Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

What a Price Premium Buys: Product-Level Evidence on Price, Tested Quality, and Durability from 743 Consumer-Test Groups in Four European Markets

Pavel Kopczyk

kopczyk@opf.slu.cz

Silesian University in Opava

Research Article

Keywords: Price–quality relationship, Product durability, Consumer testing, Build quality, Brand premium, Preregistration

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11774817/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

What a Price Premium Buys: Product-Level Evidence on Price, Tested Quality, and Durability from 743 Consumer-Test Groups in Four European Markets

Abstract

Consumers can compare prices before purchase, but they often have limited evidence about how long a product will last. This study asks whether higher-priced models perform better on laboratory measures of durability when compared with other models in the same consumer test. The dataset contains 11,943 product observations across 743 test groups from four consumer-testing organisations in the Czech Republic, Germany, the United Kingdom, and Spain (2005–2026). Hypotheses were developed on about 40% of the test groups and evaluated on the remainder after preregistration. In the confirmation sample, higher prices were associated with higher overall scores (within-test rank correlation .45). After accounting for general quality, the association with measured durability was .26 (95% CI [.15, .37], 643 product observations). Its strength relative to the association with general quality remains uncertain. Brands with higher average price ranks had fewer owner-reported faults (−.37 across 122 brand–category groups). At one French retailer, higher prices were associated with higher displayed repairability and durability scores. Price also tracked assessor-rated build quality (.31), with uncertainty about how the association differed between and within brands. The findings provide a price benchmark for evaluating what durability disclosures add to the information available to consumers.

Keywords Price–quality relationship · Product durability · Consumer testing · Build quality · Brand premium · Preregistration

Highlights

- Within tests, higher prices are associated with higher overall scores (ρ = .45)
- Price tracks measured durability net of general quality (ρ = .26)
- Brands with higher average price ranks have fewer owner-reported faults (ρ = −.37)

Introduction

Consumers can compare prices at the point of sale. They have less direct evidence about how long competing models will last. When consumers cannot readily assess an attribute before purchase, firms may have less incentive to improve it (Akerlof, 1970; Darby & Karni, 1973; Nelson, 1970). Disclosure rules address this information problem. The EU Ecodesign for

Sustainable Products Regulation establishes a framework for product-specific information requirements through digital product passports (Regulation (EU) 2024/1781, 2024). France has required a repairability index for covered products since 2021 and introduced a durability index for televisions and washing machines in 2025 (Loi n° 2020-105, 2020).

Price may already convey some information about product quality. Research using expert test scores has found a positive average association, with variation across categories (Tellis & Wernerfelt, 1987). Studies in Austria, Germany, and the United States have compared prices with overall scores within individual tests (Boyle et al., 2018; Boyle & Kim, 2024; Kirchler et al., 2010; Rroshi & Weichselbaumer, 2021). These scores combine multiple attributes. Their association with price leaves open how closely price tracks laboratory durability after accounting for general quality.

Direct comparisons of price and physical durability also show that the pattern depends on the outcome. Higher-priced washing machines performed better in earlier endurance tests, while tests of smartphones found a clearer price gradient for scratch resistance than for resistance to rain (Clemm et al., 2020; Stiftung Warentest, 2013). An earlier analysis in this project estimated the price–durability association across brand–category averages with considerable uncertainty. An association between those averages can differ from the association between individual products (Robinson, 1950).

This paper compares products within tests from dTest, Stiftung Warentest, Which?, and OCU. Each comparison uses the prices reported by one organisation for products in the same test group. That keeps the testing organisation, market, and currency fixed within a comparison. Prices may be recorded after testing, so the main analysis uses prices reported with the test or observed within 12 months of it. The study distinguishes durability measured in endurance runs and stress tests from assessor-rated build quality, which records judgements about construction and workmanship.

The study first examines how price ranks relate to overall test scores and measured durability. It then asks whether the association with assessor-rated build quality is stronger between brands or between models of the same brand. Two additional analyses examine owner-reported faults and retailer-displayed indices. The hypotheses were developed on an exploration sample covering about 40% of the test groups. The analysis plan was registered (Kopczyk, 2026) before the remaining groups were opened as the confirmation sample (Anderson & Magruder, 2017; Fafchamps & Labonne, 2017; Nosek et al., 2018).

Within tests, price was positively associated with measured durability after accounting for general quality. The estimates leave open whether the durability association is as strong as

the association with general quality. The confirmation sample also leaves uncertainty about whether assessor-rated build quality is more strongly associated with price between brands than within them. Applying the earlier, broader definition of durability to the product-level data yielded a positive association in both samples.

Background and Hypotheses

Price as a Signal of Quality

Reputation gives firms a reason to sustain quality. If a firm earns a premium on future sales, it can lose that premium when consumers learn that its products perform poorly (Klein & Leffler, 1981; Shapiro, 1983). A brand can therefore help buyers form expectations about quality before they can assess it directly (Erdem & Swait, 1998; Erdem et al., 2006).

Manufacturers also offer models at different prices within the same brand (Randall et al., 1998). The overall price–quality association can reflect differences both between brands and among the models each brand sells. H1 assesses the overall association within tests.

H1 Within a test, higher-priced products tend to receive higher overall scores.

Measured Durability and Assessor-Rated Build Quality

Durability is difficult to assess before purchase and can take years to become apparent in use (Darby & Karni, 1973; Nelson, 1970). Research on product lifetimes documents this difficulty (Bakker et al., 2014; Cox et al., 2013; Hennies & Stamminger, 2016; Wieser, 2016). Models of durable goods show that firms' incentives to invest in product life depend on the market and on how buyers value future use (Bulow, 1986; Swan, 1970; Waldman, 1996, 2003).

Price can still be associated with durability through several mechanisms. More durable components or more demanding manufacturing processes may cost more. A reputation for reliable products may support higher prices. Manufacturers may also bundle durability with features that consumers can assess before buying. Each mechanism could produce a positive association between price and measured durability. H2 estimates that association after accounting for general quality. The mechanisms and consumers' willingness to pay remain open questions.

