Australian Government
Department of the Prime Minister and Cabinet

BETA

Energy labels that make cents

Testing energy rating labels on appliances sold online

Additional analysis: mixed-effects model

June 2018

Contents

Summary

This paper presents additional analysis undertaken as part of, and foreshadowed in, the ‘Energy labels that make cents’ report [hereafter, the Energy Labels report; Commonwealth of Australia Department of the Prime Minister and Cabinet, 2018]. This paper is, in effect, an extra appendix to the report.

Following the primary analysis for the Energy Labels report, we undertook an additional exploratory analysis. This exploratory analysis comprised a mixed-effects regression incorporating product price and star rating as covariates. This approach allowed us to assess how much the probability of a consumer purchasing, adding to cart, or clicking a product changes with each additional star rating. We report here the nature of this analysis and its results.

Overall, we found that there was not a consistent pattern of results. Some evidence suggested there may be a positive effect of the existing energy label on click outcomes, leading to more clicks for higher efficiency products in comparison to no label and the alternative label. However, taken in the context of broader inconsistency of results in this exploratory analysis, we do not regard the evidence as sufficiently reliable to conclude that either of the labels were more effective than no label or each other. We do not consider the current finding to be inconsistent with the previous results generated by the primary analysis. In the absence of a clear finding emerging from this more complex analysis, we opt to defer to the findings of the simpler and more readily interpretable primary analysis.

Rationale for the current analysis

In the primary analysis of this trial, we aggregated the data for each customer. This aggregation involved constructing three binary variables for each customer which indicated whether they clicked, added to cart or purchased any product with an energy rating that was above the median of the experimental products in its respective product category. This provided a straightforward method for analysing the effect of the labels on customer engagement with higher energy efficient products, that is, by testing the difference of proportions in each experimental group. However, it also meant that we were not able to include covariates such as star rating and product price in the analysis, since covariate values can only be ascribed to individual products, and not to aggregate 'scores' that relate to multiple products for each customer.

There are different reasons for using these two covariates in an additional analysis. In the case of product price, the benefit is an increase in the precision of our estimate. This occurs because price, when entered as a covariate, controls for the variation in outcomes induced by prices differences.

In the case of star rating, the benefit is that the analysis is able to yield an estimate of the change in probability of an outcome occurring with each unit increase in star rating. This provides a more sophisticated test of our hypothesis that the effect of the labels would be 'larger in magnitude for products with higher energy efficiency ratings'.

In contrast to the main analysis, which was conducted on an aggregated dataset with one record per consumer, the current analysis was conducted on a fully disaggregated long-form dataset that included an observation for each product that each customer viewed on a filter page. Irrespective of whether a consumer's engagement with a product proceeded further than viewing the product on the filter page, our analysis included a binary outcome variable indicating whether they clicked, added to cart or purchased the product.[1]

The current analysis comprises a linear mixed-effects regression analysis. Linear mixed-effects regression is essentially similar to multiple linear regression. However, it has the added complexity of a 'random-effect component' in addition to the usual 'fixed-effects' factors, such as the experimental effect, entered into the model. This random-effect component is each customer, which is included in the model to address the clustering, and the potential correlation of outcomes, of the multiple

[1] As discussed in the Energy Labels report (page 30), this was done to avoid introducing selection bias into our estimates that would be induced by changing the base population in different ways in the different treatment groups.

observations for each customer. The interpretation of the analysis for the fixed-effects factors is essentially the same as in multiple linear regression.

Descriptive statistics

As noted in the Energy Labels report, we applied labels to 34 experimental products. These products comprised 10 refrigerators, 8 washing machines, 8 smart televisions and 8 clothes dryers. These products varied on a range of factors including price, energy efficiency and the degree to which consumers engaged with them (Table 1). In summary:

