Individual Outcomes with Group Effects:
The Impact of Gaming on Employment of American Indians

*

By

Patricia B. Reagan

Department of Economics

Ohio State University

reagan.3@osu.edu

Lung-fei Lee

Department of Economics

Ohio State University

lee.1777@osu.edu

and

Robert J. Gitter

Department of Economics

Ohio Wesleyan University

rjgitter@owu.edu

July 2007

* This research was performed under a pilot project funded by the National Institute of Child Health and Human Development through a grant to The Initiative for Population Research at the Ohio State University. Peihong Feng, Ji Li and Xiaoshu Xu provided excellent research assistance.

Abstract

American Indians living on reservations have substantially lower levels of employment and income than co-resident non-Indians. Gaming has emerged as an attempt by tribes to improve the socio-economic status of their members. This paper uses simulated maximum likelihood to estimate the effect of the tribal decision to open a gaming facility on employment of its members. Unlike difference-in-differences estimators, it does not require the assumption that the time trend in employment is the same across reservations. Opening a casino is associated with a 14 percentage point increase in the probability that the head of household is employed.

I. Introduction

American Indians[1] have a lower standard of living on average than the rest of the American population. Data from the 2000 Census of Population showed that median household income for Indians was 25 percent below the figure for the population as whole.[2] Further, Indians who chose to live on a reservation experience even lower levels of income and employment and higher levels of poverty and unemployment than Indians who reside elsewhere.[3]They also have substantially lower levels of employment and income than co-resident non-Indians. A relatively recent proposed remedy to these long-standing economic hardships has been the introduction of tribal-run gaming, incl. Gaming, which, including bingo, poker, and slot machines as well as large-scale casino gambling. members.[4]

The legality of tribal gaming was unclear until the passage of the Indian Gaming Regulatory Act in 1988. The Act permitted tribal gaming in any state that did not specifically make all gaming within the state illegal. The result was a dramatic increase in the number of gaming facilities theough the 1990s. Many tribes, however, have chosen not to open gaming facilities on their reservations, thus facilitating the estimation of the impact of gaming on various outcomes of interest. There have already been a handful of studies that have attempted to determine the impact of casinos on Indians residing nearby as well as on the local economy. Evans and Topoleski (2002) used a difference-in-difference approach and found that opening a Las Vegas style casino with slot machines and/or table games resulted in increased aggregate employment on reservations and reservation service areas.[5]

---

[1] We have chosen to use the term American Indian rather than Native American. We rely on Census data that uses the term American Indian as well as information from the National Indian Gaming Association.

[2] Median household income in 1999 for the entire US population was $41,994. Median household income was $30,693 for American Indians who belong to one of the 539 tribes designated by the Census. Alaska natives were not included in the calculation of household income for American Indians. See http://factfinder.census.gov/home/aian/sf_aian.html

[3] See Gitter and Reagan (2002.)

[4] These goals were formally stated in the Indian Gaming Regulatory Act of 1988.

[5] Evans and Topoleski also use data on Indians from the Indian Labor Force Report of the U.S. Department of Indian Affairs. Data are presented for employment in the service area of a reservation. In many cases, the number of Indians differs substantially from the number of tribal

However, half of all individuals residing on Indian reservations are non-Indians and those non-Indians have different employment patterns than the co-resident Indians (Gitter and Reagan (2002)). Reagan and Gitter (forthcoming) instead evaluate the impact of gaming on tribal members. They use a 2-stage estimator to control for the endogeneity of the decision to open a gaming facility and cluster errors in the individual outcomes equation at the reservation level.

This paper develops and implements a maximum likelihood estimator that allows for individual outcomes with group effects. The outcome of interest is the employment of householders who self-identify, or whose spouse self-identifies, with the tribe(s) controlling reservation land. The decision by the tribe to open a gaming facility is treated as a group effect. The estimator allows for correlation between the unobservables in the gaming equation and in the individual employment equation. This estimator provides an alternative to the widely used difference-in differences (DID) estimator as a means to evaluate policy interventions where policy is set at a more aggregate level than the outcomes that it affects.

We compare our ML estimates to those obtained from the commonly used DID estimator. Unlike our estimator, DID is based on a linear model, so that the outcomes of tribal members can be aggregated to the sample average outcomes in the two time periods. Consistency of the estimator requires that group and time fixed effects are additive (Angrist and Krueger (1999)). In other words, it requires that tribes that open casinos and those that do not would have had the same trends in employment had the casinos not been opened. Meyer (1995) outlines a number of other reasons why DID estimates might not lead to consistent estimates of treatment on the treated including the possibility that tribes that introduced casinos did so because they anticipated greater returns to casinos than tribes that did not introduce casinos.

The proposed ML estimator, in contrast to the DID estimator, makes no assumption that the time trends in employment would have been the same for tribes that did not open a casino and

members. This is due to the fact that members of other tribes may live in the area. To the extent that these people are not hired by the tribal run gaming facility, this would affect the calculated result of the impact of a tribe opening a casino on measures of economic success of the tribal members.

the counterfactual trend for tribes that did open a casino. Furthermore, the nonlinearity of our ML estimator allows for indirect interactions between the casino and control variables as well as allowing for endogeneity of casino choice in that group effects can change over time (Besley and Case (2000)).

The plan of the paper is as follows. In section 2, we describe the institutional background of reservations and gaming. In section 3, we describe the estimator. Sections 4 and 5, respectively, describe the data and the results. Section 6 contains concluding comments.

II. Background

The American government has changed its policy towards American Indians and reservations several times over the years. The policy of forcibly relocating American Indians to reservations during much of the Nineteenth Century evolved into one where American Indians were free to live where they chose. During the middle part of the last century the federal government encouraged Indians to leave the reservation but recent decades have seen a policy of encouraging tribal self-determination. Currently there are approximately 270 federally recognized American Indian reservations[6] and 562 federally recognized tribes[7] with some additional state recognized reservations and tribes. A few reservations are home to more than one tribe and not all tribes have a reservation.

As the federal government allowed more tribal self-determination, one response to the lack of jobs and income was the opening of a few bingo parlors on reservations in the late 1970s. After a series of court decisions left the legal status of state regulation of gaming unclear,[8] in 1988 the federal government clarified the situation by passing the Indian Gaming Regulatory Act (IGRA.) The Act set up three classes of gaming. Class I covers social games with minimal prizes and is regulated only by the tribes. It has little impact on economic development and is not treated as gaming for our purposes. Class II includes bingo, lotto and pull-tabs and is regulated only by

[6] http://www.infoplease.com/spot/aihmnations1.html

[7] Bureau of Indian Affairs at http://www.doi.gov/bureau-indian-affairs.html.

