Appendix to "Bundling to save: Analyzing package size choices in South African grocery stores"

Andrea Szabó
Economics Department, University of Houston
E-mail: aszabo2@uh.edu

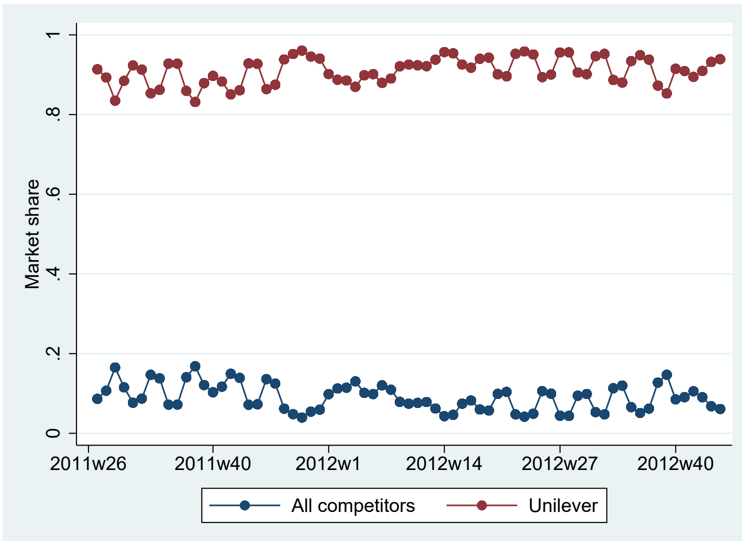
October 16, 2024

Contents

1 Additional data characteristics

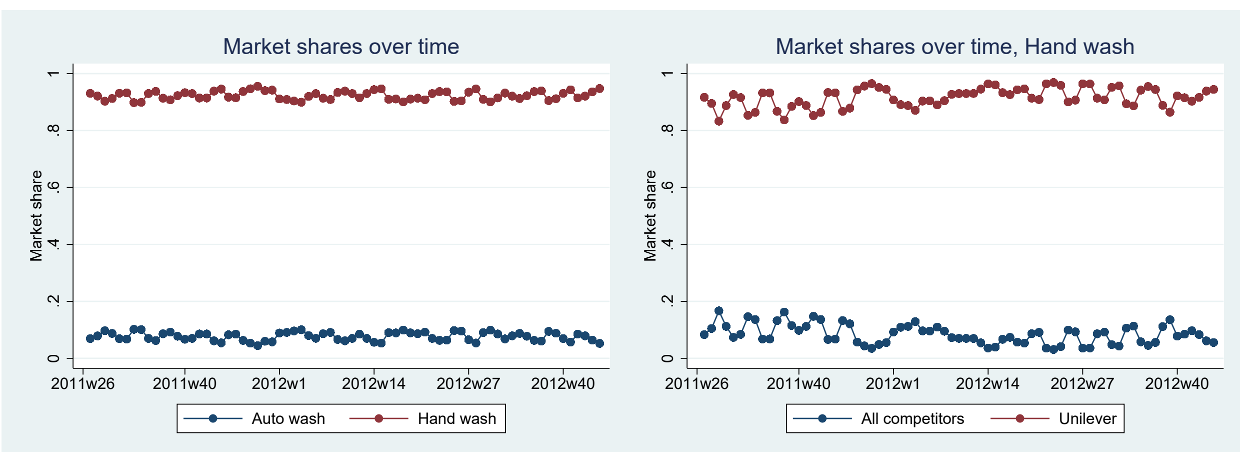
1.1 Market characteristics

Figure A.1: Market shares



Notes: Market shares based on sales value by week. Unilever products and all products by competitors.

Figure A.2: Market shares of hand wash and automatic detergents



Notes: Market shares based on sales value. Left panel: Market share of all handwash vs automatic detergents. Right panel: Market share of hand-wash detergents only, Unilever vs all competitors.

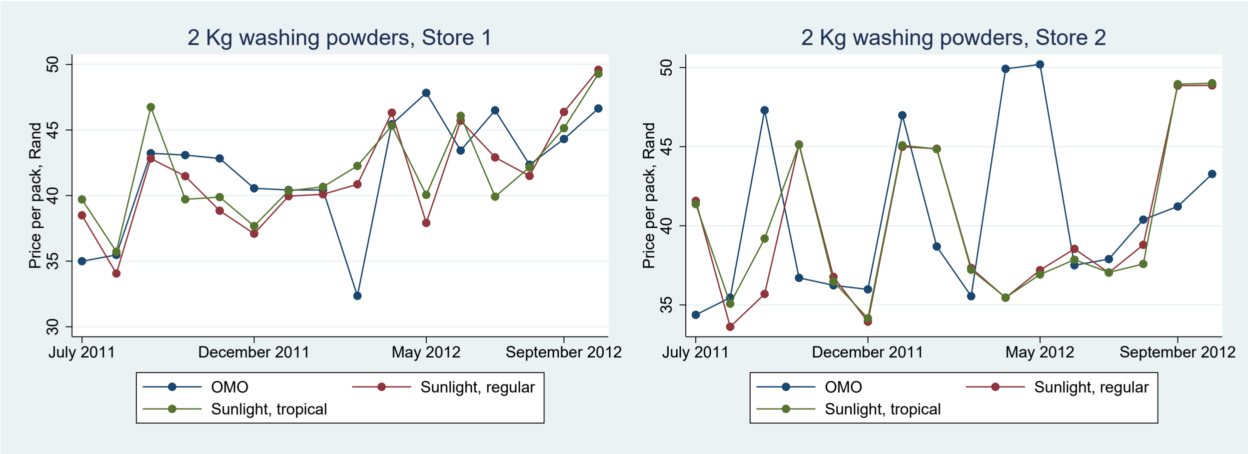
Table A.3: Availability of various brands and sizes

	Sunlight, regular	Sunlight, tropical	OMO
250g	0.94	0.81	0.95
500g	1	0.77	0.99
1kg	1	0.96	1
2kg	1	0.99	1
3kg	0.99		
5kg	0.98		

Notes: Fraction of all markets (months $\times$ stores) in the sample. N = 5255.

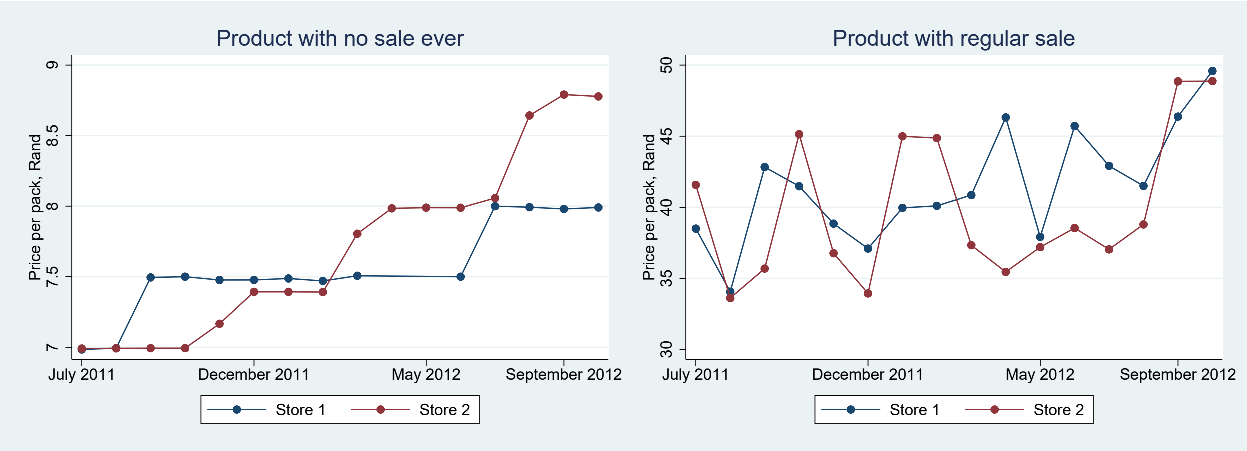
1.2 Prices

Figure A.4: Prices of 2 kg packages in selected stores



Notes: Store level monthly sales weighted prices in Rand.