Laboratories measure durability by running products through repeated cycles or subjecting them to one-off stresses such as drops and immersion. Assessor-rated build quality is a judgement about construction and workmanship. Only one product in the dataset has both ratings, so their agreement cannot be assessed. H2 concerns measured durability. Assessor-rated build quality is analysed separately.

H2a Price is positively associated with measured durability after accounting for general quality.

H2b The price association with measured durability is less than .10 below the price association with general quality, after accounting for the other measure in each case.

The comparison of assessor-rated build quality with measured robustness (H2c) is exploratory because the ratings come mainly from different organisations and categories.

Brand or Model

Reputation and brand positioning may produce differences in price and build ratings between brands. Product design and production costs may produce differences between models within a brand as well. H3 focuses on assessor-rated build quality. In the exploration sample, its partial association with price was stronger than that for measured durability. Enough brands also had multiple models in the same test to estimate a within-brand slope.

H3a The association between price and assessor-rated build quality is stronger between brands than between models of the same brand.

H3b Within a brand and test, higher-priced models receive higher build-quality ratings.

Durability in Use and as Declared

Laboratory tests measure performance under specified conditions. Which?'s reliability survey provides a separate measure of faults reported by owners, aggregated by brand and category. Scores displayed by a French retailer provide another comparison at product level.

H4 Brands with higher average price ranks within tests have fewer owner-reported faults.

H5 Higher-priced products have higher displayed repairability (H5a) and durability (H5b) scores.

Two organisations print prices that may date from long after the test. Whether this matters is described but not tested, because only those two organisations have such prices (M1).

Data

Testing Organisations and Price Sources

Four organisations publish product test scores and prices (Table 1). dTest and Stiftung Warentest report the price with the test. Which? records a "typical price" when it checks the product. OCU reports a named retailer's offer, collected on 4 August 2026.

[Table 1 about here]

Price Recovery and Timing

The study recovered dTest prices from the original sub-rating text and archived product information, using the sub-rating text when both were available. It also estimated missing publication dates from the sequence of product-page identifiers. Supporting Material, Section S1, reports the recovery and validation checks.

The within-test analysis uses local-currency prices. The cross-market comparison converts dated prices to euros at the European Central Bank's monthly average exchange rate, then adjusts them to 2024 prices with the local consumer price index (Eurostat HICP, or ONS CPI for the United Kingdom). The study excluded 270 prices flagged as invalid.

Analysis Sample and Test Groups

The sample covers household appliances and consumer electronics. An observation enters the sample if it has the testing organisation's overall score, a valid price, and an eligible product category (Fig. 1). dTest organises products by test topic, so the category comes from the test name.

[Fig. 1 about here]

The database contained duplicate Which? results for five sets of reviews. This affected 646 results for mobile phones, sound bars, built-in ovens, printers, and smartwatches also filed as fitness trackers. The study retained one copy per result, giving priority to the exploration sample when the two copies fell in different samples and otherwise keeping the lower batch number.

The full sample contains 11,943 product observations across 743 test groups, covering 611 brands and 50 categories. Of those groups, 530 contain at least five observations and the median contains eleven. Test groups are the units of sample assignment and resampling. For Stiftung Warentest, a group is a recorded batch, which can be one round of a published comparison. For Which? and OCU, it is a batch within a test year. For dTest, it is a test series, which can span several years (63 of 222 series do). The main analysis compares dTest

products within their series and publication year. Undated products form a separate comparison within each series.

Overall Score and Study Measures

The *overall score* is the testing organisation's own verdict, expressed on a 0–100 scale. The organisations report purchase prices separately from their scores. Which? explicitly excludes price, Stiftung Warentest reports price and quality independently, and OCU gives a separate value-for-money award (Compra Maestra). dTest prices appear as unrated text. OCU includes a survey-based brand reliability score for 391 of its 2,365 products. Which? says it incorporates running costs in its television scores, and the corpus captures television energy use.

In the exploration sample, a check reconstructed each overall score from the organisation's sub-ratings within each test. Median R^2 was .99 for dTest and OCU, .95 for Which?, and .91 for Stiftung Warentest. Adding price increased median R^2 by .01 or less. For Which?, the association between price and the portion of the score left unexplained by its five-star sub-ratings was positive (median partial correlation .23). Rounding of the sub-ratings is one possible explanation.

General quality (*C*) is the study's average across the available dimensions of performance, energy, safety, noise, ease of use, features, and comfort. It excludes durability, repairability, and brand verdicts. It includes printing costs when rated by Which? or dTest.

Measured durability (*D*) is the mean of normalised ratings from laboratory endurance runs and one-off stress tests. Examples include motor endurance, scaling over repeated use, drop tests, and immersion. Measured durability is available for 1,538 product observations.

Assessor-rated build quality is the mean of sub-ratings of construction and workmanship, such as "zpracování," "Verarbeitung," and "Build quality." It is available for 1,889 product observations. One observation has both measured durability and assessor-rated build quality.

The three study measures exclude purchase-price labels. The classifier treats temperature stability in refrigerators, freezers, and ovens and motion handling in televisions as performance. It treats leak protection in Stiftung Warentest washing-machine tests as safety. Figure 2 shows the mix of durability ratings by organisation. For three quarters of the dTest products with a durability rating, that rating is an assessment of build quality.

[Fig. 2 about here]

Owner-Reported Faults and Retailer Scores

For H4, the Which? reliability survey reports the share of owners who experienced a fault for each brand and product category.

For H5, the study captured 1,644 product pages from the French retailer Fnac (fnac.com) between 17 September and 23 October 2026, across 13 product categories. Each page shows a price and, where the product carries them, the French repairability and durability scores on a 0–10 scale, listed among the product characteristics. A repairability score appeared on 502 pages, mostly laptops (371), and a durability score on 460, mostly televisions (401). The collection followed Fnac's listings to the last page for laptops, tablets, televisions, and top-loading washing machines, finding 94–99% of the products Fnac reported, and took the first page of 20 listings for eight other categories. H5 uses the pages with a positive price and a score, in categories with at least five such pages: in the confirmation sample, 300 listings in five categories for repairability and 236 in three categories for durability. Supporting Material, Section S4, gives the details.