- *Average product price*: The products ranged in price from $333 for the cheapest clothes dryer to $3,736 for the most expensive smart TV. Dryers and washing machines were generally cheaper than smart TVs or fridges.
- *Energy efficiency*: The products ranged from 1.5 stars to 6 stars although the range was more limited for individual product categories: 1.5-2 stars for dryers, 2.5-4 stars for refrigerators, 2-4 stars for washing machines, and 4-6 stars for televisions.
- *Avoided losses*: The amount that consumers could 'avoid losing' in lifetime running costs – in comparison to the equivalent one-star product – ranged from $135 to $2,010. However, with the exception of clothes dryers, the avoided losses were always greater than $500 and were usually large relative to the sale price for the product. (For example, for all but four non-dryer products, the avoided losses comprised 70 per cent or more of the sale price.)
- *Views of filter pages*: Of the approximately 39,000 consumers who visited an experimental filter page, most (27,200) viewed washing machines. By contrast, 8,700 viewed fridges, 4,300 viewed dryers and 2,000 viewed televisions.
- *Clicks, add to carts and purchases*: Of those who viewed a filter page, the proportion of consumers who clicked on a product, to view more information about it, ranged from 1.1% to 19.4%, with a median of 5.0%. The proportion who added a product to cart ranged from 0% to 3.7%, with a median of 0.6% and the proportion who purchased a product ranged from 0% to 1.1% with a median of 0.2%. In general, consumers showed the most interest in dryers and the least in smart TVs.

Method of analysis

Trial participants. Consistent with the approach of the primary analysis, the dataset used for the current analysis included all consumers who viewed a filter page on the Appliances Online website during the trial period (n=39,614).

Outcomes. For each product viewed by each consumer, we recorded three binary outcome variables: whether they clicked on the product, whether they added the product to cart, and whether they purchased the product.

Observations per consumer. For each consumer, we recorded observations for each product they viewed in a long form dataset i.e. separate records for each product that each consumer viewed. The number of observations for each consumer varied depending on which of the four product category filter pages they had viewed (10 for fridges, 8 otherwise, unless they viewed filter pages for more than one product category). For the purpose of the analysis, it was assumed that if a consumer viewed a particular filter page, then they had viewed all products on this page.[2]

Covariates. We used two covariates in our analyses, average product price[3] and star rating. We centered both variables and interacted each with the treatment indicator as per Lin (2013). We centered product price on the mean and we centered star rating on the median. The latter was done to deal with skew in the distribution of star rating across the products and to be consistent with our primary analysis in which we separated products into high and low around the median. For each product category, these values were:

- Refrigerators: $1,231.90 and 3.25 stars.
- Washing machines: $597.05 and 3 stars.
- Televisions: $1,317.74 and 5 stars
- Clothes dryers: $416.66 and 1.75 stars

For the analysis of all products, the median star rating was 3.25. For reasons outlined below, product price was not included as a covariate in the analysis of all products.

Consumer behaviour is influenced by the relative price of close substitutes. Different product categories (for example, fridges and smart TVs) are not close substitutes, but the different sub-categories (for example, top load and front load washing machines) are. In one category, a certain price may be considered expensive relative to close substitutes, whereas in another category the same price may be considered relatively inexpensive compared to close substitutes. Thus, for analysis

[2] The exact number of people recorded for each product's filter page views differs due to a minor data collection issue which affected less than 0.5% of the sample. In some instances, the log data recorded a click, add to cart or purchase of a product being made without recording an observation for events that necessarily preceded such occurrences, such as product views, clicks or add to carts respectively. There was no systematic pattern of missingness in this data so we have excluded it.
[3] The price of each experimental product varied on up to a daily basis. For the purpose of this analysis, we calculated the average price of each product across the length of the trial and used it as a fixed factor for each product.

that includes price as a covariate, it is not appropriate to combine different product categories.

Model. Consistent with the approach of the primary analysis, the current analysis compared each of the three experimental groups (Control, Standard Label, and Alternative Label) to each other in order to test our hypotheses that the labels would outperform the control and the alternative label would outperform the standard label. In addition, we also tested the two label groups pooled together in comparison to the control group, as was done in the primary analysis for statistical considerations, namely an issue with sample size. These experimental comparisons were conducted for each of the three outcomes.

Given the constraints incurred in the use of product price as a covariate noted above, we conducted the above-described analysis for each of the four product categories separately. The fixed-effects factors for each model comprised experimental group, star rating, product price, the multiplicative interactions of star rating with experimental group, and the multiplicative interaction of product price with experimental group. Hereafter, each of these sets of models are referred to as product category sets.

