Local Price Variation and the Income Elasticity of Demand for Lottery Tickets

Thomas A. Garrett
Department of Economics
University of Mississippi
University, MS 38677
tgarrett@olemiss.edu
662-915-5829

Natalia A. Kolesnikova
Department of Economics
University of Mississippi
University, MS 38677
natalia@olemiss.edu
662-915-1589

Abstract

The cost of living varies as much across locations as it does over time. We demonstrate the importance of considering locational cost-of-living differences in empirical models of the demand for state lotteries. Previous research has shown that the nominal-income elasticity of demand for lottery tickets is less than one, suggesting that individuals and geographic regions with lower incomes tend to have a greater percentage of their income allocated toward lottery ticket purchases than do wealthier individuals and geographic regions. We first provide a conceptual framework that reveals that real-income elasticities generally will be different from nominal-income elasticities. We then re-estimate traditional cross-sectional models of lottery demand using a sample of metropolitan statistical areas. We find that the magnitude of income elasticity estimates is smaller when local cost-of-living is omitted from empirical models, especially in the case of instant lottery games.

Local Price Variation and the Income Elasticity of Demand for Lottery Tickets

I. Introduction

Research has shown that individuals and geographic regions with lower incomes tend to have a greater percentage of their income allocated toward lottery ticket purchases than do wealthier individuals and geographic regions.[1] As a result, the distributional burden of expenditures on lottery tickets is generally characterized as regressive.[2] The distributional burden of lottery ticket expenditures is traditionally determined by estimating the income elasticity of demand for lottery tickets, with a value less (greater) than one indicating regressivity (progressivity). To determine the income elasticity of demand for state lotteries, researchers typically estimate an equation where lottery ticket sales is a function of income and other demographic characteristics. Many studies use cross-sectional data, and the unit of observation is a zip code, a city, a county, or a state because individual-level data are not usually available. As a result, the "change in income" comes from the variation in nominal income across locations with no regard for price-level differences across locations.

We argue here that the empirical models of lottery demand described above should include some measure of the local price level in order to capture differences in purchasing power across locations. It is standard procedure to adjust monetary variables (such as income, revenue, GDP) for inflation when comparing them over time. The point of such an adjustment is to ensure that differences in monetary variables over time reflect only differences in purchasing power. Yet, it is common practice to compare only nominal incomes in comparisons of locations across space at a given time despite the fact that price-levels, and thus the cost of living, across

---

[1] See, for instance, Clotfelter and Cook (1987, 1989); Scott and Garen (1994); Farrell et al. (1999); Price and Novak, (1999); Forrest et al. (2000); and Garrett and Coughlin (2009).
[2] One should be cautious about inferring "regressivity" from an analysis using aggregated data.

locations are as varied as they are across time. For example, the Consumer Price Index (CPI) in 1985 (107.6) and 2000 (172.2) yields the same cost of living difference as the cost-of-living difference between Amarillo, Texas (91.6) and Los Angeles, California (146.8) in 2000. It should be of little surprise that the cost of living varies widely across locations. According to data from the 2000 Census, the median price of a house in San Francisco was five times greater than the median price of a house in Pittsburgh. Significant cross-city variation in housing prices exists even after adjusting for the quality of housing (Gabriel and Rosenthal, 2004; Chen and Rosenthal, 2008). In addition, various cost-of-living indexes indicate that prices of other consumption goods (groceries, utilities, transportation, health services) vary across locations as well.[3]

Several recent studies have demonstrated the importance of considering local cost of living. Black et al. (2009, 2014) show that differences in local prices—namely, housing prices—help to explain significant differences in the college wage premium across cities and the evolution of income inequality in the United States. Moretti (2011) emphasizes the importance of distinguishing between nominal and real earnings and shows that at least 22 percent of the increase in the college wage premium over the past 30 years is explained by spatial differences in the cost of living. Albouy (2009) investigates the unequal burden of federal taxation across cities that results from the differences in locational cost of living and wages. The cost of living has also been found to be an important determinant in the demand for children (Black et al., 2013).

Our objective in this paper is to estimate empirical cross-sectional models of lottery demand that have been established in the literature (Mikesell 1989; Hansen 1995; Price and Novak 1999, 2000; Garrett and Marsh 2002; Ghent and Grant 2010), but contribute to these models by accounting for cost-of-living differences across locations (the cross-sectional units of

---

[3] We discuss these cost-of-living indexes later in the paper.

observation) and examine how estimates of the income elasticity of demand change when cost-of-living differences are considered. We first demonstrate conceptually that the failure to consider differences in the cost of living when estimating lottery demand may yield an income elasticity of demand that is quite different from that obtained when the cost of living is considered. As we will show later using a sample of Metropolitan Statistical Areas (MSAs), our income elasticities of demand based on nominal incomes and nominal lottery sales are generally lower than the income elasticities that consider cost-of-living differences across MSAs. To see the importance of considering cost-of-living differences, assume as an extreme example that the demanded quantity of lottery tickets is the same in two cities but income *and* the cost of living in city 1 are twice as high as those in city 2. If we ignore the cost of living when calculating the income elasticity of demand, we would conclude that lottery demand in this society is absolutely income inelastic. This is, however, an erroneous conclusion as we really cannot say anything about the income elasticity of lottery demand by observing these two cities since there is no variation in purchasing power between the two cities. We further motivate this observation in the next section of the paper that discusses our conceptual framework.

