + 1/3N(0,0.5^2)$ $+1/3N(1,1)$	$Beta(2,2)$	$0.5N(0,1)+$ $0.5N(2,0.5^2)$
Mean	0	0	0.5	1
Variance	1	1.417	0.05	1.625
Skewness coefficient	0	0	0	-0.543
Kurtosis coefficient	3	3.353	2.143	2.402
Multimodal?	Unimodal	Trimodal	Unimodal	Bimodal

Using the data generating process described earlier, I plot in Figures 1 and 2 the weighting functions $w_d(\alpha,\rho)$ and $w_l(\alpha,\rho)$, respectively, when $T=3$.[3] I also plot in Figures 3 and 4 the weighting functions $w_l(\alpha,\rho)$ and $w_{FE}(\alpha,\rho)$, respectively, when $T\to\infty$. In all these figures, the black solid curve represents the correct weighting function that should be used. The other colored curves represent the weighting functions for fixed values of $\rho\in\{-2,-1,0,1,2\}$. The plots indicate that the distortion brought about by incorrect weighting is quite severe regardless of the size of T. All the plots share a common feature except for the Beta-distributed fixed effect – the incorrect weighting function tends to place higher weight on negative values of α as the value of ρ becomes larger. This means that cross-sectional units with some unobservable index α below the median get higher weight when there is positive state dependence. Similarly, cross-sectional units with unobservable index α above the median get higher weight when there is negative state dependence. Note that this effect is more pronounced for the case of positive state dependence. Another noteworthy aspect of the figures is that the incorrect weighting function can oversmooth the modes of the correct weighting function that are not located around negative values of the support of the fixed effect distribution, whether or not T is fixed or $T\to\infty$.

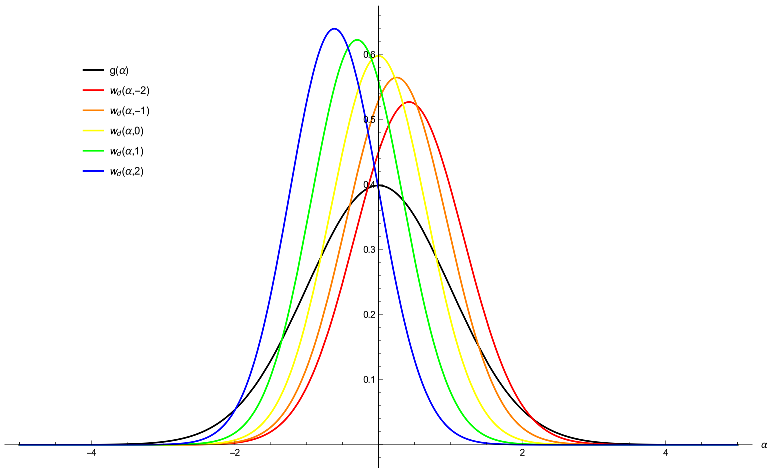
2.3 CON: The large-n limit of IV estimators may be found outside the identified set.

Chernozhukov et al. (2013) propose nonparametric bounds to estimate the identified set for the average marginal effect Δ. These nonparametric bounds are relatively easy to compute since they involve difference of means and counts for specific subsets of the data. Their method applies for any value of T. I focus on the case where $T=3$ and I perform the calculations based on the DGPs described in Subsection 2.2.3. In Figure 5, I plot[4] the large-n limits of the AH estimators (in blue for $\widehat{\gamma}_{AHd}$ and green for $\widehat{\gamma}_{AHl}$) and large-n limits of the nonparametric bounds proposed by Chernozhukov et al. (2013) (in

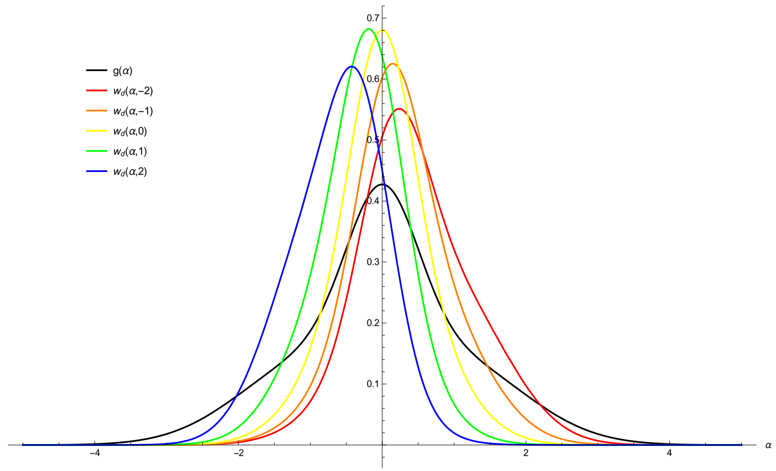
[3] I used the notation $w_l(\alpha,\rho,y_0)$ in the previous subsection. However, the DGP I study is such that $y_0\perp\alpha$.

[4] A Mathematica notebook containing the calculations is available upon request.

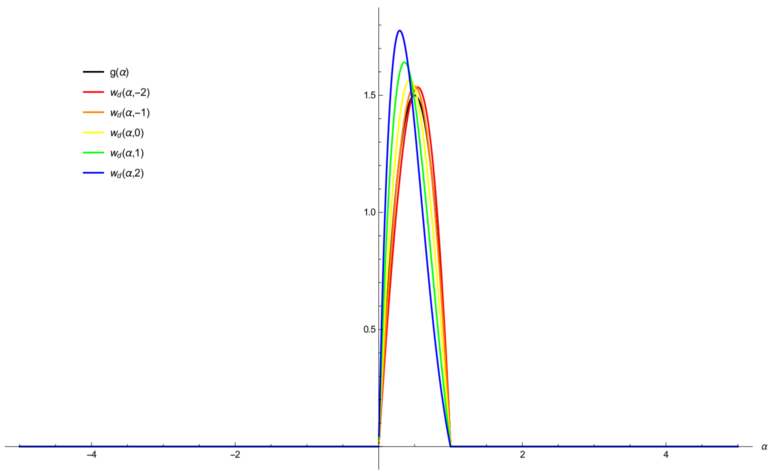
Figure 1: Weighting function $w_d(\alpha, \rho)$ under different distributions for the fixed effects for $T = 3$



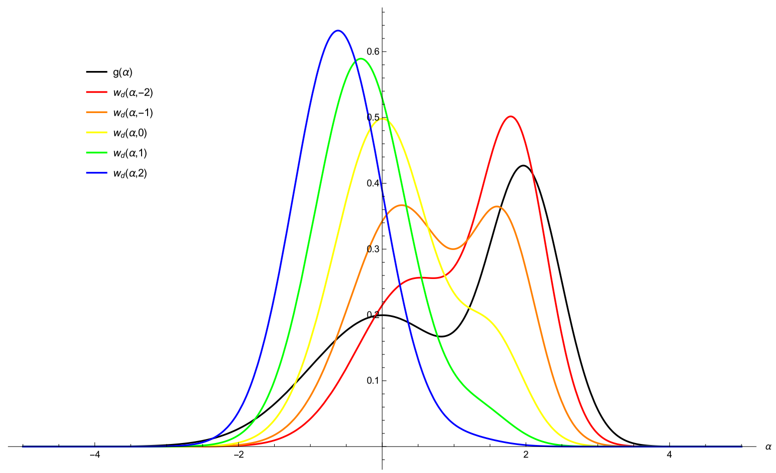
$\alpha_i \sim N(0, 1)$



$\alpha_i \sim 1/3 N(-1, 1) + 1/3 N(0, 0.5^2) + 1/3 N(1, 1)$

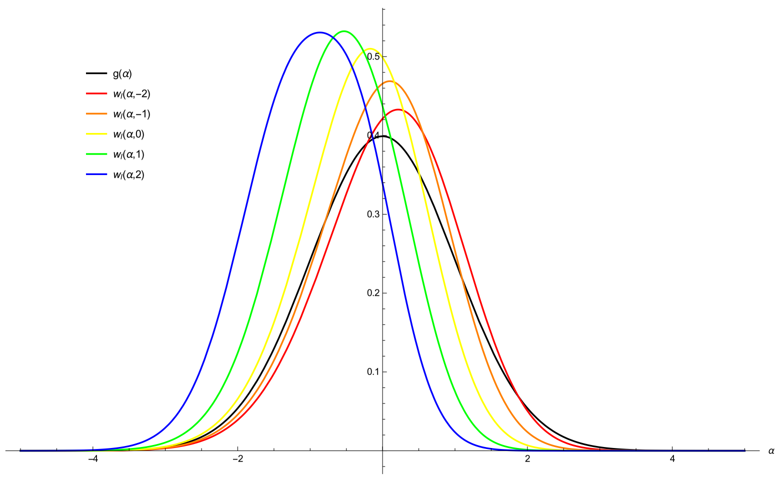


$\alpha_i \sim Beta(2, 2)$

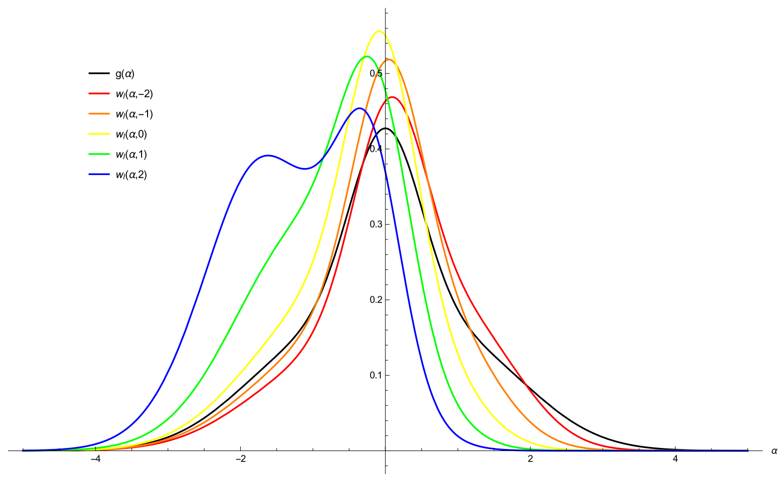


$\alpha_i \sim 0.5N(0, 1) + 0.5N(2, 0.5^2)$

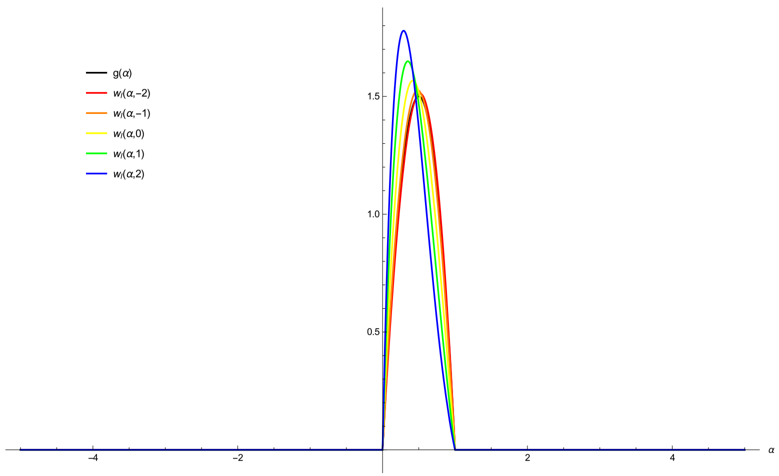
Figure 2: Weighting function $w_l(\alpha, \rho)$ under different distributions for the fixed effects for $T = 3$