Figure A.5: Evolution of prices with and without frequent promotions

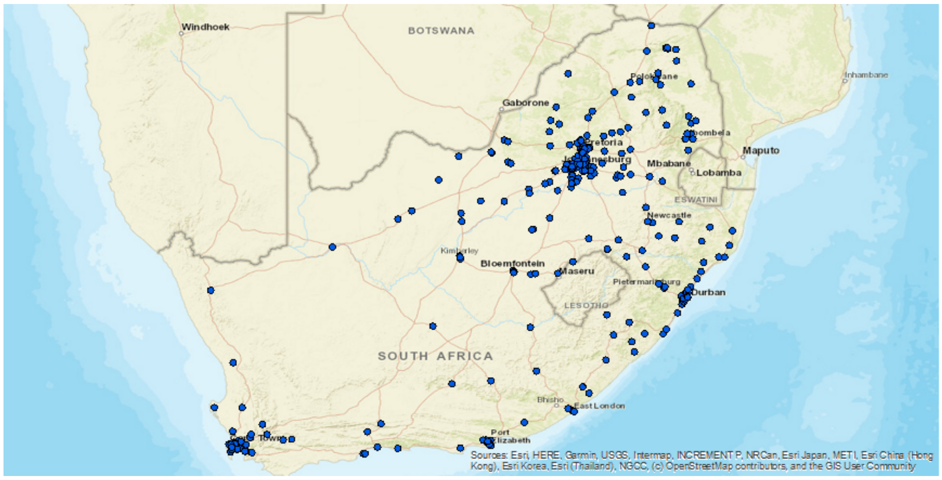


Notes: Store level monthly sales weighted prices in Rand. Left panel shows Sunlight, regular 250 g. Right panel shows Sunlight, regular 1 kg.

Figure 2shows the evolution of prices for two distinct products over time in two selected stores. The first product (Sunlight 250 g) had only 1 week of temporary promotion during the study period and the second (Sunlight 1 kg) had the most frequent promotions. The left panel of Figure A.5 shows that prices of the product with little promotion stay steady over a longer period in a given store, but they can be more than 10 percent higher in a different store. In both stores, we see an increasing price trend over the 16 month period, although the timing of the price increases is different. Figure **??** showed the distribution of prices for these same two products. The product with little temporary promotion has a smaller variation both across stores in a given month (upper panels) and over the period (lower panels).

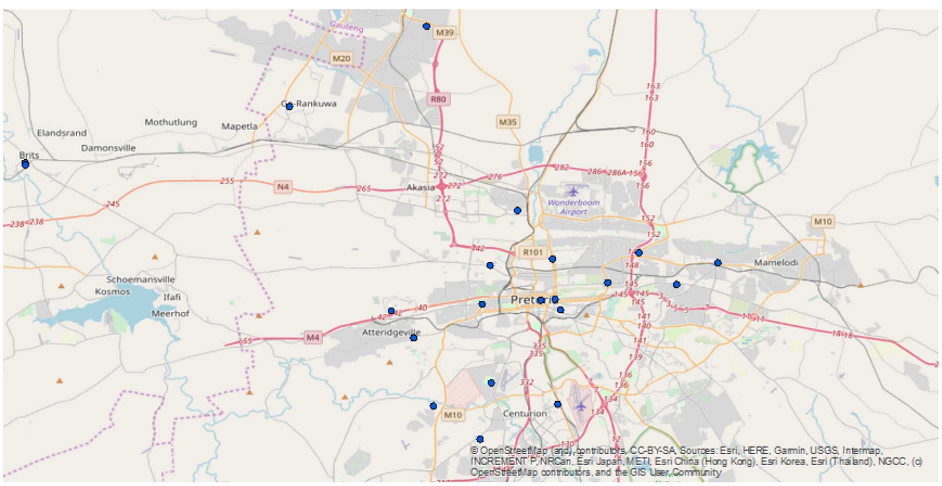
1.3 Store characteristics

Figure A.6: Location of stores in the sample



Notes: Locations based on GPS coordinates of the stores, collected from www.shoprite.co.za

Figure A.7: Location of the stores in the sample around Pretoria



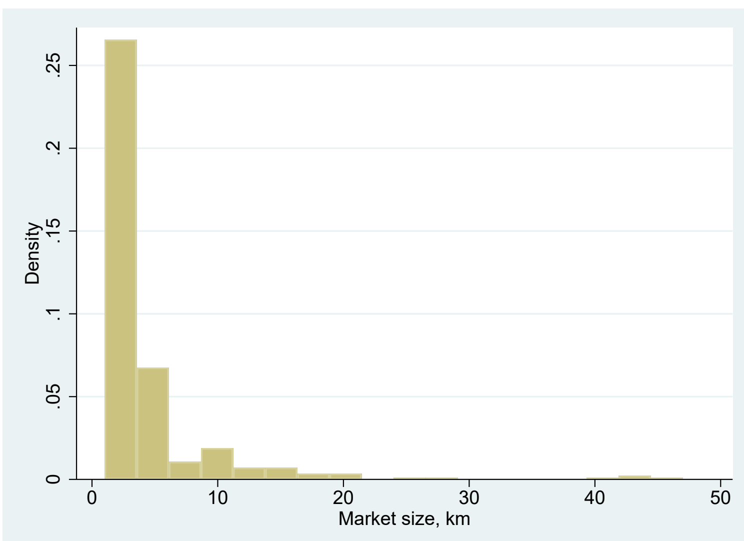
Notes: Locations based on GPS coordinates of the stores, collected from www.shoprite.co.za

Table A.8: Other store characteristics

	Percent
Low-income area	17.27
Middle-income area	43.64
High-income area	39.09
In a shopping mall	8.79
In city centre	24.85
Sunday closing time	
Not open	2.12
13	18.18
14	26.67
15	13.33
15:30	1.82
16	3.33
17	16.06
18	2.42
19	3.33
20	12.42
21	0.30

Notes: Area income categories: Source: Unilever. Other characteristics based on store locator information at www.shoprite.co.za

Figure A.9: Distribution of the market radius



Notes: Distribution of market radius corresponding to each store.

1.4 Census Data

The paper uses dataset 4.6. Household goods from the "Census 2011: Community Profiles" CD. The data is accessed using SuperCross, a software provided by the South African Census. The dataset has appliance ownership information which includes washing machine and car, besides basic household characteristics such as type of main dwelling, urban or rural location, gender and race of the household head and annual household income.

Table A.10: Household demographics

	South Africa	Store markets
Annual household income		
Mean	9092.11	10410.27
Urban area	68.03	84.48
Male household head	57.61	59.89
Owns car and washm	22.40	28.73
Owns no car or washm	55.81	46.44
Owns washm only	12.02	14.79
Owns car only	9.77	10.04
Population group of household head		
Black African	77.71	69.7
White	10.96	13.49
Other	11.33	16.8
Type of dwelling		
House	66.97	67.41
Flat/apartment	12.45	18.3
Other (Informal dwelling, shack in backyard)	20.58	14.28
N	10,261,921	3,012,142

Notes: Based on 2011 South African Census. Households on store markets are identified based on their distance from a store. See Section 6.1. for details. Income is annual household income in Rand. Other variables are percentages of total.

2 Details of the survey

2.1 Sampling

The survey was entirely funded by the University of Houston. It was approved by the Human Subject Committee of the University of Houston, and was conducted in accordance with the standards of that institution regarding the ethical treatment of human subjects (Protocol number: 2626). Participation in the survey was voluntary and respondents could stop participating in the survey at any time. Only adults between the ages of 18 and 65 were asked to participate.

Surveys were collected from 300 households. For logistical reasons, sampling had to be restricted to a single metropolitan area. I chose the area around Pretoria because of the diverse socio-economic characteristics of its population.

I first took all the stores in my dataset located within 20 miles from Pretoria (25 stores). I then extended this area 5 miles to the north to include more rural areas, resulting in a total of 27 stores. For marketing reasons, Unilever categorizes the stores into living standard measure (LSM) areas. Of these 27 stores, 4 are located in LSM areas 1-4 (low), 15 stores in LSM areas 5-6 (middle) and 8 stores in LSM areas 7-10 (high). One of these stores was closed at the time of the survey due to damage from a tornado. Of the remaining 26 stores, I randomly selected a store from each of the three LSM groups. I selected the sample of households to be surveyed around each of these 3 stores as follows.

For each store, I randomly selected 5 of the 10 closest small area layers of the 2011 South African Census. Surveyors were provided maps of each of these 5*3 areas. From each map, they selected an intersection, and starting from there interviewed 5 households in each direction. Specifically, surveyors visited every 5th house in each direction, subject to the constraint that the final sample had to be stratified based on dwelling type recorded in the Census ("house," "flat," and "informal/other"). Households to be interviewed were selected to match as closely as possible the corresponding fraction of each dwelling type from the census.

