American Economic Review: Papers & Proceedings 100 (May 2010): 480–484
http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.480

Investigating Income Effects in Scanner Data: Do Gasoline Prices Affect Grocery Purchases?

By Dora Gicheva, Justine Hastings, and Sofia Villas-Boas*

There is much discussion in the popular press about how consumers adjust their purchase decisions for items from lattes and restaurant meals to which type of meat to purchase for dinner during times of rising fuel prices.[1] While analysts ascribe declines in retail sector profits when fuel prices rise to changes in demand elasticity, most empirical analysis of consumer choice for such daily items abstracts from intertemporal income effects. Instead, fuel prices are used in demand estimation as exogenous shifters of production costs, and therefore valid instruments for identifying demand parameters. Though introspection and popular press suggest that sharp changes in fuel costs may shift price sensitivity in nonfuel purchases through an income effect, little empirical work has been done to estimate or quantify this effect.[2]

In this paper we use sharp changes in gasoline prices to estimate the impact that short run changes in disposable income have on measures of consumer price sensitivity at the grocery store. We use weekly store level scanner data from 180 West Coast grocery stores for products (UPCs) in frequently purchased food categories. We find evidence that consumers adjust to higher gasoline prices by substituting within a category towards products that are on sale (i.e., on promotion): the fraction of purchases from sale items increases significantly with gasoline prices. The effect is generally stronger at stores serving lower income families. Additionally, we find that the quantity weighted price paid for products decreases when gasoline prices increase; consumers are able to save money on groceries by shifting purchases towards promotional items.

Because gasoline expenditures during this period rise one for one with gasoline prices and because gasoline prices have been shown to follow a random walk (Dora Gicheva, Justine S. Hastings, and Sofia B. Villas-Boas 2007; Jonathan E. Hughes, Christopher R. Knittel, and Daniel Sperling 2008; and Patrick Kline 2008), we interpret these findings as a short run income effect. Our results suggest that, in addition to increasing production costs, rising fuel prices lower profit margins by increasing competitive pressure on retail firms as consumers become more price sensitive to compensate for lost income devoted to increased fuel expenditures.

I. Data and Regression Analysis

Gasoline prices have increased dramatically several times over the past five years. This volatility has been particularly prominent in California markets where run-ups in gasoline prices are often more severe than in other regions of the country.[3] From 2000 through

* Gicheva: Department of Economics, Yale University, 37 Hillhouse Ave., New Haven, CT 06511 (e-mail: dora.gicheva@yale.edu); Hastings: Department of Economics, Yale University, 37 Hillhouse Ave., New Haven, CT 06511 (e-mail: justine.hastings@yale.edu); Villas-Boas: Department of Agricultural and Resource Economics, UC Berkeley, 207 Giannini Hall, #3310, Berkeley, CA 94720 (e-mail: sberto@berkeley.edu). We thank Daron Acemoglu, Severin Borenstein, Keith Chen, J.P. Dube, Matthew Gentzkow, Eric Hurst, George Judge, Aviv Nevo, Sharon Oster, Jon Quinn, Sergio Rebelo, Fiona Scott Morton, Jesse Shapiro, Matt Turner, and participants at the Yale Macroeconomics Workshop, the Yale Industrial Organization Lunch and the New York Federal Reserve Workshop for helpful comments. Eric Chyn and Sarah Johnston provided outstanding research assistance. This project funded by the Institution for Social and Policy Studies at Yale.

[1] For example, see articles titled "Are Frappuccino Woes or Frugality To Blame for Starbucks's Stumble?" from the *Wall Street Journal* on August 4, 2006, and "Full Tanks Put Squeeze on Working Class," in the *New York Times* on May 13, 2006.

[2] Julie Berry Cullen, Leora Friedberg, and Catherine Wolfram (2005) use data from the Consumer Expenditure Survey to test if poor families decrease food expenditures when home heating fuel prices rise.