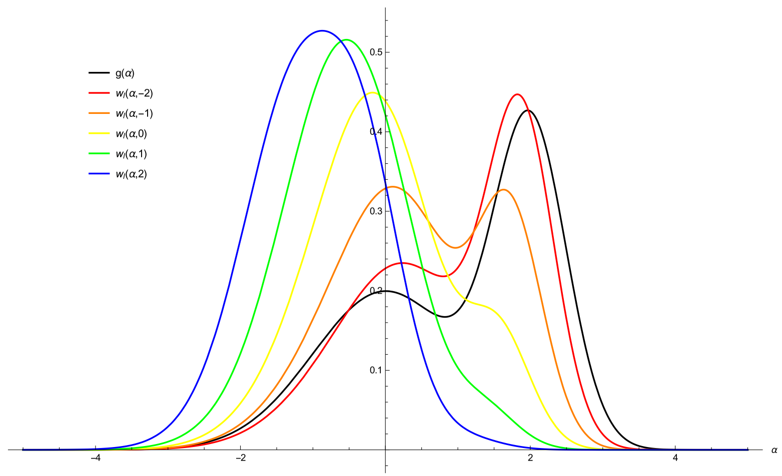
$\alpha_i \sim N(0, 1)$



$\alpha_i \sim {}^{1}\!/_{3}N(-1, 1) + {}^{1}\!/_{3}N(0, 0.5^2) + {}^{1}\!/_{3}N(1, 1)$

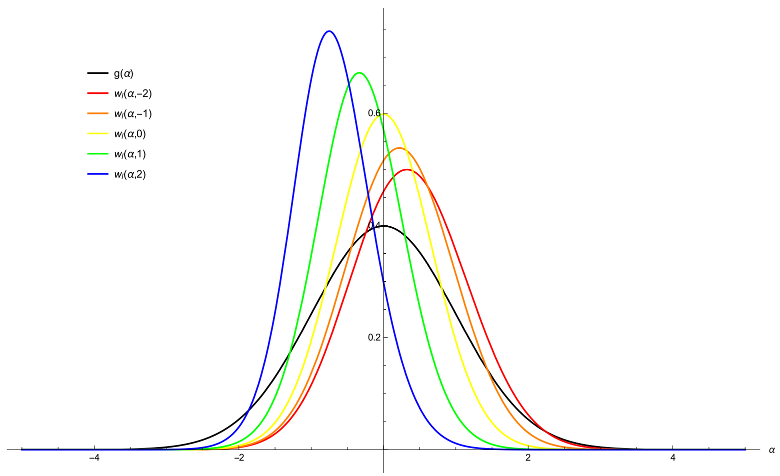


$\alpha_i \sim Beta(2, 2)$

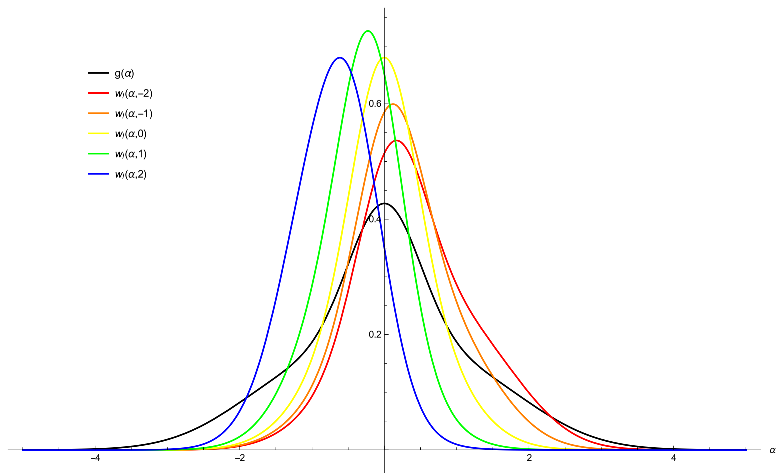


$\alpha_i \sim 0.5N(0, 1) + 0.5N(2, 0.5^2)$

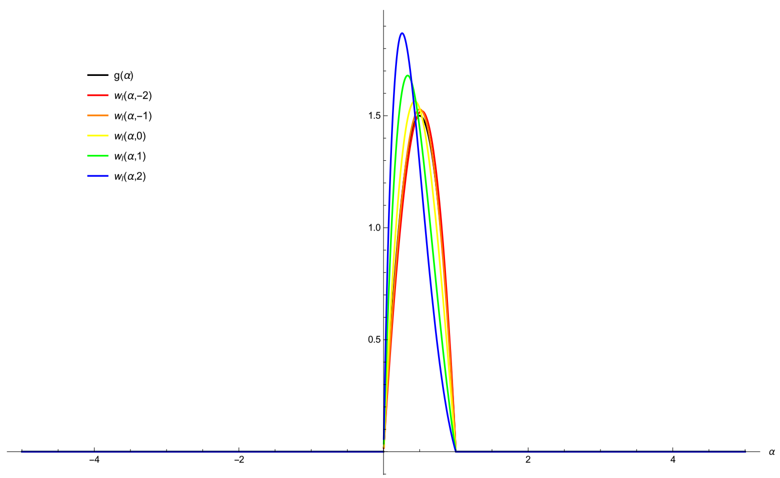
Figure 3: Weighting function $w_l(\alpha, \rho)$ under different distributions for the fixed effects for $T \to \infty$



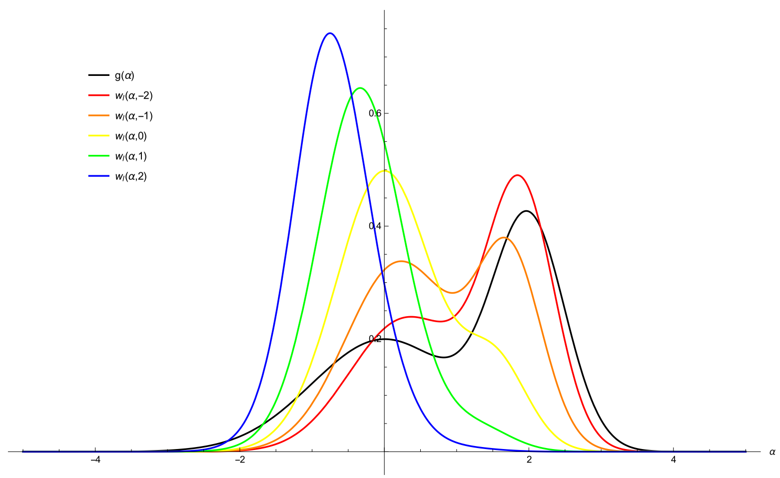
$\alpha_i \sim N(0,1)$



$\alpha_i \sim {}^{1}/{3}N(-1,1) + {}^{1}/{3}N\left(0, 0.5^2\right) + {}^{1}/{3}N(1,1)$

