Accounting for Geographic Variation in the Cost of Living: An Application to State-Lottery Demand

Local Cost of Living and Demand Estimation: An Application to State Lotteries

Thomas A. Garrett
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442
garrett@stls.frb.org
314-444-8601

Natalia A. Kolesnikova
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442
natalia.a.kolesnikova@stls.frb.org
314-444-8549

Abstract

The cost of living varies as much across regions as it does across time, but researchers have only begun to acknowledge the importance of controlling for regional differences in the cost of living when conducting cross-sectional analyses. We demonstrate the importance of considering regional cost-of-living differences by using empirical models of demand for state lotteries. Previous research on state lotteries has shown that the nominal-income elasticity of demand for lottery tickets is less than 1, suggesting expenditure on lottery tickets is regressive. We re-estimate traditional models of lottery demand using a sample of metropolitan statistical areas, but we also control for cost-of-living differences across these areas. We find that, in accordance with our conceptual framework, estimated income elasticities are larger when we control for local cost-of-living variation – that is, the regressivity of state lotteries is overstated when local cost-of-living variation is omitted from empirical models of lottery demand. Our results are robust to several measures of the cost of living, including housing price indexes.

Local Cost of Living and Demand Estimation: An Application to State Lotteries

I. Introduction

It is standard procedure to adjust nominal incomes for inflation when comparing them over time. Yet, it is common practice to compare only nominal incomes in comparisons across locations at a given time despite the fact that cost-of-living differences across 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 regions. 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 regions as well.[1]

Several recent studies have demonstrated the importance of considering local cost of living. Black et al. (2009) show that differences in local prices—namely, housing prices—help to explain significant differences in the college wage premium across cities: The college wage premium is relatively lower in more expensive cities. 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

---

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

that results from the differences in regional 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.,

forthcoming).

We argue that considering differences in local cost of living is also important when

estimating demand equations, especially when the data are at an aggregate level.   In this paper,

we consider the demand for lottery tickets, which is a topic of important policy debate. Research

has shown that lower-income individuals spend a greater percentage of their income on lottery

tickets than do wealthier individuals.[2] As a result, the distributional burden of expenditures on

lottery tickets is generally characterized as regressive. Although state taxes such as sales taxes

and excise taxes are also regressive, the distributional burden of lottery expenditures has received

much more critical attention given the revenue maximization objective of state-lottery agencies

and the moral opposition of some groups toward state-sponsored gambling.

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 1

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. Most 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 cost-of-living differences across locations. Based on

the findings of Black et al. (2009), Albouy (2009), Moretti (2011), and Black et al.

---

[2] 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).

(forthcoming), we argue that the empirical exercise should include some measure of cost of living in order to capture differences in purchasing power across regions.

In this paper, we demonstrate both conceptually and empirically that the failure to consider differences in the cost of living across metropolitan statistical areas (MSAs) when estimating lottery demand yields an income elasticity of demand that is too low. That is, the regressivity of state lotteries is overstated when regional differences in the cost of living are not taken into account. 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* consumption prices in city 1 are twice as high as those in city 2. If we ignore consumption prices 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.

Our paper proceeds in several steps. In Section II we provide the conceptual framework for the inclusion of local cost of living in demand equations. Our framework has broad applications, as it can be applied to the estimation of any demand equation using cross-sectional or panel data. We discuss our data in Section III of the paper. The results are presented and discussed in Section IV, and the final section is reserved for concluding comments.

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. Thus, the estimation must rely on

aggregate-level data. The income elasticity of demand for lottery tickets (β_1) is traditionally

obtained by estimating the following equation

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

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 variables such as

education levels, unemployment rates, the age of the population, poverty rates, and lottery game

characteristics.[3] Given the aggregated unit of observation, the equation relates the differences in

incomes across regions to the differences in lottery sales across regions. Importantly, the cost of

living in region 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

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). Equation (2) then can be used to estimate the real-

income elasticity of demand for lottery tickets:

$$\ln(X_i) = \beta_2 \cdot \ln((Y/COL)_i) + \theta_2 \cdot \mathbf{Z}_i + \varepsilon_i, \qquad (2)$$

where COL is the local cost of living and Y/COL is real income. The estimated coefficient β_2 is

real-income elasticity of lottery demand.

