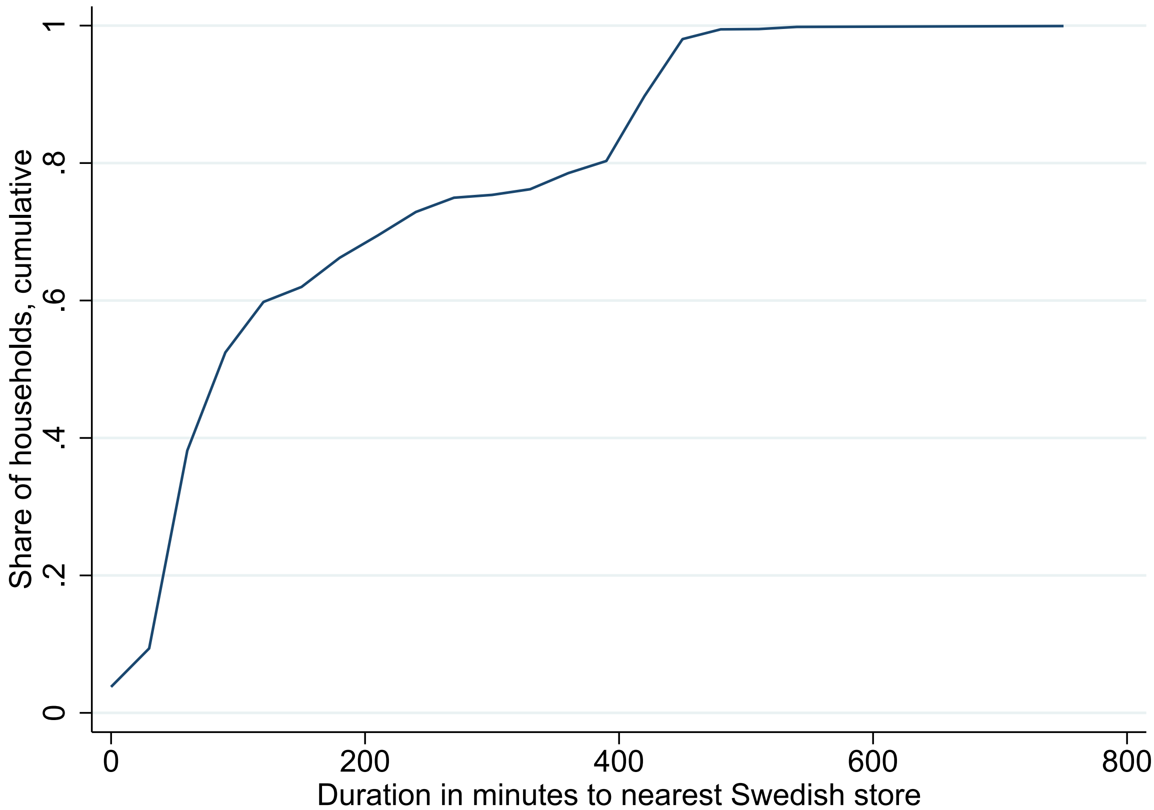
Online appendix for: Hump-shaped cross-price effects and the extensive margin in cross-border shopping