II. Conceptual Framework

One of the main obstacles facing researchers in estimating the income elasticity of demand for lottery tickets is the lack of individual-level data. As a result, the estimation must rely on aggregate-level data. However, it is important to realize that an "income elasticity of demand" obtained from aggregate data is conceptually different from the textbook definition of income elasticity of demand that is based on individual preferences and utility maximization. Thus, an aggregate measure of income elasticity cannot be assumed to represent income

elasticity of demand at the individual level. Despite this arguably misleading terminology, the previous studies (and our study as well) find that an aggregate income elasticity estimate remains useful for examining the response of regional sales with respect to changes in regional income as long as interpretation at the individual level is not made.

With cross-sectional data in hand, the income elasticity of demand for lottery tickets (β_1) has traditionally been obtained by estimating the following equation (see Mikesell 1989; Hansen 1995; Price and Novak 1999, 2000; Garrett and Marsh 2002; Ghent and Grant 2010)

$$\ln(X_i) = \beta_1 \cdot \ln(Y_i) + \theta_1 \cdot \mathbf{Z}_i + \varepsilon_i, \tag{1}$$

where i typically represents either a zip code, a county, or a state, depending on the study; X_i is per-capita lottery sales; Y_i is per-capita income; and the matrix $\mathbf{Z}$ includes other variables that capture demographics and lottery game characteristics. These studies use lottery sales (revenues) rather than the number of tickets sold as our dependent variable since data on the number of tickets sold are generally not available. But, ticket revenue and the number of tickets sold are similar since most tickets cost \$1.[4] Also, the price of lottery tickets is generally omitted from equation (1) because there is little or no variation in ticket prices across units of observation.[5] Given the aggregated unit of observation, the equation relates the differences in incomes across locations to the differences in lottery sales across locations. Importantly, the cost of living in location i is typically not considered in equation (1).

Of course, because the cost of living varies significantly across locations, the same nominal income does not imply the same purchasing power across locations. One way to

---

[4] Some instant ("scratch-off") lottery tickets do cost more than \$1, thus creating a difference between revenues and tickets sold.

[5] We refer to the dollar price of a lottery ticket, not the expected return, which is calculated using the probability of winning each prize. Since we do not have game-level data, we are unable to include the expected return of lottery tickets in our equations. We could compute an aggregate measure of expected return using aggregated (state-level) data on prize payouts, administrative costs, and payments to retailers. However, this aggregated measure of expected value will get picked up in our state-level fixed effects (discussed later). We thus follow the existing literature and explicitly model the demand for lottery sales rather than the demand for the quantity of tickets.

address the issue is to consider real income, as is common practice in making comparisons over time. We argue that similar adjustments must be made when comparing incomes across space. In this case, we calculate real income as the ratio of nominal income and a measure of the local cost-of-living, following Moretti (2011). A similar adjustment is made for lottery sales. Equation (2) then can be used to estimate the real-income elasticity of demand for lottery tickets

$$\ln(X_i/COL_i) = \beta_2 \cdot \ln(Y_i/COL_i) + \theta_2 \cdot \mathbf{Z}_i + \varepsilon_i, \tag{2}$$

where COL_i is the local cost of living in location i, Y_i/COL_i is real income for location i, and X_i/COL_i is real lottery sales for location i. The estimated coefficient β_2 is real-income elasticity of lottery demand.

It is important to understand the difference between equation (1) and equation (2). In equation (2), income now reflects the purchasing power of income across locations, and lottery sales now are presented in comparable dollars across locations. Note that this adjustment is analogous to the time series case — just as we would not compare nominal incomes in 1950 and 2000 without adjusting for changes in the price level over time, we should not compare nominal incomes in, say, New York City and Omaha, Nebraska without adjusting for differences in the cost of living. We want to have comparable measures of the purchasing power of income across locations, just as in time series analysis where we want to have comparable measures of the purchasing power of income across time.

Similarly, comparing sales of a good across locations also requires adjusting for the local price level, as what matters is the cost of the good relative to all other goods. Another way to think of this is that the purchasing power of the sales revenues from a good (which is equal to consumer expenditures on the good) is different between two locations. Again think of New

York and Omaha — $10,000 in sales revenue in New York has less purchasing power than $10,000 of sales revenue in Omaha.

Because purchasing power is what matters, a cross-sectional regression should therefore adjust monetary variables for price differences across locations to ensure that the monetary variables are comparable across all observations in the sample. If no adjustment for price differences across locations is made the interpretation of the estimated effect of one monetary variable (e.g., income) on another (e.g., sales) becomes less clear as the estimated effect (coefficient) is based on units of observation that are not directly comparable across locations. This is analogous to the time series case where nominal units of observations are not directly comparable across time due to temporal price-level differences.

Although the above discussion highlighted the importance of controlling for cost-of-living differences across locations, what remains is to think conceptually about why estimates of the nominal elasticity (β_1) and the real income elasticity (β_2) may be different. The following exercise provides insight into the potential difference in the nominal-income elasticity of demand estimated from equation (1) and the real-income elasticity of demand estimated from equation (2). For simplicity, let there be only two locations L and H — location L has low nominal income (Y_N^L) and location H has high nominal income (Y_N^H). Let nominal sales for good X (say, lottery tickets) in location L and location H be X_N^L and X_N^H, respectively, with $X_N^L < X_N^H$ because of (assumed) normality. Calculating the difference in (logged) nominal sales for the two locations $d(X_N) = \ln X_N^H - \ln X_N^L$ and the difference in (logged) nominal incomes $d(Y_N) = \ln Y_N^H - \ln Y_N^L$ for the two locations yields the percentage change in sales between the two locations and the percentage change in incomes between the two locations, respectively.[6] Then, $d(X_N)$ divided by

---

[6] Recall that for any variable Z, $\%\Delta Z \approx \ln(Z_i) - \ln(Z_j) = d(Z)$ for small changes in Z.

