Dylan McKenna
Mat Olson
Alec Recinos

What are the Determinants of State Lottery Revenue?

(1) Introduction

Lotteries are a way for governments to raise funds for various projects and programs, same as taxes. In the literature, lotteries are frequently analyzed as *voluntary taxation* and found to be particularly regressive at their inception and possibly onwards (Miyazaki et al. 1998, Wyett 1991). Since lottery revenues are meant to bolster state tax revenues and lotteries have the potential to be more regressive than conventional income, property, or sales taxes, there is good reason to analyze state lotteries to see what characteristics generate more revenue.

Using a dataset from Professor Jon Rork, we look into the determinants of state lottery revenues per capita. Our state characteristic regressors fall into four categories: lottery characteristics, fiscal controls, political controls, and demographic information. We have a strongly balanced panel across the 48 states (excluding Alaska and Hawaii) over the 35 years from 1967 to 2002. We estimate a fixed effects model for the panel data, since what we would like to find out is how each of our dependent variables generally affects state lottery revenue. Fixed effects will allow us to estimate the general effect, as it provides each state with the same coefficient. Differences between state lottery revenues are then relegated to the constant intercept term.

(2) Modeling Method

We give credit to Alm et al. for their 1993 paper examining what factors affect the probability that a state would create a lottery—some factors incorporated into their hazard model, including the political controls dummies, are used in our model under similar justifications.

We assume that each state's lottery revenues are a function of state-level characteristics (X_t):

$$LotteryRevenue_{it} = X_t + \tau_t + \psi_t + \varepsilon_{it} \tag{1}$$

where $LotteryRevenue_{it}$ represents the lottery revenues from state *i* at time *t*, τ_t is a year fixed effect, ψ_i is a state fixed effect, and ε_{it} is the error term ~ N(0, 1).

There are certain state characteristic variables that we cannot safely assume enter the equation in year *t* without introducing the problem of endogeneity. These variables are lottery payout (the percentage of prizes *paid* to recipients) and administration costs. We theorize that payout should have an effect on lottery participation and thus revenue, and in addition to our endogeneity concern it shouldn't be included in year *t* because complete payout information on that year cannot influence a lottery participant's decision until at least year *t+1*. Thus, we include a single lag of payout and administration costs for year *t* and one lag:

$$LotteryRevenue_{it} = X_t + Payout_{t-1} + Admin_t + Admin_{t-1} + \tau_t + \psi_t + \varepsilon_{it} \quad (2)$$

2.1 State Characteristic Regressors

Selecting variables for inclusion on the right hand side of the equation was not a simple task. There are many variables that could affect a state's lottery revenues, and our full dataset included dozens of variables for all the states, from 1967 to 2002. Confronted with this wealth of information, we were especially careful to pick categories of variables we were interested in, and then chose within those categories while careful to avoid problems like collinearity.

Lottery Characteristics: The lottery component of our dataset includes sales figures for various types of game, prizes and payout information, administration costs, and how much money went into other state funds. Prizes are different from payouts—with most or all large jackpots, winners are given two collection options. First, they can choose to receive a lump sum payment at the time of winning, lower than the publicized jackpot total. The alternative is annuity payouts: smaller checks received once a month over several months. While we do not have lump sum versus annuity information for big jackpots in our dataset, we do think it is more appropriate to use (a lag of) payout as opposed to prizes because it is common knowledge that bigger lottery jackpots are not as large advertised. Large lottery winnings are also subject to federal withholding and state taxes, information that lottery participants are also assumed to have. We run a parallel regression using prizes as a kind of validity check.
We used the sales figures variables to create dummy variables for each game type where dummies were not already included in the dataset and used the dummies in our regression. We opted to use the dummies rather than the sales figures for a number of reasons. First, we thought it would be more pertinent to examine the effect of a game's *presence* on total lottery revenue than to look at a game's popularity as captured by sales. Second, the variable *yrslott* for number of years with an active state lottery works in a similar manner. In addition to the specific game dummies we included a dummy for whether or not the state had a daily drawing game.

Fiscal Controls: Since lotteries function to bolster state tax revenues, some information about a state's taxing must be included as controls. We include the sales tax rate, the top indices marginal tax rate, and how much federal aid the state receives. We do not include a variable for how much money from the lottery goes to state funds. A state would discontinue its lottery if it were not worth operating (no states remove their lottery in our dataset) and it seems highly unlikely that knowing how much lottery revenue aids the state would influence the behavior of a lottery participant (there are much more efficient ways to support state endeavors if that was the primary motivator for playing).

