How Lucky are Lucky Stores?

Anil Doshi and Joseph Gavinlertvatana

April 29, 2010

1 Introduction

Guryan and Kearney (2008) (henceforth referred to as GK) find that stores that sell a winning lottery ticket experience a prolonged increase in sales of that lottery. GK calls this phenomenon the "lucky store effect". Through further demographic analysis, GK demonstrate that it is more pronounced in poorer, less educated regions and regions with more elderly people. We reanalyze the data and find two deviations from their conclusion. First, after matching on demographics, the lucky store effect disappears, calling into question their demography-based results. Second, assuming there is in fact a demographic effect, the lucky store effect may not be magnified in elderly populations after controlling for income or education, as GK argue.

1.1 Summary of GK's Data and findings

The dataset consists of lottery ticket sales and wins at the retail store level in the state of Texas from January 2000 to June 2002. The dataset is comprised of 24,400 retailers in 1,386 cities, or 3,660 zip codes. The dataset contains lottery data from three games–Lotto Texas, Texas Two Step and Cash Five. During the span of the dataset, there were 68 Lotto Texas winners, 55 Texas Two Step winners and 571 Cash Five winners. Our analysis will coincide with GK's work by focusing on the Lotto Texas game.

GK primary conclusions are:

- A store experiences an increase in sales in the weeks following a winning ticket sale–GK calls this the **lucky store effect**.

- The lucky store effect is magnified in zip codes with high elderly populations, high levels of poverty, and undereducated populations.

1.2 Organization of Paper

Our results reveal two areas in GK's analysis that further investigation. First, in Section 2, we show that by matching treated with control units on zip code demographics, the lucky store effect disappears. Second, in Section 3, after controlling for level of poverty and education, we find that the lucky store effect is not magnified among the elderly population.

In addition to our contrary results, we undertake a number of other analyses that support GK's findings. In some of their models, GK drop 70% of treated observations. We demonstrate in Section 4 that Including these observations and replicating their analysis does not change their conclusions. In Section 5, we perform a matching analysis on previous sales and find that the lucky store effect remains intact. We take into consideration the possibility for serial correlation in Section 6. After taking autoregression into account, we find similar results to GK. In Section 7, we explain a technical error made in GK's analyis. Correcting for the error actually lends additional support to GK's findings. Last in Section 8, we end with some concluding remarks about the notion of "lucky stores" and how they relate to various cognitive biases introduced by GK.

2 Matching on Demographics

GK's investigation of demographic traits of lucky stores invite a matching process to arrive at a comparison of stores in similar demographic areas, where the treatment has sold a winning lottery ticket and the control has not. When we match on demographic data, we find that the lucky store effect is eliminated. We match on all control variables including date, sales history, and demographics such as population and racial composition. As shown in Table 1, the matching process arrives at a highly balanced control dataset.[1]

The most imbalanced covariate is date. We include date in our mean differences calculation to take the imbalance into account.[2] The matched dataset provides 41 treatment observations (14 omitted) and 147 control observations (1,807,814 omitted). Estimating the

[1] We experience an issue where the L-statistic for each covariate is 0, whereas the L-statistic for the model is 1.000.

[2] We do not model log sales based on all the covariates in the match because of the success of the matching model and because the model was based on a matrix not of full rank.

Table 1: Summary Balance Information from Demographic CEM

	Difference	L1-Stat	Min	0.25	0.5	0.75	Max
date	-2.1917	0.0000	0.0000	0.0000	0.0000	0.0000	7.0000
Population	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Median household inc	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Median family inc	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% white	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% non-white	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% black	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% hispanic	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% age >64	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% educ <HS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% HS <educ <college	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% educ >college	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% public assistance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
% poverty	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
lnsales-1	0.0043	0.0000	0.0388	-0.0090	-0.0220	-0.0352	0.2705
dwintix-5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
lnsales-5	0.0014	0.0000	0.0000	-0.0303	0.0587	0.0680	0.0224
dwintix-10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
lnsales-10	0.0075	0.0000	0.2421	-0.0382	-0.0356	0.0698	0.1747

mean difference using our matched dataset and controlling for date, we find a result that is not significant. The point estimate is 0.1686 with a p-value of 0.2445. We ran the same mean difference calculation on the k-to-k matched data and produced a result that was, again, not significant. On the k-to-k data, the point estimate is 0.2149 with a p-value of 0.1911.