[8] For background on the key cases of Seminole Tribe of Florida v Butterfield and California v. Cabazon and Morongo Bands of Mission Indians see Evans and Topoleski (2002).

the tribe and federal government. Class III covers all other forms of gaming including casino-style wagering. This type of gaming also requires the tribe to negotiate a compact with the state government and 21 states have set up such a compact.[9] The IGRA gave rise to a rapid increase in the level of Indian gaming, from \$300 million dollars in 1989 to \$15.9 billion in 2003.[10]

Gaming facilities offered a potential avenue to improve the lot of American Indians living on a reservation. Even though gaming might offer a means to increased levels of income and higher employment rates, some tribes decided that the expected costs outweighed the potential benefits and chose not to open a gaming facility. In 1989 immediately after the passage of the IGRA, Indian gaming was in its infancy. Very few tribes had gaming operations and almost all tribes had to wrestle with the decision of whether to open a gaming facility.

III. Econometric Model

The econometric model for our empirical study is a discrete choice model with group decisions and individual outcomes. The individual outcomes are correlated with the group decision. As discrete choices involve latent utility functions, the econometric model shall be specified with latent variables with observed discrete indicators:

$$z_l^* = x_l \alpha + u_l, \qquad l = 1, \cdots, L, \tag{3.1}$$

and

$$y_{li}^* = x_{li}\beta_1 + z_l\beta_2 + v_l + w_{li}, \qquad i = 1, \cdots, n_l, \tag{3.2}$$

where u_l and v_l may be correlated, but w_{li}'s are mutually independent and are independent of u_l and v_l. L is the total number of groups and n_l is the number of observations from the lth group. The binary indicator z_l is 1 when $z_l^* > 0$, 0 otherwise. Similarly, $y_{li} = 1$ if $y_{li}^* > 0$, 0 otherwise.

[9] National Indian Gaming Association (2004), page 5.

[10] 1989 figure from Library of Congress Senate Rpt.104-241 - Amending the Indian Gaming Regulatory Act
http://thomas.loc.gov/cgiin/cpquery/?&db_id=cp104&r_n=sr241.104&sel=TOC_24256&
and 2003 figure from National Indian Gaming Association (2004.)

The disturbances u_l, v_l and w_{li}'s are assumed to be normally distributed; the variances of u_l and w_{li}'s are one (normalization for discrete choices) and the covariance of u_l and v_l is σ_{uv}. These imply that

$$y_{li}^* = x_{li}\beta_1 + z_l\beta_2 + \sigma_{uv}(z_l^* - x_l\alpha) + \epsilon_l + w_{li}, \tag{3.2$'$}$$

where ϵ_l is $N(0, \sigma_v^2 - \sigma_{uv}^2)$ and it is independent of z_l^*. These equations have an error component structure. If z_l^* was observable, this equation would be a binary choice panel setting with error components.

Due to the discrete nature of the observed outcomes and the correlation among group decision and individual outcomes, we consider the method of maximum simulated likelihood estimation for the model. We shall briefly summarized our estimation approach in this section.[11] The Appendix provides more details and compares it to standard DID models.

For each group l, let $\bar{Y}_l = (y_{l1}, \cdots, y_{l,n_l})$ and $\bar{X}_l$ consist of all exogenous variables x_{li} and x_l's. The joint probability of decision of the group and outcomes for members in the group is

$$P(\bar{Y}_l, z_l|\bar{X}_l) = \int\int \prod_{i=1}^{n_l} P(y_{li}|z_l, z^*, \bar{X}_l, \epsilon) f(\epsilon) f(z^*|z_l, \bar{X}_l) d\epsilon dz^* \cdot P(z_l|\bar{X}_l), \tag{3.3}$$

where $f(\epsilon)$ is the density of ϵ, which is $N(0, \sigma_v^2(1-\rho_{uv}^2))$, and

$$\begin{aligned} P(y_{li}|z_l, z_l^*, \bar{X}_l, \epsilon) &= y_{li}\Phi(x_{li}\beta_1 + z_l\beta_2 + \sigma_{uv}(z_l^* - x_l\alpha) + \epsilon_l) \\ &\quad + (1 - y_{li})[1 - \Phi(x_{li}\beta_1 + z_l\beta_2 + \sigma_{uv}(z_l^* - x_l\alpha) + \epsilon_l)]. \end{aligned}$$

This joint probability function can be evaluated by stochastic simulation as

$$P_m(\bar{Y}_l, z_l|\bar{X}_l) = \frac{1}{m}\sum_{j=1}^{m}\prod_{i=1}^{n_l} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, \epsilon_l^{(j)}) \cdot \{z_l\Phi(x_l\alpha) + (1 - z_l)[1 - \Phi(x_l\alpha)]\}, \tag{3.4}$$

where $z_l^{*(j)}$ is a random value generated from the (truncated) density $f(z_l^*|z_l, \bar{X}_l)$ and $\epsilon_l^{(j)}$ is a random value generated from $f(\epsilon)$.

[11] For a recursive simultaneous system with two equations for a cross sectional data, where z_l^* in the first equation is an observable continuous dependent variable, Smith and Blundell (1986) and Rivers and Vuong (1986) have suggested a two-step probit estimation method based on the conditioning equation (3.2)'. Our estimation method is also based on conditioning equations but generalizes to models where the dependent variable of the first equation can be a limited dependent variable instead of a continuous one. For the generalization, simulation methods will be required.

In our application of the method, as some n_l's are large, the product of probabilities over members' outcome probabilities can be small. In order to avoid the possible underflow problem in computation, a recursively weighting schedule originated in Lee (2000) can be adapted. This results in the following simulated log probability

$$\begin{aligned} &\ln P_m(\bar{y}_l, z_l|\bar{X}_l) \\ =& \sum_{i=1}^{n_l} \ln[\frac{1}{m}\sum_{j=1}^{m} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, \epsilon_l^{(j)})\omega_{l,i-1}^{(j)}] + z_l \ln\Phi(x_l\alpha) + (1-z_l)\ln(1-\Phi(x_l\alpha)), \end{aligned} \tag{3.5}$$

where

$$\omega_{li}^{(j)} = \frac{P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, \epsilon_l^{(j)})\omega_{l,i-1}^{(j)}}{\frac{1}{m}\sum_{j=1}^{m} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, \epsilon_l^{(j)})\omega_{l,i-1}^{(j)}}, \tag{3.6}$$

for $i = 1, \cdots, n_l$, with $\omega_{l0}^{(j)} = 1$ for all $j = 1, \cdots, m$ to start the recursion. The $\omega_{li}^{(j)}$'s for each l are weights proportional to m because $\frac{1}{m}\sum_{j=1}^{m}\omega_{li}^{(j)} = 1$ for each $i = 1, \cdots, n_l$.