$d(Y_N)$ yields the nominal-income elasticity. If we plot the two nominal income observations and their corresponding nominal sales observations in logarithmic form, the slope of the line between these two points is the estimated nominal-income elasticity of demand. This is shown in Figure 1, with the nominal-income elasticity denoted as η_N.

[Figure 1]

Now consider that the cost of living (COL) differs in the two locations. What happens to the income elasticity when we present income and sales in real terms? Real income for each location can be computed as nominal income divided by COL, so $Y_R^L = Y_N^L / COL^L$ and $Y_R^H = Y_N^H / COL^H$. Let $COL^L < 1$ and $COL^H > 1$.[7] It is thus the case that $Y_R^L > Y_N^L$ and $Y_R^H < Y_N^H$. Furthermore, $\ln Y_R^L > \ln Y_N^L$ and $\ln Y_R^H < \ln Y_N^H$, as shown on the horizontal axis of Figure 1. The percentage change in real income is $d(Y_R) = \ln Y_R^H - \ln Y_R^L$.

The same conversion to real dollars can be done for sales of good X. So we will have $d(X_R) = \ln X_R^H - \ln X_R^L \equiv \ln(X_N^H / COL^H) - \ln(X_N^L / COL^L)$. The real-income elasticity is then $d(X_R)/ d(Y_R)$. Notice, however, that whether the real-income elasticity is larger or smaller than nominal-income elasticity depends on how large the percentage change is sales $d(X_R)$ is. As Figure 1 illustrates, consider two possible outcomes (denoted and 1 and 2) for the percentage change in real sales: $d(X_{R1}) = \ln X_{R1}^H - \ln X_{R1}^L$ or $d(X_{R2}) = \ln X_{R2}^H - \ln X_{R2}^L$, where $d(X_{R1}) > d(X_{R2})$. We can plot the four observations (in logarithmic form) for the real sales of good X, recognizing that the high real income location Y_R^H is associated with real sales of X_{R1}^H *or* X_{R2}^H and that the low real income location Y_R^L is associated with real sales of X_{R1}^L *or* X_{R2}^L.

---

[7] This assumption is for graphical simplicity. Intuitively, this implies that one location has cost of living that is below the national average and another location has cost of living that is above national average. This is analogous to the time series case where the value of a price-index (e.g., CPI) is below the index's base year value (100 for CPI in 1983) for units of observation earlier than the base year (pre 1983 in the case of the CPI) and above the index's base year value for observations later than the base year (post 1983 in the case of CPI).

The slope of each line represents one of two possible values for the real income elasticity (η_{R1} and η_{R2}) for good X, but η_{R1} is greater than η_N while η_{R2} is less than η_N. To see this, first compare η_{R1} and η_N. From the figure, it is clear that $d(X_{R1})/d(X_N) > d(Y_R)/d(Y_N)$. Rearranging terms yields $d(X_{R1})/d(Y_R) > d(X_N)/d(Y_N)$, or $\eta_{R1} > \eta_N$. Similarly, now compare η_{R2} and η_N. From the figure, $d(X_{R2})/d(X_N) < d(Y_R)/d(Y_N)$. Rearranging terms yields $d(X_{R2})/d(Y_R) < d(X_N)/d(Y_N)$, or $\eta_{R2} < \eta_N$.[8]

The main point from Figure 1 is that the real-income elasticity of demand can be larger or smaller than the nominal-income elasticity of demand depending upon the relative differences in the percentage change for real income and real sales. That is, adjusting for the cost of living alters the percentage changes between the two sales and the two income observations.

It is not difficult to show that the nominal-income elasticity of demand is equal to the real-elasticity of demand if and only if the ratio of expenditure on a good to income is constant across locations.[9] In other words, we can ignore differences in the costs of living when estimating the income elasticity of demand only when the income share spent on a good (lottery tickets, in our case) is the same in every location.[10] Since it is unlikely that this condition holds, the estimated nominal- and real-income elasticities will be different.[11] Just how different the elasticities are is an empirical question, which we explore in the remainder of the paper.

---

[8] The above analysis assumes a constant change in real income for expositional simplicity, but in reality this too will vary across observations. The conclusions in that case are identical to those presented in the text.

[9] For two locations, it can be shown using a bit of algebra that the equality of nominal-income and real-income elasticities, $\dfrac{(X_1 - X_2)/X_1}{(Y_1 - Y_2)/Y_1} = \dfrac{(X_1/COL_1 - X_2/COL_2)/(X_1/COL_1)}{(Y_1/COL_1 - Y_2/COL_2)/(Y_1/COL_1)}$, implies $X_1/Y_1 = X_2/Y_2$. It is trivial to extend this to the n-location case. We ignore another solution where the cost of living is the same in all locations since it is a trivial case.

[10] Notice that this in turn implies that both nominal-income elasticity of demand and real-income elasticity of demand are equal to one.

[11] In our sample (described later), this condition does not hold: The minimum and maximum expenditure shares for instant ("scratch-off") lottery sales are 0.11% and 0.71%, respectively; the minimum and maximum expenditure shares for online (e.g., Lotto) lottery sales are 0.08% and 0.77%, respectively.

III. Data

To estimate the distributional burden of state lotteries, one would ideally prefer data at the level of the individual. However, no nationally representative sample of individuals exists for the United States that provides information on individuals' lottery expenditures, geographic location, incomes, and demographic characteristics.[12] Thus, we follow the majority of the existing literature on lottery demand and use cross-sectional data at the most disaggregated unit of observation available—in our case, Metropolitan Statistical Areas (MSAs).