Political Controls: Lotteries are seen to function as a regressive tax, as the poor spend a disproportionately higher amount of their income on lottery tickets. Regressive policies tend to be decried by the left and accepted by the right. Due to the regressive nature of the lottery, we decide to include variables that would account for the political regime of the state. Following the methods of Rork and Alm, we include whether all three state political powers (governor's office, senate, and house) were controlled by the republicans or the democrats, and whether it was an election year. Having the state controlled politically by one party might make the lottery more or less successful, depending on the current acceptance of regressive policies. Alternatively, if one party controlled all three offices, they might overlook the lottery and focus elsewhere, since they are able to pursue their goals in any field. We expect the election year variable to have little to no effect, but still included it because if an election were taking place one year, it might bring the lottery into focus and cause a change in policy.

Demographic Information: Demographic information was included so that we could account for the actual players of the lottery and purchasers of the games. Not all people are equally likely to play the lottery, so we needed to investigate and control for the different demographics between the states. We included in our regression the percent of the state population that was between 25 and 64 (the age range most likely to play the lottery) (Barnes et al. 2010), the percent of the population that had graduated high school, the percent of the population that had graduated college, and the unemployment rate. We did not include any variables that accounted for minority populations. We concluded that likelihood to play the lottery was largely class, not race, related and so any race variables would likely have an effect solely through their proxy for class.

(3) Results and Checks

The results of our regression are presented below, in Table 1. Our selected final regression is the first, which uses a lag of lottery payout as opposed to lottery prizes. The reasoning behind our selection, that it is common knowledge that lottery payouts are not as large as the advertised price and that using prizes creates some distortion, was previously explained. The distortion can perhaps most clearly be seen in the difference in the coefficients of *dpbsales*. In the first regression, the coefficient is negative and significant at the 5% level. This is largely due to historical reasons. Powerball was first created as a joint lottery among small states who could not afford to offer their own large jackpot. As a result, the coefficient is negative because it captures the differences between small and large states. We corrected for this as best we could by using per capita lottery revenue, instead of gross lottery revenue, but the effect still lingered. The coefficient is positive, but both economically and statistically insignificant in the second regression. Using the advertised prize amount removes the differences between the states, but by doing so obscures that information.

Table 1: Results of Fixed Effects Regression Analysis

	pclott_rev	pclott_rev
L.payout	-0.376	
	(0.112)**	
admin_costs	0.963	0.916
	(0.039)**	(0.035)**
L.admin_costs	-0.126	-0.531
	(0.073)	(0.065)**
yrslott	5.807	3.775
	(0.394)**	(0.361)**
dinstixsales	-5.979	-12.734
	(5.098)	(3.726)**
dpulltabsales	48.798	17.617
	(6.447)**	(5.885)**
dpbsales	-19.875	1.820
	(6.042)**	(5.495)
dbgsales	7.154	-9.229
	(7.939)	(7.141)
dvlt	98.103	100.969
	(6.985)**	(6.193)**
A_lotto	4.467	-1.943
	(2.390)	(2.119)
ddaily	-3.265	10.936
	(6.751)	(6.082)
per_age_25_64	366.540	436.744
	(131.049)**	(115.472)**
urate	-0.095	-0.648
	(0.648)	(0.580)
top_ind_MTR	0.551	0.163
	(0.651)	(0.582)
sales_rate	-1.410	-1.305
	(1.958)	(1.733)
state_GDP	-0.000	-0.000
	(0.000)**	(0.000)**
all3d	-8.995	-1.215
	(2.845)**	(2.570)
all3r	-12.916	-12.431
	(3.383)**	(3.023)**
elecyr	0.048	-0.232
	(2.259)	(2.019)
pci	0.002	0.002
	(0.001)**	(0.001)**
fed_aid	0.002	0.001
	(0.000)**	(0.000)**
Pct__High_Shool	-1.832	-2.055
	(0.440)**	(0.394)**
Pct__College	0.220	1.624
	(0.887)	(0.794)*
L.lottery_prizes		0.132
		(0.007)**
_cons	-54.105	-81.479
	(47.638)	(41.718)
R^2	0.76	0.81
N	1,484	1,484