Given that conducting the analyses in separate product categories reduces the sample size and thereby the statistical power of the analyses, we also conducted the analysis using data for all products combined, but for the reasons noted above, without price and its interaction with the experimental group as a covariate. Hereafter this set of models is referred to as the combined products set.

Altogether, the combinations of factors we were interested in testing led to the construction of 60 models. Specifically, each of these models was a linear mixed-effects regression model[4], with the above-specified fixed-effects factors and a random intercept for each consumer, which was used to account for variance in engagement between different consumers. All analyses were conducted using the 'multilevel mixed-effects models' package in Stata version 15.0.

In each of the models, the component which is relevant to our hypothesis (that the effect of the labels would be larger in magnitude for products with higher energy efficiency ratings) is the interaction of the experimental group with star rating. The coefficients for this term indicate the difference in differences between the two experimental groups in terms of the percentage point change in probability of the outcome occurring with each unit increase in star rating. Where the change in probability is more positive (or less negative) in the treatment group than the control

[4] We also conducted logistic mixed-effects regression for each as a robustness check. This produced a similar pattern of results except in some instances where the logistic model failed to converge due to perfect prediction among some covariate strata. Thus we do not report the results of the logistic models.

group, then the coefficient is positive, and vice versa for negative coefficients. Simply put, positive coefficients indicate a beneficial effect of the labels on energy-efficient purchasing behaviour, and negative coefficients indicate an undesirable effect.

Results

We identified no clear pattern of findings in the results. We describe below the results and our reasoning for this interpretation.

From each of the models, we have extracted only the relevant output and presented them in Tables 2-6. Each cell of the tables presents the beta coefficient of the experimental group by star rating interaction term, its associated p-value and the model intercept for each of the models in the combined product set (Table 2) and each of the four product category sets (Tables 3-6).

Generally, the coefficients are very small due to the low prevalence of events. That is, the actual number of products purchased, clicked or added to cart is very small in comparison to the number of products that could have been engaged with had every consumer purchased, clicked or added to cart every product they viewed on a filter page.

There were two models for which the interaction term reached the conventional cut-off for statistical significance (alpha<0.05)[5]. These were both for the click outcome in the combined products set, comparing the standard energy rating label to the no-label control condition (p=0.013) and to the alternative label condition (p=0.002) as shown in Table 2. They suggest the existing energy label was yielding a greater proportion of clicks on higher star-rated products than in the no-label control and the alternative label conditions.

There is, however, a broader inconsistency in the results which suggest this finding may be unreliable. Specifically, the direction of the effects of the labels is frequently inconsistent across outcomes within the same set of products and experimental group comparisons, and moreover, this inconsistency in of itself is not consistent across the different product category sets. It is possible that these inconsistencies reflect chance deviations in the differences between groups. Furthermore, if we were to apply corrections for multiple comparisons, given the large number of tests we conducted, we would be even more conservative in our interpretation of the reliability of the findings. Therefore, our interpretation is that these findings should not be regarded as sufficiently reliable to conclude there is a true effect for these two specific outcomes.

[5] We are aware that there is a lively academic debate about the merits of testing for ‘statistical significance’, the appropriateness of conventional thresholds such as p<0.05 (or any thresholds at all), and even the use of p-values generally. See, in particular, the ‘The American Statistical Association Statement on Statistical Significance and P-Values’ (Wasserstein and Lazar, 2016).

Conclusion

The current analysis did not yield further insights into the effects of energy rating labels on consumer behavior. We do not consider the current finding to be inconsistent with the previous results generated by the primary analysis. In the absence of a clear finding emerging from this more complex analysis, we opt to defer to the simpler and more readily interpretable findings from the primary analysis.