Surveyors recorded the GPS coordinate and a detailed description of the selected houses. Based on this information, surveyors visited the same houses during the second round of the survey.

2.2 Purchase, consumption, and inventory data

Out of the 300 respondents, 91.3 % typically buy powdered detergents only.[1] This is very close to the market share from 2013 (see Figure A.2).

I have 575 observations with purchase, inventory and consumption information. To infer whether the package currently in inventory was purchased during the past month, I compute the sum of the current inventory and consumption, and if this is smaller than the package size then I assume that the detergent was purchased more than a month ago. In this case, I assign "no purchase" to the current month for the given household. Otherwise, the household's purchase is the package they showed to the surveyor.

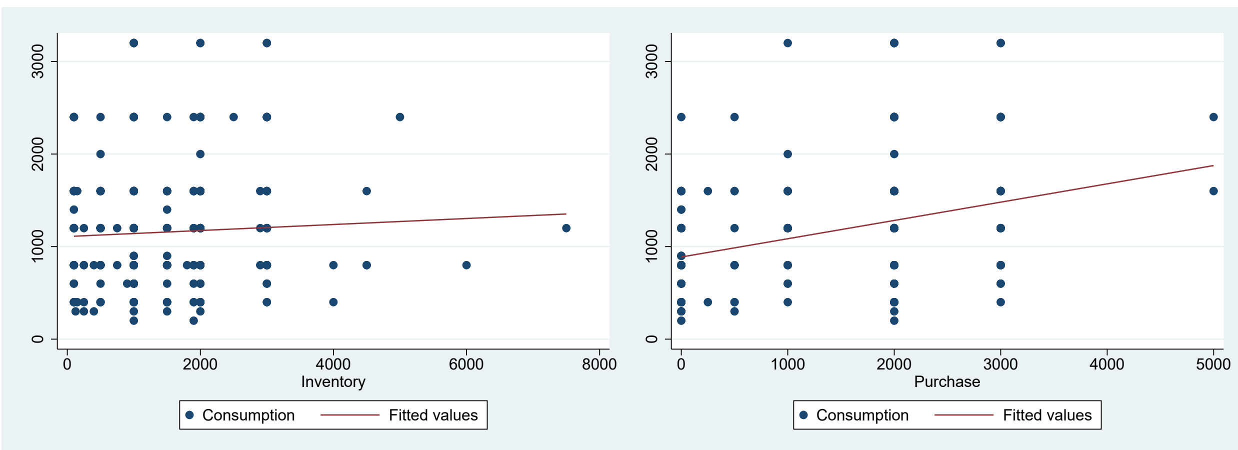
Based on the data, 32.87% of the households did not purchase detergent during the

[1] Only 4 respondents stated that they typically buy liquid detergents and only 1 said that they typically use bar soap instead of detergent. 21 additional respondents buy a combination of powdered detergent and either liquid or bar soap.

current month. This percentage is the highest (37.89%) for the highest LSM area. These are also the households who are somewhat more likely to purchase larger packages both in the survey and in the scanner data.

Two patterns are visible in the data. First, reported consumption is not correlated with inventory at home. This makes sense since the households are unlikely to use more detergent just because they have a new package at home, or do fewer loads because there is less detergent left in the package. Second, there is a positive, statistically significant correlation between consumption and purchase size. Households who tend to buy larger packages consume more on average. Figure A.11 shows both of these relations in the data.

Figure A.11: Consumption as a function of purchase and inventory in the survey

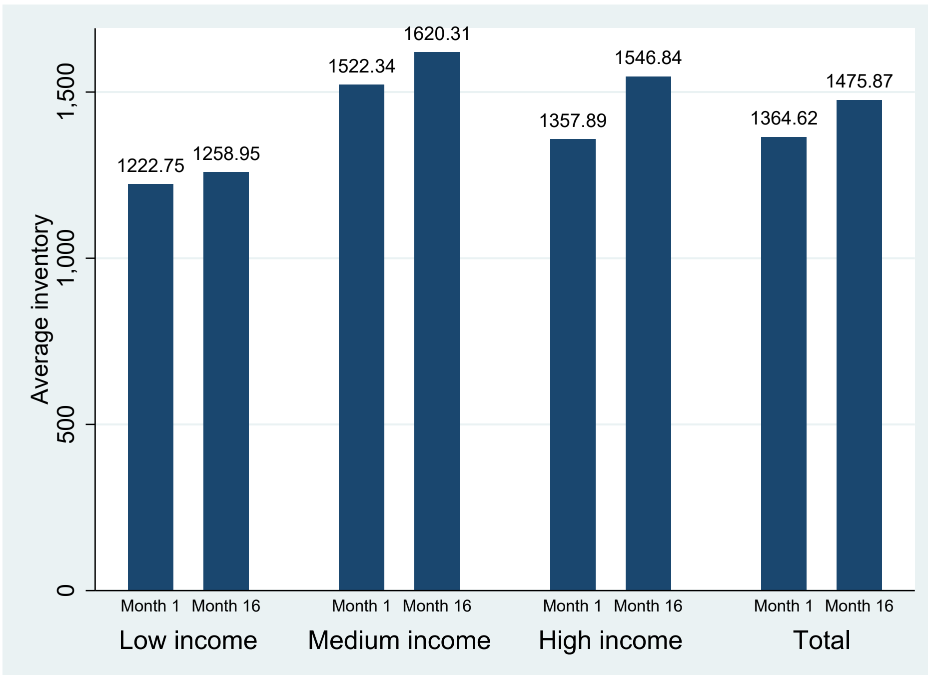


Consequently, I do not assume in the dynamic model that consumption depends on inventory directly. Instead I assume that consumption depends on household characteristics, including income of the area. The model also takes into account that current consumption cannot be larger then current inventory. This means that although the model assumes that a specific household has a preset consumption level (which changes only with a random consumption shock), it is still able to predict substantially lower consumption levels if inventory not met.

To use the survey data in the dynamic programming problem I do the following. Each observation of consumption inventory and purchased package size is randomly assigned to the model's simulated individuals based on the package size variable. That is, once the purchased size is drawn based on the market shares for simulated individuals for each market, survey data is randomly matched based on package size. This is done separately for markets in three income areas.

Another noteworthy feature of the survey is that average inventory during the first and the second round of the survey is not statistically different. This is the case for the average across all households or across household groups. Note that there is a 16-month difference between the first and the second round of the survey, which is the exactly the same time period I observe in the scanner data. Figure A.12 plots mean inventory across the round of surveys.

Figure A.12: Average inventory of households, by area



Notes: p-values of Month 1 vs. Month 16 differences: 0.7520, 0.5018, 0.1353, and 0.1394, respectively. Inventory is given in grams.

Figure A.12 implies that average consumption is the same as the average of the purchased quantity over the 16 month period. I use this information in computing the dynamic parameters of the consumer. Specifically, I first draw a sequence of 16 monthly purchased quantities based the observed market shares and I compute the average consumption based on the simulated purchase.

3 Details of the dynamic estimation

Markets and simulations. The dynamic estimation uses markets with no bundling opportunities, as well as markets that feature bundling opportunities. I drop only markets that do not have all brands and sizes to keep the consumer's choice set constant (430 out of 5255 markets). Finally, I only use markets where I have at least 16 consecutive time periods. This can happen because a few stores did open during the period and/or the store did not carry all sizes during the 16 month period.

This results in 528, 1248 and 992 markets, respectively, for each LSM group. From the static part of the estimation, I have 400 simulated consumers on each of these markets. To reduce the computational complexity of the dynamic estimation, I restrict attention to a random sample of markets and consumers. To solve the dynamic programming problem, I randomly draw 400 markets, and 50 consumers from each. For the dynamic estimation, I use all markets, with 50 consumers from each.

For each individual, for each market, the model predicts individual choice probabilities for each possible package size. Since I have 100 draws for consumption shock, I average predicted choice probabilities across these options when computing $Q(\theta_h)$.

Individual purchase, consumption, and inventory. For the dynamic programming problem, one needs to know purchased package size, consumption, and inventory at the individual level. I simulate purchases based on the observed market shares in the data. The survey data provides information on the joint distribution of inventory and consumption conditional on purchase. I draw inventory and consumption pairs for each (simulated) individual from this distribution. Observing this joint distribution in the survey helps identify the parameters of the flexible polynomial of state variables used to approximate the value function.