[3] California requires its own formulation of gasoline to meet California Air Resources Board emissions standards. This formulation is not required in other regions of the

2005, California gasoline markets experienced several large spikes in gasoline prices, with prices rising and falling by over 25 percent on several occasions, in a pattern most likely exogenous to other factors that affect household income or household product preferences over time. Using data from the Consumer Expenditure Survey, Gicheva, Hastings, and Villas-Boas (2007) find that gasoline expenditures rise one for one with gasoline prices during this period. Since the average Californian spent about five percent of income on gasoline in 2002, and since gasoline prices have been shown to follow a random walk, these changes in gasoline prices may translate into small but significant changes in permanent income available for expenditures in other categories.[4]

We have access to weekly store level data for a sample of 180 grocery stores from a retail chain in California. The retailer is a standard grocery store chain and has stores in a broad range of socioeconomic neighborhoods. For each of the stores we have weekly UPC level data for all items within four product categories: Family Cold Cereal, Family Yogurt, Fresh Chicken, and Refrigerated Orange Juice (hereafter cereal, yogurt, chicken, and orange juice, respectively). The data include the total unit quantity of each product sold, the total gross revenue, the total revenue net promotional discounts, and the total weight sold where needed (for example, pounds of meat where price is measured in dollars per pound).[5] We use these variables to construct the average gross price per week for each UPC, the average price net of discounts per week, and the total volume sold for each UPC in each week.[6] We match weekly average gasoline prices for Los Angeles to weekly measures of consumer purchase behavior in each category, and we use membership card data with attached information on customer income levels to create measures of the income level of each store's customer base.[7]

If income effects are important, we would expect to see that when gasoline prices are high, consumers purchase a higher fraction of products on sale, and that the quantity weighted net price paid per unit falls. To test this hypothesis, we run regressions of the following form, separately for each of our four product categories:

$$\ln(y_{jt}) = \alpha_j + \beta \ln(gasprice_t) + \gamma' \mathbf{X}_{jt} + \varepsilon_{jt} \tag{1}$$

where y_{jt} denotes either the fraction of sales in a category at store j in week t that come from promotional items or the quantity weighted price paid for items purchased in that category, store, and week combination. We control for store level fixed effects, α_j, as well as regional time trends, regional monthly dummies, holiday fixed effects, the fraction of UPCs in each category that are on sale in week t at store j, and its square. All of these controls are included in the vector $\mathbf{X}_{jt}$. We allow for first-order autocorrelation in the error terms, ε_{jt}.[8]

Table 1 presents coefficients on log gasoline prices from regressions of the form (1).[9] The first panel of results is for cereal, and the first column presents results from regression specification pooled across all stores, while the following columns present by quartiles of the

country, separating California to some degree from gasoline supply in the rest of the nation.

[4] Gicheva, Hastings, and Villas-Boas (2007) report Dickey-Fuller test statistics of −0.978, and the MacKinnon approximate p-value for the unit root test of 0.7613. Kline (2008) also finds that oil prices follow a random walk.

[5] In each category, we account for different container sizes when calculating prices. For example, yogurt is in price per six-ounce serving and chicken is in price per pound. For further details, please see Gicheva, Hastings, and Villas-Boas (2007).

[6] Because the grocery retailer changes promotions and sales on a weekly basis, the aggregated data yield the correct prices and promotional discounts for each weekly observation.

[7] The prices used are the Energy Information Administration's weekly average price of regular unleaded reformulated gasoline in Los Angeles, CA. The average gasoline price in Los Angeles is a good approximation for local prices that customers at our stores face but is constant across stores, avoiding potential local endogeneity between gasoline prices and grocery sales (i.e., in one neighborhood, gasoline prices are particularly high, causing customers to buy gasoline and groceries in an adjacent neighborhood). For more details on the retail scanner data and summary statistics on customer demographics, please see Gicheva, Hastings, and Villas-Boas (2007).

[8] Since we have a very long time series, the bias introduced from autocorrelation in the fixed effects model is negligible (Cheng Hsiao 1986).