Table 1: List of appliances that formed part of the trial

Product	Type	Brand	Model	Star rating	Mean price $	Avoid losing $	View	Click (% of view)	Add to cart (% of view)	Purchase (% of view)
Televisions	Smart TV	TCL	32P1S	6	382	502	2,023	140 (6.9%)	20 (1.0%)	4 (0.2%)
	Smart TV	LG	55UH652T	6	1,317	1,219	2,022	60 (3.0%)	11 (0.5%)	1 (0.1%)
	Smart TV	TCL	50E5900US	5	752	877	2,020	79 (3.9%)	11 (0.5%)	1 (0.1%)
	Smart TV	Samsung	UA75JU6400	5	3,736	2,010	2,023	38 (1.9%)	2 (0.1%)	0 (0%)
	Smart TV	Samsung	UA40KU6000	4	1,050	801	2,023	54 (2.7%)	3 (0.2%)	0 (0%)
	Smart TV	Samsung	UA50KU6000	4	1,319	756	2,023	59 (2.9%)	10 (0.5%)	1 (0.1%)
	Smart TV	Samsung	UA55KU6000	4	1,358	910	2,023	71 (3.5%)	7 (0.35%)	1 (0.1%)
	Smart TV	Samsung	UA60KU6000	4	2,148	1,080	2,023	23 (1.1%)	0 (0%)	0 (0%)
Refrigerators	Bottom mount	Samsung	SRL458ELS	4	1,171	1,820	8,724	357 (4.1%)	24 (0.3%)	11 (0.1%)
	Bottom mount	Samsung	SRL457MW	4	988	1,820	8,725	484 (5.6%)	54 (0.6%)	16 (0.2%)
	Bottom mount	Westinghouse	WBE5300SARH	3.5	1,293	1,681	8,725	421 (4.8%)	46 (0.5%)	17 (0.2%)
	Multi-door	Fisher & Paykel	RF522ADUSX5	3.5	2,365	1,666	8,724	186 (2.1%)	20 (0.2%)	5 (0.1%)
	Multi-door	Westinghouse	WHE5200SA-D	3	1,568	1,418	8,725	222 (2.5%)	26 (0.3%)	9 (0.1%)
	Multi-door	Electrolux	EQE6207SD	3	2,083	1,612	8,725	199 (2.3%)	18 (0.2%)	4 (0.1%)
	Multi-door	Samsung	SRF583DLS	2.5	1,742	1,214	8,726	165 (1.9%)	6 (0.1%)	1 (0.01%)
	Top mount	Samsung	SR318LSTC	3.5	665	1,345	8,720	932 (10.7%)	136 (1.6%)	43 (0.5%)
	Top mount	Samsung	SR254MW	3.5	551	1,213	8,720	767 (8.8%)	101 (1.2%)	42 (0.5%)
	Top mount	Haier	HRF224FW	2.5	477	786	8,724	570 (6.5%)	55 (0.6%)	10 (0.1%)

Table 1 (continued): List of appliances that formed part of the trial

Product	Type	Brand	Model	Star rating	Mean price $	Avoid losing $	View	Click (% of view)	Add to cart (% of view)	Purchase (% of view)
Washing Machines	Front load	Bosch	WAW28460AU	4	1,119	1,866	27,242	2,074 (7.6%)	264 (1.0%)	74 (0.3%)
	Front load	Bosch	WAE22466AU	4	628	1,679	27,246	948 (3.5%)	129 (0.5%)	47 (0.2%)
	Front load	Samsung	WW75J4233GW	4	616	1,839	27,243	1,312 (4.8%)	215 (0.8%)	74 (0.3%)
	Front load	LG	WD12021D6	3.5	687	1,415	27,236	1,166 (4.3%)	145 (0.5%)	41 (0.2%)
	Front load	LG	WD1200D	3.5	571	1,437	27,237	1,429 (5.3%)	223 (0.8%)	46 (0.2%)
	Front load	Euromaid	WM7	3	477	1,278	27,228	2,383 (8.8%)	538 (2.0%)	176 (0.7%)
	Top load	Samsung	WA80F5G4DJW	2.5	578	1,088	27,242	1,644 (6.0%)	317 (1.2%)	111 (0.4%)
	Top load	Samsung	WA65F5S2URW	2	456	583	27,241	1,451 (5.3%)	274 (1.0%)	78 (0.3%)
Clothes Dryers	Vented	Fisher & Paykel	DE5060M1	2	419	216	4,267	769 (18.0%)	143 (3.4%)	39 (0.9%)
	Vented	Fisher & Paykel	DE4060M1	2	351	161	4,269	830 (19.4%)	158 (3.7%)	48 (1.1%)
	Vented	Electrolux	EDV6051	2	509	250	4,279	194 (4.5%)	21 (0.5%)	6 (0.1%)
	Vented	Euromaid	DM4KG	2	333	161	4,281	401 (9.4%)	47 (1.1%)	23 (0.5%)
	Vented	Fisher & Paykel	DE6060G1	2	665	250	4,280	485 (11.3%)	63 (1.5%)	18 (0.4%)
	Vented	Midea	MDV07	2	414	294	4,280	307 (7.2%)	56 (1.3%)	18 (0.4%)
	Vented	Simpson	39P400M	1.5	353	135	4,278	249 (5.8%)	30 (0.7%)	6 (0.1%)
	Vented	Simpson	39S500M	1.5	486	184	4,278	306 (7.2%)	30 (0.7%)	4 (0.1%)