Richard Friberg   Frode Steen   Simen A. Ulsaker

March 24, 2020

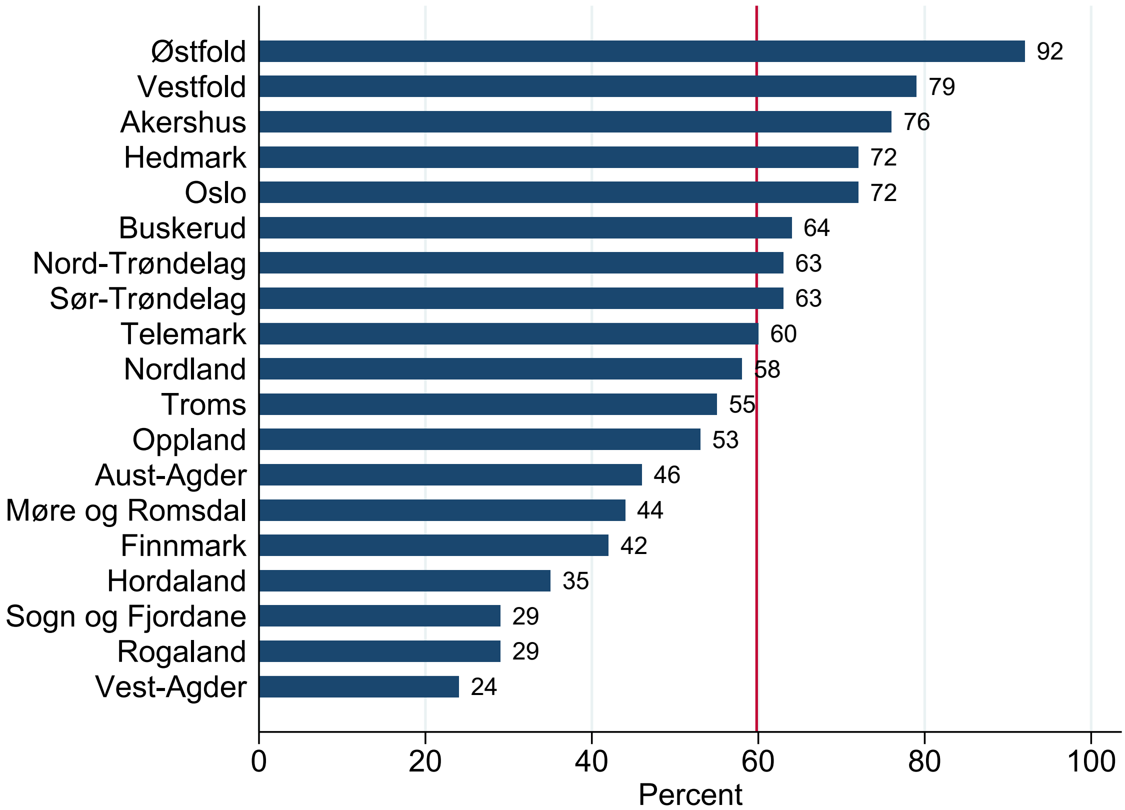
A.1 Cumulative distribution of Norwegian population by driving duration to nearest Swedish grocery store.

Figure A.1: Cumulative distribution of driving duration to nearest Swedish grocery store.



A.2 Share of border shoppers by Norwegian regions

Figure A.2: Share of border shoppers



Note: The figure shows the county-level proportions of the population that have shopped groceries in Sweden during the last 12 months. The red line indicates the national average of 59 %. Numbers based on survey responses from Norwegian respondents. Survey undertaken 22-27 February 2018, n=1009.

A.3 Cross-regional variation in Norwegian prices

In the empirical analysis we have used a national price index to measure the price level in Norwegian stores. A possible concern is that there could be regional price differences that are not accounted for when we use such a national price index. In particular, one might expect that prices in stores close to Sweden would be systematically different from prices further from the border. On the other hand, anecdotal evidence suggests that Norwegian grocery chains to a large degree impose uniform national prices.

To provide some empirical evidence on this question, we have obtained transaction level data from a random sample of the members of NG's frequent buyer program for the year 2016. We have used this data to compute average prices at the product level for different chains and regions (defined by the same bins of driving duration to Sweden as in the main analysis). We use products from the same categories as in the main analysis (meat, cheese, soda and sweets) and keep only products for which we have observations in all months in all of the chain-region pairs. We then regress the logarithm of the price on month, chain, and region dummies. As reported in Table A.1, there is no indication that prices vary with the distance to Sweden.