[9] The quartiles are of the distribution of median customer level income across stores in our sample, with cutpoints of less than \$50,000, between \$50,000 and \$69,500, between \$69,500 and \$90,500, and greater than \$90,500.

TABLE 1—RELATIONSHIP BETWEEN PERCENT SOLD ON SALE AND GAS PRICES

Dependent variable: ln(percent of sales from promotional items)	All stores (1)	Stores in income quartile 1 (2)	Stores in income quartile 2 (3)	Stores in income quartile 3 (4)	Stores in income quartile 4 (5)
Adult cereal: ln(gas price)	0.190	0.269	0.170	0.179	0.154
	(0.012)**	(0.029)**	(0.021)**	(0.023)**	(0.025)**
Dep. variable mean	0.65	0.66	0.66	0.64	0.62
Yogurt: ln(gas price)	0.252	0.360	0.234	0.283	0.164
	(0.040)**	(0.085)**	(0.076)**	(0.079)**	(0.078)*
Dep. variable mean	0.50	0.53	0.51	0.51	0.47
Chicken: ln(gas price)	0.491	0.548	0.522	0.475	0.445
	(0.055)**	(0.129)**	(0.110)**	(0.111)**	(0.091)**
Dep. variable mean	0.60	0.63	0.61	0.59	0.58
Fresh orange juice: ln(gas price)	0.103	0.075	0.103	0.103	0.131
	(0.007)**	(0.016)**	(0.013)**	(0.014)**	(0.014)**
Dep. mean	0.83	0.84	0.83	0.83	0.82
Observations	27,540	6,426	7,344	6,885	6,885
Number of stores	180	42	48	45	45

Notes: Standard errors in parentheses. Residuals allowed to follow a first-order autoregressive process. Each cell reports the coefficient and standard error on $\ln(\text{gasprice}_t)$ from the regression specified in equation (1). The dependent variable is the log fraction of total sales in each category at store *j* in week *t* that are attributable to UPCs that were offered on promotion. Right-hand side variables are: store fixed effects, monthly dummies interacted with regional dummies, time trends interacted with regional dummies, holiday dummies, the fraction of items on promotion in week *t* in store *j*, and its square. Holiday dummies include separate dummies by year for major holidays and the week before and after the holiday if it falls on a weekend (Thanksgiving, Christmas, New Year's and Fourth of July). We adjust prices of all items to account for differences in container size when calculating prices.

**Significant at the 1 percent level.

*Significant at the 5 percent level.

customer income distribution.[10] The coefficient on gasoline prices is positive and significant, indicating that a 100 percent increase in gasoline prices results in a 19 percent increase in the fraction of cereal purchases coming from promotional items. This is a substantial effect, implying that underlying consumer price sensitivity and product purchase decisions change substantially in response to changes in gasoline prices. The coefficient is largest for stores serving patrons in the lowest quartile of the income distribution, and decreases with median patron income as we would expect. A similar pattern holds for yogurt. The fraction of items purchased on promotion increases by an average of 25 percent in response to a 100 percent increase in gasoline prices, with the effect falling by more than half as we move from the lowest to the highest income stores (0.360 to 0.164). Purchases in the chicken category display a similar pattern across income quartiles, but the overall effect is much larger. A 100 percent increase in gasoline prices increases the fraction of purchases coming from sale items by 49.1 percent. This may be due to the overall higher price of chicken relative to cereal or yogurt. We find the smallest percentage effect in the orange juice category, and the opposite pattern across income quartiles. This may be because an easy substitute for this category is frozen or shelf-stable juices, causing those in lower income brackets to further substitute between, instead of just within, category.

Table 2 presents results from regressions of the form (1) with the quantity weighted net price paid for cereals sold as the dependent variable. The regression specification is identical to that

[10] We use all cereals in this category, but drop cereals that appear very infrequently (for example holiday or themed versions of cereals that appear for only a short duration). We adjust the prices of cereals to account for differences in box sizes, standardizing the prices so that they are comparable across boxes.