Table 2: Interaction of star rating with experimental group for all products

	Control Group vs Energy Rating Label	Control Group vs Alternative Label	Energy Rating Label vs Alternative Label	Control Group vs Either Label (pooled)
Click	$3.0x10^{-3}$ (p=0.013) Intercept: $5.8x10^{-2}$	$-7.1x10^{-4}$ (p=0.56) Intercept: $5.8x10^{-2}$	$-3.7x10^{-3}$ (p=0.002) Intercept: $5.5x10^{-2}$	$1.1x10^{-3}$ (p=0.27) Intercept: $5.8x10^{-2}$
Add to cart	$2.3x10^{-4}$ (p=0.63) Intercept: $8.2x10^{-3}$	$-1.6x10^{-5}$ (p=0.97) Intercept: $8.2x10^{-3}$	$-2.5x10^{-4}$ (p=0.61) Intercept: $8.6x10^{-3}$	$1.1x10^{-4}$ (p=0.79) Intercept: $8.2x10^{-3}$
Purchase	$-8.6x10^{-5}$ (p=0.75) Intercept: $2.4x10^{-3}$	$-7.9x10^{-5}$ (p=0.77) Intercept: $2.4x10^{-3}$	$5.2x10^{-6}$ (p=0.99) Intercept: $2.6x10^{-3}$	$-8.1x10^{-5}$ (p=0.73) Intercept: $2.4x10^{-3}$

Note: results are reported as the beta coefficient for the interaction term in each model, the associated p-value and the intercept of the model, which occurs at the median star-rating (3.25) due to centring. A positive effect indicates the treatment group was superior to the control, or in the case of the alternative vs. standard label that the alternative was superior, and vice versa for a negative effect. Other fixed-effects factors from each model, which are not reported, include experimental group and star rating.

Table 3: Interaction of star rating with experimental group for refrigerators

	Control Group vs Energy Rating Label	Control Group vs Alternative Label	Energy Rating Label vs Alternative Label	Control Group vs Either Label (pooled)
Click	$6.3x10^{-3}$ (p=0.063) Intercept: $5.3x10^{-2}$	$4.0x10^{-3}$ (p=0.25) Intercept: $5.3x10^{-2}$	$-2.3x10^{-3}$ (p=0.49) Intercept: $4.8x10^{-2}$	$5.1x10^{-3}$ (p=0.081) Intercept: $5.3x10^{-2}$
Add to cart	$4.7x10^{-4}$ (p=0.69) Intercept: $5.8x10^{-3}$	$-5.9x10^{-4}$ (p=0.62) Intercept: $5.8x10^{-3}$	$-1.1x10^{-3}$ (p=0.36) Intercept: $5.5x10^{-3}$	$-6.7x10^{-5}$ (p=0.95) Intercept: $5.8x10^{-3}$
Purchase	$-4.5x10^{-5}$ (p=0.95) Intercept: $1.9x10^{-3}$	$9.8x10^{-5}$ (p=0.88) Intercept: $1.9x10^{-3}$	$1.4x10^{-4}$ (p=0.83) Intercept: $1.8x10^{-3}$	$2.7x10^{-5}$ (p=0.96) Intercept: $1.9x10^{-3}$

Note: results are reported as the beta coefficient for the interaction term in each model, the associated p-value and the intercept of the model, which occurs at the median star-rating (3.25) due to centring. A positive effect indicates the treatment group was superior to the control, or in the case of the alternative label vs. standard label that the alternative was superior, and vice versa for a negative effect. Other fixed-effects factors for each model, which are not reported, include experimental group, star rating, product price and the multiplicative interaction of product price with experimental group.