The following exercise provides insight into the difference in the nominal-income

elasticity of demand estimated from equation (1) and the real-income elasticity of demand

---

[3] The price of lottery tickets is generally omitted from equation (1) because most lottery tickets cost $1. Two issues are worth clarifying. First, we refer to the dollar price of a lottery ticket, not the expected return, which is calculated using the probability of winning each prize. Second, although some instant ("scratch-off") lottery tickets cost more than $1, we do not have game-level sales data. We thus follow the existing literature and explicitly model the demand for lottery sales rather than the demand for the quantity of tickets.

estimated from equation (2). Suppose there is a set of regions each having different nominal income (Y_N). Let region L and region H be two representative regions where region L has low nominal income (Y_N^L) and region H has high nominal income (Y_N^H). The quantity of X purchased in each region is X^L and X^H, respectively, with $X^L < X^H$ because of (assumed) normality. The relationship between Y_N and X is denoted by line Y_N in Figure 1.

Now consider that the cost of living (COL) differs across regions and region H has a higher cost of living than region L ($COL^H > COL^L$). For simplicity, let $COL^L < 1$ and $COL^H > 1$.[4] Then real income in region L is $Y_N^L / COL^L = Y_R^L$ and real income in region H is $Y_N^H / COL^H = Y_R^H$. It is thus the case that $Y_R^L > Y_N^L$ and $Y_R^H < Y_N^H$. That is, real income is greater than nominal income in low-income regions and real income is lower than nominal income in high-income regions. The relationship between real income and X is denoted by the line Y_R in Figure 1. It is evident from Figure 1, using discrete notation, that $(\Delta Y_N / Y_N) > (\Delta Y_R / Y_R)$ between X^L and X^H. For any X and Y, income elasticity (η) is calculated as $\eta = (\Delta X / X)/(\Delta Y / Y)$. Thus, $\eta_N < \eta_R$; the nominal-income elasticity of demand is less than the real-income elasticity of demand. Therefore, the income elasticity (β_1) from equation (1) using nominal income should be less than the income elasticity (β_2) from equation (2) using real income.

[Figure 1]

If the researcher wishes to disentangle the effects of nominal income (holding COL constant) and COL, equation (2) may not be appropriate. As a result, estimating the following version of equation (2) might be more suitable:

$$\ln(X_i) = \beta_3 \cdot \ln(Y_i) + \alpha_3 \cdot \ln(COL_i) + \theta_3 \cdot \mathbf{Z}_i + \varepsilon_i, \tag{3}$$

where β_3 is the estimate of the nominal-income elasticity of demand for lottery tickets while

---

[4] The assumption that $COL^L < 1$ and $COL^H > 1$ is for graphical simplicity. Both regions could have a COL>1 or a COL <1. This would simply change the y-intercept of Y_R and Y_N, with no change in their slopes. As long as $COL_H > COL_L$, the slope of Y_R in Figure 1 will be less than the slope of Y_N.

controlling for differences in the local cost of living. Notice that equations (2) and (3) would be identical if $\beta_3 = -\alpha_3$. Equation (3) is more flexible as it does not require the coefficients on income and cost of living to be of equal magnitude.