The log likelihood function for the sample with L groups is

$$\begin{aligned} \ln\mathcal{L} =& \sum_{l=1}^{L}\sum_{i=1}^{n_l} \ln[\frac{1}{m}\sum_{j=1}^{m} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, \epsilon_l^{(j)})\omega_{l,i-1}^{(j)}] \\ &+ \sum_{l=1}^{L}\{z_l \ln\Phi(x_l\alpha) + (1-z_l)\ln(1-\Phi(x_l\alpha))\}. \end{aligned} \tag{3.7}$$

The second component of the log likelihood function represents the decisions of the groups. The correlations of a group's decision and outcomes of its members are captured by the weights ω's. These weights are functions of unknown parameters. Our estimation method is a maximum simulated likelihood approach.[12]

IV. Data

In order to estimate a two-stage model of the tribal decision to open a gaming facility and its subsequent impact on individual local tribal members, we need a data set that has four features: (i) Location and the tribal affiliation or affiliations of all Indian Reservations in the lower 48 states; (ii) Identification of which reservations had gaming facilities as of 2000; (iii) Aggregate

[12] For readers who are interested in simulation estimation methods, the text book by Gouriéroux and Monfort (1996) is a good one to begin with.

demographic and socio-economic status of tribal members living on or near Indian Reservations in 1989, just after the passage of the IGRA in 1988 when the tribal decision to open gaming facilities became widespread; (iv) Data on the individual outcomes and characteristics of local tribal member observed some years after the passage of the IGRA to evaluate the economic consequences of the tribal choices regarding gaming on local tribal members. Although no single conventional data set possesses all of these features, we have constructed one that does.

There is no detailed specific census of American Indians living on reservations.[13] The 1990 and 2000 Public Use Microdata Sample (PUMS) 5% samples, however, provide us with detailed economic and demographic data on individuals. A question on race allows us to identify tribal affiliations of American Indians and the tribe of the householder and spouse. Data from the PUMS does not allow us to identify whether or not the individual lives on a reservation let alone which reservation. Within the data set we can identify the Public Use Microdad Area (PUMA), a geographic area of at least 100,000 people. As we have information on the county in which a reservation is located, we were able to contruct a measure of whether someone lived in a PUMA with a reservation. If there was a reservation in any of the constituent counties, the whole PUMA was treated as having a reservation. If a county had serval PUMAs, we coded the entire county as having a reservation.

The 1990 and 2000 PUMAs were redefined based on population changes. Since we require a cross-walk between location of tribal members in 1990 and 2000, we used the lowest common denominator of PUMA aggregates between the decennial censuses as our common geographic aggregate to which a reservation was assigned. Strictly speaking we identify a PUMA aggregate and identify whether or not a reservation is located in that PUMA aggregate.

The Census does not provide information on the tribal affiliation or affiliations of Indian Reservations. We used various sources to determine the tribe that had the reservation.[14] If

[13] The Bureau of Indian Affairs has employment and population data by tribal reservation and Evans and Topoleski (2002) used this data set in their paper. The data set, however, does not have demographic and income data.

[14] The Census does not assign a specific tribe to a reservation. In some cases, such as the Blackfeet Reservation in Montana, it is easy to determine the tribe that has a reservation. In many cases,

the American Indian race (tribal) code of a householder or spouse from PUMS matched that of a reservation in the PUMA aggregate, then the householder was said to belong to the tribe associated with that reservation and is said to be a tribal member.[15] Our sample is therefore limited to the heads of households who identify, or whose spouse identifies, themselves as belonging to the tribe associated with that reservation. [16]

We used data from the National Indian Gaming Association to identify the location of gaming facilities in the United States along with the tribe that operated the facility (National Indian Gaming Association, 2000). The list of gaming operations includes bingo parlors (Class II facilities) as well as casinos that offer slot machines and or/ table games (Class III). We determined the county in which the facility was located and, based on the location and tribe, assigned it to a reservation and an aggregate PUMA.[17]

There were 103 reservations that had data on tribal members in both the 1990 and 2000 Census.[18] We model the decision by a tribe to open a gaming facility on its reservation as a

---

such as the Big Pine Reservation in California, we had to go to tribal websites and use such sources as Tiller (1995.) Oklahoma has few reservations, per se, but several Tribal Jurisdictional Statistical Areas. We counted these as reservations as well as Tribal Designated Statistical Areas in other states and state reservations with gaming facilities.

[15] An individual is identified as a member of the tribe if they self-report the tribal ethnicity of the tribe of the local reservation. The Decennial Census does not collect data on whether or not a person is a registered member of a particular Indian tribe. Instead it collects data on ethnicity, allowing individuals to select from a number of different Indian tribes. For ease exposition, we misuse the word tribal member to denote ethnic identification with an Indian tribe.

[16] There has been a great deal of publicity about two casinos in Connecticut; Foxwoods run by the Mashantucket Pequots and Mohegan Sun run by the Mohegans. Both of these tribes are smaller than the median size of tribes opening up casinos, but due to their proximity to New York, Boston and other population centers have been able to maintain extremely large and successful casinos. (See Evans and Topoloski, 2002.) It should be noted that these tribes were among the combined tribal codes we were not able to separate out and were excluded from our study. Neither Foxwoods nor Mohegan Sun was included among the gaming facilities in our data.

[17] In some cases a tribe had more than one actual reservation in a PUMA, with some of them having a gaming operation and some not. In these cases we classified the PUMA as having a reservation with a gaming operation.

[18] We excluded reservations located in areas where the Census data only listed Indians as Other Tribe or Tribe Not Specified. For example, if a person was an Indian but not a member of one of the larger tribes, they were coded as Other Tribe or Tribe Not Specified. For these people we could not determine if the respondent was a member of the tribe of that for reservation. For example, both Cree and Assinibone are classified as Other Tribe. If we were to include Other Tribes in our analysis, we would use the characteristics of Cree in examining whether or not the Assinibone

function of the mean socio-economic characteristics of the householders of local tribal members or their spouses with data from the 1990 Census. The variables in the gaming equation include the natural log of 1989 mean per capita household income of local tribal members, the fraction of local tribal members who are unemployed, and an indicator for MSA.[19] Further, we included an indicator variable for Navajo Nation reservations because the Dine have a tradition that discourages gambling.[20]

The means of aggregated data from the 1990 Census used in the gaming facility equation are presented in Table 1. Of the 103 reservations in the sample, more than 76% ($n = 79$) had opened a gaming facility by 1999. Non-gaming reservations had higher mean per capita income and lower unemployment rates than reservations with gaming facilities, although these differences are not statistically significant.[21] Gaming reservations were less likely to be located in metropolitan areas than non-gaming reservation, but the difference was not statistically significant. Almost 30% of the non-gaming reservations were affiliated with the Navajo tribe, whereas 1% of gaming reservations were Navajo.

The outcome sample consists of 17,305 householders who report, or whose spouse reports, the ethnicity of the local reservation. Data are taken from the 2000 PUMS and income levels reflect incomes in 1999. There were 168 observations on average for each reservation, but the standard deviation was large. Twenty-three reservations had less than 20 observations in 2000 and the most populous reservation had 2,231 observations.

The means and definitions of the variables used in the employment and out of the labor force

decided to open up a gaming facility as we cannot distinguish between the two groups with Census data. We excluded observations coded Other Tribes and Tribe Not Specified from our analysis for this reason.

[19] Number unemployed divided by the population, i.e., not the unemployment rate.