Table A.1: Cross-region variation in Norwegian prices

	ln(Price)
$30 < \text{Duration} < 60$	0.00029
	(0.00723)
$60 < \text{Duration} < 90$	0.00012
	(0.00733)
$90 < \text{Duration} < 120$	0.00048
	(0.00731)
$120 < \text{Duration} < 150$	0.00072
	(0.00724)
$150 < \text{Duration} < 180$	0.00182
	(0.00702)
Constant	3.20581
	(0.00692)
Joint test duration groups (p-value)	0.99986
Observations	3888
Number of products	18
Month FE	Yes
Chain FE	Yes
EAN number FE	Yes

Note: The dependent variable is the natural logarithm of the average monthly price at the region-chain level. We use data from three different chains within the NG umbrella. The eight chain formats used in the main analysis are nested within these three chains. The sample period is the year 2016. Clustered standard errors are reported in parentheses.

A.4 Regression with short period and COICOP data

Table A.2: COICOP – short period

	Meat	Cheese	Soda	Sweets
Duration $< 30 \times \ln(P^N/P^S)$	-1.541	-0.784	0.048	-1.457
	(0.172)	(0.169)	(0.228)	(0.240)
$30 <$ Duration $< 60 \times \ln(P^N/P^S)$	-1.730	-0.914	-0.042	-1.962
	(0.293)	(0.224)	(0.201)	(0.197)
$60 <$ Duration $< 90 \times \ln(P^N/P^S)$	-0.982	-0.452	0.243	-1.075
	(0.179)	(0.182)	(0.193)	(0.114)
$90 <$ Duration $< 120 \times \ln(P^N/P^S)$	-0.608	-0.456	0.379	-0.666
	(0.214)	(0.183)	(0.218)	(0.139)
$120 <$ Duration $< 150 \times \ln(P^N/P^S)$	-0.496	-0.565	0.368	-0.604
	(0.267)	(0.203)	(0.224)	(0.332)
$150 <$ Duration $< 180 \times \ln(P^N/P^S)$	-0.528	-0.770	0.080	-1.461
	(0.274)	(0.204)	(0.234)	(0.341)
Constant	4.517	5.724	7.867	6.588
	(0.792)	(0.591)	(0.475)	(0.424)
Observations	21166	21166	21168	21167
R^2	0.470	0.432	0.335	0.389
Month number FE	Yes	Yes	Yes	Yes
Store format FE	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes

Note: This table reports results from an estimation of the model specified in Equation 1. Monthly price indexes are calculated based on COICOP and the sample period is 2014-2016. The standard errors reported in parentheses are clustered at the municipality level.

A.5 Alternative bins for duration

As a robustness exercise Table A.3 presents results from the same specification as the benchmark specification in Table 5 but with more finely defined duration bins, in this case 20-minute bins. As seen the humpshaped pattern emerges also in this specification with the strongest effects found 40-60 minutes away from the closest Swedish store.

Table A.3: Demand regression: Finer delination of duration bins

	Meat	Cheese	Soda	Sweets
Duration $<$ 20 $\times$ $\ln(P^N/P^S)$	-1.210	-0.807	-2.321	-0.691
	(0.152)	(0.168)	(0.252)	(0.727)
20 $<$ Duration $<$ 40 $\times$ $\ln(P^N/P^S)$	-1.256	-0.859	-2.224	-0.957
	(0.134)	(0.137)	(0.166)	(0.200)
40 $<$ Duration $<$ 60 $\times$ $\ln(P^N/P^S)$	-1.434	-0.916	-2.341	-1.385
	(0.223)	(0.111)	(0.159)	(0.191)
60 $<$ Duration $<$ 80 $\times$ $\ln(P^N/P^S)$	-0.937	-0.605	-2.101	-0.752
	(0.113)	(0.082)	(0.164)	(0.147)
80 $<$ Duration $<$ 100 $\times$ $\ln(P^N/P^S)$	-0.447	-0.397	-1.967	-0.226
	(0.141)	(0.096)	(0.136)	(0.167)
100 $<$ Duration $<$ 120 $\times$ $\ln(P^N/P^S)$	-0.219	-0.520	-1.855	0.059
	(0.178)	(0.089)	(0.149)	(0.185)
120 $<$ Duration $<$ 140 $\times$ $\ln(P^N/P^S)$	-0.384	-0.627	-1.947	-0.169
	(0.188)	(0.126)	(0.170)	(0.238)
140 $<$ Duration $<$ 160 $\times$ $\ln(P^N/P^S)$	-0.776	-0.812	-2.191	-0.835
	(0.255)	(0.136)	(0.178)	(0.343)
160 $<$ Duration $<$ 180 $\times$ $\ln(P^N/P^S)$	0.033	-0.676	-2.088	-0.413
	(0.235)	(0.121)	(0.184)	(0.395)
Constant	4.831	6.330	9.054	6.766
	(0.722)	(0.665)	(0.598)	(0.522)
Observations	34389	34389	34392	34391
R^2	0.477	0.442	0.332	0.390
Month number FE	Yes	Yes	Yes	Yes
Store format FE	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes

