STATE LOTTERY REVENUE:
THE IMPORTANCE OF GAME CHARACTERISTICS

THOMAS A. GARRETT
Federal Reserve Bank of St. Louis

RUSSELL S. SOBEL
West Virginia University

Previous studies have found that lottery ticket sales are influenced by certain socioeconomic characteristics of the population. The authors extend the state lottery literature by examining how lottery game characteristics, such as the overall expected value, the top prize, and the total number of combinations, influence ticket sales after controlling for socioeconomic variables. They perform an empirical analysis on a unique set of data that include information for 135 online lottery games in the United States. The results show that ticket sales are significantly influenced by the size of the top prize and the odds of winning it, but ticket sales are not significantly affected by the expected value of the lower prizes. The out-of-sample predictive power of the model is then tested using a real-world example of changes to the prize and odds structure of PowerBall.

Keywords: lottery; ticket sales; state lottery revenue; top prize

Thirty-eight states and the District of Columbia now offer state-sponsored lotteries. Although state lotteries provide entertainment for millions of people nationwide, the primary goal of a lottery is to maximize lottery revenues for the state. According to Clotfelter and Cook (1989, 11), “Lottery agencies are not merely acting out of a liberal respect for consumer sovereignty. They are engaged in a well-focused quest for increased revenues.” In fact, the operating papers governing most state lotteries generally cite revenue maximization as the goal of the bureaucracy charged with administering the lottery. In 2002, 31%
of total lottery ticket sales, or roughly $13 billion, was generated in net lottery revenue in the United States.²

A fairly sizable body of literature has explored the relationship between the socioeconomic characteristics of a state’s population and lottery ticket sales.³ These studies generally find that lottery sales are higher for individuals who belong to minority groups, individuals with little or no college education, residents of urban areas, and individuals between the ages of 45 and 65. Of all socioeconomic characteristics studied, income has received the most attention. The coefficient on the income variable lends insight into whether the revenue burden of the state lottery is regressive, proportional, or progressive. Most studies have found that overall lottery expenditures tend to be regressive.⁴

These previous studies have provided insight into how the socioeconomic characteristics of state residents explain differences in lottery sales across states. We build on this previous literature by exploring how lottery game sales also depend on game characteristics such as the value of the top prize, the chances of winning, and the overall expected value of the lottery. A few previous studies have examined the importance of game characteristics on predicting lottery revenues for a given state lottery. The first of these studies was conducted by Vrooman (1976). Focusing on the New York lottery, Vrooman included the expected value of the lottery and the probability of winning the top prize in a regression to predict lottery sales. Vrooman found that both the expected value of the lottery and the probability of winning the top prize were not significant determinants of lottery ticket expenditures in New York. In a later study of the Washington state lottery, Thiel (1991) found that worsening the odds of winning the jackpot led to an increase in lottery revenues, as more players were attracted to the larger jackpots that result from fewer lottery winners. Similarly, Scoggins (1995) examined the Florida lotto and found that lottery expenditures, and thus net revenues to the state, will be increased by allocating a larger percentage of sales to the top prize. The significance of game characteristics in these studies of individual lotteries supports our conjecture that differences in game characteristics should be important in explaining differences in sales across lottery games.
Our objective is to examine how the prize and odds structure of lottery games influence lottery ticket sales. In accordance with the stated mission of state lottery agencies, several studies have shown that lottery agencies do indeed structure their lottery games to maximize revenue. In this study, we wish to see which prize and odds structure across states and games provides the greatest revenue to state lottery agencies. We use a set of data on prize payouts, odds of winnings, and lottery game sales for 135 online lottery games in the United States to develop several measures of game characteristics to include in our empirical analysis. Our model is estimated on the full sample of lottery games and also on subgroups of games by type (lotto vs. smaller prize games, for example). Our results show that the amount of the top prize and the odds of winning it are significant in explaining differences in sales across U.S. lottery games. However, after controlling for these characteristics of the highest prize, we find that the expected value of lower prize tiers has no significant influence on lottery sales. We find that the explanatory power of our models that include game characteristics far exceeds the explanatory power achieved by using only socioeconomic characteristics. Finally, we find support for the out-of-sample predictive power of our model using a change to PowerBall.