As long as COL is not included in the regression equation, the income elasticity estimate in equation (1) will be biased. The direction of the bias is determined by the product of the sign of α_3 and (in part) the sign of the correlation between income and the cost of living (Greene, 2000; Jargowsky, 2005).[5] With respect to the correlation between nominal income and COL, our data reveal a positive and statistically significant correlation of 0.52 between MSA income and prices: Higher-income cities have higher prices. Regarding the relationship between COL and X in equation (3), the coefficient α_3 will be positive (negative) if lottery tickets and the composite good are gross substitutes (complements). Because lottery tickets are a very small part of individuals' consumption bundles, lowering the purchasing power by increasing the cost of living leads to a reduction in lottery tickets sales, suggesting α_3 is negative.[6]

Therefore, a negative (and statistically significant) α_3 and a positive correlation between nominal income Y and COL suggests that estimating equation (1), as has been done in previous research, will result in an income elasticity of demand that is biased downward. Estimating equation (3), which accounts for regional price variation, should yield an income elasticity that is larger than that obtained from equation (1). Because equations (2) and (3) are similar, we should find that the income elasticities from these equations are similar and are both larger than the nominal-income elasticity estimated from equation (1).

---

[5] The bias in β_3 is also determined by the correlation between income and the other variables in **Z.** The average correlation between income and our other independent variables is 0.18. Because income is positively correlated with local prices (see text) and, on average, with the other independent variables, there is no ambiguity in the final determination of the degree of bias once the sign of α_3 is known.
[6] The majority of studies have found that lottery tickets are normal goods.

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.[7] 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.[8] The COLI measures relative price levels for consumer goods and services and has been used in previous studies to capture price variation across locations.[9] The price index for each MSA is interpreted as a percentage of the average for all

---

[7] 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 1.

[8] See www.coli.org/ for a description of how the price indexes are calculated and for a list of all commodities included in each price index. The Council for Community and Economic Research is formerly known as ACCRA (American Chamber of Commerce Research Association).

[9] See, for example, Cebula and Coombs (2008), Nonnemaker et al. (2009), and Moretti (2011).

urban areas (where the average is set to 100).[10] The COLI is a weighted average of price indexes

of six different categories: groceries, housing, utilities, transportation, health care, and

miscellaneous good and services.[11] 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.[12] We then contacted state lottery agencies

and obtained county-level sales data for the year 2000. Lottery sales at the county-level were

then summed to arrive at lottery sales at the MSA-level.

We also obtained MSA-level sales for instant lottery games ("scratch-offs") and online

lottery games (e.g., Lotto, Mega Millions, Powerball) separately in a manner identical to that of

total lottery sales (the sum of instant game sales and online game sales) to explore the role of

local prices in explaining the demand for different lottery products.[13] In addition, considering

different types of lottery games gives us more ways to test our hypothesis. Research has shown

that the income elasticities of demand for online games and instant games can be quite different,

---

[10] Several issues involving the price data are discussed in the Appendix.
[11] 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.
[12] See http://www.census.gov/population/www/metroareas/lists/historical/99mfips.txt.
[13] 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 on television only several times a week.

with the former having a higher income elasticity of demand (even suggesting progressivity).[14]

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. In particular, 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.[15] Lottery-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.[16] 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 is included to capture 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

---

[14] Research by Mikesell (1989), Oster (2004), and Garrett and Coughlin (2009) suggests that large-jackpot online lottery games may attract wealthier players than lower-jackpot instant games; as a result, the income elasticity of demand for some online games can be greater than for instant lottery games. These studies find evidence that online games can be progressive. 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.

[15] 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.

[16] 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.

and Mega Millions) generate the largest jackpots, and thus states 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).

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.[17] The MSAs used in the analysis are listed in the Appendix.

IV. Empirical Results

We estimate equations (1), (2), and (3) using per capita total lottery sales, instant lottery sales, and 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.[18] Because we wish to obtain the income elasticity of demand, lottery sales and per capita income are converted to natural logarithms before estimation. We also transform COLI into natural logarithms.

For each game category we compare the income elasticity estimates from equation (1) with the elasticity estimates from equations (2) and (3). As our previous discussion suggests, we expect to find that the income elasticity coefficient from equation (1) is smaller than the income elasticity coefficients from equations (2) and (3). This would support our hypothesis that local prices play an important role in explaining cross-sectional differences in lottery demand, and that

---

[17] 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.