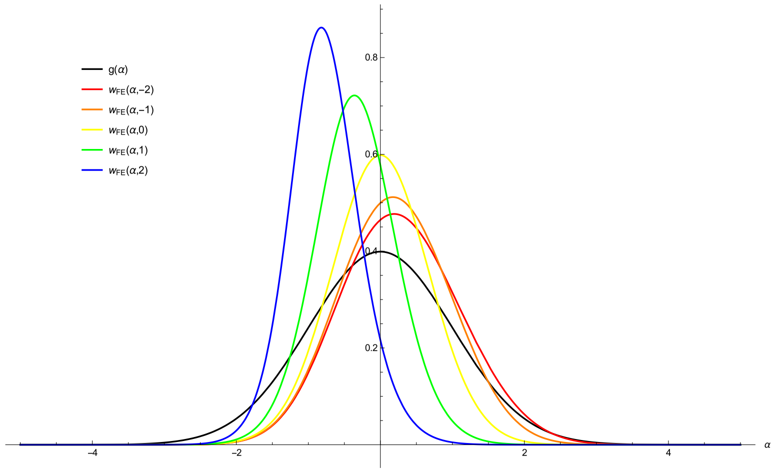


$\alpha_i \sim Beta(2,2)$

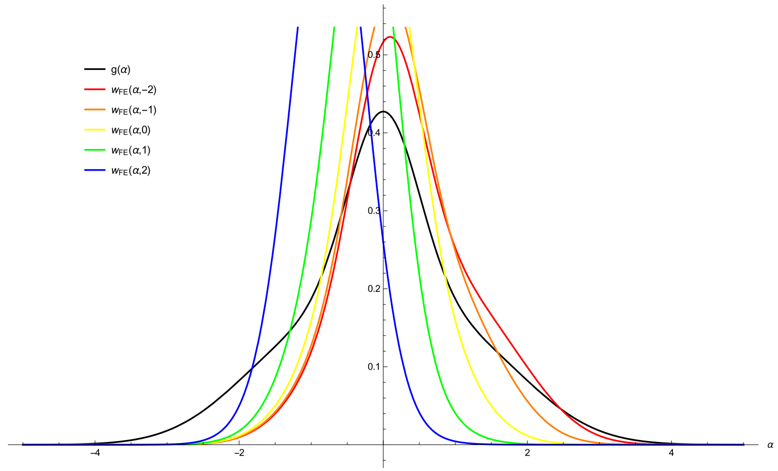


$\alpha_i \sim 0.5N(0,1) + 0.5N\left(2, 0.5^2\right)$

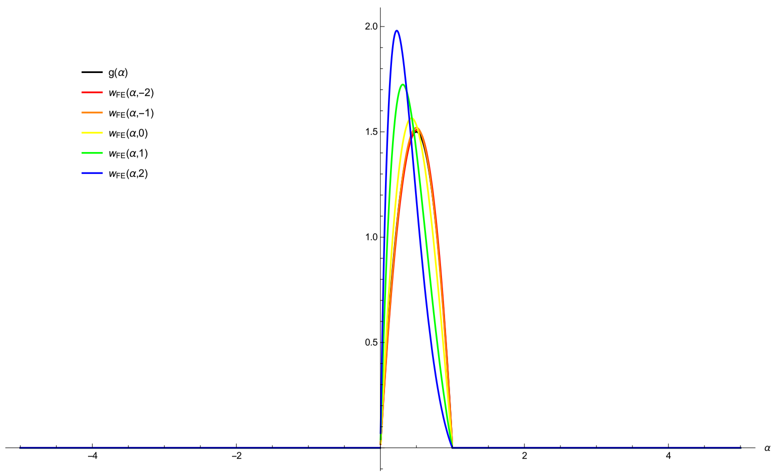
Figure 4: Weighting function $w_{FE}(\alpha, \rho)$ under different distributions for the fixed effects for $T \to \infty$



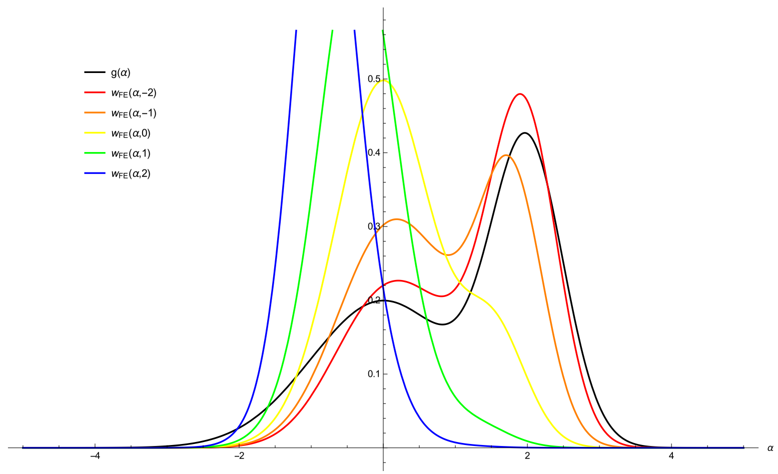
$\alpha_i \sim N(0,1)$



$\alpha_i \sim {}^1\!/\!{}_3 N(-1,1) + {}^1\!/\!{}_3 N(0, 0.5^2) + {}^1\!/\!{}_3 N(1,1)$



$\alpha_i \sim Beta(2,2)$



$\alpha_i \sim 0.5N(0,1) + 0.5N(2, 0.5^2)$

red for the lower bound $\hat{\Delta}_l$ and orange for the upper bound $\hat{\Delta}_u$) evaluated at different values of $\rho \in [-2, 2]$. I also calculate the true Δ (in black) using the true distribution of (y_0, α).

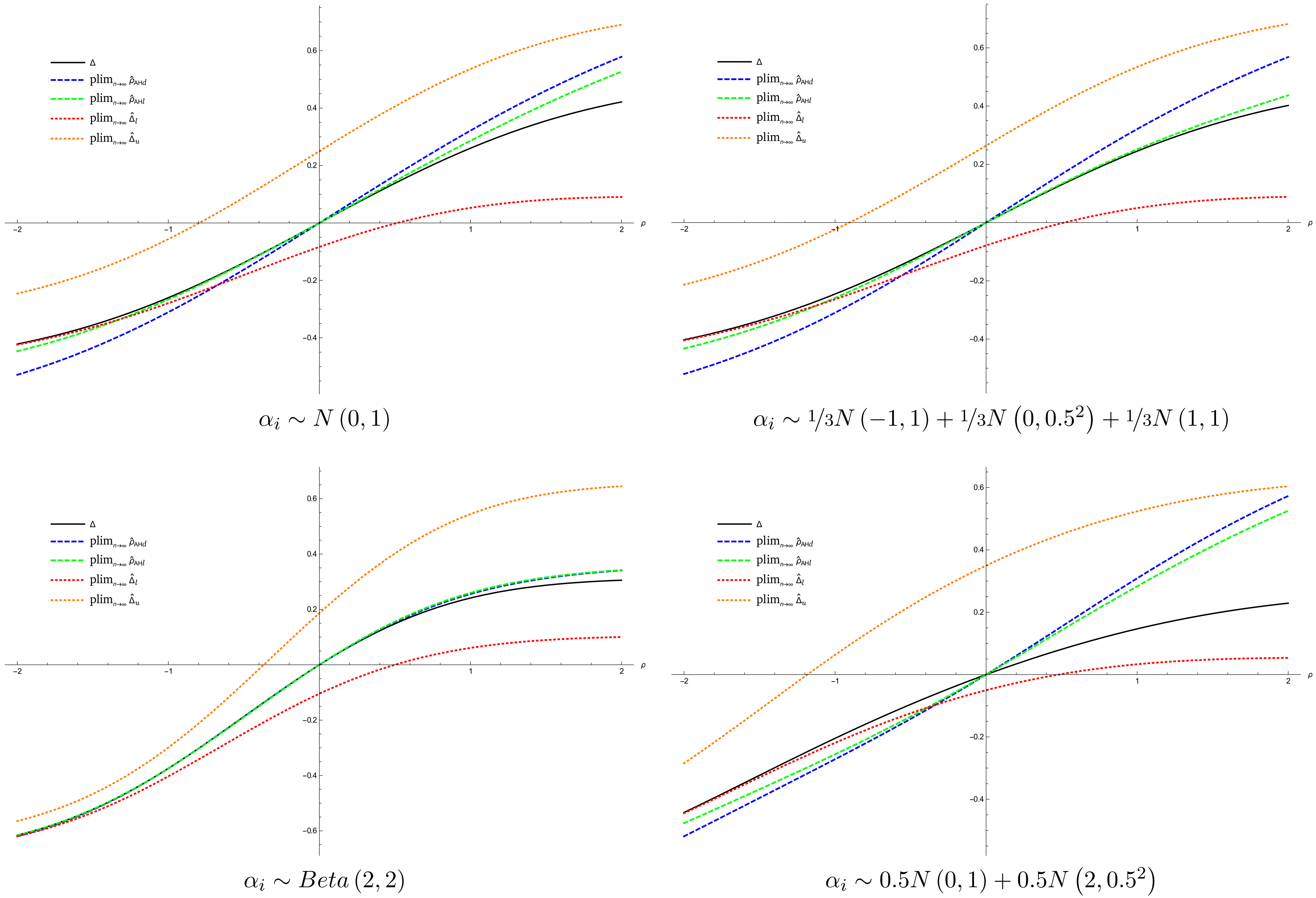
Even in the benchmark case where $\alpha_i \sim N(0, 1)$, both the large-n limits of the AH estimators are larger than Δ when $\rho > 0$. Further note that for relatively strong negative state dependence, both these large-n limits are outside the identified set. This result is practically relevant because negative state dependence has been found in the literature on scarring effects (see references in Torgovitsky (2015)). We have similar results for the case where $\alpha_i \sim {}^1\!/\!_3 N(-1, 1) + {}^1\!/\!_3 N(0, 0.5^2) + {}^1\!/\!_3 N(1, 1)$. It appears that the large-n limit of the AH estimator using levels as instruments is doing quite well, relative to the standard normal case. For $\alpha_i \sim Beta(2, 2)$, the large-n limits of the AH estimators are practically the same as Δ and both can be found in the identified set. The key seems to be the bounded support for the fixed effect, which is $[0, 1]$. For $\alpha_i \sim 0.5N(0, 1) + 0.5N(2, 0.5^2)$, both the large-$n$ limits of the AH estimators nearly coincide and are much larger than Δ even for less persistent state dependence.

Finally, Chernozhukov et al. (2013) show in their Theorem 4 that the identified set for Δ shrinks to a singleton as $T \to \infty$. Thus, it becomes more likely that the large-T limits in (4), (5), and (6) would be outside the identified set.

2.4 PRO/CON: IV estimators may be used to nonparametrically test the null of no first-order state dependence.

The good news is that the IV estimators are able to estimate a zero average marginal effect, if it were the truth. This observation may allow us to construct a nonparametric test of the hypothesis that $\Delta = 0$ or, equivalently, $\rho = 0$. In particular, one does not need to impose any assumptions about the unknown joint distribution of (y_0, α) and the inverse link function H to implement the test. In the case of a predetermined binary regressor, one also does not need to impose any assumptions about the model representing the feedback effect of past values of y on the current value of the predetermined binary regressor. Furthermore, this nonparametric test may be applied directly when T is fixed.

Figure 5: Large-n limits of the AH estimators under different distributions for the fixed effects



More importantly, it does not require us to find which pairs of binary sequences will have relative odds equal to 1 under the null[5]. To implement this Wald test, default routines in Stata or any other statistical software can easily be used without any modifications.