THE EXPECTED VALUE AND EXPECTED PAYOUT OF A LOTTERY TICKET

The expected value of purchasing a lottery ticket is negative to the lottery player because the expected prize payout on a ticket is less than the dollar cost of purchasing the ticket. The expected value of purchasing a lottery ticket, however, can differ substantially across lottery games because of differences in the expected prize payout across different lottery games. In this section, we derive expressions that allow us to calculate the expected prize payout on a given lottery ticket from the general prize structure of each game.

Although the computation of expected prize payout for a lottery ticket is a basic summation of each possible prize amount multiplied by the probability of the prize, the presence of pari-mutuel lottery prizes requires a modification of the simple formula because the prize
amounts are not fixed. This section discusses pari-mutuel and fixed prizes and presents the derivations and final expression for our expected prize payout variable that considers both fixed and pari-mutuel prize lottery games.

The state determines the structure of the lottery by setting the prize amounts, number of combinations, and the probabilities of winning each prize. There are two general ways in which prize amounts may be set by the state. The first type is a fixed prize for which the prize amount won by the player is a certain amount that is independent of the number of other players winning that prize and of the outcomes on other prizes. The second type of prize is a pari-mutuel prize for which the prize amount depends on how many other players win that same prize and the total dollar amount wagered for the lottery drawing. For a pari-mutuel prize lottery, the state allocates a fixed percentage of ticket sales to each prize level and then splits this total amount among all winners of that prize. Some lottery games (such as daily number games) have all fixed prize amounts, but most large lotto games have a pari-mutuel top prize and fixed lower prizes.

Because lottery games can have pari-mutuel prize amounts, the question is how to determine the expected prize payout of a ticket for a lottery that includes pari-mutuel prizes because the prize amounts are not fixed. Consider an all pari-mutuel lottery game that offers three prize tiers, with 25% of total sales allocated to pay top prize (jackpot) winners, 15% of total sales allocated to pay second prize winners, and 10% of sales allocated to pay third prize winners. It is clear from this example that the state will allocate 50% of total sales to prizes. Intuitively, it can be seen that across all players, half of the money wagered is returned as winnings, and thus the average player will have an expected prize amount equal to 50 cents per dollar wagered. We now turn to a more formal derivation of the fact that the expected prize payout value of this lottery game is simply the share of sales allocated to all prizes.

From an aggregate perspective, the lottery’s expected total prize payout ($TP$) to all players on a given drawing is by definition equal to the sum across all $N$ tickets sold of the expected prize payout ($EP$) for each ticket ($n$).
Because ex ante all tickets have equal expected prize payout values \((EP_n = EP\) for all \(N\) tickets), the right-hand side of the equation simply reduces to the expected total payout being equal to the number of tickets sold, \(N\), times the expected prize payout for a single ticket \((EP)\). Given that all prizes are won, in the case of an all pari-mutuel lottery game, the total prize payout for the state is not random at all but is a fixed share of the amount wagered on the drawing. However, it is possible that on any given drawing for a pari-mutuel lottery, there is no winner in one or more of the prize tiers. In most cases, this money is distributed across other prizes, but for unwon top prizes, the state lottery generally rolls this amount over to the next drawing. This does cause the expected total prize payout to vary slightly from drawing to drawing when it occurs. However, the average expected payout value of a ticket across all drawings that year (both the ones in which there was no winner and the drawing with the eventual winner) remains unaffected. Thus, the variable we are deriving here in this section may be more properly defined as the average expected payout value for the lottery. In our later empirical work, we incorporate both this average value and a measure of its variance.

Continuing with the derivation, for mathematical simplicity, we assume each ticket costs $1 and that \(N\) tickets are sold. Letting \(S\) denote the share of the lottery revenue for the drawing that is allocated to all prize pools, then the state’s total payout for a drawing will be equal to \(S \cdot N\). Thus, equation (1) may be rewritten as

\[
S \cdot N = N \cdot EP. \tag{2}
\]

Dividing both sides by \(N\) gives

\[
S = EP. \tag{3}
\]

Equation (3) demonstrates the intuitive result that the expected prize value of each ticket \((EP)\) is simply the share of the ticket revenue devoted to the prize pools. When a lottery game consists of some prizes that are fixed and some prizes that are pari-mutuel, the total expected prize payout value is simply the sum of the expected prize payout
value of the fixed prize games and the expected prize payout value of the pari-mutuel prize games.