[18] 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).

the failure to include local cost of living in models of lottery demand can result in income elasticity estimates that are biased downward, thus providing an inaccurate estimate of the distributional burden of lottery ticket expenditures.

The empirical results from our models of lottery demand (total sales, instant sales, and online sales) are shown in Table 2. All equations were estimated by GLS using White's heteroskedasticity-corrected standard errors.[19] The results presented in columns (1), (4), and (7) of Table 2 are from equation (1), which includes only nominal income. The key feature of these "traditional" models is that they omit the local cost of living measure, thus failing to capture how differences in the purchasing power of income influence lottery sales across MSAs. Our nominal-income elasticity estimates from equation (1) for total sales and instant-lottery sales are 0.62 and 0.27, respectively, and fall within the range of estimates obtained in earlier studies. Notice that online lottery sales are found to be progressive, but the estimated nominal-income elasticity of demand of 1.11 is not significantly different than 1 (thus suggesting proportionality). This finding is similar to that of Garrett and Coughlin (2009), who used county-level data for several states, and Perez and Humphreys (2011), who used individual-level data from the Spanish lottery.

[Table 2]

Next, we estimate equation (2) where the key independent variable is real personal per capita income. The results from this exercise are reported in columns (2), (5), and (8) of Table 2. As expected, the real-income elasticity of demand is higher than the nominal-income elasticity of demand in each case. In other words, lotteries appear to be less regressive when we consider real income than when we consider nominal income. For total lottery sales, the estimated nominal-

---

[19] 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.

income elasticity (column 1) is 0.62 and implies regressivity, whereas the estimated real-income elasticity (column 2) is much higher (0.76) and not statistically different from 1, which suggests expenditures on lotteries are proportional. The nominal-income elasticity for instant lottery sales is 0.27 and is not statistically different from zero (column 4), whereas the real-income elasticity for instant lottery sales (column 5) is 0.65 and is statistically significant. For online lottery sales, the nominal-income elasticity (column 7) is 1.11, whereas the real-income elasticity (column 8) is 1.14. Online lottery sales appear to be slightly progressive, although both coefficients are not statistically different from 1.

Finally, we estimate equation (3), which includes both nominal per capita income and COLI as separate independent variables. We find that COLI has a negative and statistically significant influence on all three categories of lottery sales (columns 3, 6, and 9 in Table 2). This supports our hypothesis that lower real-income due to higher consumption prices leads to a reduction in lottery ticket sales. The coefficients on COLI are interpreted as meaning that an MSA with a cost of living that is 10 percent higher is expected to have lottery ticket sales 7 to 14 percent lower, ceteris paribus.

Our theoretical discussion suggested that if local cost of living is a negative and significant determinant of lottery sales, then models of lottery demand that fail to consider the local cost of living should produce nominal-income elasticities of demand that are biased downward. The downward bias in the nominal-income elasticity of demand for each category of lottery sales is evident in Table 2. For total lottery sales, the nominal-income elasticity of demand increases from 0.62 to 0.75 (an increase of 21 percent) when COLI is included in the model. As a result, lotteries in our sample of MSAs change from regressive when local prices

13

are omitted to not statistically different than proportional when local prices are included.[20] For

instant lottery sales, the income elasticity changes from 0.27 and not statistically different from

zero when COLI is omitted to 0.49 and significantly different from zero when COLI is included.

Finally, online lottery sales are slightly progressive in both models based on the income elasticity

point-estimates (not statistically different than 1), but the degree of progressivity is higher when

COLI is included. The income elasticity of demand for online tickets increases from 1.11 to

1.20, an increase of 8 percent.