After solving the nested dynamic programming problem (for a given vector of dynamic parameters), I simulate over time the purchase (and inventory) decision of the consumers. There is no need to discretize either the consumption or the inventory levels. The maximum potential inventory is set to 50% higher than the highest observed inventory.

Outside option. In a typical BLP application the outside option is only a normalization, but the case here is different. In the dynamic problem, the outside option corresponds to a consumer not purchasing any detergent. To better approximate the share of no purchase, I compute the fraction of surveyed consumers who did not purchase detergent in the given month (these values are similar in both rounds of the survey). I normalize the observed market shares using this average, keeping the relative share of the outside good across markets constant.

4 Additional results

Table A.13: Correlation between bundling opportunities and market characteristics

	250 g	500 g	1 kg	2 kg	3 kg	5 kg
Low income area	0.037	0.045	0.008	0.016	0.009	0.008
	(0.028)	(0.028)	(0.016)	(0.017)	(0.017)	(0.015)
Middle income area	0.018	0.021	0.003	-0.002	-0.002	-0.006
	(0.016)	(0.016)	(0.011)	(0.012)	(0.011)	(0.009)
Mall	-0.017	-0.016	0.006	-0.015	-0.013	-0.015
	(0.021)	(0.022)	(0.012)	(0.015)	(0.015)	(0.012)
City center	0.025	0.019	-0.006	-0.012	-0.002	-0.005
	(0.016)	(0.017)	(0.010)	(0.013)	(0.013)	(0.010)
Sunday hours	0.010	0.009	-0.001	0.001	-0.006	-0.000
	(0.016)	(0.017)	(0.012)	(0.020)	(0.014)	(0.013)
HH share black	0.297	0.271	-0.081	0.007	0.051	0.059
	(0.071)	(0.070)	(0.036)	(0.037)	(0.039)	(0.035)
HH share white	-0.069	-0.068	-0.117	-0.074	0.004	0.015
	(0.086)	(0.081)	(0.040)	(0.056)	(0.055)	(0.054)
HH share flat	0.263	0.177	-0.059	0.067	0.035	0.083
	(0.088)	(0.082)	(0.076)	(0.099)	(0.072)	(0.061)
HH share house	0.163	0.101	-0.123	-0.030	-0.058	0.040
	(0.086)	(0.080)	(0.079)	(0.099)	(0.073)	(0.057)
HH share male HH head	0.512	0.506	-0.169	0.152	0.233	0.286
	(0.120)	(0.125)	(0.074)	(0.094)	(0.091)	(0.081)
HH share urban	-0.049	-0.044	-0.014	-0.053	-0.070	-0.059
	(0.040)	(0.039)	(0.026)	(0.027)	(0.029)	(0.025)
HH share no car or washm	-0.193	-0.205	-0.105	-0.043	-0.018	0.050
	(0.133)	(0.127)	(0.067)	(0.080)	(0.074)	(0.062)
HH share washm only	-0.391	-0.382	-0.198	-0.168	-0.020	-0.008
	(0.175)	(0.169)	(0.100)	(0.129)	(0.123)	(0.111)
HH share car only	-0.428	-0.294	0.017	-0.103	-0.024	-0.042
	(0.337)	(0.338)	(0.182)	(0.238)	(0.224)	(0.209)
Adj. R^2	0.26	0.24	0.38	0.33	0.40	0.49
Adj. R^2 controls only	0.25	0.23	0.38	0.32	0.40	0.49
N	14,189	14,483	15,548	15,696	5,199	5,167

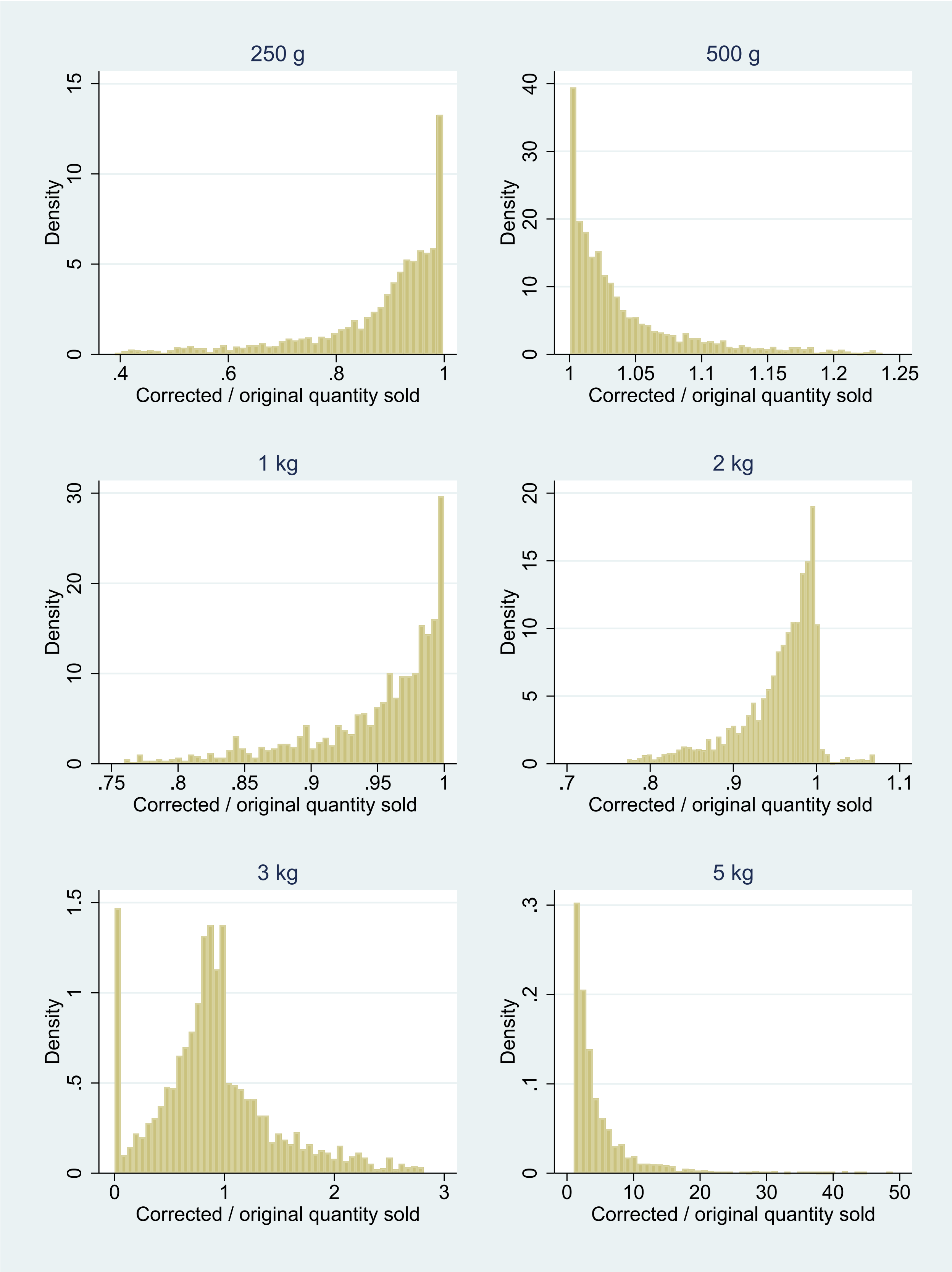
Notes: The dependent variable in each regression is an indicator for the presence of bundling opportunities. All regressions control for month and state fixed effects, market size, distance to closest store and number of other stores in the market.