Method

Comparing Products Inside a Test

The primary analyses for H1–H3 use contemporaneous prices. A price qualifies if the organisation reported it with the test or recorded it within 12 months before or after the test. Each analysis includes only observations with all required variables and comparison groups with at least five such observations.

Each product's price rank shows where its price falls among the products in the same comparison. The study ranks scores in the same way. Ranks range from −0.5 to +0.5, with the lowest value at −0.5 and the highest at +0.5 when there are no ties at the ends. Tied observations receive the average of their ranks. The within-test correlation pools the ranks from all eligible comparisons (Fig. 3). Ranking within each comparison makes scores and prices comparable across organisations and currencies. H2 uses partial correlations to measure the association between price and one score after accounting for the other score.

[Fig. 3 about here]

Between-Brand and Within-Brand Comparisons

H3 splits each product's price rank into its brand's average price rank in the same comparison and its distance from that average (Bell & Jones, 2015; Mundlak, 1978). The slope on the brand average (β_B) measures the association between brands. The slope on the distance from

the brand average (β_W) measures the association between models of the same brand. Only brands with at least two models in a comparison identify the within-brand slope. Both price components are centred within the comparison. Both slopes use the same rank scale.

Comparing Owner-Reported Faults and Retailer Scores

For H4, the analysis averages each brand's within-test price ranks within each product category, using the four organisations' prices regardless of date. It retains brand–category groups with at least two priced products and a Which? survey match. Within each category with at least three matches, the study ranks the brands by their average price rank and by the share of owners who reported a fault, then pools those ranks across categories.

For H5, the unit is a Fnac listing. The analysis ranks the price and displayed score within each category with at least five eligible listings, then pools the ranks across categories.

Inference

Confidence intervals and one-sided *p*-values come from 20,000 bootstrap draws. The primary analyses resample entire test groups, including entire dTest series. H4 resamples brand–category groups within their categories, and H5 resamples products within their categories. For both, the study also reports intervals from resampling whole categories. Because the same brands appear in several categories, an exploratory check for H4 also resamples whole brands. Each estimate uses its own random stream. The hypothesis test and the reported interval for that statistic use the same draws.

H1 tests whether the correlation exceeds .10, the smallest effect of interest (Lakens, 2017). H2b tests whether the difference between the price–durability and price–quality partial correlations exceeds −.10. The study applies the Holm adjustment separately to the primary hypotheses (H1–H3) and secondary hypotheses (H4–H5). It classifies adjusted *p*-values from .045 to .055 as inconclusive because the decision that close to .05 can depend on the bootstrap draws.

Split Sample and Preregistration

The data were frozen and split on 31 October 2026. A fixed rule hashes each test group's identifier with a published salt, targeting a 40% exploration sample and a 60% confirmation sample. The same rule assigns brand–category groups and Fnac products by their identifiers. The exploration sample contains 291 test groups and 4,697 product observations. The confirmation sample contains 452 test groups and 7,246 product observations. The rule also assigns tests published after registration.

The hypotheses were developed using the exploration sample. The analysis plan, including its hypotheses, estimators, decision rules, and code, was registered on the Open Science Framework on 11 November 2026 (Kopczyk, 2026). The code opened the confirmation sample only after the hash of the registered plan had been recorded. The hypothesis decisions come from a single analysis of the frozen confirmation data using the registered code.

After the confirmation sample was opened, the descriptive comparison with the earlier study was adapted to run on that sample. Three descriptive or exploratory analyses were added: a breakdown of H5 by category, an H4 interval that resamples whole brands, and the shares of within-test pairs in which the higher-priced product scored better. None of them changes a preregistered decision. Supporting Material, Section S3, lists the deviations from the registered plan.

The plan discloses two superseded statistics computed on all the data before the split. The study chose measured durability as the primary outcome for H2 after exploring a broader definition, which remains in the sensitivity analysis. H3 changed from measured durability to assessor-rated build quality after both measures had been explored. The decomposition for measured durability is reported descriptively. The inconclusive band was added on the day of registration because H4's exploration-sample *p*-value (.053) was sensitive to the bootstrap draws. All confirmation-sample decisions fall outside that band.

Robustness

A specification curve (Simonsohn et al., 2020) shows how the estimates for H1 and H2 change under alternative price timing rules, test groupings, sample thresholds, transformations, durability definitions, and organisation coverage. Two further checks remove the 185 products that appear in both samples (173 under another organisation) and restrict Which? and OCU prices to within six months of the test. Supporting Material, Section S2, gives the full grid and counts for both samples.

Use of Generative AI

Claude (Anthropic) assisted with writing and testing the data-preparation and analysis code and checking the reproducibility of the reported estimates. Claude and Codex (OpenAI) assisted with editing the text for readability. The author is responsible for the study design, interpretation, and final text.

Results

Table 2 shows estimates from both samples and the preregistered decisions based on the confirmation sample. Five of the eight hypotheses were supported (H1, H2a, H4, H5a, and H5b). The confirmation sample provided insufficient evidence to support H2b, H3a, and H3b.

[Table 2 about here]

Price and Overall Test Scores (H1)

Higher prices were associated with higher overall scores (Fig. 4, Table 3). In the confirmation sample, the within-test rank correlation was .45, 95% CI [.41, .49], for 5,053 products in 232 tests. The exploration sample also gave .45. H1 was supported (Holm-adjusted $p < .001$ for a correlation above .10). The correlation is positive for every organisation, from .41 for dTest to .51 for OCU. It barely moves when prices of any date are used (.43), when products present in both samples are dropped (.45), or when the prices of Which? and OCU must lie within six months of the test (.44). Across the 72 distinct specifications, it ranges from .42 to .52. Put another way, in 69% of within-test pairs with different prices (95% CI [67%, 73%]), the higher-priced product had the higher overall score, counting ties as half. Large tests contribute most pairs (one OCU test of 323 cooker hoods supplies 42%), but the share is 66–73% within each organisation.