We would like to have an index similar to the one provided by the Bureau of Labor Statistics (BLS) — the Consumer Price Index for All Urban Consumers (CPI-U). The CPI-U represents a weighted price index of a basket of goods and services. Unfortunately, the BLS produces local indexes for only 27 MSAs. More importantly, even these local indexes are not suitable for comparing living costs across areas as they measure only how much prices have changed over time in a given MSA.

Instead, we use the ACCRA Cost of Living Index (COLI) provided by the Council for Community and Economic Research.[13] The COLI measures relative price levels for consumer goods and services and has been used in previous studies to capture price variation across locations (Cebula and Coombs, 2008; Nonnemaker et al., 2009; Moretti, 2011). The price index

---

[12] Scott and Garen (1994) use individual-level data on lottery expenditures for the state of Kentucky. Kearney (2005) uses survey data from the BLS Consumer Expenditure Survey and the National Opinion Research Council to explore the impact of lotteries on consumers' expenditures on other goods. Perez and Humphreys (2011) use survey data on expenditures on the Spanish lottery. The survey data used by Kearney (2005) and Perez and Humphreys (2011) cannot, unfortunately, be used to estimate models of lottery demand since the surveys did not provide information on lottery expenditures, geographic identifiers, or continuous measures of income (only income ranges). Perez and Humphreys (2011) use micro data on the Spanish lottery and find income elasticities greater than one.

[13] We use the ACCRA index because it is widely known and has been used extensively in previous literature. However, there is debate over the "best" price index in terms of capturing the cost of living (Carrillo, Early, and Olsen; 2012). While we acknowledge this debate, our purpose here is to simply demonstrate that income elasticity estimates can be quite different when the cost of living is considered. See footnote #23 for a comparison of results with an alternative measure of the cost of living. For a description of how the ACCRA price indexes are calculated and for a list of all commodities included in each price index, see www.coli.org. The Council for Community and Economic Research is formerly known as ACCRA (American Chamber of Commerce Research Association).

for each MSA is interpreted as a percentage of the average for all urban areas (where the average is set to 100).[14] The COLI is a weighted average of price indexes of six different categories: groceries, housing, utilities, transportation, health care, and miscellaneous good and services.[15] Each category includes many goods but it should be noted that the basket for COLI is smaller than the one used by the BLS for computing CPI-U. The big advantage of COLI is that it allows comparison between MSAs at a given point of time and is available for a larger number (over 300) of MSAs. Thus, we use an MSA as a unit of analysis because it provides the lowest level of aggregation for which local price data are available.

Lottery-ticket sales at the MSA-level are not readily available and therefore had to be constructed from county-level lottery sales data. To do so, we first obtained a list of all counties within each MSA (for which we had obtained local price data) by using the 2000 Census MSA boundary definitions and component (county) names.[16] We then contacted state lottery agencies and obtained county-level sales data for instant lottery games ("scratch-offs") and online lottery games (e.g., Lotto, Mega Millions, Powerball) for the year 2000 and then summed sales at the county level to arrive at online lottery sales at the MSA-level and instant lottery sales at the MSA-level. Examining online games and instant games separately allows us to explore the role of local prices in explaining the demand for different lottery products, as well as providing us with additional tests of our hypothesis.

Online games and instant games are considered different lottery products because online lottery games offer much higher jackpots than instant lottery games and the potential frequency of play for online games is less than that of instant games as drawings for online games are aired

---

[14] Several issues involving the price data are discussed in the Appendix.
[15] The weights for the individual indexes are as follows: Groceries – 12.5 percent, Housing – 29.8 percent, Utilities – 9.9 percent, Transportation – 10.7 percent, Health care – 4.1 percent, and Miscellaneous good and services – 32.9 percent. The sum of the weights does not equal 100 due to rounding.
[16] See http://www.census.gov/population/www/metroareas/lists/historical/99mfips.txt.

on television only several times a week. Research has shown that the income elasticities of demand for online games and instant games can be different.[17] Descriptive statistics for COLI and lottery sales are shown in Table 1.

[Table 1]

We follow the past literature and include several economic, demographic, and game characteristic variables in our models of lottery demand. The economic and demographic variables we include are per capita personal income, population density, and the percentage of the population with a bachelor's degree or higher.[18] Game characteristics include the age of the lottery in years, an indicator dummy variable for whether the state participates in multi-state lottery games, the number of years the state has participated in multi-state lottery games, and an indicator variable for whether the state has commercial casino gambling.[19] The values for these variables are the same for each MSA in a state since the lottery is state-wide. The age of the lottery captures the differences in each state lottery's life cycle (Mikesell, 1994). Because we are comparing different state lotteries in different stages of their life-cycles, the expected sign on age is ambiguous. Multi-state games (e.g., Powerball) generate the largest jackpots, and thus states

---

[17] Ghent and Grant (2010) use data on the income distribution rather than income levels and find that the degree of regressivity can differ by the type of lottery game. Despite differences in the income elasticity of demand for instant games and online games, most research finds that both types of games have an income elasticity of demand less than or equal to (statistically) one.

[18] Personal income, population density, and the percentage of the population with a bachelor's degree or higher were gathered from the 2000 U.S. Census. Studies have considered other economic and demographic variables as well, including the unemployment rate, the age of the population, and the percent of the population living below the poverty level. A confounding issue in modeling lottery demand is that many of these variables tend to be highly correlated with each other as well as with income and education. We included the unemployment rate, the age of the population, and the percent of the population living below the poverty level in our initial models of lottery demand, but overall the resulting coefficients were not statistically significant and did not affect the final results and conclusions. As a result, our final models include only income and education. The results from our initial models with the additional economic and demographic variables are available upon request.