Table A.14: Correlation between bundling opportunities and market characteristics

	Low income	Medium income	Mall	Centre	Sunday	Black	Flat
250 g	0.048	-0.005	-0.036	0.035	0.014	0.153	-0.022
	(0.024)	(0.015)	(0.018)	(0.018)	(0.015)	(0.032)	(0.036)
	0.25	0.25	0.25	0.25	0.25	0.26	0.25
500 g	0.053	-0.006	-0.035	0.029	0.012	0.147	-0.040
	(0.024)	(0.015)	(0.018)	(0.019)	(0.016)	(0.031)	(0.037)
	0.23	0.23	0.23	0.23	0.23	0.24	0.23
1 kg	0.004	-0.002	0.015	-0.008	0.000	-0.043	0.077
	(0.012)	(0.009)	(0.010)	(0.010)	(0.012)	(0.019)	(0.020)
	0.38	0.38	0.38	0.38	0.38	0.38	0.38
2 kg	0.020	-0.012	-0.012	-0.010	0.001	0.014	0.065
	(0.014)	(0.011)	(0.013)	(0.013)	(0.019)	(0.021)	(0.027)
	0.32	0.32	0.32	0.32	0.32	0.32	0.32
3 kg	0.018	-0.011	-0.018	-0.002	-0.006	0.033	0.044
	(0.014)	(0.010)	(0.013)	(0.012)	(0.013)	(0.018)	(0.026)
	0.40	0.40	0.40	0.40	0.40	0.40	0.40
5 kg	0.022	-0.012	-0.019	-0.002	0.002	0.057	-0.004
	(0.013)	(0.009)	(0.010)	(0.010)	(0.012)	(0.018)	(0.024)
	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	House	**Male HH**	**Urban**	**White**	**No car or washm**	**Washm only**	**Car only**
250 g	0.009	0.113	-0.081	-0.038	0.139	-0.412	0.889
	(0.039)	(0.103)	(0.037)	(0.033)	(0.042)	(0.096)	(0.250)
	0.25	0.25	0.25	0.25	0.25	0.25	0.25
500 g	0.011	0.143	-0.079	-0.033	0.135	-0.433	0.911
	(0.039)	(0.107)	(0.035)	(0.035)	(0.041)	(0.094)	(0.254)
	0.23	0.23	0.23	0.23	0.23	0.24	0.23
1 kg	-0.086	-0.085	-0.012	0.006	-0.040	-0.017	0.041
	(0.026)	(0.067)	(0.022)	(0.021)	(0.029)	(0.058)	(0.150)
	0.38	0.38	0.38	0.38	0.38	0.38	0.38
2 kg	-0.084	0.023	-0.050	-0.004	0.029	-0.149	0.291
	(0.032)	(0.080)	(0.022)	(0.024)	(0.033)	(0.065)	(0.165)
	0.32	0.32	0.32	0.32	0.32	0.32	0.32
3 kg	-0.072	0.059	-0.056	0.000	0.048	-0.158	0.309
	(0.030)	(0.076)	(0.021)	(0.024)	(0.028)	(0.063)	(0.158)
	0.40	0.40	0.40	0.40	0.40	0.40	0.40
5 kg	-0.019	0.063	-0.061	-0.020	0.081	-0.159	0.333
	(0.026)	(0.071)	(0.019)	(0.024)	(0.026)	(0.053)	(0.147)
	0.49	0.49	0.49	0.49	0.49	0.49	0.49

Notes: Univariate regressions of bundling opportunities for different sizes on market characteristics. All regressions control for month and state fixed effects, market size, distance to closest store and number of other stores in the market.

Figure A.15: Ratio of corrected and original quantities sold



Notes: 330 markets.

Table A.16: Parameter estimates: static demand

	250g	500g	1kg	2kg	3kg	5kg
Linear parameters						
Price / 100	-97.290	-53.601	-36.182	-15.916	-10.113	-15.526
	(13.923)	(17.745)	(6.917)	(1.402)	(1.171)	(5.507)
Omo	0.492	0.371	0.614	1.109		
	(0.183)	(0.247)	(0.349)	(0.172)		
Sun	0.313	0.543	0.358	0.377	-3.877	2.630
	(0.151)	(0.222)	(0.256)	(0.138)	(5.294)	(2.335)
Constant	-3.798	-0.917	-2.136	0.917		
	(4.222)	(2.119)	(4.850)	(1.189)		
Low-income area	-0.148	0.164	-0.106	-0.108	-0.269	0.246
	(0.114)	(0.140)	(0.188)	(0.112)	(0.110)	(0.105)
Middle-income area	0.029	0.105	-0.062	-0.053	-0.116	0.163
	(0.086)	(0.102)	(0.135)	(0.069)	(0.070)	(0.081)
Mall	-0.235	-0.107	-0.044	-0.196	0.089	-0.297
	(0.094)	(0.108)	(0.179)	(0.094)	(0.117)	(0.115)
Center	-0.222	-0.147	-0.140	0.073	-0.088	-0.113
	(0.051)	(0.052)	(0.085)	(0.070)	(0.088)	(0.074)
Open Sunday	-0.052	-0.229	-0.012	0.122	0.059	0.096
	(0.058)	(0.048)	(0.089)	(0.056)	(0.066)	(0.080)
Neighboring stores	-4.808	-1.073	0.398	2.041	2.480	1.867
	(1.159)	(0.805)	(1.742)	(1.216)	(0.965)	(1.026)
Dist. to nearest store	-0.749	0.001	0.478	0.988	0.674	0.819
	(0.229)	(0.203)	(0.381)	(0.282)	(0.239)	(0.260)
HH share flat	-0.193	0.061	-0.745	-1.648	-2.172	-0.919
	(0.407)	(0.390)	(0.556)	(0.519)	(0.504)	(1.196)
HH share house	-0.306	0.094	0.004	-1.117	-1.195	-1.806
	(0.299)	(0.347)	(0.605)	(0.455)	(0.344)	(0.601)
HH share White	1.601	1.467	2.555	0.877	0.656	0.970
	(0.475)	(0.379)	(0.970)	(0.315)	(0.332)	(0.416)
HH share Black	-0.758	-0.419	-0.472	-0.213	-0.008	0.229
	(0.188)	(0.208)	(0.372)	(0.204)	(0.214)	(0.228)
HH share urban	0.320	0.263	0.113	-0.415	-0.693	-0.423
	(0.199)	(0.195)	(0.222)	(0.278)	(0.198)	(0.193)
Omo × HH share Black	0.110	0.679	0.204	0.201		
	(0.118)	(0.175)	(0.171)	(0.087)		
Sunlight x HH share Black	0.647	1.113	1.014	1.012		
	(0.107)	(0.176)	(0.155)	(0.074)		
Omo × HH share urban	-0.302	-0.222	-0.208	-0.746		
	(0.151)	(0.212)	(0.325)	(0.163)		
Sun × HH share urban	-0.320	-0.066	0.022	-0.258		
	(0.132)	(0.189)	(0.234)	(0.127)		
Omo × Low-inc area	-0.042	-0.350	-0.083	-0.273		
	(0.118)	(0.148)	(0.196)	(0.113)		
Sun × Low-inc area	0.098	-0.313	0.071	-0.112		
	(0.098)	(0.139)	(0.139)	(0.086)		
Omo × Mid-inc area	-0.077	-0.140	-0.113	0.009		
	(0.081)	(0.101)	(0.118)	(0.066)		
Sun × Mid-inc area	-0.043	-0.132	-0.096	0.014		
	(0.073)	(0.095)	(0.102)	(0.054)		

Cont'd on next page

Table A.16 cont'd

	250g	500g	1kg	2kg	3kg	5kg
Non-linear parameters						
Price × male	19.839	22.998	8.600	0.388	-0.446	6.251
	(8.907)	(15.893)	(2.759)	(1.261)	(0.973)	(5.303)
Constant × income	6.094	1.818	1.636	1.575	5.871	1.558
	(4.059)	(1.736)	(0.962)	(1.165)	(4.681)	(2.502)
Constant × no car or washm	4.237	2.175	4.919	-1.799	-5.297	-1.570
	(1.619)	(0.870)	(2.673)	(1.093)	(256.806)	(4.780)
Constant × washm only	4.027	2.701	6.896	1.831	2.682	1.538
	(1.346)	(0.793)	(2.651)	(0.675)	(1.688)	(1.202)
Constant × car only	-2.050	-6.024	5.490	2.276	2.925	1.396
	(61.514)	(693.013)	(2.613)	(0.778)	(1.198)	(0.817)
J	1.305	1.613	0.095	1.156	4.137	0.501
p-value	0.521	0.446	0.758	0.561	0.530	0.779
Newey-West D	48.441	51.073	52.501	20.960	29.922	33.820
p-value	0.000	0.000	0.000	0.001	0.000	0.000
N (market × products)	2798	2888	3193	3230	1049	1029
Unique markets	1064	1088	1088	1088	1049	1029
Unique months	16	16	16	16	16	15
Unique stores	326	326	326	326	324	325

Notes: Parameter estimates from the BLP model. Standard errors robust to heteroskedasticity and intra-market correlation in parentheses. All specifications contain province and quarter fixed effects. J is the overidentification test statistic with corresponding p-value. Newey-West D is a likelihood ratio test for the null hypothesis that the nonlinear parameters are jointly 0 with the corresponding p-value.