[20] According to Navajo legend, gambling once led to their enslavement and gambling is seen as a potential vehicle for re-enslavement. (See Valerie Taliman "Dine rejection of casinos a cultural choice." Indian Country Times December 26, 2002.)

[21] In calculating the mean level of income for tribal members on a reservation, one observation had a mean level of income in 1990 of $47,000 (1999 dollars.) This was due to the small number of tribal members in the sample and along with some of them having a very high level of income. We eliminated this observation and it had virtually no impact on our results.

equations are presented in Table 2. Fifty-six percent of the householders were employed, 6 percent were unemployed and 38 percent were out of the labor force. Sixty-eight percent of the sample lived on a reservation with a gaming facility, although 76 percent of reservations had gaming facilities.[22] Nineteen percent of the sample resided in an MSA. Thirty percent of the sample had less than 12 years of education; 33 percent had 12 years of education or its equivalent; 28 percent had some post-secondary education; only 9 percent were college graduates.

V. Results

Table 3 presents our results on the effect of opening a casino between 1989 and 1999 and individual employment in 1999. The first column presents univariate Probit estimates of single equations for the probability that a tribe opened a casino and individual employment. There is no attempt to control for the endogeneity of casino. The estimated marginal effect of opening a casino on the probability of employment is about 9%.

The second column of Table 3 presents the ML estimates that jointly estimate the probability that a casino was opened and the probability of employment. The ML estimates for tribal decision to open a casino are qualitatively similar to the univariate model. The probability that a casino will be opened increases with the unemployment rate up to 27%, which encompassed the relevant range of the data. Reservations associated with the Navaho Nation were less likely to have opened a casino.

The ML estimates of the employment equation are similar to the univariate estimates, with the exception of the coefficients on college graduation and casino. The univariate results indicated that college graduates were less likely than individuals with a high school education or some college to be employed. The ML estimates imply that the employment probability increases with each of the education categories. The marginal effect of opening a casino on the probability of employment is estimated to be 14%, which is a 50% increase in the estimated effect of casinos on

[22] This is consistent with the fact that by far the most populous reservation was the Navajo reservation in the Four Corners region, which did not have a gaming facility.

employment compared to the univariate model. The downward bias in the univariate model is not surprising given that the estimate of the correlation coefficient between error terms in the casino and employment equations is negative. The tribes that were more likely to open a casino are the ones with members who are less likely to be employed for unobserved reasons. The correlation coefficient is -0.8.

In results not reported, we estimated a probability of unemployment comparable to those estimated for employment and found no statistically significant effects of opening a casino for either the univariate or the ML specification. To confirm that all of the effects of opening a casino are between a reduction of OLF and employment, we present comparable estimates for OLF in Table 4. ML estimates of the casino equation are very similar to those reported in Table 3 and are therefore omitted from Table 4.

The first column of Table 4 presents the results from a univariate Probit model of OLF. The estimated marginal effect opening a casino on the probability of OLF is -8% in the univariate model. The ML estimates suggest that the probability of being out of the labor force decreases by 11% for Indians living on reservations that opened a casino. The estimated correlation between the error terms in the casino and OLF equations is 0.89. Again we find that the univariate estimate of the effect of casinos on labor force status is biased towards zero. Tribal members living on reservations that opened a casino for unobserved reasons are more likely to be OLF for unobserved reasons.

Table 5 presents the DID estimates for the employment and OLF equations. The standard errors on almost all coefficients are greater than those found using ML. The point estimates of the effect of casinos on employment are closer to zero than our ML estimates and statistically insignificant. There are a number of possible reasons for the differences between the two sets of estimates. It is well known that a causal interpretation of the DID results can be undermined by different time trends between reservations with and without casinos. There may also be endogeneity in the casino choice because we found negative correlation between the unobservables in the casino and employment equations. Tribes that expected low future employment rates may have responded with an increased likelihood of establishing a casino. This may tend to bias the DID estimates towards zero. Another source of bias can be omitted interactions between casinos and explanatory variables, which are indirectly accounted for in the nonlinear ML estimator.

VI. Conclusions

American Indians, who live on reservations, have substantially lower levels of employment and income than co-resident non-Indians. Gaming has emerged as an attempt by tribes to improve the socio-economic status of their members. Since passage of Indian Gaming Regulatory Act of 1988, the number of tribes opening gaming facilities has increased. This paper develops and implements a maximum likelihood estimator of the effect of the tribal decision to open a gaming facility on the individual employment outcomes of its members. This estimator provides an alternative to the widely used difference-in-differences (DID) estimator as a means to evaluate policy interventions where policy is set at a more aggregate level than the outcomes that it affects. Unlike DID estimators that are based on aggregation of a linear probability model with strong maintained assumptions about the independence of time and reservation fixed effects, the nonlinear estimator proposed in this paper is based on underlying joint normality of the tribal decision to open a casino and subsequent individual choices regarding labor force status. We found substantial negative correlation between the unobservables in the employment and casino equations, providing evidence that tribes with otherwise low employment rates were more likely to open a casino. Our estimates suggest that opening a casino increases the probability of employment by about 14% and decreases the probability of being out of the labor force by about 8%. Single equation and DID estimates were biased towards zero.

We have motivated the estimator by estimating the effect of the introduction of Indian gaming facilities on the employment of tribal members. However, our estimator can be applied to any situation in which there is a policy decision made at a group level and interest lies in its implications for individual outcomes. DID estimators yield biased estimates of the effects of group decisions when the group decision is related to expectations about future outcomes so that the time trends in outcomes vary systematically by group. Some examples include the effect on African American drop out rates of court ordered school desegregation (Guryan, (2004)) and the effect of state ”no-fault” divorce laws on divorce rates (Friedberg (1998)). In both instances the unobservables governing the timing and introduction of these policies are likely to be correlated with the outcomes.

We have also motivated the estimator using a discrete individual outcome. The approach can be extended to continuous outcomes, where there is reason to believe that the assumptions underlying DID might be violated because group policies are related to group specific time trends. Examples include state anti takeover laws which may increase managerial discretion and their effect on firms' employee compensation (Bertrand and Mullainathan (1999)).

Appendix. Econometric Models and Estimation

A. The Linear Probability Model and DID Estimation

For the model with (3.1) and (3.2), it is possible to have a simple method for the estimation of (3.2) by the method of instrumental variables if a certain linear probability specification is adopted for the disturbances w_{li}. As in the general setting, we assume that the disturbances w_{li} are independent of (u_l, v_l) in a group given all the exogenous variables $x_{l1}, \cdots, x_{ln_l}$ and x_l in the group. Let X_l denote the vector of all exogenous variables in the group l, for simplicity. Even though z_l is endogenous, (3.2) can be specified with a linear probability form if $(-w_{li})$ is assumed to be uniformly distributed with the support $[0, 1]$.