We produce a result that is not significant, in contrast with GK's findings. We conclude that the analysis conducted by GK contained bias that was reduced by employing the matching methodology. It is not clear that neighborhoods with certain racial composition or wealth levels are more susceptible to the lucky store effect. Our findings support the hypothesis that the lucky store effect is the result of cognitive biases, such as the "hot hand" bias, rather than being based on socio-economic factors. Moreover, our analysis does not support the notion that certain socioeconomic groups are more susceptible to the bias than others.

3 Interaction Between Lucky Store and Age

GK claim that the lucky store effect is magnified among certain populations by using one interaction variable per regression. To test this model specification, we assume there is a socioeconomic effect, as GK argue. Instead of just having one interaction variable in each regression, we tested using two interaction variables as shown in the results summarized by Table 2. The first three columns replicate the regressions run in GK's paper with one interaction in each. The final two columns represent our models that include two interactions. We use first differences (detailed in Section 6) and control for poverty and education. The interaction between winning store and age is no longer significant when controlling for poverty or education.

Table 2: Lucky Store Effect and Demographics

	GK (1)	GK (3)	GK (4)	HS*over 65	Over 65*poverty
1-week lag	0.086	0.135	0.098	-0.002	0.026
	(0.066)	(0.086)	(0.065)	(0.094)	(0.092)
*(percent HS dropout)	1.037	-	-	0.946	-
	(0.241)	-	-	(0.251)	-
*(percent over 65)	-	1.348	-	0.760	0.065
	-	(0.550)	-	(0.572)	(0.584)
*(percent poverty)	-	-	1.661	-	1.503
	-	-	(0.398)	-	(0.422)

Our results suggest that a population with high percentage of elderly people may not exhibit the lucky store effect if it is already highly educated or has high income, while an undereducated and poor population may exhibit the lucky store effect regardless of age. Thus, from a public policy point of view, efforts to attenuate the lucky store effect can be prioritized towards impoverished or undereducated populations, rather than those with higher age.

4 Including Dropped Treatments

Even though the original data include information on 68 stores, approximately 70% (47 to 48 stores) of them are dropped by GK during data processing prior to each regression. The results reported in the paper included the effect of only 20 to 21 stores out of a possible 68

stores. We check whether omitting 70% of the treated observations introduces bias to the results.

There were four independent variables for the regressions looking at substitution effects (Table 2c in GK):

1. Whether a store sold a winning ticket
2. Log lotto sales
3. Log sales of non-lotto games
4. Whether the store has a winning store within a one mile radius

The binomial variable of whether a store sold a winning ticket (1) contained no missing values. Since the dependent variable is log lotto sales, we feel it is be unreasonable to fill in values for log lotto sales (2). Since log sales of other games (3) is a control variable and not a central piece in the analysis, we fill in missing values with the median for all stores as a preliminary test. Also, as it is highly unlikely for two stores within a mile of each other to sell a winning lotto ticket in the same week, we fill zeros for any missing variables of whether the store has a winning store within a one mile radius (4). We only fill in missing data for winning stores and ignore control non-winning stores because the large sample size of the control units.

As a result, 61 to 67 out of 68 observations are included in our analysis. We rerun the regressions to test GK's substitution effects and found that the coefficients do not change significantly.

We tested the sensitivity of filling in (3) log sales of non-lotto games with median values by rerunning the original regressions on the original dataset and the larger dataset. We find no differences between the coefficients of interest reported in our respective models. Thus, we conclude that even though dropping 70% of the treated observations in analysis has the potential to create bias, including them do not impact the conclusions of the analysis.

5 Matching on Sales History

We perform a number of matching methods to test model dependence. We limited the matching exercise to predicting the effect of having sold a winning ticket one week ago[3].

[3]In GK's terminology, we focus on the 1-week lag of selling a winning ticket

Our matching models include nearest neighbor matching on sales history with exact date matching, full matching on sales history with exact date matching and Coarsened Exact Matching. We then run the regressions as specified by GK using the matched data. The effect of the lucky store effect varies but does not change substantively from their results. We devote more time towards using CEM as a matching method as described below.