Table 4: Interaction of star rating with experimental group for washing machines

	Control Group vs Energy Rating Label	Control Group vs Alternative Label	Energy Rating Label vs Alternative Label	Control Group vs Either Label (pooled)
Click	$2.3x10^{-3}$ (p=0.26) Intercept: $6.1x10^{-2}$	$7.9x10^{-4}$ (p=0.70) Intercept: $6.1x10^{-2}$	$-1.5x10^{-3}$ (p=0.45) Intercept: $5.8x10^{-2}$	$1.6x10^{-3}$ (p=0.38) Intercept: $6.1x10^{-2}$
Add to cart	$7.7x10^{-5}$ (p=0.93) Intercept: $1.1x10^{-2}$	$6.3x10^{-4}$ (p=0.47) Intercept: $1.1x10^{-2}$	$5.5x10^{-4}$ (p=0.53) Intercept: $1.1x10^{-2}$	$3.5x10^{-4}$ (p=0.64) Intercept: $1.1x10^{-2}$
Purchase	$4.8x10^{-4}$ (p=0.32) Intercept: $3.3x10^{-3}$	$5.0x10^{-4}$ (p=0.30) Intercept: $3.3x10^{-3}$	$2.3x10^{-5}$ (p=0.96) Intercept: $3.2x10^{-3}$	$4.9x10^{-4}$ (p=0.24) Intercept: $3.3x10^{-3}$

Note: results are reported as the beta coefficient for the interaction term in each model, the associated p-value and the intercept of the model, which occurs at the median star-rating (3) due to centring. A positive effect indicates the treatment group was superior to the control, or in the case of the alternative label vs. standard label that the alternative was superior, and vice versa for a negative effect. Other fixed-effects factors for each model, which are not reported, include experimental group, star rating, product price and the multiplicative interaction of product price with experimental group.

Table 5: Interaction of star rating with experimental group for televisions

	Control Group vs Energy Rating Label	Control Group vs Alternative Label	Energy Rating Label vs Alternative Label	Control Group vs Either Label (pooled)
Click	$5.0x10^{-3}$ (p=0.21) Intercept: $3.2x10^{-2}$	$-1.3x10^{-3}$ (p=0.73) Intercept: $3.2x10^{-2}$	$-6.3x10^{-3}$ (p=0.12) Intercept: $4.1x10^{-2}$	$1.8x10^{-3}$ (p=0.60) Intercept: $3.2x10^{-2}$
Add to cart	$7.0x10^{-4}$ (p=0.64) Intercept: $2.6x10^{-3}$	$4.2x10^{-4}$ (p=0.74) Intercept: $2.6x10^{-3}$	$-2.8x10^{-4}$ (p=0.87) Intercept: $7.1x10^{-3}$	$5.6x10^{-4}$ (p=0.66) Intercept: $2.6x10^{-3}$
Purchase	$1.2x10^{-4}$ (p=0.82) Intercept: $3.0x10^{-4}$	$1.1x10^{-4}$ (p=0.80) Intercept: $3.0x10^{-4}$	$-3.5x10^{-6}$ (p=0.99) Intercept: $9.5x10^{-4}$	$1.2x10^{-4}$ (p=0.80) Intercept: $3.0x10^{-4}$

Note: results are reported as the beta coefficient for the interaction term in each model, the associated p-value and the intercept of the model, which occurs at the median star-rating (5) due to centring. A positive effect indicates the treatment group was superior to the control, or in the case of the alternative label vs. standard label that the alternative was superior, and vice versa for a negative effect. Other fixed-effects factors for each model, which are not reported, include experimental group, star rating, product price and the multiplicative interaction of product price with experimental group.