Table A.17: Inclusive values

	250 g	500 g	1 kg	2 kg	3 kg	5 kg
ω_{t-1}^{250}	0.927	-0.02	-0.043	-0.061	0.000	0.113
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.005)
ω_{t-1}^{500}	0.018	0.976	-0.013	0.139	-0.097	-0.041
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.003)
ω_{t-1}^{1}	-0.001	-0.031	0.941	0.233	-0.142	-0.05
	(0.004)	(0.004)	(0.004)	(0.009)	(0.008)	(0.004)
ω_{t-1}^{2}	-0.029	0.012	0.035	0.647	0.191	0.138
	(0.006)	(0.006)	(0.006)	(0.013)	(0.013)	(0.006)
ω_{t-1}^{3}	-0.036	-0.052	-0.067	0.201	0.783	0.068
	(0.004)	(0.004)	(0.004)	(0.010)	(0.010)	(0.004)
ω_{t-1}^{5}	0.05	0.019	0.007	0.247	-0.03	0.775
	(0.006)	(0.006)	(0.006)	(0.011)	(0.011)	(0.008)
Constant	0.071	0.068	0.069	0.212	-0.066	-0.179
	(0.007)	(0.007)	(0.008)	(0.011)	(0.012)	(0.008)
Adj. R^2	0.94	0.97	0.94	0.86	0.93	0.87
N	1,712,000	1,712,000	1,712,000	1,712,000	1,712,000	1,712,000

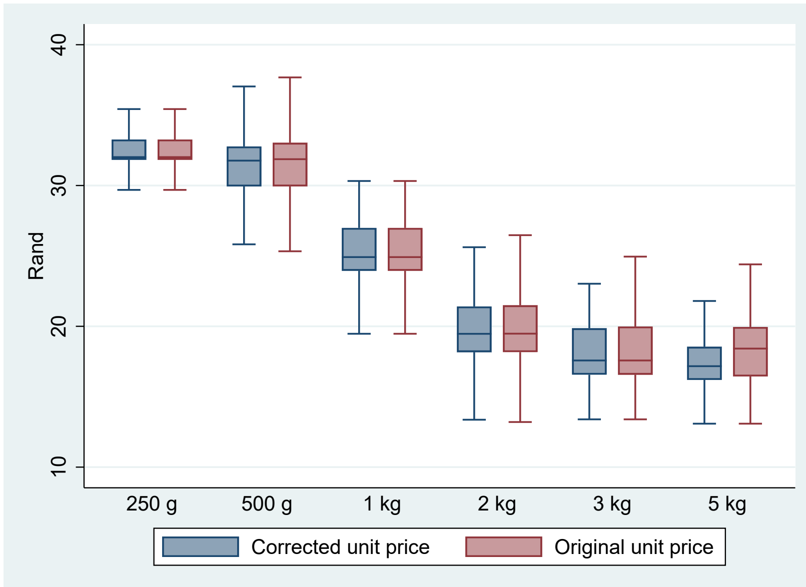
Notes: Estimates of the inclusive value process. The explanatory variables are lagged values of the inclusive value of every package size.

Table A.18: Fit of different inclusive value specifications

	250 g	500 g	1 kg	2 kg	3 kg	5 kg
Income area						
Low-income area	0.94	0.97	0.93	0.84	0.93	0.89
Middle-income-area	0.94	0.97	0.94	0.85	0.94	0.87
High-income area	0.93	0.96	0.94	0.87	0.92	0.86
Ownership status						
No car or washm	0.74	0.89	0.82	0.81	0.72	0.79
Washm only	0.82	0.9	0.81	0.84	0.74	0.79
Car only	0.75	0.88	0.8	0.81	0.7	0.78
Car and washm	0.79	0.89	0.8	0.84	0.71	0.77
Alternative specifications						
Sum of five additional lags	0.95	0.98	0.96	0.88	0.95	0.89
Second lag added	0.94	0.97	0.95	0.86	0.94	0.88

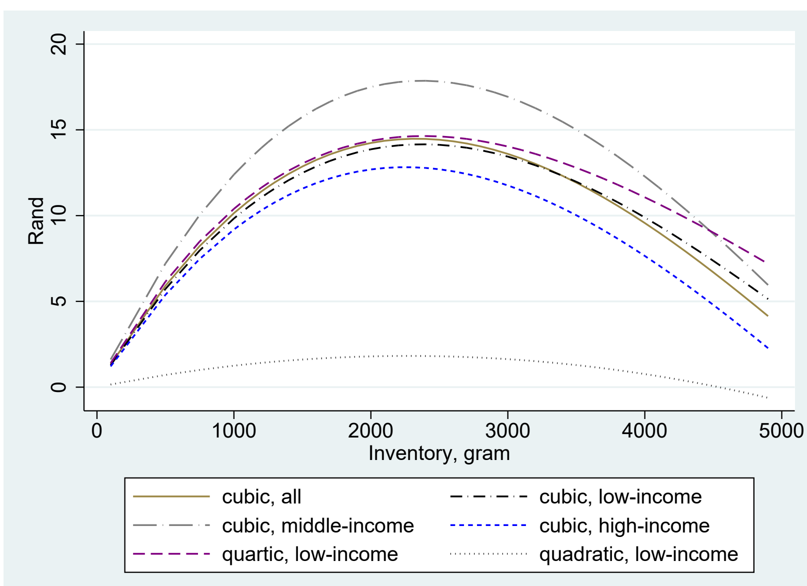
Notes: Adjusted R^2 values from different specifications of the inclusive value process. The top panel estimates separate processes by income area. The middle panel estimates separate processes by car/washing machine ownership. On the bottom panel, the first row includes the sum of five additional lags (t-2 to t-6) and the last row includes one additional lag (t-2).

Figure A.19: Average unit price by package size



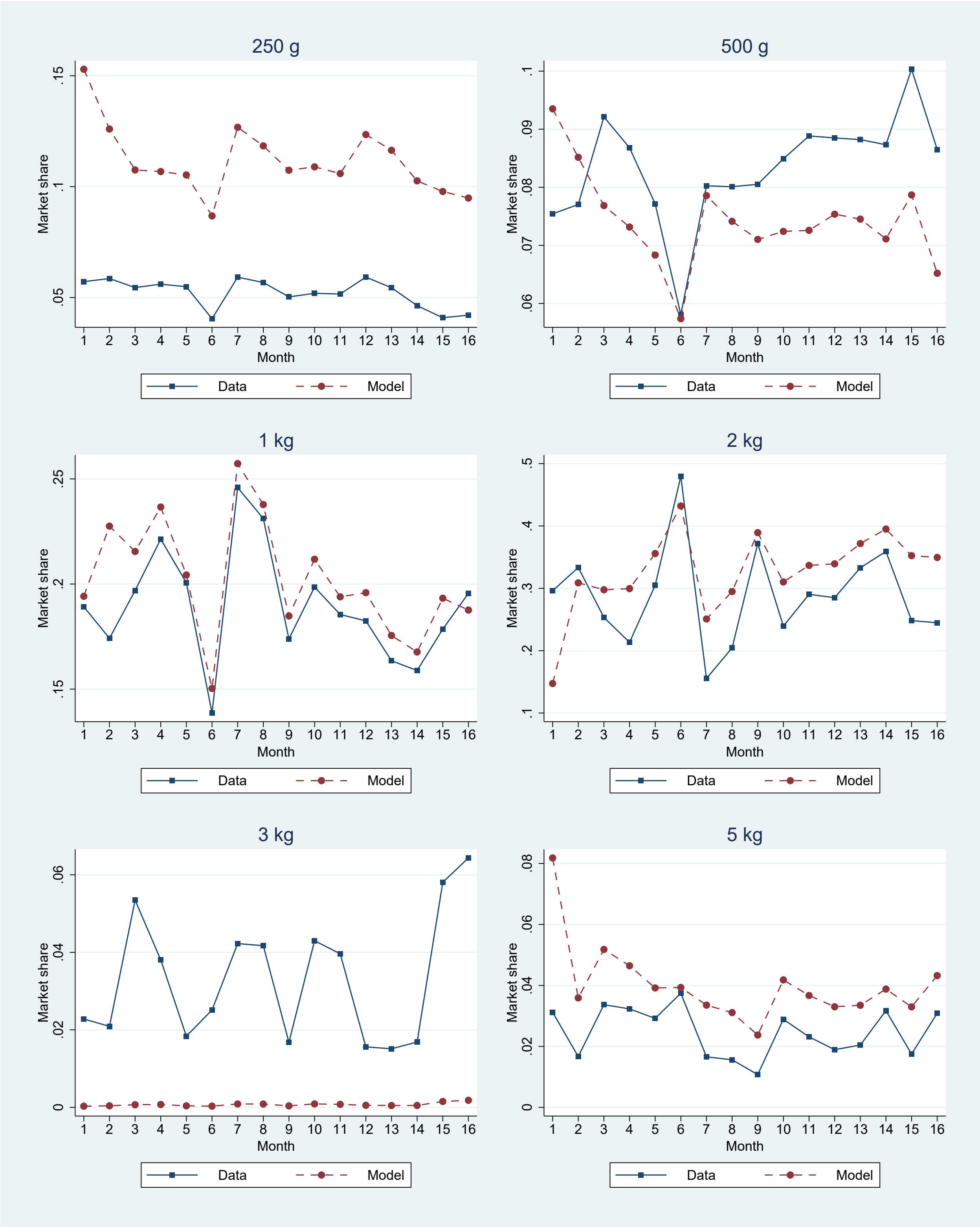
Notes: Original unit prices refer to prices observed in the data. Corrected refers to prices corrected for bundling opportunities as described in the paper.