As $y_{li} = 1$ is determined by $y_{li}^* > 0$, which is equivalent to $x_{li}\beta_1 + z_l\beta_2 + v_l > (-w_{li})$, hence

$$P(y_{li} = 1|X_l, z_l, v_l) = x_{li}\beta_1 + z_l\beta_2 + v_l.$$

Note that this probability is conditional on the endogenous variable z_l as well as v_l; so it is a well defined probability.

Define the error $\eta_{li} = y_{li} - (x_{li}\beta_1 + z_l\beta_2 + v_l)$. Thus, we can write the linear probability equation for y_{li} as

$$y_{li} = x_{li}\beta_1 + z_l\beta_2 + v_l + \eta_{li}.$$

Because

$$\begin{aligned} E(\eta_{li}|X_l, z_l, v_l) &= E(y_{li}|X_l, z_l, v_l) - (x_{li}\beta_1 + z_l\beta_2 + v_l) \\ &= P(y_{li} = 1|X_l, z_l, v_l) - (x_{li}\beta_1 + z_l\beta_2 + v_l) \\ &= 0, \end{aligned}$$

it follows that $E(\eta_{li}|X_l) = 0$, i.e., η_{li} has zero mean conditional on all exogenous variables. However, η_{li}'s have heteroscedastic variances. Because η_{li} takes only two distinct values, conditional on X_l, z_l and v_l, $var(\eta_{li}|X_l, z_l, v_l) = P(y_{li} = 1|X_l, z_l, v_l)(1 - P(y_{li} = 1|X_l, z_l, v_l))$. Hence,

$$\begin{aligned} var(\eta_{li}|X_l) &= E[var(\eta_{li}|X_l, z_l, v_l)|X_l] + var[E(\eta_{li}|X_l, z_l, v_l)|X_l] \\ &= E[(x_{li}\beta_1 + z_l\beta_2 + v_l)(1 - (x_{li}\beta_1 + z_l\beta_2 + v_l))|X_l], \end{aligned}$$

which is a function of x_{li} and x_l, so the variance of η_{li} conditional on exogenous variables is heteroscedastic across members in a group.

Define the overall disturbance $\epsilon_{li} = v_l + \eta_{li}$ of a member in a group. Then, the linear probability equation for y_{li} is

$$y_{li} = x_{li}\beta_1 + z_l\beta_2 + \epsilon_{li}.$$

As ϵ_{li} has the component v_l, ϵ_{li} is correlated with z_l, so z_l remains as an endogenous variable in the above linear probability model.

The outcomes of individuals in a group can be aggregated to obtain the sample average outcome of the group with the linear probability model. The difference in difference approach requires sample average outcomes with at least two time periods.

For a period t, denote $\bar{y}_{lt} = \frac{1}{n_l}\sum_{i=1}^{n_l} y_{li,t}$ where $y_{li,t}$ is the outcome y of an individual i in the group l at time t. Similarly, groups averages $\bar{x}_{lt}$, etc., are defined . At each t, the individual outcome equations can be aggregated into

$$\bar{y}_{lt} = \bar{x}_{lt}\beta_1 + z_{lt}\beta_2 + v_l + u_t + \bar{\eta}_{lt},$$

for $l = 1, \cdots, L$; $t = 1, \cdots, T$, where u_t is a time dummy variables invariant across groups. The time dummy captures the macroeconomic environment at each time period. So there are group fixed effects as well as time fixed effects in this equation. The equations with various groups and time periods form a panel data model. As z_{lt} may be endogenous, it can be correlated with both v_l and u_t but is uncorrelated with $\bar{\eta}_{lt}$ as z_{lt} is uncorrelated with $\eta_{li,t}$ for all i.

As the source of endogeneity of z_{lt} is its correlations with both v_l and u_t, the difference in difference method is to eliminate both the group and time fixed effects, so that the remaining equations can be estimated by the simple least square method. Define the average outcome across all groups at time t as $\bar{\bar{y}}_{.t} = \frac{1}{L}\sum_{l=1}^{L}\bar{y}_{lt}$, the average outcome of a group across time as $\bar{\bar{y}}_{l.} = \frac{1}{T}\sum_{t=1}^{T}\bar{y}_{lt}$, and the overall average is $\bar{\bar{y}}_{..} = \frac{1}{LT}\sum_{t=1}^{T}\sum_{l=1}^{L}\bar{y}_{lt}$. Similarly, averages are defined for other variables and disturbances. By eliminating the fixed effects, we have the within equation:

$$(\bar{y}_{lt} - \bar{\bar{y}}_{.t} - \bar{\bar{y}}_{l.} + \bar{\bar{y}}_{..}) = (\bar{x}_{lt} - \bar{\bar{x}}_{.t} - \bar{\bar{x}}_{l.} + \bar{\bar{x}}_{..})\beta_1 + (z_{lt} - \bar{z}_{.t} - \bar{z}_{l.} + \bar{\bar{z}}_{..})\beta_2 + (\eta_{lt} - \bar{\eta}_{.t} - \bar{\eta}_{l.} + \bar{\bar{\eta}}_{..}),$$

for $l = 1, \cdots, L$; $t = 1, \cdots, T$. The β's can then be estimated by the least squares applied to the within equation. This method is known to be exactly the same as the least squares estimation applied to the panel data by estimating jointly all the fixed effects v_l's and u_t's, and β's.

The difference in difference approach has its original intuition from the case with two time periods, say, the first and second periods. In the first period, the policy z_{l1} is not available, $z_{l1} = 0$ for all l, but the policy decision of z_{l2} is operational in the second period. With two periods, the fixed effect v_l for each group l can be eliminated by taking difference across time as

$$(\bar{y}_{l2} - \bar{y}_{l1}) = (\bar{x}_{l2} - \bar{x}_{l1})\beta_1 + z_{l,2}\beta_2 + (u_2 - u_1) + (\bar{\eta}_{l2} - \bar{\eta}_{l1}), \quad l = 1, \cdots, L.$$

The remaining time effects can then be eliminated by taking the difference of the difference from the group average, which gives

$$(\bar{y}_{l2} - \bar{y}_{l1}) - (\bar{\bar{y}}_{.2} - \bar{\bar{y}}_{.1}) = [(\bar{x}_{l2} - \bar{x}_{l1}) - (\bar{\bar{x}}_{.2} - \bar{\bar{x}}_{.1})]\beta_1 + (z_{l2} - \bar{\bar{z}}_{.2})\beta_2 + [(\bar{\eta}_{l2} - \bar{\eta}_{l1}) - (\bar{\bar{\eta}}_{.2} - \bar{\bar{\eta}}_{.1})],$$

for $l = 1, \cdots, L$. The difference in difference estimate is the least squares estimate applied to this equation. This is exactly the fixed effects least squares estimate because $\bar{y}_{l2} - \bar{\bar{y}}_{.2} - \bar{\bar{y}}_{l.} + \bar{\bar{y}}_{..} = \frac{1}{2}[(\bar{y}_{l2} - \bar{y}_{l1}) - (\bar{\bar{y}}_{.2} - \bar{\bar{y}}_{.1})]$.