Unfortunately, this good news will be dampened by three considerations. First, the nonparametric test cannot be applied directly when T is large. Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003) have already shown that the asymptotic distribution of the FE and GMM estimators are not centered at zero under a large-n, large-T asymptotic scheme. It is unclear how one should construct the bias correction for the GMM estimator when there is model misspecification. Galvao and Kato (2014) have shown how to conduct inference for the pseudo-true parameter β_0 in (7). As discussed previously, this pseudo-true parameter has economic content when $\rho = 0$ because it represents a zero average marginal effect of first-order state dependence. The procedure that performed well in their Monte Carlo experiments is that of bias-correcting the FE estimator with the half-panel jackknife and applying cross-sectional bootstrap to a pivotal statistic, which in our context is a simple t-statistic. Unfortunately, the rate of convergence is $\sqrt{n}$ rather than $\sqrt{nT}$ even if the procedure is justified under a large-n, large-T asymptotic scheme. The rate of convergence depends on whether $\mathbb{E}\left(\widetilde{y}_{i,t-1}\varepsilon_{it}|\alpha_i\right)$ is equal to zero or not. In our context, we have

$$\begin{aligned}
\mathbb{E}\left(\widetilde{y}_{i,t-1}\varepsilon_{it}|\alpha_i\right) &= \left[H\left(\alpha+\rho\right)-\beta_0\right]\Pr\left(y_{t-1}=1|\alpha\right)\left(1-\Pr\left(y_{t-1}=1|\alpha\right)\right) \\
&= \left[H\left(\alpha+\rho\right)-\beta_0\right]\left[\frac{H\left(\alpha\right)}{1-H\left(\alpha+\rho\right)+H\left(\alpha\right)}\right]\left[\frac{1-H\left(\alpha+\rho\right)}{1-H\left(\alpha+\rho\right)+H\left(\alpha\right)}\right]
\end{aligned}$$

which is not equal to zero even if $\beta_0 = 0$ or, equivalently, $\rho = 0$.

Second, the nonparametric test might not have good power properties. It can be shown that for the large-n, fixed-T case,

$$\sqrt{n}\left(\widehat{\gamma}_j - ME_j\left(\rho\right)\right) \xrightarrow{d} N\left(0, V\left(\widehat{\gamma}_j, \rho\right)\right), \quad j \in \left\{AHd, AHl\right\},$$

[5]See Chay and Hyslop (2014) for an application of this idea but in the context of a descriptive analysis rather than a formal test.

where $ME_j(\rho)$ is the large-n limit of the corresponding estimator $\widehat{\gamma}_j$. Consider the test of the null that the average marginal effect of first-order state dependence is equal to zero, i.e. $\Delta = 0$. We reject the null at the $100\alpha\%$ level whenever $\left|\sqrt{n}\widehat{\gamma}_j / \sqrt{V(\widehat{\gamma}_j, 0)}\right| > z_\alpha$, where z_α satisfies $\Phi(z_\alpha) = 1 - \alpha$. As a result, the power function derived for local alternatives that satisfy $\sqrt{n}\rho_n = h$ can be written as

$$
\begin{aligned}
\Pr_{\rho_n \neq 0}\left(\left|\frac{\sqrt{n}\widehat{\gamma}_j}{\sqrt{V(\widehat{\gamma}_j, 0)}}\right| > z_\alpha\right) &= 1 - \Phi\left(\frac{z_\alpha\sqrt{V(\widehat{\gamma}_j, 0)} - \sqrt{n}\left(ME_j(\rho_n) - ME_j(0)\right)}{\sqrt{V(\widehat{\gamma}_j, \rho_n)}}\right) + o(1) \\
&= 1 - \Phi\left(\frac{z_\alpha\sqrt{V(\widehat{\gamma}_j, 0)} - \sqrt{n}\rho_n ME_j'(\rho_n) + o(\sqrt{n}\rho_n)}{\sqrt{V(\widehat{\gamma}_j, \rho_n)}}\right) + o(1) \\
&\overset{n\to\infty}{\to} 1 - \Phi\left(z_\alpha - h \cdot \frac{ME_j'(0)}{\sqrt{V(\widehat{\gamma}_j, 0)}}\right)
\end{aligned}
$$

Ultimately, comparing different tests (based on different estimators of the zero average marginal effect) depends on the ratio $ME_j'(0) / \sqrt{V(\widehat{\gamma}_j, 0)}$. Note that the numerator of this ratio depends on the value of the weighting functions evaluated at $\rho = 0$. Expressions for this ratio for the two AH estimators are as follows:

$$
\begin{aligned}
\frac{ME_{AHd}'(0)}{\sqrt{V(\widehat{\gamma}_{AHd}, 0)}} &= \frac{\int H'(\alpha) H(\alpha)(1 - H(\alpha)) g(\alpha)\ d\alpha}{2\int H(\alpha)(1 - H(\alpha))^2 g(\alpha)\ d\alpha} \\
\frac{ME_{AHl}'(0)}{\sqrt{V(\widehat{\gamma}_{AHl}, 0)}} &= \frac{\int H'(\alpha) H(\alpha)(1 - H(\alpha)) g(\alpha)\ d\alpha + \int H'(\alpha) H(\alpha)(1 - H(\alpha)) f(\alpha, 1)\ d\alpha}{2\int H(\alpha)^2 (1 - H(\alpha)) g(\alpha)\ d\alpha + 2\int H(\alpha)(1 - H(\alpha))^2 f(\alpha, 1)\ d\alpha}
\end{aligned}
$$

Despite having these closed forms, it is not possible to compare the asymptotic relative efficiency of the tests based on the two AH estimators. In the large-n, large-T case and provided that some bias-correction is made, it is worth noting that the numerator of the ratio $ME_j'(0) / \sqrt{V(\widehat{\gamma}_j, 0)}$ is equal across all $j \in \{AHd, AHl, FE, GMM\}$ because

$$
ME_j'(0) = \int H'(\alpha) w_j(\alpha, 0)\ d\alpha
$$

and

$$
w_d(\alpha, 0) = w_l(\alpha, 0) = w_{FE}(\alpha, 0) = w_{GMM}(\alpha, 0).
$$

The latter can be seen readily from the previous subsections. Thus, the asymptotic relative efficiency of the various tests will ultimately depend on which of the estimators have the lowest asymptotic variance. I leave this issue to further research.

I conduct a small Monte Carlo study comparing the power properties of these two AH estimators in the large-n, fixed-T case. Figures 6 and 7 contain plots of the estimated power curves based on 10000 replications of each of the DGPs in Subsection 2.2.3. The estimated power curves are computed over a equally spaced grid for $\rho \in [-1, 1]$ and for $n \in \{100, 500, 2500\}$. Both AH estimators have good size control. From a practical point of view, researchers should use $\widehat{\gamma}_{AHl}$ because $\widehat{\gamma}_{AHd}$ may not exist for very small values of n. Both figures feature low empirical rejection rates (below 50%) when $\rho \in [-0.25, 0) \cup (0, 0.25]$. A very noteworthy aspect of the figures is the situation where $\alpha_i \sim Beta\,(2, 2)$. In Figure 5, we find that the true average marginal effect was approximated quite well, yet the estimated power curves indicate poor power properties compared to the other distributions for unobserved heterogeneity.

Finally, the nonparametric test becomes more complicated to construct once you include other regressors beyond the lagged binary dependent variable or the predetermined binary regressor. The analytical properties of the resulting estimators once additional regressors are accounted for are beyond the scope of this paper.

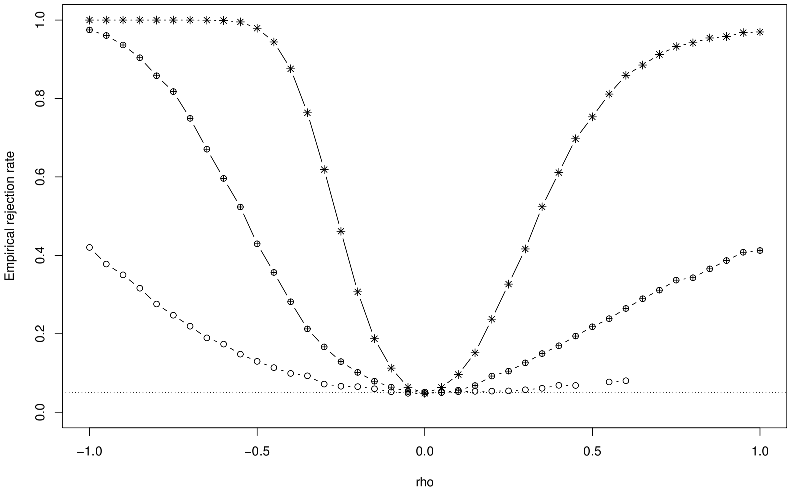
3 Empirical illustration

Using one of the empirical applications in Chernozhukov et al. (2013) on female labor force participation and fertility, I now illustrate why IV estimation of the dynamic LPM with fixed effects is a situation where the cons can outweigh the pros. I use Stata 12 for this illustration using the most convenient and popular procedures as possible.

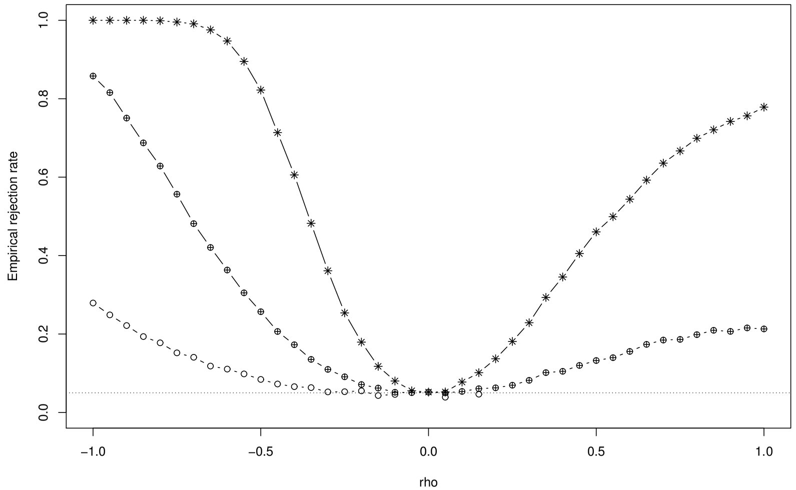
They estimate the following model using complete longitudinal data on 1587 married women selected from the National Longitudinal Survey of Youth 1979 and observed for three years – 1990, 1992, and 1994:

$$LFP_{it} = \mathbf{1}\left(\beta \cdot kids_{it} + \alpha_i \geq \epsilon_{it}\right).$$