Figure A.20: Estimated inventory cost



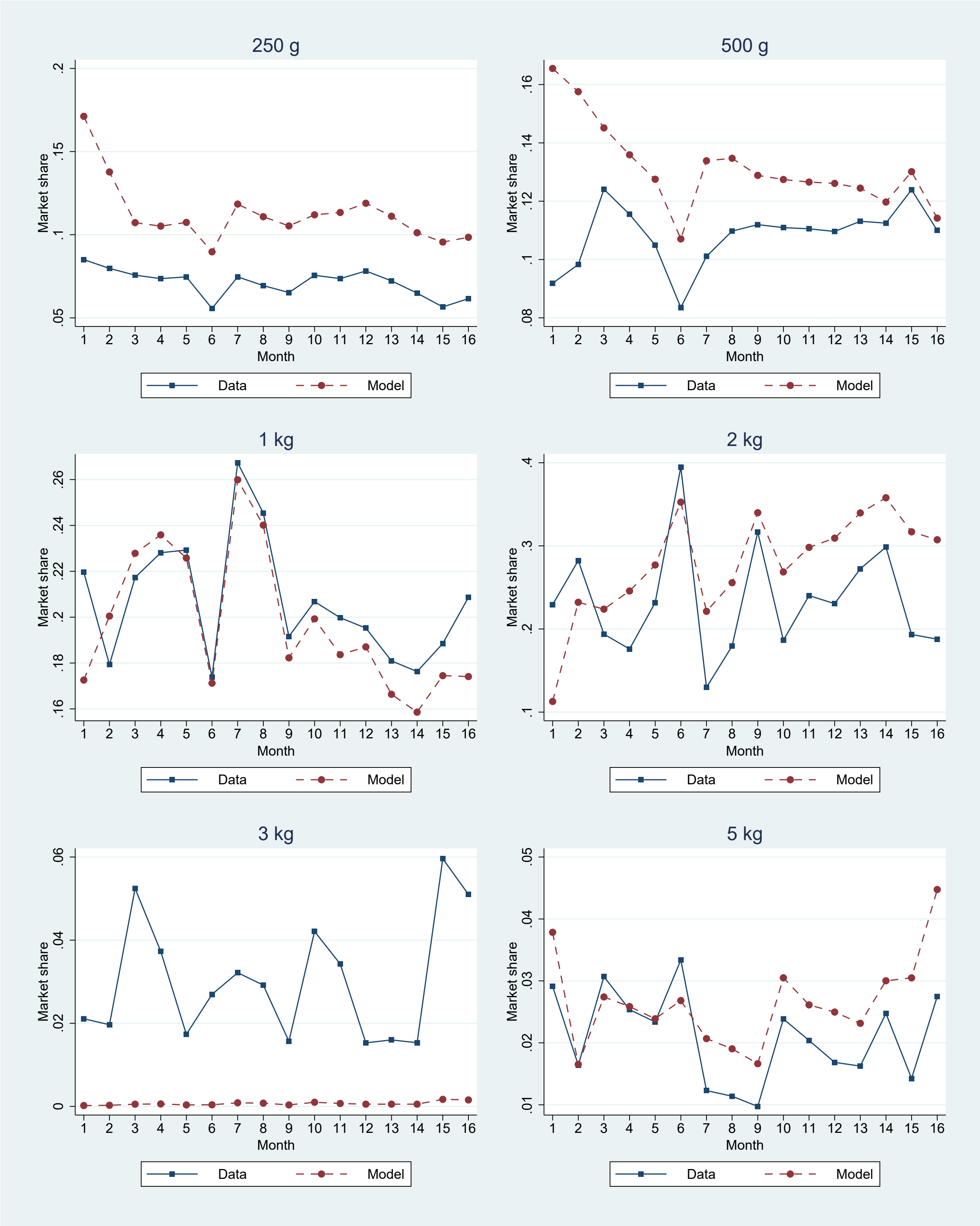
Notes: Inventory costs predicted by the dynamic model. Cubic specifications correspond to columns (1) - (4) in Table 7. Quartic and quadratic specifications are for low-income areas. See the text for details.

Figure A.21: Model fit, Middle-income area



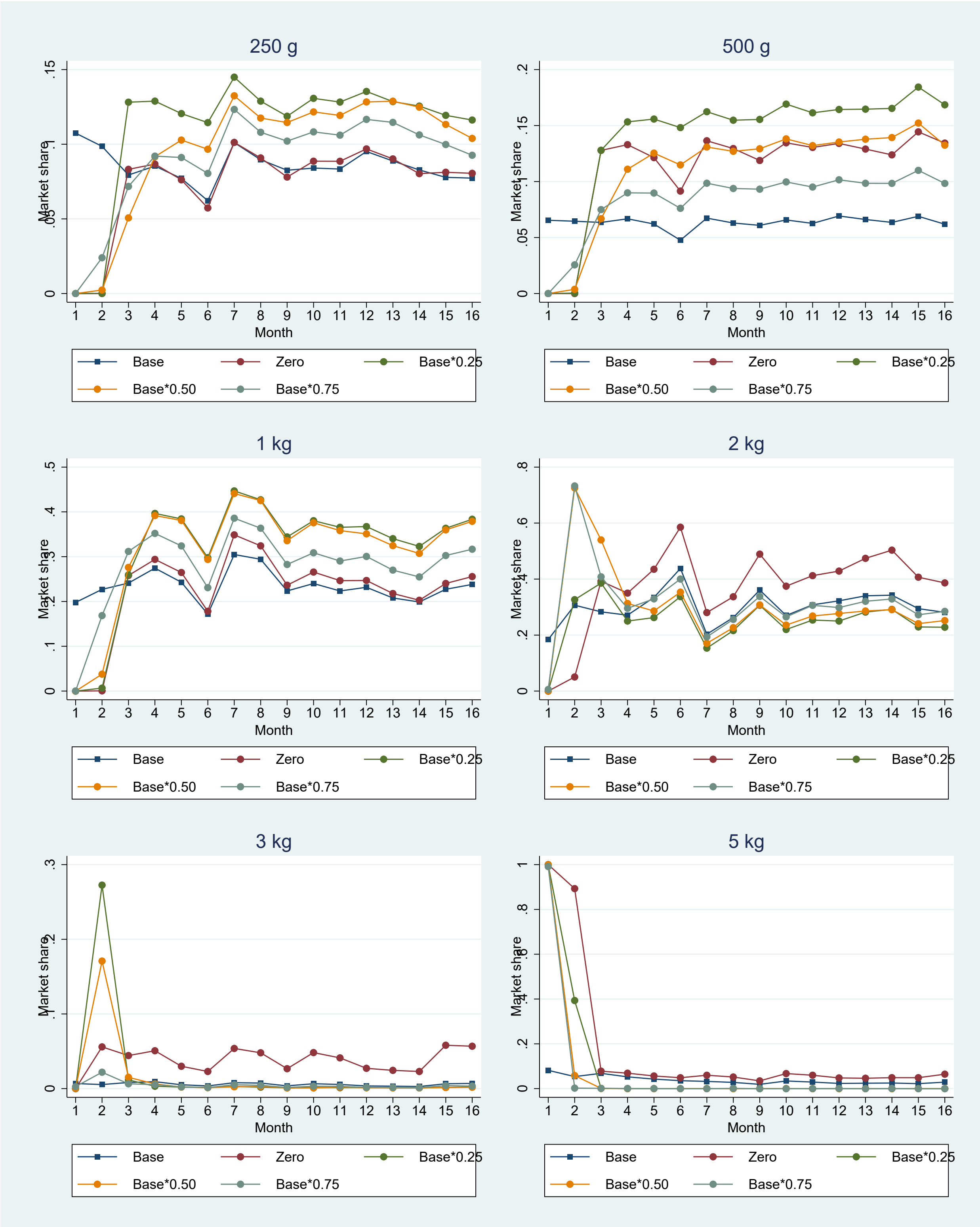
Notes: Market shares of different sizes observed in the data and predicted by the model, Middle-income area, estimation sample