The expected prize payout value of an all fixed prize lottery games is given by

$$EP = \sum_{k=1}^{K} P_k \cdot W_k = \sum_{k=1}^{K} \left( \frac{c_k}{C} \right) \cdot W_k,$$

where $k$ represents one of $K$ total prizes, $P_k$ is the probability of winning prize $k$, and $W_k$ is the dollar amount of prize $k$. The probability of winning prize $k$ is found by dividing the number of combinations that win prize $k$, $c_k$, by the total number of combinations, $C$, available to the lottery player, so that $P_k = c_k/C$.

For lottery games that have some prizes that are pari-mutuel and some prizes that are fixed, the expected prize payout value of a ticket is the sum of the expected prize value across the fixed prizes plus the expected prize value across the pari-mutuel prizes.

$$EP = EP_{fixed} + EP_{pari-mutuel}. (5)$$

Now, substituting the results of equations (3) and (4) into equation (5) gives

$$EP = \sum_{fixed} P_k \cdot W_k + S = \sum_{fixed} \left( \frac{c_k}{C} \right) \cdot W_k + S,$$

where, for lottery games with some or all pari-mutuel prizes, $S$ again represents the total percentage of ticket revenue devoted to all pari-mutuel prizes. For a lottery game offering only fixed prize amounts, computing the expected prize payout value remains the same as in equation (4) because $S$ would equal zero in equation (6). Lottery games having only pari-mutuel prizes will have an expected prize payout value equal to $S$, as seen by equation (3). Finally, those lottery games having both fixed and pari-mutuel prizes will have an expected prize payout value equal to the expected prize payout value of the fixed prize amounts plus the total percentage of ticket sales devoted to the pari-mutuel prize amounts.
U.S. ONLINE LOTTERY GAMES

The expected prize payout value for our sampling of lottery games is computed using data on lottery game sales and prize and odds structures for 135 lotto, daily number, and nightly keno games available in the United States during 1995. Here we briefly describe each type of game used in our analysis.

Lotto games typically offer pari-mutuel, multi-million-dollar jackpots, with odds of winning the jackpot set at several million to one. These games require that players match 5 of 5 or 6 of 6 numbers out of a given range of numbers to win the top prize, which is usually a pari-mutuel prize amount. If a player’s numbers match those drawn, he or she wins the top prize (or a portion of the pari-mutuel top prize if there is more than one winner). Lotto games have the added feature of a rollover, meaning that if there is no jackpot winner, the prize money allocated to the jackpot is added to the jackpot prize money for the next drawing. Some lotto games are available in numerous states. These multistate games involve several states participating in the same lottery game. Due to the administrative costs of lottery games, states having a low population, and thus lower sales on average, cannot offer multi-million-dollar jackpots. But if several states pool their sales, these large jackpots can be offered. The multistate games available in 1995 included PowerBall, Tri-West Lotto, Tri-State Card Cash, and Tri-State Megabucks.

Daily number games, sometimes called three-digit and four-digit games, are offered 6 or 7 nights a week. For the three-digit game, a player picks a number between 000 and 999 and wins a top fixed prize of $500 (in most states) if his or her three-digit number matches the three-digit number drawn. Similarly, for the four-digit game, a player picks a number between 0000 and 9999 and wins a top fixed prize of $5,000 if his or her four-digit number matches the four-digit number drawn.