5.1 Coarsened Exact Matching (CEM)

We performed an extensive sensitivity analysis of the results using matched data on both the original smaller treatment dataset and the our larger sample dataset with nearly all treatment observations intact. The smaller dataset contained 251,597 control candidates and 17 wins. The larger dataset contained 722,443 control candidates and 61 wins. The results of the sensitivity analysis demonstrate a tradeoff between coarseness of the matching bounds and significance of the results, particularly in the smaller sample.

For both datasets, we ran user-specified CEM defining increasingly finer bounds for all log sales covariates, ranging from divisions of 1.00 log sales to 0.05 log sales.[4] Each matched dataset was then regressed against the logged sales to estimate the causal effect of selling a winning lottery ticket on the dependent variable. Covariates that did not match adequately were included in the regression. The controls include date, lagged lotto sales, lagged sales of non-lotto games, for five periods in addition to the treatment variable. Given the large number of control observations in this case, and to further test the robustness of the treatment size, we then performed a k-to-k match. The process further refines the CEM previously conducted by matching the "closest" control (as measured by Euclidean distance) match to each treatment, such that there is only one matched control observation to each treatment. Last, given that the user-defined divisions were effectively "rounding" of the log sales figures, we also ran an automated CEM and a related k-to-k match and the linear estimate models.

Table 3 summarizes the results from this sensitivity analysis. The range of effect of selling a winning ticket on sales ranges from 24.86% to 47.38%. Using the smaller dataset, the maximum number of treatments matched in any CEM model is 16, compared with the larger dataset where a maximum of 59 treatments match. With increasingly finer divisions in the user-specified matching for log sales controls, the number of matched treatments declines for both datasets. For example, at divisions of 0.5 and 0.25 log sales, the number of matched observations decline from 11 to 2 in the small dataset and from 50 to 33 in the large dataset.

[4]Results for 0.05 log sales division omitted from Table 3 because there were 0 matches in both datasets.

Table 3: Summary Balance Information from Demographic CEM

	Control	Treatment	Estimate	p-value	CI Min	CI Max
Small Dataset						
Automated CEM						
Matched	113	13	0.3680	0.0000	0.2307	0.5054
k-to-k	13	13	0.2678	0.0078	0.1008	0.4347
1.0 Divisions						
Matched	1879	15	0.3377	0.0000	0.2371	0.4383
k-to-k	15	15	0.2558	0.0153	0.0699	0.4417
0.75 Division						
Matched	345	16	0.4738	0.0000	0.3661	0.5815
k-to-k	16	16	0.4148	0.0027	0.1785	0.6512
0.5 Division						
Matched	80	11	0.2583	0.0000	0.1512	0.3653
k-to-k	11	11	0.1501	0.1790	-0.0518	0.3520
0.25 Division						
Matched	2	2	n/a	n/a	n/a	n/a
k-to-k	2	2	n/a	n/a	n/a	n/a
Large Dataset						
Automated CEM						
Matched	10922	52	0.3520	0.0000	0.2901	0.4140
k-to-k	52	52	0.3718	0.0000	0.2653	0.4783
1.0 Divisions						
Matched	66069	56	0.3643	0.0000	0.3021	0.4265
k-to-k	56	56	0.3495	0.0000	0.2432	0.4558
0.75 Division						
Matched	24870	59	0.4125	0.0000	0.3491	0.4758
k-to-k	59	59	0.4099	0.0000	0.3006	0.5192
0.5 Division						
Matched	9346	50	0.3293	0.0000	0.2678	0.3908
k-to-k	50	50	0.3433	0.0000	0.2311	0.4554
0.25 Division						
Matched	806	33	0.3505	0.0000	0.2602	0.4407
k-to-k	33	33	0.4226	0.0006	0.1971	0.6492

To benchmark the user-defined CEM models, we ran a standard automated CEM model and its respective regression. The results are in line with the conclusions of the GK paper and are useful in summarizing the sensitivity analysis. The smaller dataset produced 13 matched treatment wins (4 omitted) with 113 (251,484 omitted) controls in the many-to-k matching. The mean difference between treatment and control was 36.80% and 26.78% for the many-to-k and k-to-k regressions, respectively. In the larger dataset, there were 52 treatment wins (9 omitted) matched to 10,922 controls (711,521 omitted). The mean difference was 35.20% and 37.18% for many-to-k and k-to-k, respectively. While slightly discounted from the effect found in GK, the magnitude and significance of the effect are largely the same.