[Fig. 4 about here]

[Table 3 about here]

Price and Measured Durability (H2)

In the confirmation sample, higher-priced products ranked higher on measured durability after accounting for general quality (partial rank correlation .26, 95% CI [.15, .37], across 643 products in 29 tests). Of those products, 333 were smartphones or headphones. The corresponding estimate in the exploration sample was .20. The confirmation result supported H2a (Holm-adjusted $p < .001$, Fig. 5 and Table 4).

In the same confirmation sample, the price association with general quality after accounting for durability was .33 (95% CI [.17, .46]). The estimated difference was −.07 (95% CI [−.26, .15]), compared with −.20 in the exploration sample. The preregistered test asked whether the difference was above −.10. The result provided insufficient evidence to support H2b (Holm-adjusted $p = .37$). The relative strength of the two associations remains uncertain. Among the 643 products with measured durability, and without accounting for general quality, the higher-priced product of a pair had better measured durability in 45% of pairs, worse in

25%, and the same in 30%: 60% counting ties as half (95% CI [55%, 65%]). In the same pairs, it had better general quality in 64% (65% counting ties as half, 95% CI [60%, 71%]).

The estimate was .30 when dTest was excluded. Across the 71 distinct confirmation specifications for measured durability, estimates ranged from .17 to .31 and every interval excluded zero (Fig. 6). The estimate was .31 for endurance runs (307 products in 17 tests) and .19 for one-off stress tests (95% CI [.04, .34], 400 products). The exploration interval for stress tests included zero.

Assessor-rated build quality had a partial correlation with price of .31 after accounting for general quality (95% CI [.20, .41], 801 products in 41 tests). The exploration estimate was .42. dTest supplied 667 of the 801 build ratings, so differences between measures can also reflect differences in organisations and categories. The exploratory contrast between build ratings and stress tests (H2c) was .11 (one-sided $p = .10$). The estimate using all durability labels together, as in the earlier analysis, was .29.

[Fig. 5 about here]

[Table 4 about here]

[Fig. 6 about here]

Build Quality Between and Within Brands (H3)

The between-brand slope for assessor-rated build quality exceeded the within-brand slope in the exploration sample. In the confirmation sample, the between-brand slope was 0.36, 95% CI [0.25, 0.46], and the within-brand slope was 0.19, 95% CI [−0.01, 0.39] (Fig. 7, Table 5). Their difference was 0.17, 95% CI [−0.04, 0.37]. The confirmation results provided insufficient evidence for a stronger association between brands (H3a, Holm-adjusted $p = .12$) or a positive association within brands (H3b, Holm-adjusted $p = .096$, unadjusted $p = .032$).

The other outcomes are descriptive. For measured durability, the confirmation slopes were 0.45 between brands and 0.04 within brands. The difference was 0.42 (95% CI [0.19, 0.62]), compared with 0.17 (95% CI [−0.06, 0.48]) in the exploration sample. The registered H3 tests apply to assessor-rated build quality. For general quality, the slopes were similar in both samples (0.37 between brands and 0.38 within brands in the confirmation sample).

[Fig. 7 about here]

[Table 5 about here]

Owner-Reported Faults (H4)

Brands with higher average within-test price ranks had lower shares of owners reporting a fault to Which?. In the confirmation sample, the within-category correlation was −.37, 95% CI [−.53, −.20], across 122 brand–category groups in 16 categories. The interval stayed below zero when whole categories were resampled ([−.54, −.20]). The result supported H4 (Holm-adjusted $p < .001$). Because 27 of the 48 brands appear in more than one category, and a brand's price rank and fault share tend to agree across its categories, an exploratory check resampled whole brands: the interval widened to [−.54, −.13] but stayed below zero. The exploration estimate was −.21 and inconclusive. Using price ranks from Which? tests only gave −.46 across 94 groups.

Scores Displayed on Fnac (H5)

Higher-priced listings had higher displayed repairability and durability scores, supporting H5a and H5b. The within-category correlation for repairability was .23, 95% CI [.12, .34], across 300 listings, including 220 laptops. Resampling whole categories gave an interval of [.03, .26]. For the displayed durability score, the correlation was .39, 95% CI [.27, .50], across 236 listings, including 212 televisions. The estimates were positive for televisions (.38), tablets (.59), and top-loading washing machines (.32).

The French mandatory durability index covers televisions and washing machines (Ministère de la Transition écologique, 2025). The confirmation sample nevertheless includes 12 tablet listings with a field labelled "Indice de durabilité". Their archived records contain no scoring sheets or component scores, so the source and calculation of those values are unverified. The result therefore describes the scores as displayed.

Contemporaneous Prices or Not (M1, Descriptive)

Pooled, the price–score association is .45 for prices observed within 12 months of the test and .38 for later prices (1,868 products in 117 tests). Every later price comes from Which? or OCU, so this comparison is between organisations as much as between price dates. Inside Which? and OCU, contemporaneous prices give .50 (1,636 products) and later prices .38. The difference, .13, 95% CI [.01, .23], suggests that a price observed long after the test tracks the score somewhat less closely. In the exploration sample, the same difference was .08 and its interval included zero.

The Same Product in Two Markets

In total, 240 products are priced in two markets within 12 months of each other (276 pairs). The median price gap is 24%. British typical prices are 14% above Spanish retailer offers for the same product (205 pairs). Czech test prices are above both (Spain to Czech Republic 0.745, 22 pairs; United Kingdom to Czech Republic 0.818, 28 pairs). Each organisation's price has a different basis, so these gaps mix market and method and are reported only as a description. This comparison involves no score and uses both samples.