Keno is a fixed prize game that made its debut into the lottery portfolio of many states during the 1990s. The “keno computer” randomly chooses 20 numbers between 1 and 80. The player can choose to play any 1 of 10 number games and hopes to match numbers drawn by the
computer. For example, if a player chooses to play the 10-number game, he or she selects 10 numbers between 1 and 80 on the playslip. Twenty numbers between 1 and 80 are then randomly drawn and displayed on a monitor. If some or all of the 10 numbers selected by the player match the numbers drawn by the computer, the player wins. In some states, keno can be played at bars, restaurants, and even gas stations and is offered every 5 minutes on video monitors between the hours of roughly 6:00 a.m. and 2:00 a.m. Other states offer a single keno drawing several times per week.\textsuperscript{12}

**EMPIRICAL METHOD AND DATA**

The expected prize payout value of a lottery incorporates information about every prize amount and the odds of winning each prize. However, a substantial portion of a lottery’s expected prize payout value is from the top prize, despite the remote odds of winning it. It is these top prize amounts that are generally used to advertise the lottery, and it is these jackpots that are material for news stories when roll-overs make them grow to large amounts. Because of the potential for the top prize component to have a distinct influence on sales relative to the value from the lower prizes, we also include the top prize amount and the total number of combinations as variables in our regression analysis. Descriptive statistics for all variables used in the analysis are presented in Table 1.

Although the three game-characteristic variables are our main focus, we include several standard demographic variables that have been shown to be important determinants of lottery sales in other studies. We include these variables to control for other characteristics that may affect lottery sales and to ensure that our game-characteristic variables are not capturing the impact of any omitted variables. In addition, we include the number of lottery retail outlets available in each state to control for the availability of lottery tickets.

Our analysis involves an estimation of the following model of lottery sales:

\[
PC\ Sales_i = \beta_0 + \beta_1 \cdot \left( \frac{1}{EP_i} \right) + \beta_2 \cdot Top\ Prize_i + \beta_3 \cdot Total\ Combinations_i + \beta_4 \cdot Variance_i + \Gamma_i + \epsilon_i,
\]
where \( PC_{Sales} \), is per capita lottery sales for game \( i \), \( Top\ Prize \), is the top prize for game \( i \), \( Total\ Combinations \), is the total number of combinations for game \( i \), and \( \Gamma_i \) is a vector of demographic variables corresponding to lottery game \( i \)'s state. Because we are estimating an inverse demand curve for lottery games, the price of a lottery ticket can be represented by \( 1/E_{Pi} \), which is the reciprocal of the expected prize payout value of lottery game \( i \). It is very important to note that the inclusion of the top prize and the number of combinations as separate variables affects the interpretation of the coefficient on \( E_{Pi} \). The coefficient now reveals the marginal impact of increasing the expected prize payout value of the lottery, holding constant the value and odds of winning the top prize. Thus, this coefficient is a reflection of how an increase in the expected prize payout value from the lower prize tiers affects lottery sales. The coefficients on \( Top\ Prize \), and \( Total\ Combinations \), reveal how players respond to changes in the prize or odds structure of the top prize, holding constant the overall expected prize payout value of the lottery. Based on the findings of Thiel (1991) and Scoggins (1995), we expect a positive coefficient on \( Top\ Prize \), and a negative coefficient on \( Total\ Combinations \). Because only one combination wins the jackpot, the odds of winning the top prize decline with a larger number of combinations. Restructuring a lottery game’s top

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per capita sales</td>
<td>21.117</td>
<td>22.999</td>
<td>0.141</td>
<td>157.961</td>
</tr>
<tr>
<td>Expected prize payout value</td>
<td>0.502</td>
<td>0.041</td>
<td>0.24</td>
<td>10.661</td>
</tr>
<tr>
<td>Variance</td>
<td>97,893</td>
<td>280,411</td>
<td>243.44</td>
<td>2,472,286</td>
</tr>
<tr>
<td>Top prize</td>
<td>698,385</td>
<td>1,231,003</td>
<td>500</td>
<td>5,873,452</td>
</tr>
<tr>
<td>Total combinations</td>
<td>9,937,384</td>
<td>18,912,271</td>
<td>1,000</td>
<td>54,979,155</td>
</tr>
<tr>
<td>% Ages 45-65</td>
<td>17.233</td>
<td>2.154</td>
<td>13.663</td>
<td>23.663</td>
</tr>
<tr>
<td>% Nonwhite</td>
<td>15.537</td>
<td>13.012</td>
<td>1.421</td>
<td>70.346</td>
</tr>
<tr>
<td>% College</td>
<td>19.941</td>
<td>5.379</td>
<td>2.1</td>
<td>33.3</td>
</tr>
<tr>
<td>% Urban</td>
<td>69.576</td>
<td>15.632</td>
<td>32.2</td>
<td>100</td>
</tr>
<tr>
<td>Per capita income</td>
<td>23,292.9</td>
<td>3,594.1</td>
<td>17,714</td>
<td>33,435</td>
</tr>
<tr>
<td>Per capita lottery retail outlets</td>
<td>0.00089</td>
<td>0.00224</td>
<td>0.00058</td>
<td>0.00167</td>
</tr>
</tbody>
</table>