[19] Lottery game characteristic data were obtained from Lafluer (2009). A list of the states with commercial casino gaming was provided by the American Gaming Association (www.americangaming.org). Garrett and Sobel (1999) and Kearney (2005) found that lottery game characteristics such as the top prize, the variance of prize payouts, and the skewness of prize payouts explain differences in game-specific sales. The game characteristics we use are in accordance with other studies that have used aggregated lottery sales data instead of game-specific data. We should note that multi-state games tend to have the largest top prize and the greatest variance and skewness of prize payouts, which Garrett and Sobel (1999) and Kearney (2005) have shown to be determinants of lottery sales.

that participate in these games are expected to have higher sales (Garrett and Sobel, 1999; Kearney, 2005). The casino dummy variable captures any effects of competition between casino gaming in the state and the state lottery (Elliot and Navin, 2002; Garrett and Coughlin, 2009). Descriptive statistics for these variables are also shown in Table 1.

We conduct our analysis using data on lottery sales, local prices, personal income, and demographic and game characteristics for 111 MSAs for the year 2000. The sample size and year of study were dictated by the greatest availability of local price and lottery sales data.[20] The MSAs used in the analysis are listed in the Appendix.

IV. Empirical Results

We estimate equations (1) and (2) using per capita instant lottery sales and per capita online lottery sales as our dependent variables. All equations contain the aforementioned economic, demographic, and game characteristic variables, as well as a set of state dummy variables to capture potential heterogeneity across states.[21] Because we wish to obtain the income elasticity of demand, lottery sales and per capita income are converted to natural logarithms before estimation.

For each game category we compare the nominal-income elasticity estimates from equation (1) with the real-income elasticity estimates from equation (2). As our previous discussion suggests, we expect to find that the income elasticity coefficient from equation (1) is different than the income elasticity coefficient from equation (2). This would support our hypothesis that local prices play an important role in explaining cross-sectional differences in

---

[20] We are restricted to estimating a cross-section rather than a panel for two reasons. First, the geographic definition of many MSAs has changed over time. Second, and more important, the price indexes for each MSA are constructed relative to the average of all MSA prices and therefore cannot be compared over time.

[21] We retain the multi-state lottery game variables in the instant lottery game regressions to capture any substitutability or complementarity between instant and online lottery games (Clotfelter and Cook, 1989).

lottery demand, and that the failure to include local cost of living in models of lottery demand

can result in income elasticity estimates that are biased, thus providing an inaccurate estimate of

the distributional burden of lottery ticket expenditures.

The empirical results from our models of lottery demand for instant sales and online sales

are shown in Table 2 and Table 3 respectively. All equations were estimated by GLS using

White's heteroskedasticity-corrected standard errors.[22] The results presented in columns (1) and

(2) of Table 2 and Table 3 are from equation (1), which estimates the nominal-income elasticity

of demand excluding control variables (column 1) and including control variables (column 2).

The key feature of these "traditional" models is that they omit local cost of living, thus failing to

capture how differences in the purchasing power of income influence lottery sales across MSAs.

Consider the regression results from equation (1) with control variables, which are shown

in column (2) of each of the two tables. For instant lottery sales (Table 2), we estimate a

nominal income elasticity of 0.268 that is not statistically different than zero. The estimated

income elasticity for online games (Table 3) is 1.106 and is statistically different than zero, but it

is not statistically different than one. This finding is similar to that of Mikesell (1989) who used

county-level data for Illinois, and Perez and Humphreys (2011) who used individual-level data

for Spain.

[Table 2]

[Table 3]

Next, we estimate equation (2) where the dependent variable is real sales per capita and

the key independent variable is real per capita income. The regressions results are presented in

column (3) and column (4) of Table 2 and Table 3 where control variables are both excluded

---

[22] To ensure our results were not influenced by price outliers, we omitted from the sample those MSAs with the highest commodity prices (e.g., New York, Los Angeles, San Francisco) and re-estimated all the regressions. The results from the regressions that omitted the highest-priced cities were nearly identical to those presented here.

(column 3) and included (column 4). As expected based on our earlier conceptual framework, the real-income elasticity of demand is different than the nominal-income elasticity of demand in each case.[23] Specifically, the estimated income elasticity is larger when we consider real income than when we consider nominal income, especially for instant lottery sales.

Consider specific regression results. For instant lottery sales, the nominal-income elasticity is 0.268 (column 2 of Table 2) and is not statistically different than zero, whereas the real-income elasticity for instant lottery sales (column 4 of Table 2) is 0.817 and is not statistically different than 1. For online lottery sales, the nominal-income elasticity estimate (column 2 of Table 3) is 1.106 and the real-income elasticity (column 4 of Table 3) is 1.260, both of which are statistically different than zero and not statistically different than one.

In agreement with our predictions, the results reveal the nominal-income elasticity of demand that omits local cost of living can be quite different from the real-income elasticity which considers the local cost of living. We find that for instant sales and online sales, the real-income elasticity is greater than the nominal-income elasticity in each case. The difference in the nominal and real elasticities is more pronounced for instant lottery sales. The local cost of living appears to play an important role in determining the income elasticity of demand for lottery tickets.

V. Concluding Comments

A growing body of literature argues that empirical modeling of cross-sectional data should account for geographic variation in the cost of living, much in the same way that time-

---

[23] Because the nominal and real income elasticities are each estimated in separate regressions, we cannot directly test for their equality (say, using a pairwise t-test). We thus refer to significant quantitative differences rather than statistically significant differences when claiming coefficients are different. In several cases below we compare coefficients based on each coefficient's statistical difference from one or from zero which provides additional evidence on significant quantitative differences in the coefficient estimates.

series data are frequently adjusted for inflation in order to make accurate comparisons over time. Accounting for differences in purchasing power across locations when using cross-sectional data is just as important as accounting for differences in purchasing power across time.