TABLE 2—RELATIONSHIP BETWEEN LN QUANTITY-WEIGHTED GROCERY PRICE AND GAS PRICES, SELECTED PRODUCT CATEGORIES

Dependent variable: ln(percent of quantity-weighted price paid)	All stores (1)	Stores in income quartile 1 (2)	Stores in income quartile 2 (3)	Stores in income quartile 3 (4)	Stores in income quartile 4 (5)
Adult Cereal: Coef. on ln(gas price)	−0.049	−0.066	−0.036	−0.058	−0.038
	(0.007)**	(0.016)**	(0.013)**	(0.013)**	(0.013)**
Dep. mean	3.10	3.03	3.06	3.11	3.20
Yogurt: Coef. on ln(gas price)	−0.072	−0.084	−0.085	−0.051	−0.075
	(0.010)**	(0.021)**	(0.019)**	(0.020)**	(0.020)**
Dep. mean	0.73	0.71	0.72	0.74	0.76
Chicken: Coef. on ln(gas price)	−0.103	−0.075	−0.095	−0.153	−0.09
	(0.022)**	(0.046)	(0.044)*	(0.044)**	(0.041)*
Dep. mean	2.37	2.05	2.32	2.39	2.70
Fresh orange juice: Coef. on ln(gas price)	−0.109	−0.101	−0.116	−0.106	−0.11
	(0.008)**	(0.018)**	(0.016)**	(0.016)**	(0.016)**
Dep. mean	3.10	3.03	3.08	3.12	3.17
Observations	27,540	6,426	7,344	6,885	6,885
Number of stores	180	42	48	45	45

Notes: Standard errors in parentheses. Residuals allowed to follow a first-order autoregressive process. Each cell reports the coefficient and standard error on ln($gasprice_t$) from the regression specified in equation (1). The dependent variable is the log of the quantity-weighted price index for purchased products in each category at store *j* in week *t* calculated using prices inclusive of promotional discounts. Right-hand-side variables are: store fixed effects, monthly dummies interacted with regional dummies, time trends interacted with regional dummies, holiday dummies, the fraction of items on promotion in week *t* in store *j*, and its square. Holiday dummies include separate dummies by year for major holidays and the week before and after the holiday if it falls on a weekend (Thanksgiving, Christmas, New Year's and Fourth of July). We adjust prices of all items to account for differences in container size when calculating prices.

**Significant at the 1 percent level.

*Significant at the 5 percent level.

in Table 1, with "the log of quantity weighted net prices" instead of "the log of percent of items sold on promotion" as the dependent variable. The results show that the quantity weighted net price falls significantly when gasoline prices increase. If gasoline prices increase by 100 percent, the quantity weighted price paid by consumers falls on average by 5–10 percent. For example, the average quantity weighted price paid per box of cereal is approximately $3, implying that consumers decrease their overall cereal expenditures by 15 cents per box. In general, comparing the estimated effects on prices to those on fraction of purchases from sale items within each category, the largest savings are generally in the income quartiles where substitution towards promotional products was largest.

II. Interpretation and Conclusions

The results from tables 1 and 2 suggest that there is a substantial consumer response to increased fuel prices, with consumers substituting significantly towards sale items from full price items in a range of retail grocery products when gasoline prices rise. We conducted several robustness checks in Gicheva, Hastings, and Villas-Boas (2007) to test this interpretation. First, we examined how retail prices themselves respond to increased fuel prices. In the Cereal category, for example, we found that shelf prices are unchanged by fuel prices, but that increased fuel prices result in slightly higher prices net of discounts.[11] When we controlled for this price index in our main regressions, we found very similar results. We conclude that in retail grocery, any price adjustments to input costs

[11] Gicheva, Hastings, and Villas-Boas (2007) report a five percent increase in net prices as a result of a 100 percent increase in gasoline prices, which is similar to other estimates of cost-based increase in PPI and CPI resulting from fuel price increases (see, e.g., Chinkook Lee 2002).