We emphasize two main takeaways from the results. First, there is effectively no trade-off between fineness of user-defined divisions and significance of winning as a predictor of log sales in the smaller dataset. Second, a potentially significant caveat of our matching methodology pertains to the L-statistic of balance. Each covariate of each model generated a L-stat of 0.00, implying a perfect overlap between the control and treatment distributions for that covariate. However, each model generated a model L-stat of 1.00, implying a complete separation of the treatment and control histograms.

6 First Differences

GK use sales history to predict current sales, which creates correlation among weekly errors in the regression. Despite this, they use OLS, a framework in which the residuals are assumed to be independent and identically distributed. To investigate the impact of their methodology, we use first differences[5] to reduce the reliance on the model assumption of independent and identically distributed residuals and to allow for sequential dependence. By applying first differences and dropping observations where stores sell a winning ticket this week but none last week (i.e. the 1-week lag variable is equal to -1), the effects reported by GK are unchanged.

7 Typographical Error

GK find some evidence that stores within the same zip code as a winning store experienced increased sales **prior** to the sale of the winning ticket, which is highly counterintuitive. Winning tickets should not be predictable. Second, while the winning store did not experience

[5]See Goldberger (2003) Chapter 10 for a discussion on the first differencing methodology.

increased sales prior to the sale of the winning ticket, stores in the same zip code did. If this is true and there is indeed a ramp-up in sales prior to the winning ticket being sold, one could argue that any increase in sales after the winning ticket is sold could be a carryover from the build-up prior to the sale of the winning ticket.

Table 4: Amended coefficients and standard errors from Typographical Correction

	GK Coef	GK SE	Corrected Coef	Corrected SE
5-week lead	0.019	0.010	-0.003	0.011
4-week lead	0.024	0.010	-0.027	0.010
3-week lead	0.033	0.009	0.010	0.010
2-week lead	0.001	0.009	-0.008	0.010
1-week lead	-0.015	0.009	-0.022	0.009

However, we find a typographical error in their data processing leading up to this analysis. By correcting the error, the increase in sales is mitigated as shown in Table 4. Thus, we find stronger support for GK's hypothesis. By showing that there is no increase in sales in stores within the same zip code as the store selling the winning ticket before the winning ticket is sold, there is a more stark contrast between the winning store and nearby stores after the sale of the winning ticket.

8 Concluding Remarks

We investigate GK's paper that studies the effect of selling a winning lottery ticket on sales of retail store. GK concluded that the sale of a winning ticket increased sales for that retailer of that lottery game by 38% in the following week and that the effect dissipated over time. Through additional analysis they conclude that the effect is particularly evident in elderly, poorer, and undereducated neighborhoods. Our investigation provides support for some of their claims and refutes others. The magnification of the lucky store effect is not found among elderly neighborhoods, after controlling for income and education. We also conduct an extensive matching sensitivity and find that the impact of selling a winning lottery ticket is positive and significant–ranging from a 24.86% to 47.38% increase in sales.

GK's conclusions of the impact of lucky stores across demographies may have significant policy impact. In this analysis, we make a sharp departure from GK. Rather than targeting communities of elderly people, we suggest that efforts to attentuate the lucky store effect

may be better directed towards low income and low education populations rather than to those of high age. Lastly, our matching analysis results in no significance of lucky stores once the data is appropriately matched on the socioeconomic covariates.

References

[1] Goldberger, Arthur S. 2003. *A Course in Econometrics*.

[2] Guryan, Jonathan and Kearney, Melissa S. 2008. "Gambling at Lucky Stores: Empirical Evidence from State Lottery Sales." *American Economic Review* 98:1, 458-473.

[3] Iacus, Stefano M., King, Gary and Porro, Giuseppe. 2009. CEM: Coarsened Exact Matching. *Journal of Statistical Software* (forthcoming).