Discussion

Price and Overall Test Scores

In the confirmation sample, the within-test rank correlation between price and overall score was .45 (95% CI [.41, .49], 5,053 product observations across 232 test groups). This agrees in direction with Kirchler et al. (2010), who reported an average rank correlation of .30 across 208 Austrian consumer tests, and Boyle et al. (2018), who reported a mean correlation of .36 across 50 US durable-goods categories with both national and store brands. The present estimate is numerically higher than those averages and falls between Kirchler et al.'s estimates for video, camera and music products (.42) and for computers and telephones (.58). Both earlier studies summarised separate test correlations, while the present study correlates ranks pooled across products. Sample composition matters as well. Boyle and Kim (2024) found that the price–quality association in durable goods was stronger in categories that included store brands than in categories with only national brands.

Rroshi and Weichselbaumer (2021) found a positive association in German consumer tests. Among 2,503 durable products, the estimated price for a grade-3 product was lower than for a grade-1 product, the best grade, by an amount equivalent to 22% of the mean price in its test. Their outcome was relative price and their comparison was between overall grades. The results together indicate that higher-priced models often receive better overall test scores. Within-test comparisons still leave open which aspects of quality account for this association.

Price and Measured Durability

After accounting for general quality, the price–durability partial rank correlation was .26 (95% CI [.15, .37], 643 product observations across 29 test groups), supporting H2a. In the same sample, the corresponding association with general quality after accounting for durability was .33 (95% CI [.17, .46]). The paired difference was −.07 (95% CI [−.26, .15]). The interval spans both a durability association more than .10 weaker than the general-quality association

and a stronger durability association. The test provided insufficient evidence to support H2b (Holm-adjusted $p = .37$).

Two earlier laboratory comparisons help interpret the positive association. In Stiftung Warentest's 2003–2012 washing-machine endurance tests, by the equivalent of ten years of use, machines in the under-€550 bracket had about three times the cumulative failure rate of those above €700 (Stiftung Warentest, 2013, as reproduced in Alfieri et al., 2018, Fig. 3). In a study of 108 smartphones, the highest scratch-resistance grade went to 5 of 34 phones below €240 (15%) and 38 of 40 above €550 (95%). All 108 phones received good or very good ratings in the rain test (Clemm et al., 2020). The price contrast can thus vary even across tests on the same products.

These studies compared performance across price brackets. The present study relates price ranks to measured durability within each test while accounting for general quality. The partial correlation was .31 for endurance runs (95% CI [.20, .42], 307 observations) and .19 for one-off stress tests (95% CI [.04, .34], 400 observations). These descriptive estimates draw on partly different products and protocols. Outside appliances and electronics, Guo et al. (2025) found that jeans priced at £140 and £22.50 ranked first and second for durability among nine pairs of women's jeans. Even where the average association is positive, individual products can perform quite differently from what their prices suggest.

The Earlier Brand-Level Conclusion

The earlier brand-level analysis estimated a partial correlation of .12 between price and durability after accounting for general quality (95% CI [−.05, .24], 216 brand–category groups). Its durability measure combined build ratings with measured durability. Figure 8 and Table 6 show what happens when the current data are analysed using that broader measure and selected choices from the earlier analysis. The confirmation comparisons are preregistered as descriptive.

Averaging the current products by brand and category gave .47 in the exploration sample and .36 in the confirmation sample. The estimates ranged from .33 to .43 when the analysis used alternative durability classifications, included brand-reliability verdicts, or ranked prices across markets and years. For the groups that also had the earlier catalogue-based brand prices, replacing the tested products' own prices changed the estimate from .27 to .14 in exploration (49 groups) and from .51 to .45 in confirmation (73 groups). The original pipeline was only partially reconstructed, so the source of the earlier imprecise estimate remains unresolved.

[Fig. 8 about here]

[Table 6 about here]

Build Ratings and Measured Durability

The association with price was stronger for assessor-rated build quality than for measured durability in the exploration sample (.42 and .20). The estimates were closer in confirmation (.31 and .26). The comparison is descriptive because the two ratings come mainly from different organisations and categories. An assessor may pick up visible differences in materials or construction. The analysis has no record of which cues, such as brand or price, assessors could see. Endurance and stress-test scores depend more directly on measured product performance.

Implications for Consumer Policy

Within tests, price provided some information about measured durability after accounting for general quality. The association of .26 leaves substantial variation in durability ranks unexplained. For a single purchase, the higher-priced of two products in a test had better measured durability in 45% of pairs and worse in 25%. The owner survey also showed fewer reported faults for brands with higher average price ranks. That comparison uses brand–category groups and a separate outcome.

At Fnac, prices were positively associated with the displayed scores. Televisions accounted for most durability-score observations and laptops for most repairability-score observations. The French laptop repairability formula rewards a lower ratio of spare-parts prices to the price of the new product (Arrêté du 29 décembre 2020, 2020). This criterion could contribute to the observed price association, even though the equipment price used in the score can differ from the Fnac offer. The retailer comparison measures how the displayed scores vary with price.

The results give disclosure evaluations a benchmark. One question is whether a label improves prediction of independently measured durability beyond what price already predicts. A second question is whether consumers use the label and make different choices when it is available.

Limitations

These are associations inside tests, not effects of price. The data show that prices move with durability, not why: buyers' valuation, production cost, and brand reputation are not separated. The testing organisations select the products they test, usually popular models in each market,

so the results describe tested products rather than the full range on sale. The confirmation analysis for measured durability used 643 products, including 333 smartphones and headphones, so these results rest heavily on two categories and may not carry over to others, such as large appliances. Only one product has both a build rating and a measured-durability score. The prices of Which? and OCU are mostly not contemporaneous, and those of OCU exist only for products still on sale. The overall scores are the organisations' own; none includes the purchase price, but that of OCU includes a survey-based brand-reliability score. H4 rests on one survey in one market and on brand matching, and its fault shares are not adjusted for the age of the products surveyed, which the stored survey data do not record. H5 rests on one retailer, and mostly on laptops and televisions. M1 cannot separate price timing from organisation, except inside Which? and OCU. Entity resolution across organisations has not been checked by hand, which matters for the comparison across markets and for the 185 products that span the two samples. Both samples come from one frozen corpus and share brands in H4. The plan therefore also registers an out-of-time replication on tests published after registration.