For the aggregate equation with the linear probability specification, the variance of $\bar{\eta}_{lt}$ will, in general, be heteroscedastic. As $var(\bar{\eta}_{lt}|X_l) = \frac{1}{n_l^2}\sum_{i=1}^{n_l}\sum_{j=1}^{n_l} E(\eta_{lit}\eta_{ljt}|X_l)$, $E(\eta_{lit}\eta_{ljt}|X_l) = 0$ whenever $i \neq j$, and $var(\bar{\eta}_{lit}|X_l) = E[(x_{li}\beta_1 + z_l\beta_2 + v_l)(1 - x_{li}\beta_1 - z_l\beta_2 - v_l)|X_l]$, one has

$$var(\bar{\eta}_{nt}|X_l) = \frac{1}{n_l^2}\sum_{i=1}^{n_l} E[(x_{li}\beta_1 + z_l\beta_2 + v_l)(1 - x_{li}\beta_1 - z_l\beta_2 - v_l)|X_l].$$

These variances are heteroscedastic because n_l and X_l are in general different for different l.[23]

B. The Simulated Likelihood and the Probit Model

In the subsections B.1 and B.2, the simulation approach will be presented in its generality for any parameter distributions of the disturbances. The subsection B.3 provides the specific case where the disturbances are normally distributed.

[23] For nonlinear models, the D-D method can be complicated. For some recent progress on nonlinear models, see Athey and Imbens (2006).

B.1 Likelihood and Simulated Likelihood Functions of the Model

Because w_{li}'s are mutually independent in the model consisting of (3.1) and (3.2),

$$P(\bar{Y}_l|z_l, z_l^*, \bar{X}_l, v_l) = \prod_{i=1}^{n_l} P(y_{li}|z_l, z_l^*, \bar{X}_l, v_l),$$

where $\bar{Y}_l = (y_{l1}, \cdots, y_{l,n_l})$. Thus,

$$P(\bar{Y}_l|z_l, z_l^*, \bar{X}_l) = \int \prod_{i=1}^{n_l} P(y_{li}|z_l, z_l^*, \bar{X}_l, v_l) f(v_l|z_l, z_l^*, \bar{X}_l) dv_l,$$

where $f(v_l|z_l, z_l^*, \bar{X}_l)$ is the density of v conditional on $z_l, \bar{X}_l$ and z_l^*.

It follows

$$\begin{aligned} P(\bar{Y}_l, z_l|\bar{X}_l) &= P(\bar{Y}_l|z_l, \bar{X}_l) P(z_l|\bar{X}_l) = \int P(\bar{Y}_l|z_l, z_l^*, \bar{X}_l) f(z_l^*|z_l, \bar{X}_l) dz_l^* \cdot P(z_l|\bar{X}_l) \\ &= \int\int \prod_{i=1}^{n_l} P(y_{li}|z_l, z_l^*, \bar{X}_l, v_l) f(v_l|z_l, z_l^*, \bar{X}_l) f(z_l^*|z_l, \bar{X}_l) dv_l dz_l^* \cdot P(z_l|\bar{X}_l). \end{aligned}$$

With multiple integrals, this joint probability function can be evaluated by stochastic simulation. The $P(\bar{Y}_l, z_l|\bar{X}_l)$ can be estimated by

$$P_m(\bar{Y}_l, z_l|\bar{X}_l) = \frac{1}{m} \sum_{j=1}^{m} \prod_{i=1}^{n_l} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, v_l^{(j)}) \cdot \{z_l F(x_l\alpha) + (1 - z_l)[1 - F(x_l\alpha)]\},$$

where m is the number of simulated draws, $z_l^{*(j)}$ is generated from the density $f(z_l^*|z_l, \bar{X}_l)$, $v_l^{(j)}$ is a random value generated from $f(v_l|z_l^{*(j)}, \bar{X}_l)$, and $F(x_l\alpha)$ is the probability of $z_l = 1$ given x_l.

B.2 A Problem with Large n_l

In this case, in order to avoid the possible underflow problem in the evaluation of $P_m(\bar{Y}_l, z_l|\bar{X}_l)$, the following recursively defined weighting scheme is useful. Define

$$S_{li} = \frac{1}{m} \sum_{j=1}^{m} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, v_l^{(j)}) \omega_{l,i-1}^{(j)},$$

where

$$\omega_{li}^{(j)} = P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, v_l^{(j)}) \omega_{l,i-1}^{(j)} / S_{li},$$

for $i = 1, \cdots, n_l$, with $\omega_{l0}^{(j)} = 1$ for all $j = 1, \cdots, m$ to start the recursion.

The $\omega_{li}^{(j)}$'s for each l are weights proportional to m because $\frac{1}{m}\sum_{j=1}^{m}\omega_{li}^{(j)} = 1$ for each $i = 1, \cdots, n_l$. Then, it follows by construction that

$$\frac{1}{m}\sum_{j=1}^{m}\prod_{i=1}^{n_l} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, v_l^{(j)}) = \prod_{i=1}^{n_l} S_{li}.$$

Hence,

$$\begin{aligned}
&\ln P_m(\bar{y}_l, z_l|\bar{X}_l) \\
&= \ln(\{\frac{1}{m}\sum_{j=1}^{m}\prod_{i=1}^{n_l} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, v_l^{(j)})\}\{z_l F(x_l\alpha) + (1-z_l)[1-F(x_l\alpha)]\}) \\
&= \sum_{i=1}^{n_l} \ln S_{li} + z_l \ln F(x_l\alpha) + (1-z_l)\ln(1-F(x_l\alpha)) \\
&= \sum_{i=1}^{n_l} \ln[\frac{1}{m}\sum_{j=1}^{m} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, v_l^{(j)})\omega_{l,i-1}^{(j)}] + z_l \ln F(x_l\alpha) + (1-z_l)\ln(1-F(x_l\alpha)).
\end{aligned}$$

Remark: $\ln\{z_l F(x_l\alpha) + (1-z_l)[1-F(x_l\alpha)]\} = z_l \ln F(x_l\alpha) + (1-z_l)\ln[1-F(x_l\alpha)]$ because z_l is a dichotomous variable.

B.3 The Probit Model and Simulations

When the disturbances u_l, v_l and w_{li}'s are normally distributed with the variances of u_l and w_{li}'s being one under normalization, the model with (3.1) and (3.2) implies the structure in (3.1) and (3.2)$'$

$$z_l^* = x_l\alpha + u_l, \quad y_{li}^* = x_{li}\beta_1 + z_l\beta_2 + \sigma_{uv}(z_l^* - x_l\alpha) + \epsilon_l + w_{li},$$

where σ_{uv} is the covariance of u_l and v_l, and ϵ_l is $N(0, \sigma_v^2 - \sigma_{uv}^2)$. The $(z_l^* - x_l\alpha)$ in (3.2)$'$ represents a controlled variable, which captures the correlation of the group decision and individual outcomes. The (3.2)$'$ has an error component structure. If z_l^* was observable, this equation would be a binary choice panel setting with error components.