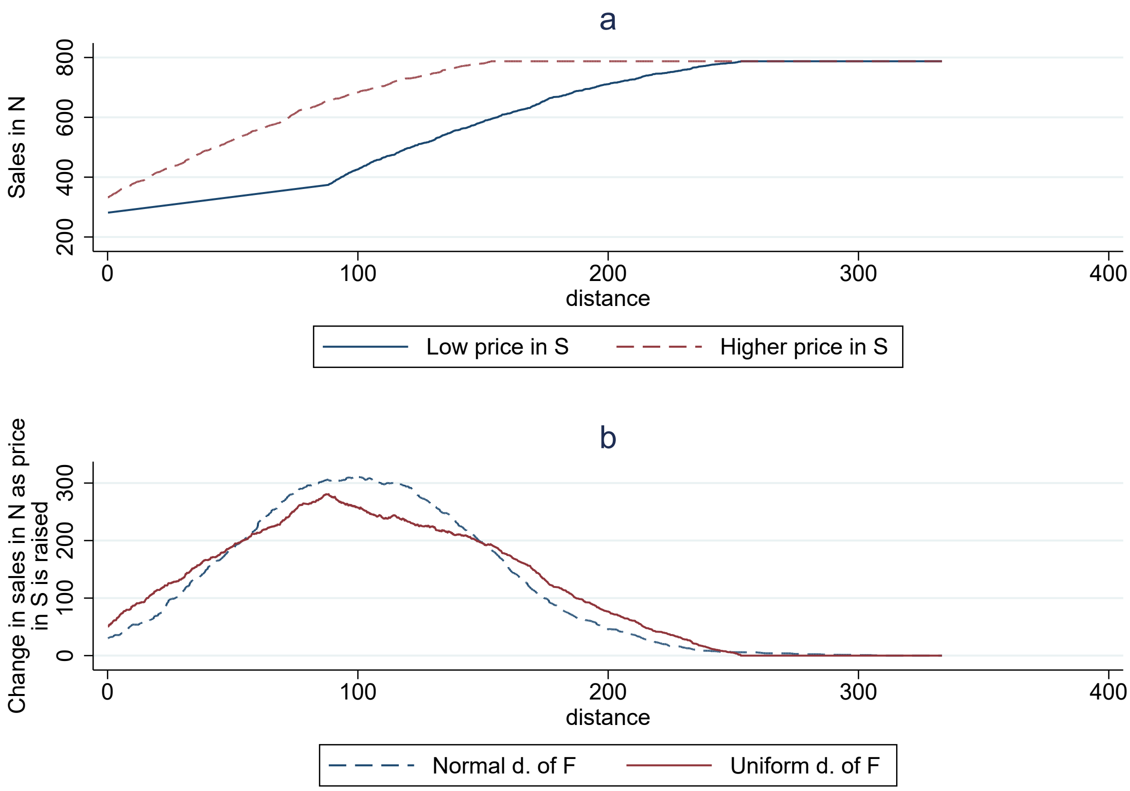
Note: This table reports results from an estimation of the model specified in Equation 1. Monthly price indexes are calculated based on COICOP and the sample period is 2012-2016. The standard errors reported in parentheses are clustered at the local labor market level.