Conclusion

Within consumer tests, higher prices were associated with higher overall scores (.45 in each sample). The association with measured durability remained positive after accounting for general quality (.26 in confirmation). A separate comparison of brand–category groups linked higher price ranks to fewer owner-reported faults (−.37, without adjustment for general quality). The relative strength of the measured-durability and general-quality associations remains uncertain, as does the difference between the two build-quality slopes. For consumer policy, the next question is how much a durability label adds to what price predicts and whether that information helps consumers choose.

Data availability

The analysis plan, the analysis code, and the results of both samples are on the Open Science Framework: the registration (https://doi.org/10.17605/OSF.IO/3KNQX) and the project it was registered from (https://osf.io/ez2p6/). The test results themselves are published by the testing organisations under their terms of use and are not redistributed.

References

Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*, 488–500. https://doi.org/10.2307/1879431

Alfieri, F., Cordella, M., Stamminger, R., & Bues, A. (2018). *Durability assessment of products: Analysis and testing of washing machines* (EUR 29487 EN). Publications Office of the European Union. https://doi.org/10.2760/115684

Anderson, M. L., & Magruder, J. (2017). *Split-sample strategies for avoiding false discoveries* (NBER Working Paper No. 23544). National Bureau of Economic Research. https://doi.org/10.3386/w23544

Arrêté du 29 décembre 2020 relatif aux critères, aux sous-critères et au système de notation pour le calcul et l’affichage de l’indice de réparabilité des ordinateurs portables. (2020). *Journal officiel de la République française*. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000042838127/

Bakker, C., Wang, F., Huisman, J., & den Hollander, M. (2014). Products that go round: Exploring product life extension through design. *Journal of Cleaner Production, 69*, 10–16. https://doi.org/10.1016/j.jclepro.2014.01.028

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods, 3*, 133–153. https://doi.org/10.1017/psrm.2014.7

Boyle, P. J., & Kim, J. (2024). The impact of store brands on intangible aspects of consumer welfare. *Journal of International Consumer Marketing, 36*(2), 151–165. https://doi.org/10.1080/08961530.2023.2215463

Boyle, P. J., Kim, H., & Lathrop, E. S. (2018). The relationship between price and quality in durable product categories with private label brands. *Journal of Product & Brand Management, 27*(6), 647–660. https://doi.org/10.1108/JPBM-09-2017-1590

Bulow, J. (1986). An economic theory of planned obsolescence. *Quarterly Journal of Economics, 101*, 729–749. https://doi.org/10.2307/1884176

Clemm, C., Berwald, A., Prewitz, C., Nissen, N. F., & Schneider-Ramelow, M. (2020). Market trends in smartphone design and reliability testing. In *Electronics Goes Green 2020+* (pp. 171–178). Fraunhofer IZM. https://prompt-project.eu/wp-content/uploads/2020/10/Clemm-et-al.pdf

Cox, J., Griffith, S., Giorgi, S., & King, G. (2013). Consumer understanding of product lifetimes. *Resources, Conservation and Recycling, 79*, 21–29. https://doi.org/10.1016/j.resconrec.2013.05.003

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of Law and Economics, 16*, 67–88. https://doi.org/10.1086/466756

Erdem, T., & Swait, J. (1998). Brand equity as a signaling phenomenon. *Journal of Consumer Psychology, 7*, 131–157. https://doi.org/10.1207/s15327663jcp0702_02

Erdem, T., Swait, J., & Valenzuela, A. (2006). Brands as signals: A cross-country validation study. *Journal of Marketing, 70*(1), 34–49. https://doi.org/10.1509/jmkg.70.1.034.qxd

Fafchamps, M., & Labonne, J. (2017). Using split samples to improve inference on causal effects. *Political Analysis, 25*, 465–482. https://doi.org/10.1017/pan.2017.22

Guo, Y., Morris, K. E., Sumner, M., & Taylor, M. (2025). A framework for measuring physical garment durability. *Cleaner and Responsible Consumption, 16*, 100245. https://doi.org/10.1016/j.clrc.2024.100245

Hennies, L., & Stamminger, R. (2016). An empirical survey on the obsolescence of appliances in German households. *Resources, Conservation and Recycling, 112*, 73–82. https://doi.org/10.1016/j.resconrec.2016.04.013

Kirchler, E., Fischer, F., & Hölzl, E. (2010). Price and its relation to objective and subjective product quality: Evidence from the Austrian market. *Journal of Consumer Policy, 33*, 275–286. https://doi.org/10.1007/s10603-010-9138-1

Klein, B., & Leffler, K. B. (1981). The role of market forces in assuring contractual performance. *Journal of Political Economy, 89*, 615–641. https://doi.org/10.1086/260996

Kopczyk, P. (2026, November 11). *What a price premium buys* [Preregistration]. OSF. https://doi.org/10.17605/OSF.IO/3KNQX

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*, 355–362. https://doi.org/10.1177/1948550617697177

Loi n° 2020-105 du 10 février 2020 relative à la lutte contre le gaspillage et à l'économie circulaire. (2020). *Journal officiel de la République française*.

Ministère de la Transition écologique. (2025, June 30). *Indice de durabilité*. https://www.ecologie.gouv.fr/politiques-publiques/indice-durabilite

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica, 46*, 69–85. https://doi.org/10.2307/1913646

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy, 78*, 311–329. https://doi.org/10.1086/259630

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*, 2600–2606. https://doi.org/10.1073/pnas.1708274114

Randall, T., Ulrich, K., & Reibstein, D. (1998). Brand equity and vertical product line extent. *Marketing Science, 17*, 356–379. https://doi.org/10.1287/mksc.17.4.356

Regulation (EU) 2024/1781 of the European Parliament and of the Council of 13 June 2024 establishing a framework for the setting of ecodesign requirements for sustainable products. (2024). *Official Journal of the European Union*, L, 2024/1781.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357. https://doi.org/10.2307/2087176

Rroshi, D., & Weichselbaumer, M. (2021). *What is in a price? Evidence on quality signaling for experience goods* (Department of Economics Working Paper No. 311). WU Vienna University of Economics and Business. https://doi.org/10.57938/0c79dc48-9599-4e10-9d52-abb8092e3689

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *Quarterly Journal of Economics, 98*, 659–679. https://doi.org/10.2307/1881782

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour, 4*, 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Stiftung Warentest. (2013, August 29). *Geplante Obsoleszenz: Gerade gekauft und schon wieder hin?* https://www.test.de/Geplante-Obsoleszenz-Gerade-gekauft-und-schon-wieder-hin-4596260-0/

Swan, P. L. (1970). Durability of consumption goods. *American Economic Review, 60*, 884–894.