Conditional on z_l^*, ϵ_l, all exogenous variables $\bar{X}_l$,

$$\begin{aligned}
P(y_{li}|z_l, z_l^*, \bar{X}_l, \epsilon_l) &= y_{li}\Phi(x_{li}\beta_1 + z_l\beta_2 + \sigma_{uv}(z_l^* - x_l\alpha) + \epsilon_l) \\
&\quad + (1-y_{li})[1 - \Phi(x_{li}\beta_1 + z_l\beta_2 + \sigma_{uv}(z_l^* - x_l\alpha) + \epsilon_l)],
\end{aligned}$$

where Φ denotes the standard normal distribution function. The log likelihood function of the model follows from those in B.1 and B.2 with $P(y_{li}|z_l, z_l^*, \bar{X}_l, v_l)$ replaced by $P(y_{li}|z_l, z_l^*, \bar{X}_l, \epsilon_l)$.

For the simulated likelihood, $z_l^{*(j)}$ will be drawn as a truncated normal variable vs the equation $z_l^* = x_l\alpha + u_l$ given z_l and x_l; and $\epsilon_l^{(j)}$ is generated as a $N(0, \sigma_v^2 - \sigma_{uv}^2)$ variable.

Consider a general procedure for generating a truncated normal variable z^* given $z = 1$ and x for $z^* = x\alpha + u$ where u is $N(0, 1)$ and $z = 1$ if $z^* > 0$; 0, otherwise.

As $z^* = x\alpha + u$, $z^* > 0$ is equivalent to $u > -x\alpha$. So, given $z = 1$, u is a truncated $N(0, 1)$ on $[-x\alpha, \infty)$, of which its distribution F_1 is

$$F_1(u) = \frac{\Phi(u) - [1 - \Phi(x\alpha)]}{\Phi(x\alpha)} = 1 - \frac{1 - \Phi(u)}{\Phi(x\alpha)}.$$

Denote $v = F_1(u)$. It follows that $v = 1 - \frac{1-\Phi(u)}{\Phi(x\alpha)}$. Hence, $\Phi(u) = [1 - (1 - v)\Phi(x\alpha)]$, and, in turn, u can be solved as a function of v:

$$u = \Phi^{-1}[1 - (1 - v)\Phi(x\alpha)] = \Phi^{-1}[1 - v^*\Phi(x\alpha)],$$

where $v^* = 1 - v$ can be taken as a uniform random variable on $[0, 1]$. So generate a uniform random variable (from computer) v^* and the computed $z^* = x\alpha + \Phi^{-1}[1 - v^*\Phi(x\alpha)]$ will be a truncated normal with the truncated normal density conditional on $z = 1$.

On the other hand, given $z = 0$, u is a truncated $N(0, 1)$ on $(-\infty, -x\alpha]$, of which its distribution F_2 is $F_2(u) = \frac{\Phi(u)}{\Phi(-x\alpha)}$ for $u \leq -x\alpha$. Denote $v^* = F_2(u)$. It follows that u can be solved as $u = \Phi^{-1}[v^*\Phi(-x\alpha)]$, where v^* can be taken as a uniform random variable on $[0, 1]$. So generate a uniform random variable (from computer) v^* and the computed $z^* = x\alpha + \Phi^{-1}[v^*\Phi(-x\alpha)]$ will be a truncated normal variable with the truncated normal density conditional on $z = 0$.

Summary

(i) Conditional on $z = 1$, $z^* = x\alpha + \Phi^{-1}[1 - v\Phi(x\alpha)]$; or

(ii) conditional on $z = 0$, $z^* = x\alpha + \Phi^{-1}[v\Phi(-x\alpha)]$, where $v \sim U[0, 1]$ a uniform random variable;

(iii) simulate ϵ_l as $N(0, \sigma_v^2 - \sigma_{uv}^2)$.

For the lth group with observations $(\bar{y}_l, z_l, \bar{X}_l)$, the simulated likelihood is

$$P_m(\bar{y}_l, z_l|\bar{X}_l) = \sum_{i=1}^{n_l} \ln[\frac{1}{m}\sum_{j=1}^{m} P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, \epsilon_l^{(j)})\omega_{l,i-1}^{(j)}] + z_l \ln \Phi(x_l\alpha) + (1 - z_l)\ln(1 - \Phi(x_l\alpha)),$$

where

$$P(y_{li}|z_l, z_l^{*(j)}, \bar{X}_l, \epsilon_l^{(j)}) = y_{li}\Phi(x_{li}\beta_1 + z_l\beta_2 + \sigma_{uv}(z_l^{*(j)} - x_l\alpha) + \epsilon_l^{(j)})$$
$$+ (1 - y_{li})[1 - \Phi(x_{li}\beta_1 + z_l\beta_2 + \sigma_{uv}(z_l^{*(j)} - x_l\alpha) + \epsilon_l^{(j)})].$$

REFERENCES

Angrist, Joshua D. and Alan B. Kreuger (1999) "Empirical Strategies in Labor Economics" in *Handbook of Labor Economics* vol. 3a, edited by O. Ashenfelter and D. Card, Amsterdam: Elsevier pp.1277-1366.

Athey, Susan and Guido W. Imbens (2006). "Identification and Inference in Nonlinear Difference-In-Differences Models." *Econometrica* 74, 431-498.

Baltagi, Badi H (1995). *Econometric Analysis of Panel Data*, New York: John Wiley and Sons.

Bertrand, Marianne and Sendhil Mullainathan (1999). "Is There Discretion in Wage Setting? A Test Using Takeover Legislation." *RAND Journal of Economics* 30, 535-554.

Besley, Timothy and Anne Case (2000). "Unnatural Experiments? Estimating the Incidence of Endogenous Policies." *Economic Journal* 110, F672-F694.

Evans, William N and Julie H. Topoleski (2002)."The Social and Economic Impact of Native American Casinos. National Bureau of Economic Research, Working Paper 9198.

Friedberg, Leora (1998). "Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data." *American Economic Review* 88, 608-627.

Gitter, Robert J. and Patricia B. Reagan (2002). "Reservation Wages: An Analysis of the Employment of American Indian Men. *American Economic Review* 92, 1160-1168.

Gouriéroux, C., and A. Monfort (1996). *Simulation-based Econometric Models*, Oxford: Oxford University Press.

Guryan, Jonathan (2004). "Desegregation and Black Dropout Rates" *American Economic Review* 94, 919-943.

Hsiao, Cheng (1986). *Analysis of Panel Data*, Cambridge: Cambridge University Press.

Lee, Lung-fei (2000). "A Numerical Stable Quadrature Procedure For The One-Factor Random-Component Discrete Choice Model." *Journal of Econometrics* 95, 117-129.

Meyer, Bruce D. (1995). "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics* 13, 151-161.

National Indian Gaming Association (2004). Indian Gaming: Final Impact Analysis, http://www.indiangaming.org/index.shtml

National Indian Gaming Association (2000). Indian Gaming Resource Directory, Washington, DC: National Indian Gaming Association.