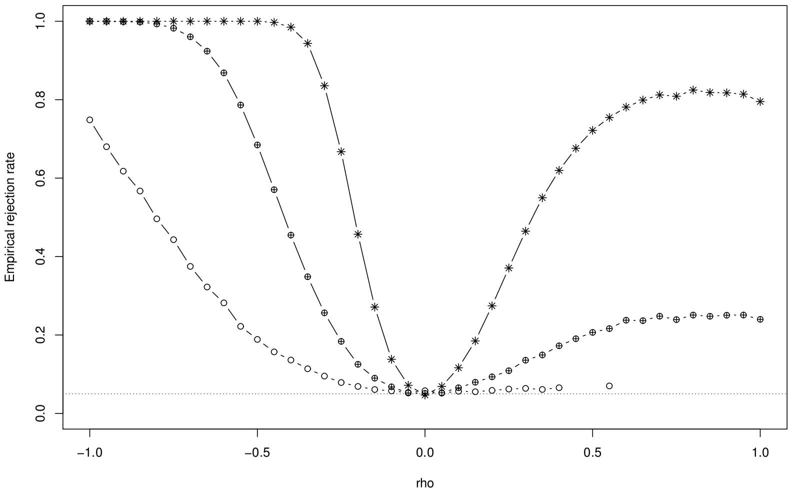
Figure 6: Power curves obtained from using $\widehat{\gamma}_{AHd}$ under different distributions for the fixed effects



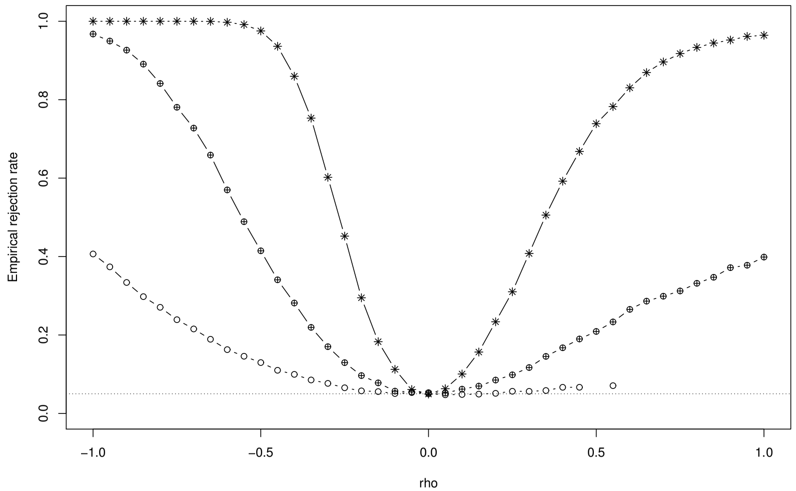
$\alpha_i \sim N(0,1)$



$\alpha_i \sim {}^1/_3 N(-1,1) + {}^1/_3 N\left(0, 0.5^2\right) + {}^1/_3 N(1,1)$



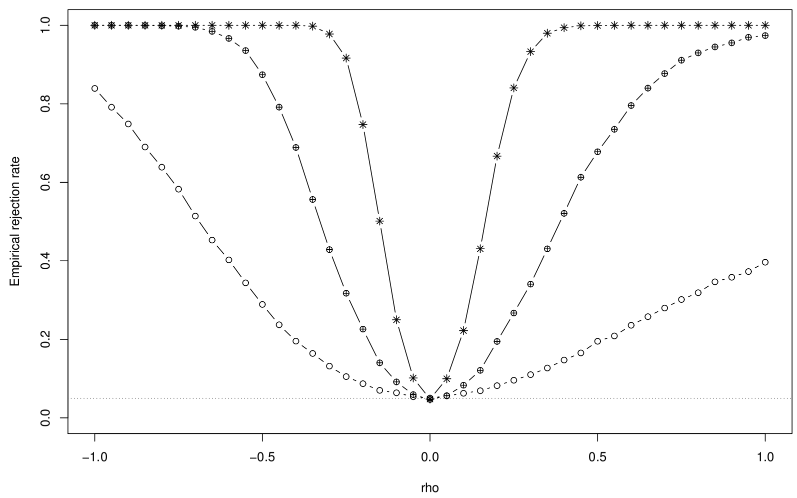
$\alpha_i \sim Beta(2,2)$



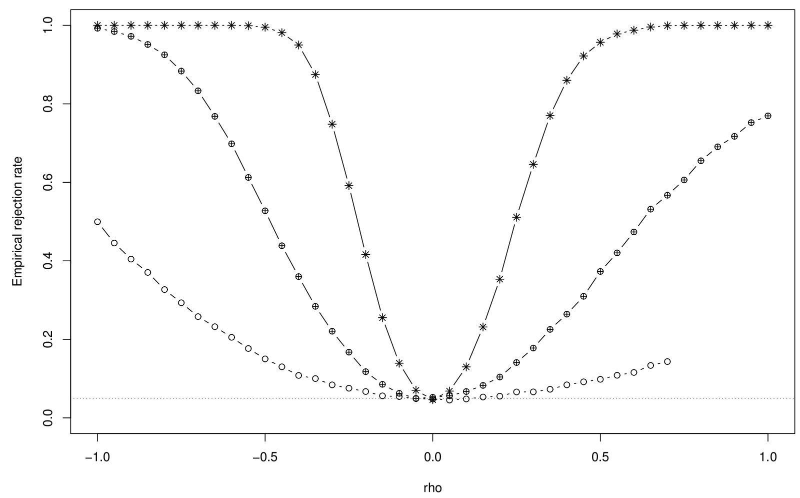
$\alpha_i \sim 0.5N(0,1) + 0.5N\left(2, 0.5^2\right)$

○ $n = 100$, ⊕ $n = 500$, * $n = 2500$; Dotted line is the 5% mark.

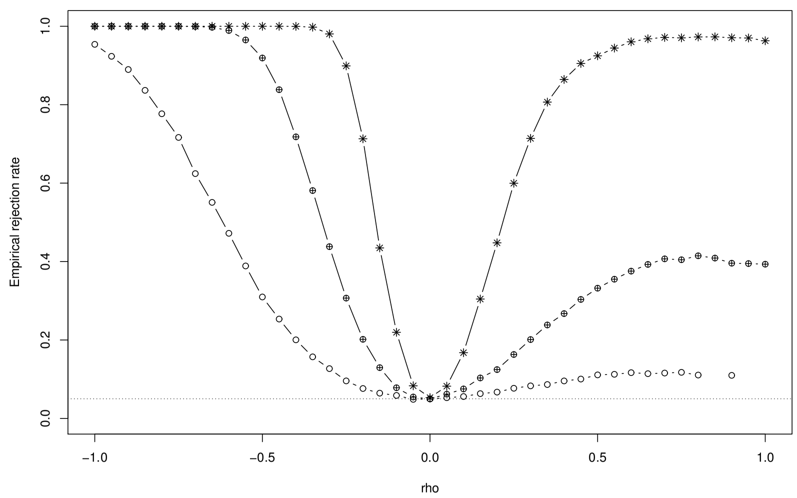
Figure 7: Power curves obtained from using $\widehat{\gamma}_{AHl}$ under different distributions for the fixed effects



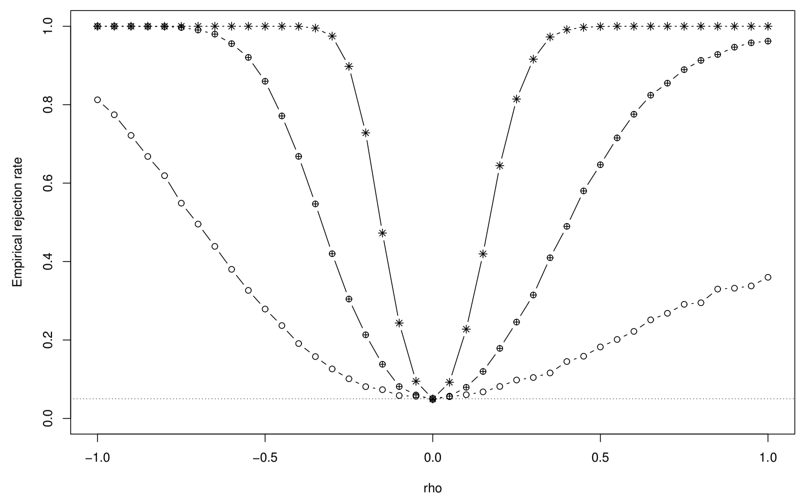
$\alpha_i \sim N(0,1)$



$\alpha_i \sim {}^1\!/\!_3 N(-1,1) + {}^1\!/\!_3 N(0, 0.5^2) + {}^1\!/\!_3 N(1,1)$



$\alpha_i \sim Beta(2,2)$



$\alpha_i \sim 0.5N(0,1) + 0.5N(2, 0.5^2)$

○ $n = 100$, ⊕ $n = 500$, * $n = 2500$; Dotted line is the 5% mark.

The parameter of interest is the average marginal effect of fertility on female labor force participation. The dependent variable is a labor force participation indicator, the regressor is a fertility indicator that takes the value 1 if the woman has a child less than 3 years old, and α_i is an individual-specific fixed effect.

Table 10: Female LFP and fertility

Procedure	Estimate of Δ	95% CI
Static NP bounds	$[-0.404, -0.043]$	$\mathcal{DU}_n = [-0.425, -0.022]^a$ $\mathcal{U}_n = [-0.429, -0.017]^a$
Static random effects probit	-0.106	$[-0.130, -0.082]^b$
FE	-0.084	$[-0.110, -0.058]^b$ $[-0.112, -0.056]^c$
First-difference OLS	-0.068	$[-0.094, -0.042]^b$ $[-0.097, -0.038]^c$

$n = 1587$, $T = 3$, fertility is treated as strictly exogenous

[a] Constructed from 1 million draws using the algorithm in Beresteanu and Molinari (2008).

[b] Obtained using the usual standard errors.

[c] Obtained using clustered standard errors.

Chernozhukov et al. (2013) computed estimates of the nonparametric bounds for the average marginal effect under the assumption that the fertility indicator is strictly exogenous (called static nonparametric bounds) and that the average marginal effect is decreasing[6] (called monotonicity) in the fertility indicator. I recompute[7] and was able to reproduce their estimated bounds. In addition, I compute FE, first-differenced OLS, and static random effects probit (using `xtprobit`) estimates of the average marginal effect under the assumption that the fertility indicator is strictly exogenous. These estimates can be found in Table 10. Notice that both the FE and first-differenced estimates can be found inside the static bounds. The estimated average marginal effect from the random effects probit is inside the static bounds, despite the very incredible assumption that the fixed effects are independent of the past, present, and current values of the fertility indicator.