A.6 Uniformly distributed fixed travel cost

One may note that we assumed that the fixed costs of cross-border travel are drawn from a normal distribution. With the bulk of individuals concentrated around the mean one might wonder if this distributional assumption is not solely responsible for generating the hump-shaped cross-price effect. To show that this is not the case we redo the calibrations with the same values as above but instead assume that fixed costs are drawn from a uniform distribution with approximately the same average and standard deviation as the normal distribution considered in Figures 8 and 9 (the uniform distribution bounded by

1 and 3 which clearly has a mean of 2 and a standard deviation of around 0.5). Panel a of Figure A.3 compares the sales in N as a function of distance for the same two levels of P_S as in the benchmark above. The further away from the border, the greater are local sales in N and an increase in P_S is associated with greater sales in N. The solid line in panel b of Figure A.3 traces out the difference between the two lines in panel a and we again note a hump-shaped pattern. For comparison the dashed line plots the benchmark case with normally distributed fixed costs which yields a more marked hump but it is also clear that a normal distribution is not necessary to generate the hump-shaped cross-price effect.

Figure A.3: A closer examination of the role of the distribution of fixed costs and for a hump-shaped demand response

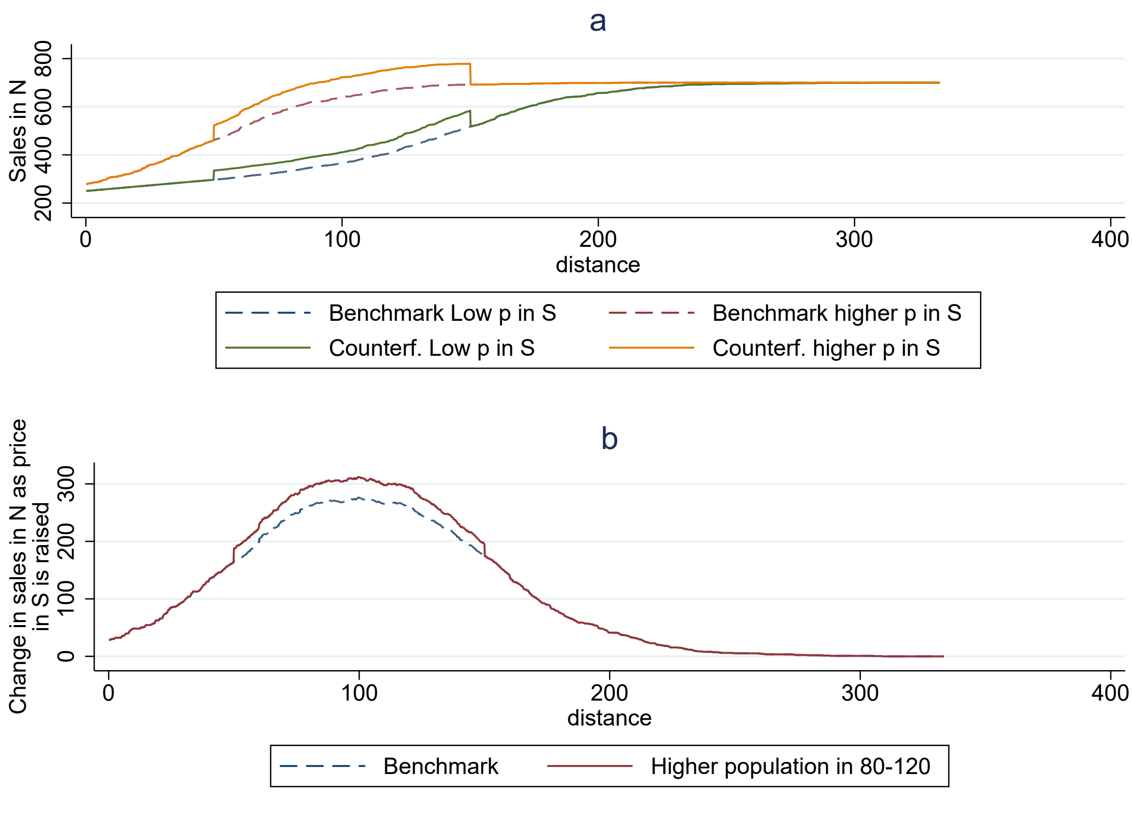


Note: Panel a shows sales volume of good N at different distances from the border (location of good S) as described in text. Parameter values in calibration: $m = 100$, $a = 15$, $b = 4$, $\theta = 0.6$ $t = 0.01$, distance increases in increments of 1/6, "low price" of $p_S = 6$ and "higher price" of $p_S = 7$. $p_N = 8$. Fixed costs at each distance are assumed to be drawn from a uniform distribution with bounded by 1 and 3. Panel b plots the difference between the two curves in a (solid line) and a comparison with same parameter values but fixed costs drawn from a normal distribution with mean 2 and standard deviation of 0.5.

A.7 Regional heterogeneity

Differences in the population density might in principle contribute to the hump-shaped pattern. The fixed costs of cross-border shopping and distance to the border interact to create the strength of the extensive margin in cross-border shopping and if a location has more consumers there is scope for a larger demand response as more consumers might be affected. Figure A.4 illustrates a simple case where the population is higher in the particular area where the extensive margin bites the most. In such a case the hump is strenghtened.

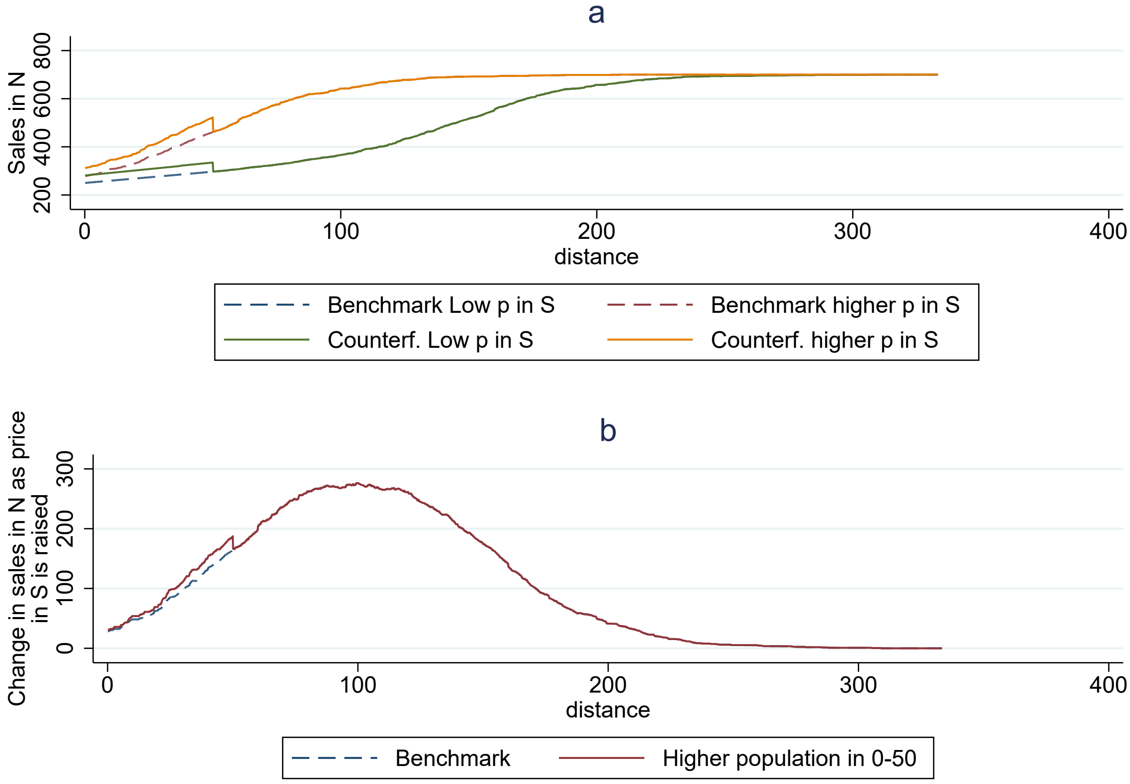
Figure A.4: Change in sales in N in response to a price change of good S and distance to S. A case with higher population in the 50-150 distance from the border.