* $p<0.05$; ** $p<0.01$

Table 2: Variable Names and Descriptions

Variable Name	Description
pclott_rev	**per capita lottery revenue** ($ per capita)
payout	payout as % of prizes
admin_costs	cost of lottery administration in year (millions of dollars)
yrslott	number of years with active lottery
dinstixsales	scratch ticket lotto dummy
dpulltabsales	pull tab lotto dummy
dpbsales	Powerball lotto dummy
dbgsales	Big Game lotto dummy
dvlt	video lotto dummy
A_lotto	number of active digit lotto games
ddaily	dummy for daily draw lotto
per_age_25_64	% of population
urate	unemployment rate
top_ind_MTR	top index marginal tax rate
sales_rate	sales tax rate
state_GDP	state GDP
all3d	governor, house, senate all democrat
all3r	governor, house, senate all republican
elecyr	election year dummy
pci	per capita income ($)
fed_aid	federal aid received by state in year
Pct__High_Shool	% completed high school
Pct__College	% completed college
lottery_prizes	prize figures

Among our lottery characteristics, *l.payout, admin_costs, yrslott, dpulltabsales, dpbsales, and dvlt* were statistically significant at the 1% significance level. Their coefficients and magnitude all correspond to our expectations. As the lagged amount of payout increases, we would expect lottery revenue to slightly decrease, and here we see it doing so by 38 cents per capita for a one percent increase. This is economically insignificant, but captures the fact that as the percent of prizes paid increases, the amount of revenue kept decreases. An increase in the administration costs by $1 million results in an increase of 96 cents per capita. This captures the scale effect of lottery administrations. Larger lotteries have larger administrations, with correspondingly larger costs, and are likely to have larger revenues as well. This effect is economically small, however. For every year the lottery exists, revenues increase by $5 per capita. As lottery administrations gain more experience, grow, and spread awareness and popularity for their games, their revenues will increase. Video lotteries and pull tab tickets are shown to be the most revenue-maximizing lottery games. Having the games increases a state's revenue by $98 per capita and $48 per capita, respectively. Both are quite economically significant amounts.

Among our demographic characteristics, *per_age_25_64,* and *Pct_High_Shool* were statistically significant at the 1% significance level. The Percentage of people aged 25-64 has the coefficient with the highest magnitude among the competing variables. We anticipated this demographic characteristic to have the most economic significance from our review of other literature. The percent of high school graduates has a coefficient of -1.832, or a 1 percent increase in the amount of high school graduates reduces lottery revenue by about $2 per capita. Literature suggests that a more educated populace is less likely to engage in lottery activity, thus this result is also consistent with our hypothesis.

In our fiscal controls we get significant coefficients for *pci, state_GDP,* and *fed aid,* all at the 1% significance level. At the same time all of the coefficients for these variables were very close to zero, suggesting to us that these variables had little to do with determining lottery revenue. Although per capita income tells quite a bit about the state's income demographic as a whole - it tells little about the income distribution and varying income gaps within the state.

The included political controls also corresponded with our expectations. The presence of an election year was statistically insignificant, but both *all3d* and *all3r* had economically and statistically significant coefficients. Interestingly, both were negative, showing that when one party has complete control in the state, lottery revenues will decrease. This strengthens our earlier hypothesis that when a party has control, they turn their focus elsewhere and are able to draw revenues from their favored, more-politically charged strategies, such as increasing taxes on the wealthy.

(4) Conclusion and Suggestions for Further Work

Our results are likely to be externally valid in similar situations within the United States. Our panel data covered all 48 continental states over 35 years, from 1967-2002. It was also a strongly balanced panel, indicating the reliability of our results. Over such a long time period, numerous political regimes passed, and technologies, public opinion, and demographic characteristics changed drastically. The data account for such variation and so the corresponding results have high external validity. It is unlikely, however, that our results would remain valid outside of the United States, even in a similar political structure, such as the European Union (where there are a number of individuals that all are regulated and governed by a central power). The data, as well as many unaccounted for regulations and factors, are specific to the U.S., so our results would likely not apply elsewhere.

The data should also be internally valid. One of the strengths of the fixed effects model is that it results in consistent estimates even with groupwise heteroskedasticity. It also provides consistent estimates even in the presence of correlation between the regressors and the error term. We attempted to minimize such correlation in order to increase the efficiency of our regression. Post-regression, we ran a Wald test which strongly confirmed the presence of groupwise heteroskedasticity. The fixed effects model, however, only takes into account variance within the individual, so the heteroskedasticity between states does not affect the regression estimates. Additionally, the Woolridge test for first-order autocorrelation was run after the regression. It concluded that autocorrelation was present in our regression. This is not a surprise, and not necessarily a problem. While running time series data, especially over a large panel, it is unlikely that all possible variables could be included. As a result, part of this autocorrelation comes about through omitted variable bias. Perhaps if we included income quantiles, instead of just per capita income, for example, the autocorrelation would be lessened.