To illustrate the results in this paper, I now treat the fertility indicator as predetermined and compute estimates of the dynamic nonparametric bounds under this as-

[6] Details as to how to construct the bounds under monotonicity can be found in the Supplemental Material to Chernozhukov et al. (2013).

[7] A Stata do-file for this entire section is available for replication upon request.

sumption. I also calculate the AH (using `ivregress`) and Arellano-Bond (using David Roodman's (2009) `xtabond2` or the built-in Stata command `xtdpd`) estimates. These estimates can be found in Table 11. Notice that the AH and Arellano-Bond estimates, which actually assume predeterminedness, are outside the dynamic bounds. The estimates are also substantially smaller (roughly 11% smaller in absolute value) than the upper bound of the set estimate.

Table 11: Female LFP and fertility

Procedure	Estimate of Δ	95% CI
Dynamic NP bounds	$[-0.386, -0.187]$	$\mathcal{DU}_n = [-0.409, -0.164]^a$ $\mathcal{U}_n = [-0.429, -0.144]^a$
AH (lagged differences)	-0.007	$[-0.131, 0.117]^b$ $[-0.140, 0.126]^c$
AH (lagged levels)	-0.022	$[-0.071, 0.028]^b$ $[-0.071, 0.027]^c$
Arellano-Bond (one-step)	-0.023	$[-0.063, 0.017]^b$
Arellano-Bond (two-step)	-0.022	$[-0.070, 0.026]^b$ $[-0.070, 0.026]^d$

$n = 1587$, $T = 3$, fertility is treated as predetermined

[a] Constructed from 1 million draws using the algorithm in Beresteanu and Molinari (2008).

[b] Obtained using the usual standard errors.

[c] Obtained using clustered standard errors.

[d] Calculated with the Windmeijer correction.

I also construct 95% asymptotic confidence intervals for each procedure. The tables provide information as to what standard errors were used. For the nonparametric bounds, I use the algorithm by Beresteanu and Molinari (2008) – which is one of the suggested procedures in Chernozhukov et al. (2013). I report two confidence intervals denoted as $\mathcal{DU}_n$ and $\mathcal{U}_n$ by Beresteanu and Molinari (2008). Note that the null hypothesis $\beta = 0$ is not rejected if we use linear dynamic panel data methods for inference, yet $\beta = 0$ is rejected when dynamic nonparametric bounds were used. Thus, the findings in Table 11 further reinforce the potential lack of power of the nonparametric test of the null hypothesis $\beta = 0$.

Finally, I calculated some diagnostics typically reported by researchers applying linear dynamic panel data methods such as a test of overidentifying restrictions and tests of weak identification. I do not report them explicitly here because their usage may be questionable but the calculations indicate that we do not reject the model and that the weak identification-robust confidence intervals are very close to the reported interval $[-0.070, 0.026]$. Thus, all these indicate "passing" marks for the dynamic LPM with fixed effects despite the non-rejection of the possibly false null $\beta = 0$.

4 Concluding remarks

I show that using IV/GMM methods to estimate the dynamic LPM with fixed effects is inappropriate as $n \to \infty$ (whether T is fixed or T diverges with n). The analytical results indicate that incorrect weighting of the individual treatment effect is the source of the problem. This incorrect weighting function has a form that depends on both the value of the fixed effect, the initial condition, their joint distribution, and the true value of the autoregressive parameter. The examples indicate that there is a tendency for this incorrect weighting function to place higher weight on fixed effect values below the median of the distribution of the fixed effects, whenever the true autoregressive parameter is positive. From these examples, believing that the IV estimators will be practically close to the true average marginal effect requires us to hold strong prior beliefs about the support of the distribution for unobserved heterogeneity. These prior beliefs are exactly what researchers wanted to avoid by using a fixed effects method.

In addition, I construct specific examples to show that the estimators may be outside the identified set even in the limit. Therefore, it is more appropriate to use the non-parametric bounds proposed by Chernozhukov et al. (2013), especially if one is unwilling to specify the form for the inverse link function and the joint distribution of the initial conditions and the fixed effects.

The large-n, large-T results I obtain are based on sequential asymptotics. Given the results in Section 2 and the Appendix, it is very likely that we should obtain similar inconsistency results based on joint asymptotics. The results in the paper point out that the direction of the asymptotic bias of the estimator for the average marginal effect cannot be obtained. This is in stark contrast with the direction of the asymptotic bias derived by Nickell (1981). The procedure used by Fernandez-Val (2009) is bias-correcting the FE estimator using an estimate of the Nickell bias. Although the Monte Carlo experiments of Fernandez-Val (2009) indicate good finite sample performance of this procedure, future work should study exactly what these corrections are doing.

For large-n, large-T settings, testing nonparametrically the point null of zero first-order state dependence or zero effect for the predetermined binary treatment would have to use the inference procedure by Galvao and Kato (2014). For large-n, fixed-T settings, the recommended procedure for now is to use the AH estimator using lagged levels as instruments. For this nonparametric test to have good power properties, a large value of n would be required as seen in the Monte Carlo simulations.

It would also be interesting to derive similar analytical results for correlated random effects models so that the results in Wooldridge (2005) and Murtazashvili and Wooldridge (2008) can be extended to the dynamic case. In the empirical application, I find that the average marginal effect from the usual random effects probit under strict exogeneity can be found within the static nonparametric bounds. Respecting the inherent nonlinearity of a discrete choice model (even under potential misspecification of both the inverse link function and the distribution of the fixed effects) may be responsible for this finding. Future work on this will be of practical interest.

Finally, an extension to the case where there are other strictly exogenous or predetermined regressors would be most welcome given that the nonparametric test is restricted only to a model without these other regressors. A possible direction would be to calculate the estimators for subsets of the data for which all the other regressors take on specific values. A kernel-based procedure may be used for regressors that are continuous. If this direction is possible, then the nonparametric test will have wider applicability.

References

Acemoglu, D, S Johnson, JA Robinson, and P Yared (2009). "Reevaluating the modernization hypothesis". *Journal of Monetary Economics* 56.8, pp. 1043–1058.

Alvarez, J and M Arellano (2003). "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators". *Econometrica* 71.4, pp. 1121–1159.

Anderson, T and C Hsiao (1981). "Estimation of Dynamic Models with Error Components". *Journal of the American Statistical Association* 76.375, pp. 598–606.

Anderson, T and C Hsiao (1982). "Formulation and estimation of dynamic models using panel data". *Journal of Econometrics* 18.1, pp. 47–82.

Angrist, JD and JS Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

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, pp. 2–16.

Arellano, M and S Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". *Review of Economic Studies* 58.2, pp. 277–97.

Arellano, M and S Bonhomme (2011). "Nonlinear panel data analysis". *Annual Review of Economics* 3, pp. 395–424.

Arellano, M and O Bover (1995). "Another look at the instrumental variable estimation of error-components models". *Journal of Econometrics* 68.1, pp. 29–51.

Beresteanu, A and F Molinari (2008). "Asymptotic Properties for a Class of Partially Identified Models". *Econometrica* 76.4, pp. 763–814. URL: http://dx.doi.org/10.1111/j.1468-0262.2008.00859.x.

Bernard, AB and JB Jensen (2004). "Why Some Firms Export". *The Review of Economics and Statistics* 86.2, pp. 561–569.

Burda, M, M Harding, and J Hausman (2015). "A Bayesian Semiparametric Competing Risk Model with Unobserved Heterogeneity". *Journal of Applied Econometrics* 30.3, pp. 353–376.

Chay, KY and DR Hyslop (2014). "Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence Using Alternative Approaches". In: *Safety Nets and Benefit Dependence.* Chap. 1, pp. 1–39.

Chernozhukov, V, I Fernández-Val, J Hahn, and W Newey (2013). "Average and Quantile Effects in Nonseparable Panel Models". *Econometrica* 81.2, pp. 535–580.

Fernandez-Val, I (2009). "Fixed effects estimation of structural parameters and marginal effects in panel probit models". *Journal of Econometrics* 150, pp. 71–85.

Galvao, AF and K Kato (2014). "Estimation and Inference for Linear Panel Data Models Under Misspecification When Both n and T are Large". *Journal of Business & Economic Statistics* 32.2, pp. 285–309.

Hahn, J (2001). "Comment: Binary Regressors in Nonlinear Panel-Data Models with Fixed Effects". 19.1, pp. 16–17.

Hahn, J and G Kuersteiner (2002). "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T Are Large". *Econometrica* 70.4, pp. 1639–1657. URL: http://dx.doi.org/10.1111/1468-0262.00344.

Hyslop, DR (1999). "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women". *Econometrica* 67.6, pp. 1255–1294.

Jiménez, G, S Ongena, JL Peydró, and J Saurina (2014). "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking?" *Econometrica* 82.2, pp. 463–505.

Lewbel, A, Y Dong, and TT Yang (2012). "Viewpoint: Comparing features of convenient estimators for binary choice models with endogenous regressors". *Canadian Journal of Economics* 45.3, pp. 809–829.

Loken, KV, M Mogstad, and M Wiswall (2012). "What Linear Estimators Miss: The Effects of Family Income on Child Outcomes". *American Economic Journal: Applied Economics* 4.2, pp. 1–35. URL: http://www.aeaweb.org/articles?id=10.1257/app.4.2.1.

Murtazashvili, I and JM Wooldridge (2008). "Fixed effects instrumental variables estimation in correlated random coefficient panel data models". *Journal of Econometrics* 142.1, pp. 539–552.

Nickell, SJ (1981). "Biases in Dynamic Models with Fixed Effects". *Econometrica* 49.6, pp. 1417–26.

Okui, R (2015). "Misspecification in Dynamic Panel Data Models and Model-Free Inferences". *Japanese Economic Review*.

Phillips, PCB and HR Moon (1999). "Linear Regression Limit Theory for Nonstationary Panel Data". *Econometrica* 67.5, pp. 1057–1111.