In this paper, we explored the role of geographic variation in cost of living in empirical models of demand for state-lottery tickets across MSAs in the United States. Previous research on the distributional burden of lottery ticket expenditures has used only nominal income and nominal lottery sales across locations. We argued that the failure to consider locational cost-of-living differences in empirical models of lottery demand may yield incorrect estimates of the income elasticity of demand. Specifically, our conceptual framework suggested that an estimated nominal-income elasticity of demand will be different than the real-income elasticity of demand, except in the very special case when the share of income devoted to lottery expenditures is the same in all locations.

In accordance with our conceptual framework, our estimated income elasticities are different when controlling for cost-of-living differences across MSAs. Our estimated real-income elasticities are larger than nominal-income elasticities, and thus reveal that the regressivity of state lotteries may be overstated when locational differences in the cost of living are ignored. Our results suggest that if individual-level data on lottery expenditures were available, then cost-of-living differences across individuals' locations should be considered in these individual-level models of lottery demand as well.[24]

---

[24] There is considerable debate in the literature as to which price index best captures the cost of living (see footnote #12). While we did not enter the debate on this issue as it would take us away from our main focus, we should note that we re-estimated our models using the new MSA-level composite price indexes developed by Carrillo, Early, and Olsen (2012), hereafter CEO. Our real income elasticity estimates using the CEO index are quite similar to the estimates we obtained using the ACCRA price index. Specifically, our income elasticity estimates using the CEO index are as follows (each significant at 5 percent or better): Real income elasticity for instant sales = 0.532; real income elasticity for online sales = 1.187. Compare these results with those in column (4) of Tables 2 and 3, respectively.

Our analysis of lottery demand across MSAs demonstrates the importance of accounting for differences in the cost of living (or prices, more generally) across geographic locations. Although we focused on the single example of the demand for lottery tickets across MSAs, the conceptual framework and its conclusions presented here are applicable to any cross-sectional analysis using monetary variables (e.g., government spending, sales, wages, etc.) at any level of geographic aggregation (e.g., zip codes, counties, cities, states).[25] Since we find that controlling for geographic variation in the cost of living matters for lottery demand, it is reasonable to believe that doing so also matters for various other issues as well, such as tax incidence, impact of intergovernmental grants on local spending, gender and racial wage gaps, income and tax inequality, and union wage premiums, to name just a few. Given that research in these areas often has important policy implications, it may be beneficial for future research to revisit previous analyses and consider that accounting for the geographic variation in the cost of living may provide alternative results and policy recommendations.

---

[25] Our results raise an issue if panel data are used, namely that the units of observation should be adjusted for differences in the cost of living across time (which is usually done) *and* differences in the cost of living across locations. Exactly how this should be done is an interesting question for future research.

Appendix

The following MSAs are used in the analysis (2000 U.S. Census definitions).

Abilene, TX	El Paso, TX	Modesto, CA	South Bend, IN
Akron, OH	Elkhart-Goshen, IN	Muncie, IN	Spokane, WA
Albuquerque, NM	Eugene-Springfield, OR	Nassau-Suffolk, NY	Springfield, IL
Amarillo, TX	Evansville-Henderson, IN-KY	New York, NY	Springfield, MO
Austin-San Marcos, TX	Fort Myers-Cape Coral, FL	Newark, NJ	St. Cloud, MN
Beaumont-Port Arthur, TX	Fort Walton Beach, FL	Oakland, CA	St. Joseph, MO
Bellingham, WA	Fort Worth-Arlington, TX	Odessa-Midland, TX	St. Louis, MO-IL
Binghamton, NY	Fresno, CA	Olympia, WA	Syracuse, NY
Bloomington, IN	Glens Falls, NY	Orlando, FL	Tacoma, WA
Bloomington-Normal, IL	Grand Rapids-Muskegon-Holland, MI	Panama City, FL	Tallahassee, FL
Boise City, ID	Houston, TX	Pensacola, FL	Tampa-St. Petersburg-Clearwater, FL
Bremerton, WA	Indianapolis, IN	Peoria-Pekin, IL	Terre Haute, IN
Brownsville-Harlingen-San Benito, TX	Jacksonville, FL	Phoenix-Mesa, AZ	Toledo, OH
Bryan-College Station, TX	Joplin, MO	Portland-Vancouver, OR-WA	Topeka, KS
Buffalo-Niagara Falls, NY	Kansas City, MO-KS	Richland-Kennewick-Pasco, WA	Tucson, AZ
Cedar Rapids, IA	Killeen-Temple, TX	Richmond-Petersburg, VA	Tyler, TX
Champaign-Urbana, IL	Lafayette, IN	Riverside-San Bernardino, CA	Visalia-Tulare-Porterville, CA
Chicago, IL	Lansing-East Lansing, MI	Roanoke, VA	Waco, TX
Cincinnati, OH-KY-IN	Las Cruces, NM	Rochester, MN	Waterloo-Cedar Falls, IA
Cleveland-Lorain-Elyria, OH	Lexington, KY	Rockford, IL	West Palm Beach-Boca Raton, FL
Columbia, MO	Lima, OH	Sacramento, CA	Wichita Falls, TX
Columbus, OH	Longview-Marshall, TX	Salem, OR	Wichita, KS
Dallas, TX	Los Angeles-Long Beach, CA	San Antonio, TX	Yakima, WA
Davenport-Moline-Rock Island, IA-IL	Louisville, KY-IN	San Diego, CA	Youngstown-Warren, OH
Daytona Beach, FL	Lubbock, TX	San Francisco, CA	Yuma, AZ
Dayton-Springfield, OH	Lynchburg, VA	Santa Barbara-Santa Maria-Lompoc, CA	
Decatur, IL	Mansfield, OH	Santa Fe, NM	
Des Moines, IA	McAllen-Edinburg-Mission, TX	Sarasota-Bradenton, FL	
Detroit, MI	Miami, FL		