A summary of our results thus far is useful. In agreement with our predictions, the results

reveal that when the local cost of living is controlled for, the income elasticity of demand

increases (and the regressivity of instant and total lottery sales decreases), as the income

elasticity is biased downward when local cost of living is omitted from the model. The overall

local cost of living has been shown to play an important role in the demand for lottery tickets and

in determining the income elasticity of demand.

As a practical matter, unfortunately, data on local cost of living might not be readily

available depending on the level of aggregation. However, if at least some measure of local cost

of living is available, then the following equation can be estimated:

$$\ln(X_i) = \beta_4 \cdot \ln(Y_i) + \alpha_4 \cdot \ln(\widetilde{COL}_i) + \theta_4 \cdot \mathbf{Z}_i + \varepsilon_i, \tag{4}$$

where $\widetilde{COL}_i$ is *some proxy* of the cost of living in region *i*.

We wish to investigate whether housing prices alone are a reasonable proxy for the cost

of living, given that housing prices are generally available for more regions and levels of

aggregation than COLI.   We want to know if equations (3) and (4) provide similar results when

equation (4) is estimated using a measure of housing prices.

---

[20] The income elasticity in column (3) of Table 2 is not statistically different from 1.

The largest component of COLI is housing prices, and the housing component of COLI is highly correlated with the composite COLI ($\rho = 0.98$). We find, not surprisingly, that equation (3) and equation (4), which includes the housing component of COLI, produce very similar income elasticities. Thus, to better evaluate the validity of housing price as a control for differences in the cost of living between regions, we need a measure of MSA housing prices from a different source. We therefore use a quality-adjusted housing price index developed by Chen and Rosenthal (2008), which we normalize to be between 0 and 1.[21]

Table 3 reports the results using this quality-adjusted house price index. The second column for each lottery category presents the results from estimating equation (4) with the quality-adjusted housing price index. To aid in comparison, the results using COLI from columns (3), (6), and (9) in Table 2 are reproduced in columns (1), (3), and (5) of Table 3. Overall, the nominal-income elasticities estimated when controlling for housing prices (columns 2, 4, and 6) appear to be similar to the nominal-income elasticities estimated when controlling for overall cost of living (columns 1, 3, and 5, respectively). Thus, regional housing prices are found to be suitable proxies for regional cost-of-living differences when data on the latter may not be available.

[Table 3]

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-series data are frequently adjusted for inflation in order to make accurate comparisons over time.

---

[21] We thank Stuart Rosenthal for generously making the index available to us.

Accounting for differences in purchasing power across regions 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 across regions and has generally found the income elasticity of lottery demand to be less than 1, suggesting that state-lottery purchases are regressive. We argued that the failure to consider regional cost-of-living differences in empirical models of lottery demand will yield incorrect estimates of the income elasticity of demand. Specifically, our conceptual framework suggested that an estimated income elasticity of demand will be too low in the absence of control for regional cost of living— that is, the regressivity of state lotteries is overstated when regional cost of living is omitted from empirical models of lottery demand.

In accordance with our conceptual framework, our estimated income elasticities are larger when controlling for cost-of-living differences across MSAs. This conclusion holds when we consider both real income and nominal income with controls for cost of living across metropolitan areas. Our estimated income elasticities are roughly 8 to 20 percent larger when we control for cost of living, and thus reveal that the regressivity of state lotteries is overstated when regional 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.

One obstacle for researchers is the availability of regional data on the cost of living. The availability of data may dictate the sample chosen by the researcher (as we have done) or require the computation of regional prices (Moretti, 2011). Although cost-of-living data are not

available at many or all regional levels (e.g., zip codes, counties, cities, MSAs), data on housing prices are available for many regional levels. We showed that the cost of housing across MSAs is a suitable proxy for the overall cost of living – our empirical results using overall cost of living and the cost of housing are quite comparable. This suggests that researchers could use the several house-price indexes that are available at numerous levels of data aggregation, such as the Federal Housing Finance Agency (FHFA) House Price Index and the CoreLogic House Price Index, to control for differences in the cost of living across regions.