Roodman, D (2009). "How to do xtabond2: An introduction to difference and system GMM in Stata". *Stata Journal* 9.1, pp. 86–136. URL: https://ideas.repec.org/a/tsj/stataj/v9y2009i1p86-136.html.

Rossi, P, G Allenby, and R McCulloch (2012). *Bayesian Statistics and Marketing*. Wiley Series in Probability and Statistics. Wiley.

Torgovitsky, A (2015). *Partial Identification of State Dependence*. Tech. rep. URL: http://ssrn.com/abstract=2564305.

Wooldridge, JM (2005). "Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models". *The Review of Economics and Statistics* 87.2, pp. 385–390.

Wooldridge, JM (2010). *Econometric Analysis of Cross-Section and Panel Data*. MIT Press.

A Some calculations for (3)

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

$$\begin{aligned}
& \Pr\left(y_{i0}=0, y_{i1}=1, y_{i2}=1, y_{i3}=0\right) \\
= & \int \Pr\left(y_{i0}=0, y_{i1}=1, y_{i2}=1, y_{i3}=0 \mid \alpha\right) g(\alpha)\, d\alpha \\
= & \int \Pr\left(y_{i3}=0 \mid y_{i0}=0, y_{i1}=1, y_{i2}=1, \alpha\right) \Pr\left(y_{i2}=1 \mid y_{i0}=0, y_{i1}=1, \alpha\right) \times \\
& \Pr\left(y_{i1}=1 \mid y_{i0}=0, \alpha\right) \Pr\left(y_{i0}=0 \mid \alpha\right) g(\alpha)\, d\alpha \\
= & \int \Pr\left(y_{i3}=0 \mid y_{i2}=1, \alpha\right) \Pr\left(y_{i2}=1 \mid y_{i1}=1, \alpha\right) \Pr\left(y_{i1}=1 \mid y_{i0}=0, \alpha\right) f(\alpha, 0)\, d\alpha \\
= & \int (1-H(\alpha+\rho)) H(\alpha+\rho) H(\alpha) f(\alpha, 0)\, d\alpha, \qquad (8)
\end{aligned}$$

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

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

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

B Some calculations for the large-T case

Recall that these estimators are given by the following expressions:

$$\widehat{\gamma}_{AHd}=\frac{\sum_{i=1}^{n} \sum_{t=3}^{T} \Delta y_{i,t-2} \Delta y_{it}}{\sum_{i=1}^{n} \sum_{t=3}^{T} \Delta y_{i,t-2} \Delta y_{i,t-1}}, \widehat{\gamma}_{AHl}=\frac{\sum_{i=1}^{n} \sum_{t=2}^{T} y_{i,t-2} \Delta y_{it}}{\sum_{i=1}^{n} \sum_{t=2}^{T} y_{i,t-2} \Delta y_{i,t-1}}, \widehat{\gamma}_{FD}=\frac{\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta y_{it} \Delta y_{i,t-1}}{\sum_{i=1}^{n} \sum_{t=2}^{T}\left(\Delta y_{i,t-1}\right)^{2}}.$$

Note that $\Delta y_{i,t-2}\Delta y_{it} = y_{i,t-2}y_{it} - y_{i,t-3}y_{it} - y_{i,t-2}y_{i,t-1} + y_{i,t-3}y_{i,t-1}$. Observe that the binary nature of y allows us to write

$$\frac{1}{T}\sum_{t=3}^{T} y_{i,t-2}y_{it} \xrightarrow{p} \lim_{T\to\infty}\frac{1}{T}\sum_{t=3}^{T}\Pr\left(y_{it}=1, y_{i,t-2}=1\right).$$

By the law of total probability, the definition of conditional probability, and calculations similar to (8), we are able to express $\Pr\left(y_{it}=1, y_{i,t-2}=1\right)$ as

$$\begin{aligned}
& \Pr\left(y_{it}=1, y_{i,t-2}=1\right) \\
= \; & \Pr\left(y_{it}=1, y_{i,t-1}=0, y_{i,t-2}=1\right) + \Pr\left(y_{it}=1, y_{i,t-1}=1, y_{i,t-2}=1\right) \\
= \; & \int H(\alpha)\left(1-H(\alpha+\rho)\right)\Pr\left(y_{i,t-2}=1|\alpha\right)g(\alpha)\,d\alpha + \int H(\alpha+\rho)^2\Pr\left(y_{i,t-2}=1|\alpha\right)g(\alpha)\,d\alpha.
\end{aligned}$$

As a result, we have

$$\frac{1}{T}\sum_{t=3}^{T} y_{i,t-2}y_{it} \xrightarrow{p} \int\left[H(\alpha+\rho)^2 + H(\alpha)\left(1-H(\alpha+\rho)\right)\right]\left[\lim_{T\to\infty}\frac{1}{T}\sum_{t=3}^{T}\Pr\left(y_{i,t-2}=1|\alpha\right)\right]g(\alpha)\,d\alpha.$$

Finally observe that $\Pr\left(y_{i,t-2}=1|\alpha\right)$ obeys a first-order nonhomogeneous difference equation. In particular, note that

$$\begin{aligned}
\Pr\left(y_{i1}=1|\alpha\right) &= \Pr\left(y_{i1}=1|y_{i0}=1,\alpha\right)\Pr\left(y_{i0}=1|\alpha\right) + \Pr\left(y_{i1}=1|y_{i0}=0,\alpha\right)\Pr\left(y_{i0}=0|\alpha\right) \\
&= \left[H(\alpha+\rho)-H(\alpha)\right]\Pr\left(y_{i0}=1|\alpha\right) + H(\alpha) \\
\Pr\left(y_{i2}=1|\alpha\right) &= \Pr\left(y_{i2}=1|y_{i1}=1,\alpha\right)\Pr\left(y_{i1}=1|\alpha\right) + \Pr\left(y_{i2}=1|y_{i1}=0,\alpha\right)\Pr\left(y_{i1}=0|\alpha\right) \\
&= \left[H(\alpha+\rho)-H(\alpha)\right]\Pr\left(y_{i1}=1|\alpha\right) + H(\alpha) \\
&\vdots \\
\Pr\left(y_{it}=1|\alpha\right) &= \left[H(\alpha+\rho)-H(\alpha)\right]\Pr\left(y_{i,t-1}=1|\alpha\right) + H(\alpha)
\end{aligned}$$

The solution to the above difference equation can be written as

$$\Pr\left(y_{it}=1|\alpha\right) = \left[H(\alpha+\rho)-H(\alpha)\right]^t\Pr\left(y_{i0}=1|\alpha\right) + \sum_{s=0}^{t-1}\left[H(\alpha+\rho)-H(\alpha)\right]^s H(\alpha).$$

Note that $|H(\alpha+\rho) - H(\alpha)| < 1$. As a result, the effect of the initial condition disappears as $T \to \infty$:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=3}^{T} \Pr(y_{i,t-2} = 1|\alpha) = \frac{H(\alpha)}{1 - H(\alpha+\rho) + H(\alpha)}.$$

Thus, we have

$$\frac{1}{T}\sum_{t=3}^{T} y_{i,t-2}y_{it} \xrightarrow{p} \int \left[H(\alpha+\rho)^2 + H(\alpha)(1 - H(\alpha+\rho))\right]\left[\frac{H(\alpha)}{1 - H(\alpha+\rho) + H(\alpha)}\right] g(\alpha)\, d\alpha.$$

Following similar calculations, we can derive the large-T limits of the other components. In particular,

$$\begin{aligned}
\frac{1}{T}\sum_{t=3}^{T} y_{i,t-2}y_{i,t-1} &\xrightarrow{p} \lim_{T\to\infty}\frac{1}{T}\sum_{t=3}^{T}\Pr(y_{i,t-1}=1, y_{i,t-2}=1) \\
&= \int H(\alpha+\rho)\left[\frac{H(\alpha)}{1 - H(\alpha+\rho) + H(\alpha)}\right] g(\alpha)\, d\alpha. \\
\frac{1}{T}\sum_{t=3}^{T} y_{i,t-3}y_{it} &\xrightarrow{p} \lim_{T\to\infty}\frac{1}{T}\sum_{t=3}^{T}\Pr(y_{it}=1, y_{i,t-3}=1) \\
&= \int H(\alpha+\rho)^3\left[\frac{H(\alpha)}{1 - H(\alpha+\rho) + H(\alpha)}\right] g(\alpha)\, d\alpha \\
&\quad + \int 2H(\alpha+\rho)H(\alpha)(1 - H(\alpha+\rho))\left[\frac{H(\alpha)}{1 - H(\alpha+\rho) + H(\alpha)}\right] g(\alpha)\, d\alpha \\
&\quad + \int H(\alpha)(1 - H(\alpha))(1 - H(\alpha+\rho))\left[\frac{H(\alpha)}{1 - H(\alpha+\rho) + H(\alpha)}\right] g(\alpha)\, d\alpha.
\end{aligned}$$

Observe that the last term $\frac{1}{T}\sum_{t=3}^{T} y_{i,t-3}y_{i,t-1}$ has the same probability limit as $\frac{1}{T}\sum_{t=3}^{T} y_{i,t-2}y_{i,t}$ as $T \to \infty$. Assembling all the results together, we have as $T \to \infty$,

$$\frac{1}{T}\sum_{t=3}^{T} \Delta y_{i,t-2}\Delta y_{it} \xrightarrow{p} -\int (1 - H(\alpha+\rho)) H(\alpha) (H(\alpha+\rho) - H(\alpha))\, g(\alpha)\, d\alpha.$$

The other large-T results now follow similar computations. In particular, we have as $T \to \infty$,[8]

$$\begin{aligned}
\frac{1}{T}\sum_{t=3}^{T}\Delta y_{i,t-2}\Delta y_{i,t-1} &\xrightarrow{p} -\int\left(1-H\left(\alpha+\rho\right)\right)H\left(\alpha\right)g\left(\alpha\right)\,d\alpha, \\
\frac{1}{T}\sum_{t=2}^{T}y_{i,t-2}\Delta y_{it} &\xrightarrow{p} -\int\frac{\left(1-H\left(\alpha+\rho\right)\right)H\left(\alpha\right)}{1-H\left(\alpha+\rho\right)+H\left(\alpha\right)}\left(H\left(\alpha+\rho\right)-H\left(\alpha\right)\right)g\left(\alpha\right)\,d\alpha, \\
\frac{1}{T}\sum_{t=2}^{T}y_{i,t-2}\Delta y_{i,t-1} &\xrightarrow{p} -\int\frac{\left(1-H\left(\alpha+\rho\right)\right)H\left(\alpha\right)}{1-H\left(\alpha+\rho\right)+H\left(\alpha\right)}g\left(\alpha\right)\,d\alpha, \\
\frac{1}{T}\sum_{t=2}^{T}\Delta y_{it}\Delta y_{i,t-1} &\xrightarrow{p} -\int\left(1-H\left(\alpha+\rho\right)\right)H\left(\alpha\right)g\left(\alpha\right)\,d\alpha, \\
\frac{1}{T}\sum_{t=2}^{T}\left(\Delta y_{i,t-1}\right)^{2} &\xrightarrow{p} -2\int\frac{\left(1-H\left(\alpha+\rho\right)\right)H\left(\alpha\right)}{1-H\left(\alpha+\rho\right)+H\left(\alpha\right)}g\left(\alpha\right)\,d\alpha.
\end{aligned}$$

Notice that the limiting quantities above do not depend on i. Therefore, as $n \to \infty$, we will obtain (4), (5), and (6) after some algebra.