MSA Price Indexes - Notes

There are several issues regarding the MSA-level price indexes that we had to address. First, the Council for Community and Economic Research's MSA cost-of-living index definitions are based on the U.S. Census's 2003 revised MSA boundary definitions, some of

which differ slightly from the 2000 Census MSA boundary definitions. In these cases, we attempted to match MSA definitions to ensure that the price indexes and other data covered the same geographic area as closely as possible. We will gladly provide a list of all matched non-identical MSAs upon request. See http://www.coli.org/surveyforms/SampleData.zip for more information.

Second, the Council for Community and Economic Research COLI are quarterly in frequency, but indexes for each quarter are not available for all MSAs. The annual price indexes used in our paper are averages of all quarters available for each MSA in 2000. There is little difference in the COLI in each MSA from one quarter to the next, and as such the findings regarding the impact of prices on lottery demand were not qualitatively different when we used price indexes for a specific quarter rather than the average of all available quarters in the year. ∎

References

Albouy, David. "The Unequal Geographic Burden of Federal Taxation." *Journal of Political Economy*, vol. 117, no. 4 (August 2009): 635-667.

Black, Dan; Natalia Kolesnikova, and Lowell Taylor. "Earnings Functions When Wages and Prices Vary by Location." *Journal of Labor Economics*, vol. 27, no. 1 (January 2009): 21-47.

Black, Dan; Natalia Kolesnikova, and Lowell Taylor. "Local Labor Markets and the Evolution of Inequality" *Annual Review of Economics*, vol. 6 (2014) forthcoming.

Black, Dan; Natalia Kolesnikova, Seth Sanders, and Lowell Taylor. "Are Children 'Normal'?" *Review of Economics and Statistics*, vol. 95, no.1, (March 2013): pp. 21–33.

Carrillo, Paul E., Dirk W. Early, and Edgar O. Olsen. "A Panel of Price Indices for Housing Services, Other Goods, and All Goods for All Areas in the United States 1982-2010." University of Virginia Working Paper, 2012.

Cebula, Richard and Christopher K. Coombs. "Recent Evidence on Factors Influencing the Female Labor Force Participation Rate." *Journal of Labor Research*, vol. 29, no. 3, 2008: 272-284.

Chen, Y. and S. Rosenthal (2008). "Local Amenities and Life Cycle Migration: Do People Move for Jobs or Fun?" *Journal of Urban Economics*, vol. 64, no. 3 (November 2008): 519-37.

Clotfelter, Charles, and Philip Cook. "Implicit Taxation in Lottery Finance." *National Tax Journal* 40 No. 4 (December 1987): 533-46.

Clotfelter, Charles, and Philip Cook. *Selling Hope: State Lotteries in America.* Cambridge: Harvard University Press, 1989.

Elliot, Donald, and John Navin. "Has Riverboat Gambling Reduced State Lottery Revenue?" *Public Finance Review* 30 No. 3 (May 2002): 235-47.

Farrell, Lisa, Edgar Morgenroth, and Ian Walker. "A Time Series Analysis of U.K Lottery Sales: Long and Short-run Price Elasticities." *Oxford Bulletin of Economics and Statistics* 61 No. 4 (November 1999): 513-26.

Forrest, David, O. David Gulley, and Robert Simmons. "Elasticity of Demand for UK National Lottery Tickets." *National Tax Journal* 53 No. 4 (Dec. 2000): 853-63.

Gabriel, Stuart and Stuart Rosenthal. 2004. "Quality of the business environment versus quality of life: do firms and households like the same cities?" *Review of Economics and Statistics* 86 no. 1 (February 2004): 438-444.

Garrett, Thomas, and Russell Sobel. "Gamblers Favor Skewness, Not Risk: Further Evidence from United States' Lottery Games." *Economics Letters* 63 No. 1 (April 1999): 85-90.

Garrett, Thomas, and Cletus Coughlin. "Inter-temporal Differences in the Income Elasticity of Demand for Lottery Tickets." *National Tax Journal*, vol. 62, no. 1, March 2009, pp. 77-99

Ghent, Linda S. and Alan P. Grant. "The Demand for Lottery Products and Their Distributional Consequences." *National Tax Journal*, vol. 63, no. 2, June 2010, pp. 253-268.

Hansen, Ann. "The Tax Incidence of the Colorado State Lottery Instant Game." *Public Finance Review*, vol. 23, no. 3 (July 1995): 385-398.

Kearney, Melissa Schnettini. "State Lotteries and Consumer Behavior." *Journal of Public Economics* 89 Nos. 11-12 (December 2005): 2269-99.

Lafleur's 2009 World Lottery Almanac, eds. Teresa LaFleur and Bruce LaFleur, TLF Publications, Boyds, Maryland, 2009.

Miksell, John. "A Note on the Changing Incidence of State Lottery Finance." *Social Science Quarterly* 70 No. 2 (June 1989): 513-21.

Mikesell, John. "State Lottery Sales and Economic Activity." *National Tax Journal* 47 No. 1 (March 1994): 165-71.

Moretti, Enrico. "Real Wage Inequality." Working Paper, Department of Economics, University of California Berkeley, 2011.

Nonnemaker, James; Eric Finklestein; Mark Engelen, Thomas Hoerger, and Matthew Farrelly. "Have Efforts to Reduce Smoking Really Contributed to the Obesity Epidemic?" *Economic Inquiry*, vol. 47, no. 2 (April 2009): 366-376.