Figure A.22: Model fit, High-income area



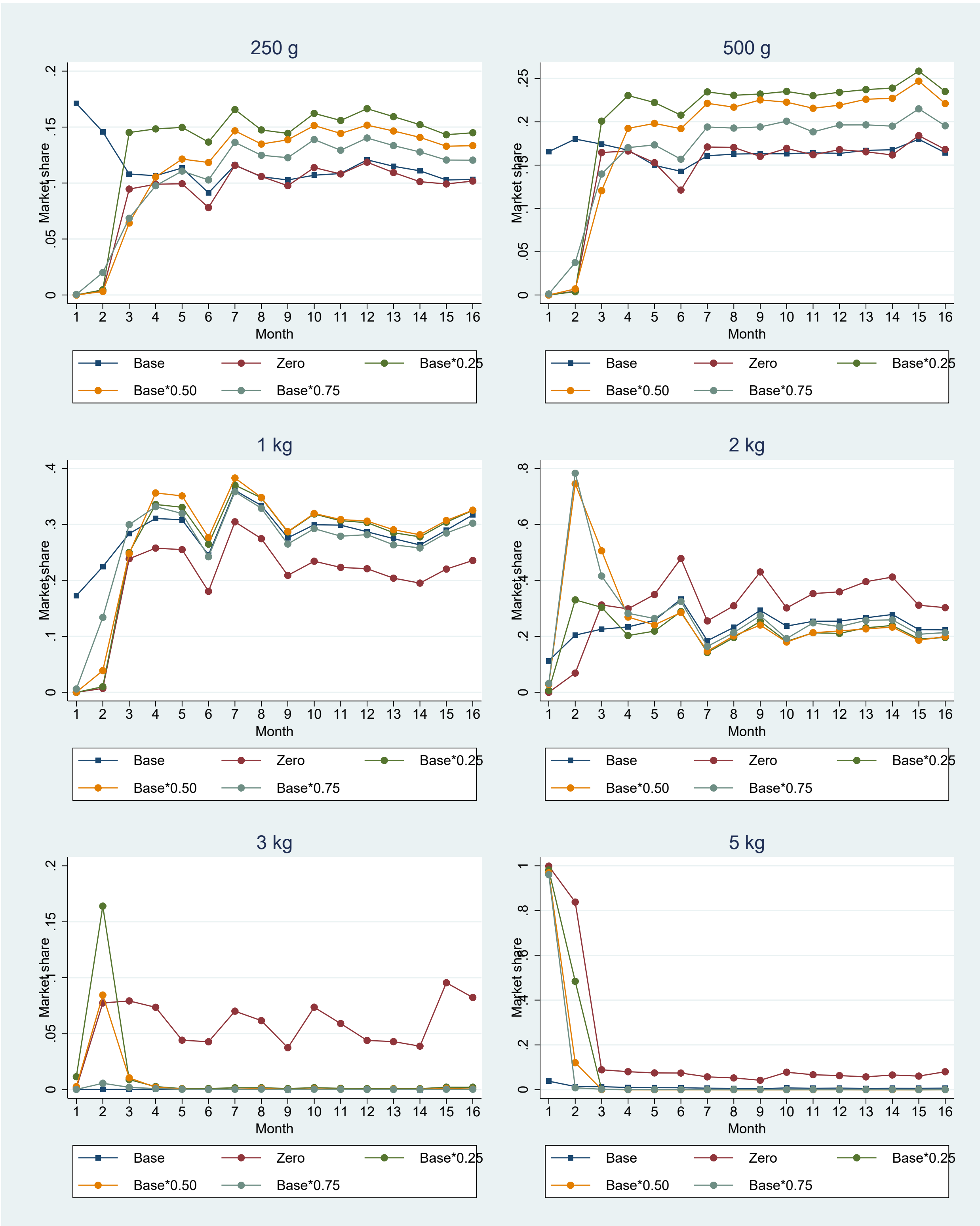
Notes: Market shares of different sizes observed in the data and predicted by the model, High-income area, estimation sample

Figure A.23: Counterfactual market shares with reduced fixed cost of purchase, low-income areas



Notes: 50 simulated consumers for each store.

Figure A.24: Counterfactual market shares with reduced fixed cost of purchase, high-income areas



Notes: 50 simulated consumers for each store.

Table A.25: Household inventory and consumption when reducing the fixed cost of purchase

	Low-income area		High-income area	
	Consumption	Inventory	Consumption	Inventory
Baseline	98.54	316.37	87.94	226.01
Base*0.75	98.54	701.52	87.94	739.76
Base*0.50	98.54	784.14	87.94	809.16
Base*0.25	98.54	860.25	87.94	887.77
Zero	98.54	1343.09	87.94	1406.75

Notes: Each row corresponds to a different scenario where the consumer faces reduced fixed cost of purchasing each size compared to the baseline case. "Zero" refers to the case with no fixed costs. Simulations span a period of 16 months, with 50 individuals per store. Consumption and inventory are measured in 10 g.

Table A.26: Counterfactual simulations

	Middle-income area					
	250 g	500 g	1 kg	2 kg	3 kg	5 kg
Consumption						
Average	35.40	55.21	80.23	86.62	84.54	88.73
Median	25	50	90.58	84.61	82.76	87.30
Inventory						
Average	14.94	20.27	71.04	221.44	115.84	260.48
Median	0	0	11.08	106.20	123.15	274.80
Purchase probability						
Average	0.50	0.48	0.47	0.36	0.20	0.16
Median	0.50	0.48	0.47	0.35	0.19	0.14
Utility level (expected)						
Average	144.20	149.64	150.64	162.45	147.62	164.63
Median	141.08	146.16	146.34	151.98	145.34	161.74

Notes: Each column corresponds to a different scenario where the consumer's choice set is restricted to the given size (or the outside option). The simulations span a period of 16 months, with 50 individuals per store. Consumption and inventory are measured in 10 g.

Table A.27: Counterfactual simulations, zero fixed cost of purchase

	Low-income area						High-income area					
	250 g	500 g	1 kg	2 kg	3 kg	5 kg	250 g	500 g	1 kg	2 kg	3 kg	5 kg
Consumption												
Average	36.66	58.95	90.11	98.56	98.57	98.57	40.00	59.78	82.31	87.32	87.33	87.33
Median	25	50	98.65	98.65	98.65	98.65	25	50	84.02	84.02	84.02	84.02
Inventory												
Average	15.27	43.97	255.67	1024.19	1217.93	2503.92	27.67	69.08	291.87	909.74	1117.48	2095.82
Median	0	0	206.51	964.19	1066.34	2258.03	0	0	260.14	880.13	1045.72	1910.69
Purchase probability												
Average	0.81	0.85	0.96	0.96	0.64	0.69	0.81	0.85	0.97	0.97	0.66	0.70
Median	0.81	0.85	0.98	0.98	0.58	0.69	0.81	0.85	0.98	0.99	0.56	0.67
Utility level (expected)												
Average	160.80	170.34	197.09	248.71	253.71	257.63	126.44	132.13	148.80	185.48	190.26	194.55
Median	159.29	169.23	194.91	246.30	251.54	255.15	122.80	128.20	145.38	182.93	187.12	191.12

Notes: Each column corresponds to a different scenario where the consumer's choice set is restricted to the given size (or the outside option). The simulations span a period of 16 months, with 50 individuals per store. Consumption and inventory are measured in 10 g.