NOTE: Based on full sample size of 135.
prize and number of combinations could lead to an increase or decrease in sales, depending on the magnitude of each variable’s impact on lottery sales.

Several studies using time-series game-level data have shown that the true payout rate of a lottery may not be equal to its expected prize payout value (e.g., the true long-run payout value) because of jackpot rollovers and variability in lottery sales over time. We consider the variance of prize outcomes for each lottery game \(i\) to capture the impact of rollovers. Those lottery games having a greater variance of prize outcomes are more likely to experience jackpot rollovers because the odds of winning the top prize become more remote. Let \(\sigma^2\) represent the variance of prize outcomes to an individual player, given by

\[
\sigma^2 = \sum_{k=0}^{K} (W_k - EP)^2 \cdot P_k = \sum_{k=1}^{K} W_k^2 \cdot P_k - EP^2. \tag{7}
\]

The first part of equation (7) shows the standard formula for computing the variance of a random variable. Here it is the sum over all prize outcomes (including the losing outcome of \(k = 0\)) of the squared difference between the prize amount and the expected prize payout value of the lottery. This is further simplified to the sum of the prize amounts squared times their associated probabilities minus the squared expected prize payout value of the lottery. Because the losing prize of zero disappears in this version, it is shown as the summation over only the winning prizes (or \(k = 1\) to \(K\) instead of \(k = 0\) to \(K\)). Again recognizing the distinction between fixed and pari-mutuel prizes, the general formula for the variance, which considers equation (6) for the expected prize amount of a single prize, \(k\), can be represented as

\[
\sigma^2 = \sum_{\text{fixed}} W_k^2 \cdot P_k + \sum_{\text{pari-mutuel}} \frac{\sigma^2_k}{P_k} - EP^2, \tag{8}
\]

where the summations in the equation are shown separately for the fixed and pari-mutuel prizes, \(s_k\) represents the share of the ticket sales allocated to the prize pool for a pari-mutuel prize \(k\), and, as before, the probability of winning prize \(k\), \(P_k\), is simply \(c_k/C\).
EMPIRICAL RESULTS

We provide in Table 2 the results from three regression specifications. The first regression consists of all lottery games in our sample. The second regression considers only the subset of lotto games, and the third regression considers only the subset of daily number and smaller prize lottery games. In addition, to examine the improvement from the previous literature’s use of only socioeconomic characteristics, each regression was also run excluding the three game-characteristic variables.

Because we are including the top prize of game $i$ as an independent variable, the potential problem of simultaneity arises between game sales and the top prize for those games having a pari-mutuel top prize. For each drawing (and especially in the case of jackpot rollovers), announcement of the jackpot will induce lottery sales, and these sales will in turn affect the top prize because the top prize is a function of lottery sales. To ensure consistent coefficient estimates, we explored whether the top prize variable should be made endogenous in each regression model. We performed a Hausman test (Hausman 1978; Godfrey 1988, chap. 5) to determine whether there is no correlation between the top prize variable and the equation error term (e.g., the top prize is exogenous). For each regression model, the computed Hausman statistic (shown in Table 2) was less than conventional critical levels from the $\chi^2$ distribution, thus suggesting no endogeneity of the top prize variable.

As shown in Table 2, the coefficient on expected prize payout value ($1/EP$) is not significant in any of the three regression specifications. Recall that this coefficient is interpreted as the impact of a change in the expected prize payout value, holding constant the top prize and number of combinations. Thus, marginal differences in the value of the lower prize amounts have no significant explanatory power in explaining differences in per capita lottery sales. This is also supported by the insignificant coefficient on variance. The implication is that lottery players appear to be fairly unresponsive to changes in the structure of the expected prize payout value from lower prize tiers.