Tellis, G. J., & Wernerfelt, B. (1987). Competitive price and quality under asymmetric information. *Marketing Science, 6*, 240–253. https://doi.org/10.1287/mksc.6.3.240

Waldman, M. (1996). Planned obsolescence and the R&D decision. *RAND Journal of Economics, 27*, 583–595. https://doi.org/10.2307/2555845

Waldman, M. (2003). Durable goods theory for real world markets. *Journal of Economic Perspectives, 17*(1), 131–154. https://doi.org/10.1257/089533003321164985

Wieser, H. (2016). Beyond planned obsolescence: Product lifespans and the challenges to a circular economy. *GAIA – Ecological Perspectives for Science and Society, 25*, 156–160. https://doi.org/10.14512/gaia.25.3.5

Tables and Figures — What a Price Premium Buys

The exploration sample was used to develop hypotheses. The confirmation sample was opened after preregistration and used to test them (https://doi.org/10.17605/OSF.IO/3KNQX).

Table 1 Organisations, prices, and the analysis frame

Organisation	Market	What the price is	Observations	Test groups	Categories	Years	Contemporaneous price
dTest	CZ	Price reported with the test ("průměrná / orientační cena")	4,743	222	37	2005–2026	4,743 (100.0%)
Stiftung Warentest	DE	"Mittlerer Onlinepreis" reported with the test	1,036	155	33	2017–2026	1,036 (100.0%)
Which?	GB	"Typical price" at the date Which? last checked the product	3,799	218	26	2012–2026	1,705 (44.9%)
OCU	ES	A named retailer's offer on 4 August 2026	2,365	148	25	2016–2026	901 (38.1%)
Total			11,943	743	50	2005–2026	8,385 (70.2%)

Note. A price is contemporaneous if the organisation printed it with the test, or if it was observed within 12 months of the test. dTest products are compared within the same series and publication year, with undated products grouped within their series.

Table 2 Preregistered hypotheses in the two samples

Hypothesis	Statistic	Exploration	Confirmation	Confirmation sample	Holm *p*	Decision
H1 Higher-priced products score better overall (correlation above .10)	ρ(price, overall score)	.446 [.394, .497]	.452 [.408, .494]	5,053 products (232 tests)	< .001	Supported
H2a Price tracks measured durability, net of general quality	partial ρ(price, D | C)	.203 [.071, .337]	.262 [.148, .370]	643 products (29 tests)	< .001	Supported
H2b Durability association is less than .10 below general quality	partial ρ(price, D | C) − partial ρ(price, C | D)	−.201 [−.432, .047]	−.066 [−.261, .149]	643 products (29 tests)	.374	Not supported
H3a Build-quality association is stronger between brands	$\beta_B - \beta_W$, assessor-rated build quality	0.243 [0.112, 0.497]	0.170 [−0.043, 0.374]	801 products (41 tests)	.120	Not supported
H3b Within brands, higher-priced models receive higher build ratings	β_W, assessor-rated build quality	0.221 [−0.044, 0.365]	0.190 [−0.011, 0.386]	801 products (41 tests)	.096	Not supported
H4 Brands with higher price ranks have fewer reported faults	ρ(brand price position, share of owners reporting a fault)	−.207 [−.443, .046]	−.371 [−.531, −.202]	122 brand–category groups (16 categories)	< .001	Supported
H5a Higher prices are associated with higher displayed repairability scores	ρ(price, repairability index)	.224 [.079, .365]	.231 [.121, .343]	300 products (5 categories)	< .001	Supported
H5b Higher prices are associated with higher displayed durability scores	ρ(price, durability index)	.337 [.207, .459]	.390 [.274, .498]	236 products (3 categories)	< .001	Supported

Note. Estimates with 95% bootstrap intervals (20,000 draws). The decision is the preregistered one, made on the confirmation sample: a one-sided bootstrap *p*-value, Holm-adjusted within the primary family (H1–H3) and the secondary family (H4, H5), below .05, and reported as inconclusive from .045 to .055. H1 tests against .10 and H2b against −.10. The other thresholds are zero. H1–H3 rank inside a test and resample tests; H4 and H5 rank inside a category and resample cells or products inside each category. β_B and β_W are slopes on the rank scale (Table 5). In the exploration sample, where the hypotheses were formed, the same rules gave: supported H1, H2a, H3a, H5a, H5b; not supported H2b, H3b; inconclusive H4.

Table 3 Price and the overall score inside a test (H1)

	Exploration	*n* (tests)	Confirmation	*n* (tests)
All four, contemporaneous prices (H1)	.446 [.394, .497]	3,038 (143)	.452 [.408, .494]	5,053 (232)
dTest	.402 [.341, .462]	1,860 (85)	.414 [.364, .461]	2,801 (129)
Stiftung Warentest	.467 [.338, .592]	296 (23)	.499 [.380, .600]	616 (45)
Which?	.479 [.331, .596]	621 (27)	.499 [.343, .612]	1,020 (38)
OCU	.667 [.486, .854]	261 (8)	.511 [.391, .617]	616 (20)
All prices, contemporaneous or not	.460 [.409, .508]	4,503 (205)	.430 [.389, .469]	6,979 (324)
Without products present in both samples	.451 [.397, .503]	2,920 (140)	.449 [.406, .489]	4,887 (227)
Which?/OCU prices within 6 months of the test	.431 [.376, .483]	2,628 (131)	.442 [.397, .483]	4,356 (210)

Note. Within-test rank correlations of price with the organisation's overall score, with 95% bootstrap intervals resampling tests (20,000 draws).