We also accounted for possible endogeneity among the dependent variable and the independent variables. *Payout* and *Admin_Costs* were both likely to affect the total amount of revenues, but they would also have been endogenous. Increased lottery payouts would likely correspond with changing revenues. To correct for the problem of endogeneity while still taking the effects of the variables into account, we introduced the lags of these variables. The previous period's administration costs would likely affect the current period's lottery revenue (by being factored into a budget, for example) but would not be endogenous. An alternative fix for endogeneity would have been to use instrumental variables. However, neither our own deliberation and analysis of the data, nor a survey of the literature resulted in an instrumental variable that met all the conditions for validity. Using insufficiently strong instrumental variables would have resulted in greater distortion of the regression, so instead we controlled for endogeneity using the method described above. A final problem that might affect internal validity would be mistakes in the dataset. Some spans of year were scaled differently than others, for example, and we had to correct for this. Additionally, some of the lottery values recorded were simply incorrect. After noticing a large outlier in the Oregon Powerball data, we contacted the lottery and requested their own data. It was different both in the amount of sales per year, and in that Oregon actually started Powerball *four years later* than the original data suggested [towards the end of 1992, see Table 3]. These discrepancies did not, however, significantly alter the results of regression. The corrected Powerball data for Oregon is used in the final regressions.

Table 3: Original and Corrected Oregon Powerball Sales Data

Year	Original	Corrected (from OR State Lottery Public Records)
1988	3.8	---
1989	1934	---
1990	33.1	---
1991	18.4	---
1992	16.9	3.4
1993	31.4	30.4
1994	36.3	37.3
1995	38.8	38.7
1996	35.1	35.9
1997	30.4	30.4
1998	37.8	37.7
1999	43.3	41.4
2000	35.7	37.4
2001	38.1	38.5
2002	---	44.3

One extension of this research would be to further investigate the characteristics of the various lottery games, especially the expected value of the payout. By factoring in the advertised payout, the taxation rate, and the odds of winning, it would be possible to group the games by these characteristics and see how it might change the impact of payout as well as people's likelihood of playing. An additional extension to the research would be to investigate the effects of cross-state or regional games. There are massive, multi-state games, such as the Powerball Lottery, that likely operate in a different manner to smaller, regional games, such as the tri-state lottery that includes Maine, New Hampshire, and Vermont. Shared lottery games might not impact all the participating states in the same way, and might create effects that are worth investigating.

Some of these games also require residency in a participating state and could provide for an interesting case-study about inter-state lottery play and travel. Our regression indicated that the more experience a state had with running a lottery, the higher the revenue it was able to generate. This finding about experience could be extended into the individual games to see how they might benefit from increased popularity, availability, and operational know-how. A certain game, could, for example, be structured in such a way that it disproportionately gained in popularity without increasing revenue accordingly. From a policy standpoint, the aspects of the games that make them popular could then be adapted to maximize revenue.

(5) Bibliography

Alm, J., M. McKee, and M. Skidmore, "Fiscal Pressure, Tax Competition, and the Introduction of State Lotteries," *National Tax Journal*, Vol. 46, No.4 (December, 1993), pp. 463-76

Anthony D. Miyazaki, Ann Hansen and David E. Sprott, "A Longitudinal Analysis of Income-Based Tax Regressivity of State-Sponsored Lotteries", *Journal of Public Policy & Marketing*, Vol. 17, No. 2 (Fall, 1998), pp. 161-172

Grace M. Barnes, John W. Welte, Marie-Cecile O. Tidwell, Joseph H. Hoffman, "Gambling on the Lottery: Sociodemographic Correlates Across the Lifespan", *Springer Science & Business Media*, (December, 2010)

Rork, Jonathan C., "Yardstick Competition in Toll Revenues: Evidence from US States," *Journal of Transport Economics and Policy*, Vol. 43, No. 1(Jan., 2009), pp. 123-139

Wyett, Todd A., "State Lotteries: Regressive Taxes In Disguise", *The Tax Lawyer*, Vol. 44, No. 3 (Spring 1991), pp. 867-883