Perez, Levi and Brad Humphreys. "The Income Elasticity of Lottery: New Evidence from Micro Data." *Public Finance Review* 39 (July 2011): 551-570.

Price, Donald, and E. Shawn Novak. "The Tax Incidence of Three Texas Lottery Games: Regressivity, Race, and Education." *National Tax Journal* 52 No. 4 (December 1999): 741-51.

Price, Donald and E. Shawn Novak. "The Income Redistribution Effects of Texas State Lottery Games." *Public Finance Review,* vol. 28, no. 1 (January 2000): 82-92

Scott, Frank, and John Garen. "Probability of Purchase, Amount of Purchase, and the Demographic Incidence of the Lottery Tax." *Journal of Public Economics* 54 No. 1 (May 1994): 121-43.

Table 1: Descriptive Statistics

	Variables in Levels		Variables in Natural Logarithms	
	Mean	Standard Error	Mean	Standard Error
Lottery Sales				
Instant, Nominal Per Capita ($)	58.44	23.87	3.986	0.410
Online, Nominal Per Capita ($)	57.96	36.43	3.886	0.584
Instant, Real Per Capita ($)	58.51	26.61	3.964	0.469
Online, Real Per Capita ($)	55.93	32.64	3.865	0.567
Other Variables				
Nominal Per Capita Income ($)	20,516	3,635	9.913	0.181
Real Per Capita Income ($)	20,005	2,937	5.286	0.164
Population Density	395.05	668.48	-------	-------
% Population with Bachelor's	24.05	6.69	-------	-------
Age of Lottery (Years)	15.88	7.56	-------	-------
Multistate (1=Yes, 0 = No)	0.595	0.493	-------	-------
Casino (1=Yes, 0 = No)	0.261	0.441	-------	-------
Age of Multistate Lottery	3.378	3.205	-------	-------
Cost of Living (COLI)	103.40	19.36	-------	-------

Note: Sample size is 111 MSAs for the year 2000. See Appendix for a list of MSAs.

Table 2: Instant Lottery Sales

Variable	Dependent Variable: Nominal Sales Per Capita (ln)		Dependent Variable: Real Sales Per Capita (ln)	
	(1)	(2)	(3)	(4)
Nominal Per Capita Income (ln)	-0.322 (0.200)	0.268 (0.236)	-------	-------
Real Per Capita Income (ln)	-------	-------	0.462 (0.302)	0.817** (0.242)
Population Density	-------	-0.061** (0.021)	-------	-0.100** (0.035)
% pop. with ≥ 4yr. college degree	-------	-0.022** (0.005)	-------	-0.033** (0.006)
Age of lottery	-------	0.049** (0.013)	-------	0.045** (0.015)
Age of multi-state lottery	-------	-0.107** (0.046)	-------	-0.113** (0.049)
Multi-state lottery dummy variable	-------	1.121** (0.317)	-------	1.189** (0.350)
Casino dummy variable	-------	-0.441* (0.223)	-------	-0.478** (0.240)
Constant	7.178** (1.992)	0.997 (2.277)	-3.084* (1.606)	-4.977** (1.132)
State dummy variables	No	Yes	No	Yes
Adjusted R^2	0.011	0.514	0.017	0.590

Notes: * denotes significance at 10 percent, ** at 5 percent or better. White's heteroskedasticity-consistent standard errors are shown in parentheses. The coefficient on population density is multiplied by 1,000. Number of observations =111. Unit of observation is MSA.

Table 3: Online Lottery Sales

Variable	Dependent Variable: Nominal Sales Per Capita (ln)		Dependent Variable: Real Sales Per Capita (ln)	
	(1)	(2)	(3)	(4)
Nominal Per Capita Income (ln)	1.094** (0.290)	1.106** (0.185)	-------	-------
Real Per Capita Income (ln)	-------	-------	0.856** (0.328)	1.260** (0.203)
Population Density	-------	0.099 (0.060)	-------	0.128** (0.058)
% pop. with ≥ 4yr. college degree	-------	-0.024** (0.004)	-------	-0.025** (0.004)
Age of lottery	-------	0.081** (0.007)	-------	0.081** (0.007)
Age of multi-state lottery	-------	-0.097** (0.021)	-------	-0.097** (0.020)
Multi-state lottery dummy variable	-------	1.529** (0.175)	-------	1.531** (0.165)
Casino dummy variable	-------	-0.854** (0.127)	-------	-0.869** (0.124)
Constant	-6.962** (2.866)	-8.328** (1.770)	-5.267** (1.733)	-8.645** (1.064)
State dummy variables	No	Yes	No	Yes
Adjusted R^2	0.107	0.828	0.053	0.820

Notes: * denotes significance at 10 percent, ** at 5 percent or better. White's heteroskedasticity-consistent standard errors are shown in parentheses. The coefficient on population density is multiplied by 1,000. Number of observations =111. Unit of observation is MSA.

Figure 1: The Relationship between Demand and Income – Real vs. Nominal Income

Natural Log of
Lottery Sales (X)

$\ln(X_N^H)$ — Slope $= \eta_N$

$\ln(X_{R1}^H)$

$\ln(X_{R2}^H)$

$\ln(X_{R2}^L)$ — Slope $= \eta_{R2}$

$\ln(X_{R1}^L)$ — Slope $= \eta_{R1}$

$\ln(X_N^L)$

$\ln(Y_N^L)$ $\ln(Y_{R1,R2}^L)$ $\ln(Y_{R1,R2}^H)$ $\ln(Y_N^H)$ Natural Log of Income (Y)