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). 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 in urban and labor economics and local public finance, such as gender and racial wage gaps, tax incidence, income 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.

17

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		Pensacola, FL	Tampa-St. Petersburg-Clearwater, FL
Bremerton, WA	Houston, TX	Peoria-Pekin, IL	Terre Haute, IN
Brownsville-Harlingen-San Benito, TX	Indianapolis, IN	Phoenix-Mesa, AZ	Toledo, OH
Bryan-College Station, TX	Jacksonville, FL	Portland-Vancouver, OR-WA	Topeka, KS
Buffalo-Niagara Falls, NY	Joplin, MO	Richland-Kennewick-Pasco, WA	Tucson, AZ
Cedar Rapids, IA	Kansas City, MO-KS	Richmond-Petersburg, VA	Tyler, TX
Champaign-Urbana, IL	Killeen-Temple, TX	Riverside-San Bernardino, CA	Visalia-Tulare-Porterville, CA
Chicago, IL	Lafayette, IN	Roanoke, VA	Waco, TX
Cincinnati, OH-KY-IN	Lansing-East Lansing, MI	Rochester, MN	Waterloo-Cedar Falls, IA
Cleveland-Lorain-Elyria, OH	Las Cruces, NM	Rockford, IL	West Palm Beach-Boca Raton, FL
Columbia, MO	Lexington, KY	Sacramento, CA	Wichita Falls, TX
Columbus, OH	Lima, OH	Salem, OR	Wichita, KS
Dallas, TX	Longview-Marshall, TX	San Antonio, TX	Yakima, WA
Davenport-Moline-Rock Island, IA-IL	Los Angeles-Long Beach, CA	San Diego, CA	Youngstown-Warren, OH
Daytona Beach, FL	Louisville, KY-IN	San Francisco, CA	Yuma, AZ
Dayton-Springfield, OH	Lubbock, TX	Santa Barbara-Santa Maria-Lompoc, CA	
Decatur, IL	Lynchburg, VA	Santa Fe, NM	
Des Moines, IA	Mansfield, OH	Sarasota-Bradenton, FL	
Detroit, MI	McAllen-Edinburg-Mission, TX		
	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, Seth Sanders, and Lowell Taylor. "Are Children 'Normal'?" *Review of Economics and Statistics*, forthcoming.

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.

20

Greene, William. *Econometric Analysis,* Saddle River, New Jersey: Prentice Hall, 2000.

Jargowsky, Paul A. "Omitted Variable Bias." *Encyclopedia of Social Measurement*, vol. 2, Kimberly Kempf-Leonard, ed., San Diego, California: Academic Press, 2005, pp. 919-924.

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.

Oster, Emily. "Are All Lotteries Regressive? Evidence from the PowerBall." *National Tax Journal* 57 No. 2 (June 2004): 179-87.

Perez, Levi and Brad Humphreys. "The Income Elasticity of Lottery: New Evidence from Micro Data." *Public Finance Review*, forthcoming 2011.

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.

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

	Mean	Standard Error	Minimum	Maximum
COLI	103.40	19.36	88.90	242.70
Lottery Sales ($)				
Per Capita Total	116.40	49.27	33.88	263.23
Per Capita Online	57.96	36.70	13.24	177.58
Per Capita Instant	58.44	23.87	20.64	131.33