Note: Panel a shows sales volume of good N at different distances from the border (location of good S) as described in text. Population 400 at each distance apart from distances 50-150 where population is 450. Parameter values in calibration: $m = 100$, $a = 15$, $b = 4$, $\theta = 0.6$ $t = 0.01$, distance increases in increments of 1/6, "low price" of $p_S = 6$ and "higher price" of $p_S = 7$. $p_N = 8$. Fixed costs at each distance are assumed to be drawn from a normal distribution with mean of 2 and standard deviation of 0.5. Panel b plots the difference between the two curves in a.

Differences in population can also weaken the hump as illustrated in A.5 where the population density is highest closest to the border. The higher population raises the demand response close to the border and thus works to diminish the hump-shaped pattern. The effect is less pronounced than in A.4 because close to the border there is less action coming from the extensive margin.

Figure A.5: Change in sales in N in response to a price change of good S and distance to S. A case with higher population in the area closest to the border.



Note: Panel a shows sales volume of good N at different distances from the border (location of good S) as described in text. Population 400 at each distance apart from distances below 50 where population is 450. Parameter values in calibration: $m = 100$, $a = 15$, $b = 4$, $\theta = 0.6$ $t = 0.01$, distance increases in increments of 1/6, "low price" of $p_S = 6$ and "higher price" of $p_S = 7$. $p_N = 8$. Fixed costs at each distance are assumed to be drawn from a normal distribution with mean of 2 and standard deviation of 0.5. Panel b plots the difference between the two curves in a.

Together Figures A.4 and A.5 suggest that demographic variables at different distance from the border may be of interest. Table A.4 presents summary statistics on the number of households per municipality, the average number of stores in the sample per municipality, number of households per store and disposable income. There are some differences but mainly it is the region in the 60-90 minute bin that sticks out, this is the region where major cities Oslo and Trondheim are located. To the extent that higher population would be associated with a stronger cross-price effect this would tend to strenghten the demand response in the region 60-90 minutes away, rather than in the 30-60 minute bin where we see the humpshape.

Table A.4: Summary statistics by driving duration to the closest Swedish store

	30	60	90	120	150	180
Households	80868.8	115861	689526	227135.2	157260.4	53099.4
	(1117.6)	(2304.6)	(47854.0)	(32170.2)	(3051.8)	(1015.7)
Income	432960	473742.4	499979.1	491055.8	466724.4	440175.5
	(19226.8)	(19177.6)	(19595.6)	(20668.7)	(19687.3)	(20106.8)
Education	21.85	22.99	28.48	24.75	24.17	20.87
	(1.010)	(0.961)	(1.221)	(0.998)	(0.852)	(0.884)
Stores	31.60	63.20	244.8	145	92.20	36.40
	(0.894)	(2.387)	(12.52)	(7.550)	(5.762)	(1.949)

Note: This table reports summary statistics by duration bins. Education, income and number of households are measured at the municipal-year level. Education is the percentage of the population in the municipality with higher education. Income is the median after tax income. To aggregate these variables to the duration bin level, we assign each municipality to the mode distance bin (across the stores in the municipality). To get yearly values at the duration bin level of education and income we take the mean across the municipalities associated with the bin. The yearly value at the bin level of the number of households is the sum of the number of households across the municipalities associated with the bin. Stores is the number of stores at the bin level in a given year. The table presents means and standard deviations (in parentheses) of the variables (across years and within bins).