The coefficients on top prize and total combinations show, in support of the findings from Thiel (1991) and Scoggins (1995), that large...
TABLE 2: Game Characteristic as Predictors of Lottery Sales

<table>
<thead>
<tr>
<th>Variable</th>
<th>All Lottery Games</th>
<th>Lotto Games</th>
<th>Small Prize Games</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.15)</td>
<td>(0.80)</td>
</tr>
<tr>
<td>1/EP</td>
<td>0.309</td>
<td>0.157</td>
<td>-0.126</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.17)</td>
<td>(0.66)</td>
</tr>
<tr>
<td>Variancea</td>
<td>-0.149***</td>
<td>-0.158***</td>
<td>-0.0056**</td>
</tr>
<tr>
<td></td>
<td>(3.12)</td>
<td>(2.91)</td>
<td>(1.84)</td>
</tr>
<tr>
<td>Top prizea</td>
<td>0.149***</td>
<td>0.158***</td>
<td>-0.0056**</td>
</tr>
<tr>
<td></td>
<td>(3.12)</td>
<td>(2.91)</td>
<td>(1.84)</td>
</tr>
<tr>
<td>Total combinationsa</td>
<td>-0.074</td>
<td>-0.190</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.95)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>% Ages 45-65</td>
<td>1.449</td>
<td>1.106</td>
<td>0.872</td>
</tr>
<tr>
<td></td>
<td>(1.32)</td>
<td>(1.11)</td>
<td>(0.69)</td>
</tr>
<tr>
<td>% Nonwhite</td>
<td>0.807**</td>
<td>0.755**</td>
<td>0.278</td>
</tr>
<tr>
<td></td>
<td>(2.24)</td>
<td>(2.05)</td>
<td>(0.80)</td>
</tr>
<tr>
<td>% College</td>
<td>0.234</td>
<td>0.258</td>
<td>-0.0068</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
<td>(0.91)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>% Urban</td>
<td>-0.074</td>
<td>-0.190</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.95)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>Per capita income</td>
<td>0.0021***</td>
<td>0.0026***</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>(3.13)</td>
<td>(4.78)</td>
<td>(1.29)</td>
</tr>
<tr>
<td>Per capita lottery outlets</td>
<td>7,967.1</td>
<td>5,330.1</td>
<td>9,605.9</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(0.63)</td>
<td>(0.92)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.387</td>
<td>0.443</td>
<td>0.154</td>
</tr>
<tr>
<td>$F$ test for $H_0$: all game slopes = 0</td>
<td>—</td>
<td>5.44***</td>
<td>—</td>
</tr>
<tr>
<td>Breusch-Pagan $F$</td>
<td>112.79***</td>
<td>172.00***</td>
<td>38.06***</td>
</tr>
<tr>
<td>Hausman endogeneity test ($\chi^2$): top prize</td>
<td>—</td>
<td>0.643</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>135</td>
<td>135</td>
<td>88</td>
</tr>
</tbody>
</table>

NOTE: The dependent variable is per capita lottery sales. Absolute t statistics are in parentheses. White’s standard errors are used for models (1) through (5). — = not applicable.

a. The coefficients are interpreted as per a 10,000-unit change in the top prize, total combinations, and variance.

*Significant at the 10% level. **Significant at the 5% level. ***Significant at the 1% level.
lotto games having a higher top prize are likely to have higher sales. Also, an increase in the total combinations available in a lotto game, which worsens the odds of winning, results in lower lottery sales. The coefficient from the lotto regression suggests that a $100,000 increase in the top prize will lead to a $1.58 increase in per capita sales, whereas increasing the total number of combinations by 1 million will lead to a 51-cent decrease in per capita lottery sales. These results for the lotto games, reported in column (4) of the table, hold in the full-sample regression but do not hold individually for the smaller prize games. This suggests that the top prize amounts are more influential in determining lottery sales for big jackpot games than for smaller prize games.