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

Table 2: Estimated Lottery Demand – With and Without Local Cost of Living

Variable	Dependent Variable: Per Capita Total Lottery Sales (ln)			Dependent Variable: Per Capita Instant Lottery Sales (ln)			Dependent Variable: Per Capita Online Lottery Sales (ln)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Nominal Per Capita Income (ln)	0.62** (3.84)	-------	0.75** (4.86)	0.27 (1.14)	-------	0.49** (2.25)	1.11** (5.97)	-------	1.20** (5.82)
Real Per Capita Income (ln)	-------	0.76** (5.32)	-------	-------	0.65** (3.03)	-------	-------	1.14** (5.85)	-------
COLI (ln)	-------	-------	-0.81** (2.05)	-------	-------	-1.41** (2.52)	-------	-------	-0.74* (1.86)
Population Density	0.06** (2.08)	0.16** (4.96)	0.16** (2.71)	-0.06** (2.89)	0.01 (0.45)	0.12 (1.62)	0.10 (1.64)	0.24** (4.07)	0.19** (2.47)
% Pop. with $\geq$ 4yr. degree	-0.02** (5.66)	-0.20** (5.53)	-0.02** (5.32)	-0.02** (3.85)	-0.25** (4.83)	-0.02** (3.19)	-0.02** (6.35)	-0.02** (4.68)	-0.02** (5.82)
Age of lottery	0.06** (7.14)	0.06** (7.56)	0.06** (7.54)	0.05** (3.85)	0.04** (3.52)	0.04** (3.96)	0.08** (10.88)	0.09** (13.10)	0.08** (12.03)
Age of multi-state lottery	-0.12** (3.90)	-0.12** (5.18)	-0.12** (5.16)	-0.11** (2.31)	-0.11** (2.72)	-0.11** (3.13)	-0.10** (4.59)	-0.09** (6.21)	-0.10** (5.65)
Multi-state lottery dummy variable	1.27** (5.91)	1.16** (6.81)	1.15** (6.27)	1.12** (3.54)	1.01** (3.68)	0.91** (3.63)	1.53** (8.72)	1.53** (10.72)	1.53** (10.13)
Casino dummy variable	-0.57** (3.87)	-0.52** (4.20)	-0.52** (4.24)	-0.44* (1.98)	-0.41** (2.03)	-0.35** (2.00)	-0.85** (6.74)	-0.86** (7.25)	-0.87 (7.22)
Constant	-2.20 (1.40)	-0.01 (0.01)	0.34 (0.14)	1.00 (0.44)	0.42 (0.41)	5.38 (1.63)	-8.33** (4.70)	-3.70** (3.63)	-5.97** (2.63)
State dummy variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.77	0.80	0.79	0.51	0.55	0.57	0.83	0.83	0.83

Notes: * denotes significance at 10 percent, ** at 5 percent or better. Absolute t-statistics are in parentheses and are based on White's heteroskedasticity-consistent standard errors.   The coefficient on population density is multiplied by 1,000. Number of observations =111. Unit of observation is MSA.

Table 3: Estimated Lottery Demand with Local Housing Price as Proxy for Local Cost of Living

Variable	Dependent Variable: Per Capita Total Lottery Sales (ln)		Dependent Variable: Per Capita Instant Lottery Sales (ln)		Dependent Variable: Per Capita Online Lottery Sales (ln)	
	(1)	(2)	(3)	(4)	(5)	(6)
Nominal Per Capita Income (ln)	0.75** (4.86)	0.83** (5.04)	0.49** (2.25)	0.59** (3.04)	1.20** (5.82)	1.21** (5.50)
COLI (ln)	-0.81** (2.05)	-------	-1.41** (2.52)	-------	-0.74* (1.86)	-------
Quality-Adjusted Housing Price Index (ln)	-------	-0.22** (4.18)	-------	-0.40** (6.05)	-------	-0.11 (1.42)
Other RHS variables included	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.79	0.79	0.57	0.61	0.83	0.82

Notes: The coefficients in columns (1), (3), and (5) are from Table 2. All regressions include the RHS variables (not shown) listed in Table 2. * denotes significance at 10 percent, ** at 5 percent or better. Absolute t-statistics are in parentheses and are based on White's heteroskedasticity-consistent standard errors. The quality-adjusted housing price index is from Chen and Rosenthal (2008). Number of observations =111. Unit of observation is MSA.

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