C Derivation of the large-n, large-T limit of the fixed effects estimator

Galvao and Kato (2014) impose assumptions A1 to A3 to derive the large-n, large-T limit of the fixed effects estimator. Assumption A1 is about independence across cross-sectional units and a mild form of time series dependence conditional on α_i. For my case, I needed to impose the assumption that the initial condition is drawn from its stationary distribution conditional on α_i, unlike the derivations for the AH estimators.

Let $\widetilde{y}_{it} = y_{it} - \mathbb{E}\left(y_{it}=1|\alpha_i\right) = y_{it} - \Pr\left(y_{it}=1|\alpha_i\right)$ for $t = 1,\ldots,T$. Assumption A2 is about the existence and boundedness of the moments of $\widetilde{y}_{it}$. These moments are guaranteed to exist and be bounded because $\widetilde{y}_{it}$ has a Bernoulli distribution with probability $\Pr\left(y_{it}=1|\alpha_i\right) \in (0,1)$. They show that the fixed effects estimator converges

[8] Some of the calculations can be found in the Appendix. Note that even with fixed n, the inconsistency is still present.

to the following pseudo-true parameter:

$$\beta_0 = \frac{\mathbb{E}\left(\widetilde{y}_{it}\widetilde{y}_{i,t-1}\right)}{\mathbb{E}\left(\widetilde{y}^2_{i,t-1}\right)}.$$

I now calculate the denominator explicitly. First, note that

$$\begin{aligned}
\widetilde{y}^2_{i,t-1} &= y^2_{i,t-1} - 2y_{i,t-1}\Pr\left(y_{i,t-1}=1|\alpha_i\right) + \left(\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)^2 \\
&= y_{i,t-1} - 2y_{i,t-1}\Pr\left(y_{i,t-1}=1|\alpha_i\right) + \left(\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)^2.
\end{aligned}$$

Taking expectations, we have

$$\begin{aligned}
\mathbb{E}\left(\widetilde{y}^2_{i,t-1}\right) &= \mathbb{E}\left[y_{i,t-1} - 2y_{i,t-1}\Pr\left(y_{i,t-1}=1|\alpha_i\right) + \left(\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(y_{i,t-1}|\alpha_i\right) - 2\mathbb{E}\left(y_{i,t-1}|\alpha_i\right)\Pr\left(y_{i,t-1}=1|\alpha_i\right) + \left(\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)^2\right] \\
&= \mathbb{E}\left[\Pr\left(y_{i,t-1}=1|\alpha_i\right) - \left(\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)^2\right] \\
&= \mathbb{E}\left[\Pr\left(y_{i,t-1}=1|\alpha_i\right)\left(1-\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)\right].
\end{aligned}$$

Note that $\mathbb{E}\left(\widetilde{y}^2_{i,t-1}\right) > 0$ and satisfies assumption A3 of Galvao and Kato (2014). As for the numerator, note that

$$\begin{aligned}
\widetilde{y}_{it}\widetilde{y}_{i,t-1} &= y_{it}y_{i,t-1} - y_{it}\Pr\left(y_{i,t-1}=1|\alpha_i\right) - y_{i,t-1}\Pr\left(y_{it}=1|\alpha_i\right) \\
&\quad + \Pr\left(y_{it}=1|\alpha_i\right)\Pr\left(y_{i,t-1}=1|\alpha_i\right).
\end{aligned} \tag{9}$$

Take the first two terms of the right hand side of (9). Applying law of iterated expectations and $\mathbb{E}\left(y_{it}|y_{i,t-1},\alpha_i\right) = \Pr\left(y_{it}=1|y_{i,t-1},\alpha_i\right)$ gives

$$\begin{aligned}
&\mathbb{E}\left(\left(y_{i,t-1} - \Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)y_{it}\right) \\
&= \mathbb{E}\left[\mathbb{E}\left(\mathbb{E}\left(\left(y_{i,t-1} - \Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)y_{it}|y_{i,t-1},\alpha_i\right)|\alpha_i\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\left(y_{i,t-1} - \Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)\mathbb{E}\left(y_{it}|y_{i,t-1},\alpha_i\right)|\alpha_i\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\left(y_{i,t-1} - \Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)H\left(\alpha_i+\rho y_{i,t-1}\right)|\alpha_i\right)\right] \\
&= \mathbb{E}\left[\left(1-\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)H\left(\alpha_i+\rho\right)\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right]
\end{aligned}$$

$$-\mathbb{E}\left[\Pr\left(y_{i,t-1}=1|\alpha_i\right)H\left(\alpha_i\right)\left(1-\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)\right].$$

The last two terms of the right hand side of (9) is equal to zero. As a result, we obtain

$$\mathbb{E}\left(\widetilde{y}_{it}\widetilde{y}_{i,t-1}\right) = \mathbb{E}\left[\left(H\left(\alpha_i+\rho\right)-H\left(\alpha_i\right)\right)\Pr\left(y_{i,t-1}=1|\alpha_i\right)\left(1-\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)\right].$$

Combining all these findings give us the final form for the pseudo-true parameter:

$$\beta_0=\frac{\mathbb{E}\left[\left(H\left(\alpha_i+\rho\right)-H\left(\alpha_i\right)\right)\Pr\left(y_{i,t-1}=1|\alpha_i\right)\left(1-\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)\right]}{\mathbb{E}\left[\Pr\left(y_{i,t-1}=1|\alpha_i\right)\left(1-\Pr\left(y_{i,t-1}=1|\alpha_i\right)\right)\right]}.$$

D Derivation of the large-n, large-T limit of the GMM estimator

The setup in Okui (2015) is as follows: Consider the situation where a scalar y_{it} has heterogenous means $\eta_i=\mathbb{E}\left(y_{it}|\alpha_i\right)$ and autocovariances $\gamma_{k,i}=\mathbb{E}\left[\left(y_{it}-\eta_i\right)\left(y_{i,t-k}-\eta_i\right)|\alpha_i\right]$. This fits the case of the dynamic LPM. In particular, if we use the results in this appendix, we have

$$\begin{aligned}
\eta_i &= \frac{H\left(\alpha_i\right)}{1-H\left(\alpha_i+\rho\right)+H\left(\alpha_i\right)},\\
\gamma_{0,i} &= \frac{H\left(\alpha_i\right)\left(1-H\left(\alpha_i+\rho\right)\right)}{\left[1-H\left(\alpha_i+\rho\right)+H\left(\alpha_i\right)\right]^2},\\
\gamma_{1,i} &= \frac{H\left(\alpha_i\right)\left(1-H\left(\alpha_i+\rho\right)\right)\left(H\left(\alpha_i+\rho\right)-H\left(\alpha_i\right)\right)}{\left[1-H\left(\alpha_i+\rho\right)+H\left(\alpha_i\right)\right]^2},\\
\gamma_{2,i} &= \frac{H\left(\alpha_i\right)\left(1-H\left(\alpha_i+\rho\right)\right)\left(H\left(\alpha_i+\rho\right)-H\left(\alpha_i\right)\right)^2}{\left[1-H\left(\alpha_i+\rho\right)+H\left(\alpha_i\right)\right]^2}.
\end{aligned}$$

Okui (2015) shows that under his Assumption 2 that the Arellano and Bond (1991) estimator, the GMM estimator based on level moment conditions proposed by Arellano and Bover (1995), and the FE estimator all converge to the same probability limit under sequential asymptotics where first $n\to\infty$ followed by $T\to\infty$. This same probability limit is given by $\mathbb{E}\left(\gamma_{1,i}\right)/\mathbb{E}\left(\gamma_{0,i}\right)$. The resulting limit is exactly the same limit I derived for the FE estimator using the result from Galvao and Kato (2014) but under joint asymptotics.

Interestingly, the probability limit under sequential asymptotics where first $n \to \infty$ followed by $T \to \infty$ for the AH estimator using levels as the instrument set is also the same under sequential asymptotics where first $T \to \infty$ followed by $n \to \infty$. Okui (2015) shows that under his Assumption 2 that the probability limit is given by

$$\begin{aligned}
\frac{\mathbb{E}\left(\gamma_{1,i}-\gamma_{2,i}\right)}{\mathbb{E}\left(\gamma_{0,i}-\gamma_{1,i}\right)} &= \frac{\int \frac{H(\alpha)\left(1-H(\alpha+\rho)\right)}{1-H(\alpha+\rho)+H(\alpha)}\left(H(\alpha+\rho)-H(\alpha)\right) g(\alpha)\, d\alpha}{\int \frac{H(\alpha)\left(1-H(\alpha+\rho)\right)}{1-H(\alpha+\rho)+H(\alpha)} g(\alpha)\, d\alpha} \\
&= \int w_l(\alpha,\rho)\left(H(\alpha+\rho)-H(\alpha)\right) d\alpha,
\end{aligned}$$

which is the same limit I obtain in (5).