We also find support for the previous findings that nonwhite population and income are significant determinants of lottery sales. However, several of the other demographic control variables are insignificant. For the full-sample and lotto subsample regressions, the $F$ test shows that the inclusion of the game-characteristic variables significantly increases the explanatory value of the model. The difference, particularly for the subsample of lotto games, is striking. The percentage of the cross-game variation in per capita sales explained by the model for lotto games jumps from about 15% to more than 52% with the inclusion of these three variables. It is clear that a substantial portion of the variation in U.S. lotto game sales can be explained by the prize structure of the games. It also points to possible omitted variable biases in previous studies of lotto expenditures that exclude any information about the game characteristics in the models. For the smaller prize games, however, there is not much improvement in explanatory power from the game-characteristic variables, and socioeconomic characteristics are significantly greater predictors of smaller game sales than they are of lotto sales. The significant and substantial difference between the estimated regressions for lotto games and for the smaller games suggests that these are fundamentally different subsamples of lottery sales revenue. This also has the important implication that other empirical research on state lotteries should avoid techniques that pool these games together.

Although our analysis is mostly concerned with the impact of the game-characteristic variables, it is worth noting that the estimated
magnitude of several socioeconomic variables differs between the lotto and smaller prize game regressions. Most notably, the income coefficient and the coefficient on nonwhite population are both larger for smaller prize games. Although further exploration would be necessary, this might appear to have potentially important implications for incidence analysis of different types of lottery games.

### A “REAL-WORLD” TEST OF OUR MODEL: EVIDENCE FROM POWERBALL

In this section, we provide evidence of the predictive power of our lotto model from column (2) of Table 2. In 1997, the Multistate Lottery Association worsened the odds of winning the PowerBall jackpot from roughly 55 million to 1 to 80 million to 1 in an attempt to generate additional revenues. How accurate is our lotto model in predicting PowerBall sales, both before and after the prize and odds changes? The results are shown in Table 3.

Before the change in game structure, actual PowerBall sales per capita in 1995 were $25.17. Using the expected prize payout value, variance, top prize, and total combinations for PowerBall in 1995, we predict 1995 per capita PowerBall sales of $27.16 using model (4) in Table 2. A year after the prize structure changed, actual 1998 per capita PowerBall sales increased to $30.67. Using the new values for expected prize payout value, variance, top prize, and total combinations, our model predicts per capita PowerBall of $29.76 in 1998. Our model thus predicts quite well the real-world behavior of PowerBall sales both before and after the prize and odds structure was changed.

<table>
<thead>
<tr>
<th>Actual Per Capita PowerBall Sales</th>
<th>Estimated Per Capita PowerBall Sales</th>
<th>Forecast Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to change in prize structure</td>
<td>$25.17</td>
<td>$27.16</td>
</tr>
<tr>
<td>After change in prize structure</td>
<td>$30.67</td>
<td>$29.76</td>
</tr>
</tbody>
</table>

NOTE: Dollar amounts are on a per-person basis. Estimates calculated using coefficients from the lotto regression model shown in column (4) of Table 2.
and further reveals the importance of considering lottery game characteristics in a model of lottery sales.

POLICY IMPLICATIONS AND CONCLUDING REMARKS

As the demand for state government services continues to increase, states are becoming pressured to find alternative sources of revenues. Many states have adopted lotteries in hopes of easing their fiscal pressures. Given the billions of dollars in lottery revenue generated each year, previous scholarly works have examined which individuals, in terms of many socioeconomic characteristics, are more likely to play the lottery. This information potentially allows states to target certain socioeconomic groups in an attempt to further increase lottery revenues.

Previous works that have examined the socioeconomic determinants of lottery ticket expenditures have overlooked one key aspect of lottery play—namely, the importance of lottery game characteristics. Using data from a sampling of online lottery games in the United States, we included the expected prize payout value, the top prize, and the total combinations from each game in a model of lottery sales. We find that game characteristics for lotto games are significant determinants of lottery sales. Our study is the first to distinguish between top prize and lower prize tiers to see how players react to changes in either. We find that, although lower prize tiers may be appropriate in terms of offering a “consolation” prize, the expected prize payout value of these lower prizes does not seem to be a significant factor in explaining variations in lottery sales across games. The results from our sample of U.S. lottery games support the findings of single game studies that have shown that large jackpots stimulate lottery play. Support for our model was further validated with a predictive real-word test using PowerBall.