Table 6: Interaction of star rating with experimental group for clothes dryers

	Control Group vs Energy Rating Label	Control Group vs Alternative Label	Energy Rating Label vs Alternative Label	Control Group vs Either Label (pooled)
Click	$1.4x10^{-3}$ (p=0.94) Intercept: $9.4x10^{-2}$	$1.6x10^{-2}$ (p=0.39) Intercept: $9.4x10^{-2}$	$1.4x10^{-2}$ (p=0.42) Intercept: $8.7x10^{-2}$	$8.6x10^{-3}$ (p=0.58) Intercept: $9.4x10^{-2}$
Add to cart	$2.2x10^{-3}$ (p=0.77) Intercept: $1.2x10^{-2}$	$7.4x10^{-3}$ (p=0.31) Intercept: $1.2x10^{-2}$	$5.2x10^{-3}$ (p=0.49) Intercept: $1.5x10^{-2}$	$4.9x10^{-3}$ (p=0.45) Intercept: $1.2x10^{-2}$
Purchase	$3.8x10^{-3}$ (p=0.35) Intercept: $2.8x10^{-3}$	$-5.8x10^{-4}$ (p=0.88) Intercept: $2.8x10^{-3}$	$-4.4x10^{-3}$ (p=0.29) Intercept: $4.7x10^{-3}$	$1.6x10^{-3}$ (p=0.66) Intercept: $2.8x10^{-3}$

Note: results are reported as the beta coefficient for the interaction term in each model, the associated p-value and the intercept of the model, which occurs at the median star-rating (1.75) due to centring. A positive effect indicates the treatment group was superior to the control, or in the case of the alternative label vs. standard label that the alternative was superior, and vice versa for a negative effect. Other fixed-effects factors for each model, which are not reported, include experimental group, star rating, product price and the multiplicative interaction of product price with experimental group.

References

COMMONWEALTH OF AUSTRALIA, Department of the Prime Minister and Cabinet. 2018. Energy labels that make cents: testing energy rating labels on appliances sold online.

LIN, W. 2013. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique. *The Annals of Applied Statistics*, 7(1), pp. 295-318.

WASSERSTEIN, R. L. & LAZAR, N. A. 2016. The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), pp129-133.

978-1-925363-48-7 Energy labels that make cents: Additional analysis mixed-effects model (DOCX)

978-1-925363-47-0 Energy labels that make cents: Additional analysis mixed-effects model (PDF)

Other uses

Enquiries regarding this license and any other use of this document are welcome at:

Managing Director
Behavioural Economics Team of the Australian Government
Department of the Prime Minister and Cabinet
Barton ACT 2600
Email: beta@pmc.gov.au

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Department of the Prime Minister and Cabinet or the Australian Government.

Research team

Professor Michael J. Hiscox was the principal investigator for this project. Other (current and former) staff who contributed to the report were: Lilia Arcos-Holzinger; Brad Carron-Arthur; Michael Corliss; Harry Greenwell; Catherine Marks; Lani Perlesz; and Professor Robert Slonim

Acknowledgments

Thank you to Appliances Online and the Department of Environment and Energy for collaborating with us on the project. In particular, special thanks to Katharina Kuehn, Tim Griffin and Felipe Ramirez Lastrico from Appliances Online, and to Renee Robinson, Dayne Thompson, Richard Miles, Trish Lavery, Angela McGregor and Melissa Hogan from the Department of Environment and Energy.

Special thanks to Dr Elizabeth Hobman for providing advice on research design.

The trial was pre-registered on the BETA website and the American Economic Association registry:

https://www.pmc.gov.au/domestic-policy/behavioural-economics/energy-labels-make-cents-randomised-controlled-trial-test-effect-appliance-energy-rating-labels

https://www.socialscienceregistry.org/trials/2052

Australian Government

Behavioural Economics Team of the Australian Government

General enquiries beta@pmc.gov.au

Media enquiries media@pmc.gov.au

Find out more www.pmc.gov.au/beta