come primarily through changes in discount rates off of shelf prices, and that even accounting for cost increases, substitution towards discounted items is a primary way that consumers decrease expenditures on retail purchases when fuel costs rise.[12]

Second, we graphed the relationship between gasoline prices and fraction of sales coming from promotional items for Cereal by plotting a smoothed nearest neighbor regression line for the residuals from equation (1) *excluding gasoline prices*, and the residuals from a regression of gasoline prices on the other right-hand-side variables in (1) for four different stores. We found graphically a positive relationship between the percent of cereal purchases coming from items on sale and gasoline price that appears fairly symmetric; fraction of purchases coming from on sale items both rises and falls with the spikes and troughs in regression adjusted gasoline prices.

If we interpret the estimated results as short run income effects and compare their magnitude to the variation in mean fraction of purchases coming from promotional items across low and high income stores, it appears that the intertemporal income effect is substantially larger than a cross-sectional income effect. This may be because in the short run, large fractions of income are committed to expenditure categories that cannot be easily adjusted (Raj Chetty and Adam G. Szeidl 2007). Therefore, income effects may occur more than proportionally in expenditure categories that represent the most flexible and lowest cost margin for income savings such as groceries and entertainment expenditures.

Overall, we find significant effects of changes in fuel prices on price sensitivity of consumers across a range of retail grocery products. These findings suggest that product substitution towards lower price products is an important component of consumption smoothing in the presence of income shocks, and that a key way in which consumers substitute is by purchasing sale items in lieu of full price items. The magnitudes of these findings suggest that fuel prices may affect both demand and supply, changing price sensitivity through short run income effects (Greg Allenby and Peter Rossi 1991), as well as shifting costs of production.

[12] This provides added evidence to the literature showing that shelf prices are very sticky, and that relevant, higher frequency price changes come primarily through changes in promotional prices (Judith A. Chevalier, Anil K. Kashyap, and Peter E. Rossi 2003; Mark Bils and Peter J. Klenow 2004; and Emi Nakamura and Jón Steinsson 2008).

REFERENCES

Allenby, Greg M., and Peter E. Rossi. 1991. "Quality Perceptions and Asymmetric Switching between Brands." *Marketing Science*, 10(3): 185–204.

Bils, Mark, and Peter J. Klenow. 2004. "Some Evidence on the Importance of Sticky Prices." *Journal of Political Economy*, 112(5): 947–85.

Chetty, Raj, and Adam Szeidl. 2007. "Consumption Commitments and Risk Preferences." *Quarterly Journal of Economics*, 122(2): 831–77.

Chevalier, Judith A., Anil K. Kashyap, and Peter E. Rossi. 2003. "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data." *American Economic Review*, 93(1): 15–37.

Cullen, Julie Berry, Leora Friedberg, and Catherine Wolfram, 2005. "Do Households Smooth Small Consumption Shocks? Evidence from Anticipated and Unanticipated Variation in Home Energy Costs." UCEI WP-141.

Gicheva, Dora, Justine Hastings, and Sofia Villas-Boas. 2007. "Revisiting the Income Effect: Gasoline Prices and Grocery Purchases." NBER Working Paper 13614.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. Cambridge, UK: Cambridge University Press.

Hughes, Jonathan E., Christopher R. Knittel, and Daniel Sperling. 2008. "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand." *Energy Journal*, 29(1): 113–34.

Kline, Patrick. 2008. "Understanding Sectoral Labor Market Dynamics: An Equilibrium Analysis of the Oil and Gas Field Services Industry." Unpublished.

Lee, Chinkook. 2002. "The Impact of Intermediate Input Price Changes on Food Prices: An Analysis of 'From-the-Ground-Up' Effects." *Journal of Agribusiness*, 20(1): 85–102.

Nakamura, Emi, and Jon Steinsson. 2008. "Five Facts About Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics*, 123(4): 1415–64.