To further explore the role of population we also estimate our benchmark regression, Equation 1 and include the number of households in the municipality as an additional control variable and report results in Table A.5.

Table A.6 finally presents the baseline regression where we control for both the number of households in the municipality and the share with higher education in the municipality.

Table A.5: Demand regressions, also controlling for population

	Meat	Cheese	Soda	Sweets
$30 \times \ln(P^N/P^S)$	-1.209	-0.655	-2.242	-0.755
	(0.094)	(0.072)	(0.188)	(0.174)
$60 \times \ln(P^N/P^S)$	-1.479	-0.830	-2.358	-1.337
	(0.213)	(0.089)	(0.169)	(0.177)
$90 \times \ln(P^N/P^S)$	-0.643	-0.528	-2.048	-0.457
	(0.118)	(0.076)	(0.137)	(0.149)
$120 \times \ln(P^N/P^S)$	-0.491	-0.460	-1.965	-0.225
	(0.126)	(0.091)	(0.157)	(0.162)
$150 \times \ln(P^N/P^S)$	-0.430	-0.512	-1.986	-0.169
	(0.200)	(0.126)	(0.180)	(0.270)
$180 \times \ln(P^N/P^S)$	-0.486	-0.645	-2.233	-0.960
	(0.223)	(0.123)	(0.197)	(0.331)
Constant	5.462	3.521	9.453	7.127
	(0.966)	(1.419)	(1.070)	(0.757)
Observations	34389	34389	34392	34391
R^2	0.471	0.443	0.330	0.388
Month number FE	Yes	Yes	Yes	Yes
Store format FE	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes
Number of households	Yes	Yes	Yes	Yes

Note: This table reports results from an estimation of the model specified in Equation 1, including the number of households in the respective municipality as a control variable. Monthly price indexes are calculated based on COICOP and the sample period is 2012-2016. The standard errors reported in parentheses are clustered at the local labor market level.

Table A.6: Demand regressions, also controlling for population and the share with higher education

	Meat	Cheese	Soda	Sweets
$30 \times \ln(P^N/P^S)$	-1.195	-0.543	-2.228	-0.666
	(0.085)	(0.080)	(0.171)	(0.150)
$60 \times \ln(P^N/P^S)$	-1.465	-0.717	-2.343	-1.246
	(0.238)	(0.111)	(0.153)	(0.194)
$90 \times \ln(P^N/P^S)$	-0.636	-0.457	-2.044	-0.425
	(0.124)	(0.082)	(0.132)	(0.152)
$120 \times \ln(P^N/P^S)$	-0.485	-0.390	-1.962	-0.196
	(0.118)	(0.096)	(0.149)	(0.168)
$150 \times \ln(P^N/P^S)$	-0.419	-0.417	-1.976	-0.105
	(0.207)	(0.123)	(0.169)	(0.276)
$180 \times \ln(P^N/P^S)$	-0.475	-0.551	-2.223	-0.896
	(0.202)	(0.120)	(0.184)	(0.293)
Constant	5.677	5.820	10.082	8.073
	(1.571)	(1.423)	(1.428)	(1.101)
Observations	34389	34389	34392	34391
R^2	0.471	0.447	0.330	0.389
Month number FE	Yes	Yes	Yes	Yes
Store format FE	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes
Number of households	Yes	Yes	Yes	Yes
Share higher education	Yes	Yes	Yes	Yes

Note: This table reports results from an estimation of the model specified in Equation 1, including the number of households and share with higher education in the respective municipality as a control variable. Monthly price indexes are calculated based on COICOP and the sample period is 2012-2016. The standard errors reported in parentheses are clustered at the local labor market level.