Reagan, Patricia B. and Robert J. Gitter (2007). "Is Gaming the Optimal Strategy? The Impact of Gaming Facilities on the Income and Employment of American Indians, *Economics Letters* 95, 428-432.

Rivers, D., and Q. Vuong (1988). "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39, 347-366.

Smith, R., and R. Blundell (1986). "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply." *Econometrica* 54, 679-685.

Tiller, Veronica E. Velarde (1995). *Tillers Guide to Indian Country: Economic Profiles of American Indian Reservations.* Albuquerque: BowArrow Publishing Company.

Table 1.

1990 Mean Characteristics of Tribal Members Living in MSA or PUMA
of Tribal Reservation Stratified by Opening of Gaming Facility-
Unit of Observation is Reservation
(standard deviations in parentheses) [t-test in square brackets]

Variable	Variable Definition	Non-Gaming Reservation Mean	Gaming Reservation Mean	Difference in Means [t-test]
Pchhinc89	Mean Per Capita Household Income in 10,000 ($1999)	1.30 (0.79)	1.13 (0.46)	0.17 [1.32]
Unemp89	Average unemployment rate in MSA or county group of reservation in 1989 (source BLS local area statistics)	5.91 (7.39)	9.43 (9.11)	-3.52 [-1.73]
Navajo	=1 if Navajo Reservation =0 if not	0.29 (0.46)	0.01 (0.11)	0.28 [4.93]
MSA	=1 if reservation located in MSA =0 if not	0.41 (0.50)	0.29 (0.45)	0.12 [1.15]
Sample Size		24	79	

Source: 1990 Census IPUMS 5% sample, aggregated to PUMA or MSA of location of tribal reservation

Table 2.

Characteristics of Householder in 2000
Householder or spouse identifies ethnically with local reservation tribe
(standard deviations in parentheses)

Employed	=1 if householder is employed in 2000 =0 otherwise	0.56 (0.49)
Unemployed	=1 if householder is unemployed in 2000 =0 otherwise	0.06 (0.24)
Olf	=1 if householder is out of labor force =0 otherwise	0.38 (0.49)
Gaming	=1 if reservation has gaming facility =0 otherwise	0.68 (0.46)
MSA	=1 if reside near reservation located in MSA =0 otherwise	0.19 (0.39)
Male	=1 if householder is male =0 otherwise	0.47 (0.50)
Msp	=1 if householder is married with spouse present =0 otherwise	0.66 (0.47
Disabled	=1 if householder is disabled =0 otherwise	0.17 (0.37)
High	=1 if highest grade completed by householder is 12 =0 otherwise	0.33 (0.47)
Some	=1 if highest grade completed by householder is between 13 and 15 =0 otherwise	0.36 (0.48)
Colgrad	=1 if highest grade completed by householder is 16 or more =0 otherwise	0.06 (0.24)
Age	Age of householder	46.05 (15.26)

Source: 2000 Census IPUMS 5% sample, unit of observation is householder; householder or spouse self identifies with tribe owning reservation in proximity to residence; Sample size=17,305; The mean number of individuals observed on each reservation is 168, ranging from a low of 2 to a high of 2231. The quintiles of the distribution of observations across the reservations occurred at 12, 46, 95 and 220. The mean number of observations in each quintile was 7.2, 32.8, 65.6, 146.8 and 606.6.

Table 3.

Comparison Between Univariate Probit and Maximum Likelihood Estimates
Employment Equation with Endogenous Regressor

	Univariate Estimates	ML Estimates
CASINO		
Constant	0.83 (0.47)	0.40 (0.42)
Mean Per Capita HH Income 1989	-0.26 (0.28)	0.16 (0.26)
Unemployment Rate 1989	9.34 (4.54)	7.91 (4.33)
Unemployment Rate Squared 1989	-17.64 (10.28)	-14.57 (10.33)
Navaho	-2.38 (0.62)	-2.32 (0.46)
Metropolitan Area	-0.12 (0.34)	0.001 (0.33)
EMPLOYED		
Constant	-1.78 (0.19)	-1.89 (0.12)
Age	0.08 (0.01)	0.09 (0.01)
Age Squared/100	-0.11 (0.01)	-0.12 (0.01)
Male Head of Household	0.31 (0.05)	0.29 (0.02)
Married Spouse Present	0.11 (0.02)	0.08 (0.02)
High School Graduate Only	0.44 (0.05)	0.32 (0.03)
Some College	0.77 (0.07)	0.60 (0.03)
College Graduate	0.37 (0.04)	0.99 (0.05)
Disabled	-0.03 (0.08)	-0.05 (0.03)
Metropolitan Area	0.04 (0.07)	0.05 (0.05)
Casino	0.24 (0.09)	0.35 (0.07)
Sigma		0.23 (0.03)
Rho		-0.80 (0.18)

Note: Single equation estimates are based on Probit models, with standard errors clustered at the reservation. The unemployment rate and its square are jointly significant in estimates.

Table 4.

Comparison Between Univariate Probit and Maximum Likelihood Estimates
Out of Labor Force Equation with Endogenous Regressor

	Univariate Estimates	ML Estimates
Constant	1.19 (0.18)	1.22 (0.12)
Age	-0.07 (0.01)	-0.08 (0.01)
Age Squared/100	0.11 (0.01)	0.11 (0.00)
Male Head of Household	-0.42 (0.04)	-0.41 (0.021)
Married Spouse Present	-0.02 (0.03)	-0.01 (0.02)
High School Graduate Only	-0.42 (0.04)	-0.31 (0.03)
Some College	-0.72 (0.05)	-0.58 (0.03)
College Graduate	-0.35 (0.05)	-0.95 (0.05)
Disabled	0.13 (0.07)	0.15 (0.03)
Metropolitan Area	0.01 (0.02)	-0.03 (0.05)
Casino	-0.22 (0.07)	-0.28 (0.07)
Sigma		0.17 (0.02)
Rho		0.89 (0.10)

Note: Univariate estimates are based on Probit models with standard errors clustered at the reservation.

Table 5.

Difference-in-Differences Estimates
Employed and Out of Labor Force Equations

	EMPLOYED	OUT OF LABOR FORCE
Constant	0.01 (0.01)	-0.01 (0.01)
Age	0.04 (0.02)	-0.03 (0.01)
Age Squared/100	-0.04 (0.02)	0.04 (0.01)
Male Head of Household	-0.34 (0.14)	0.19 (0.12)
Married Spouse Present	0.16 (0.08)	-0.06 (0.07)
High School Graduate Only	0.28 (0.14)	-0.19 (0.12)
Some College	0.35 (0.14)	-0.18 (0.12)
College Graduate	-0.11 (0.17)	0.06 (0.15)
Disability	-0.48 (0.11)	0.42 (0.10)
Casino	0.02 (0.014)	-0.01 (0.01)

Note: To correct for heteroscedasticity DID estimates use analytic weights based on number of individuals on each reservation in 2000.