Our study has implications for future (and previous) empirical work on state lotteries. The results not only suggest that empirical studies should be including measures of game characteristics to avoid omitted variable bias but also that substantial differences between
large lotto games and other types of games necessitate that empirical work consider data from these games separately.

Lotteries will undoubtedly exist into the future. But as lotteries face competition from increasing casino and Internet gambling and interstate tax competition, states interested in keeping their lotteries viable will have to devise a strategy to generate the most revenue possible. By considering the lottery game characteristics presented in this article in combination with the socioeconomic makeup of residents, states can better predict and structure their lottery games in an attempt to generate greater revenues.

NOTES

1. The stated mission for each state lottery can be found on its Web site or annual report. Several studies have examined political issues surrounding state lottery adoption. See Davis, Filer, and Moak (1992); Alm, McKee, and Skidmore (1993); and Garrett (1999).

2. Net lottery revenues are total lottery sales minus prize payouts, commissions, and administrative expenses. The main reported uses of this revenue were education (51.6%), states' general funds (24.2%), and local programs and infrastructure (13.3%). These data are potentially misleading, however, because of fungibility issues (Mikesell and Zorn 1986). Even if the revenue is earmarked toward education, for example, the state can reduce its own general fund spending on education, effectively converting the increased earmarked revenue to money in the general fund.


6. Online games are games for which a player generally gets tickets from a computer terminal after filling out a playslip. These are in contrast to instant or scratch-off lottery games, which are not considered here.

7. Here we assume for mathematical simplicity that we are considering a randomly selected ticket. If players tend to clump together on certain number combinations, then the players on these combinations will have below-average expected values, whereas players on sparsely chosen numbers will have above-average expected values. Taking this into account (by leaving the summation in the equation but dividing both sides by N) simply yields that what we derive is the average expected prize payout value of a ticket, which is precisely our goal.

8. Assuming that the jackpot is won in the next drawing, then the average expected prize amount across the two drawings is equal to the share of revenue allocated to prizes. Because we use annual data in our empirical analysis, this simplifying assumption is valid to aid us in deriving the average expected prize payout value on the game over the year. However, the expected value of a ticket can vary slightly from drawing to drawing in the event of a rollover. In our empir-
atical analysis, we incorporate a variable to control for the impact of the variance in the expected prize payout value over the year.

9. An important aspect of this equation is that the expected prize payout values computed are essentially the true expected prize payout values, as they use the net present value of discounted jackpot prizes instead of using the advertised nominal jackpot amount. For most online lottery games, the share of lottery sales earmarked for the jackpot is used to buy an annuity, which pays the advertised jackpot amount over many years. The advertised jackpot amount is thus much higher, and using this figure to compute the expected prize payout value would result in overestimation. Our use of the share amounts that are used to buy the annuities circumvents this problem.

10. Total 1995 annual sales for each lottery game are from Public Gaming International’s Public Gaming Magazine (September 1996, Volume 24, Issue 8). Prize and odds information and the frequency of offering for each game were obtained by contacting each state lottery agency. A list of all lottery games used in the analysis will gladly be provided upon request.


12. For our analysis, we only consider keno games that are offered nightly. For those states that offer keno every 5 minutes, data are not available on sales per drawing per number game. Sales for each game could have been computed by dividing total sales by the number of games, under the assumption that sales are equal for each keno game. Given the wide variance in top prize offerings for keno games ($2.00 to $100,000), it is very unlikely that all keno games experience the same level of sales.

13. The price of a lottery ticket can also be represented as $1 - EP$. This specification provides similar results.

14. The correlation between the top prize and the total number of combinations is 0.407.


16. This equation uses the following standard simplification: $\sigma^2 = E(X^2 - E[X]^2) = E(X^2) - (E[X])^2$.

17. The structure of PowerBall has recently changed. The odds of winning the top prize are currently 120 million to 1.

18. The variance of PowerBall was 180,844 before the change in game structure and 317,343 after the change in structure.

REFERENCES


---

*Thomas A. Garrett is a senior economist in the Research Division of the Federal Reserve Bank of St. Louis, Missouri.*

*Russell S. Sobel is an associate professor in the Department of Economics at West Virginia University, Morgantown.*