Table 4 Price, durability, and general quality inside a test (H2)

	n (tests)	Price and X, adjusted for C	Price and C, adjusted for X	*r*(X, C)
Exploration sample				
Measured durability (H2)	372 (19)	.203 [.071, .337]	.404 [.217, .577]	.326
Endurance runs	128 (9)	.310 [.192, .475]	.146 [−.015, .345]	.185
One-off stress tests	276 (11)	.120 [−.052, .283]	.502 [.267, .721]	.403
Measured durability, without dTest	225 (12)	.269 [.111, .405]	.406 [.123, .654]	.293
Assessor-rated build quality	631 (25)	.417 [.279, .530]	.193 [.072, .312]	.200
All durability labels	1,003 (44)	.348 [.236, .449]	.267 [.159, .372]	.246
Classifier's own durability	1,239 (55)	.274 [.167, .374]	.275 [.182, .367]	.283
Confirmation sample				
Measured durability (H2)	643 (29)	.262 [.148, .370]	.328 [.173, .460]	.261
Endurance runs	307 (17)	.306 [.200, .421]	.159 [−.063, .339]	.199
One-off stress tests	400 (15)	.194 [.040, .337]	.501 [.356, .623]	.306
Measured durability, without dTest	429 (20)	.298 [.164, .433]	.347 [.172, .480]	.229
Assessor-rated build quality	801 (41)	.308 [.202, .405]	.215 [.101, .322]	.155
All durability labels	1,444 (70)	.291 [.216, .364]	.264 [.169, .351]	.201
Classifier's own durability	1,738 (83)	.282 [.212, .347]	.258 [.178, .336]	.228

Note. Partial within-test rank correlations with 95% bootstrap intervals. X is the row's measure; C is general quality.

Table 5 Within-between decomposition of the price rank (H3)

Outcome	Within a brand (β_W)	Between brands (β_B)	Between minus within	Products (in multi-model brand groups)
Exploration sample				
Assessor-rated build quality (H3)	0.221 [−0.044, 0.365]	0.464 [0.343, 0.574]	0.243 [0.112, 0.497]	631 (348)

Outcome	Within a brand (β_W)	Between brands (β_B)	Between minus within	Products (in multi-model brand groups)
Assessor-rated build quality, controlling C	0.186 [−0.065, 0.330]	0.445 [0.312, 0.557]	0.259 [0.129, 0.494]	631 (348)
Measured durability	0.173 [−0.094, 0.430]	0.343 [0.173, 0.528]	0.170 [−0.061, 0.481]	372 (277)
Measured durability, controlling C	0.110 [−0.126, 0.327]	0.240 [0.111, 0.402]	0.130 [−0.093, 0.441]	372 (277)
General quality	0.296 [0.197, 0.383]	0.335 [0.257, 0.411]	0.039 [−0.063, 0.145]	2,911 (2,036)
Overall score	0.389 [0.301, 0.466]	0.475 [0.413, 0.537]	0.086 [−0.007, 0.184]	3,038 (2,127)
Confirmation sample				
Assessor-rated build quality (H3)	0.190 [−0.011, 0.386]	0.360 [0.255, 0.463]	0.170 [−0.043, 0.374]	801 (450)
Assessor-rated build quality, controlling C	0.167 [−0.032, 0.362]	0.343 [0.245, 0.436]	0.176 [−0.040, 0.378]	801 (450)
Measured durability	0.035 [−0.119, 0.209]	0.455 [0.322, 0.577]	0.419 [0.189, 0.616]	643 (492)
Measured durability, controlling C	−0.009 [−0.169, 0.174]	0.399 [0.279, 0.513]	0.407 [0.179, 0.602]	643 (492)
General quality	0.376 [0.292, 0.448]	0.369 [0.304, 0.435]	−0.007 [−0.077, 0.067]	4,927 (3,553)
Overall score	0.403 [0.338, 0.464]	0.475 [0.423, 0.524]	0.072 [0.004, 0.141]	5,053 (3,636)

Note. Slopes on the rank scale with 95% bootstrap intervals. β_B is the slope on the brand's mean price rank in the test; β_W the slope on a model's distance from it, identified by brands with two or more models in a test.

Table 6 Reconciliation with the predecessor's brand-level estimate (descriptive)

	Unit	Exploration	Units	Confirmation	Units
(a) This paper's products, compared inside tests	Product in test	.348 [.236, .449]	44 tests	.291 [.216, .364]	70 tests
(b) Same products, averaged	Brand × category	.474 [.356, .607]	156 cells	.356 [.238, .472]	216 cells
(c) As (b), classifier's own durability	Brand × category	.381 [.254, .513]	184 cells	.335 [.222, .449]	250 cells
(d) As (c), plus brand-reliability verdicts	Brand × category	.381 [.256, .513]	184 cells	.329 [.216, .439]	259 cells
(e) As (b), price ranked across markets and years	Brand × category	.428 [.298, .576]	146 cells	.345 [.212, .462]	210 cells
(f) Cells carrying the predecessor's brand price: that price	Brand × category	.141 [−.144, .444]	49 cells	.454 [.219, .658]	73 cells
(f′) Same cells, tested products' own prices	Brand × category	.268 [−.018, .529]	49 cells	.512 [.276, .703]	73 cells

Note. Partial rank correlation of price with durability (all durability labels, as in the predecessor), net of general quality. Row (a) is the "All durability labels" row of Table 4. Cell intervals resample cells within category; resampling whole categories, as the predecessor did, gives, for the exploration and the confirmation sample: (b) [.226, .646] and [.223, .476]; (f) [−.164, .326] and [.318, .690]; (f′) [−.004, .488] and [.377, .825]. The confirmation-sample rows were preregistered as descriptive (preregistration §7.6) and test no hypothesis.

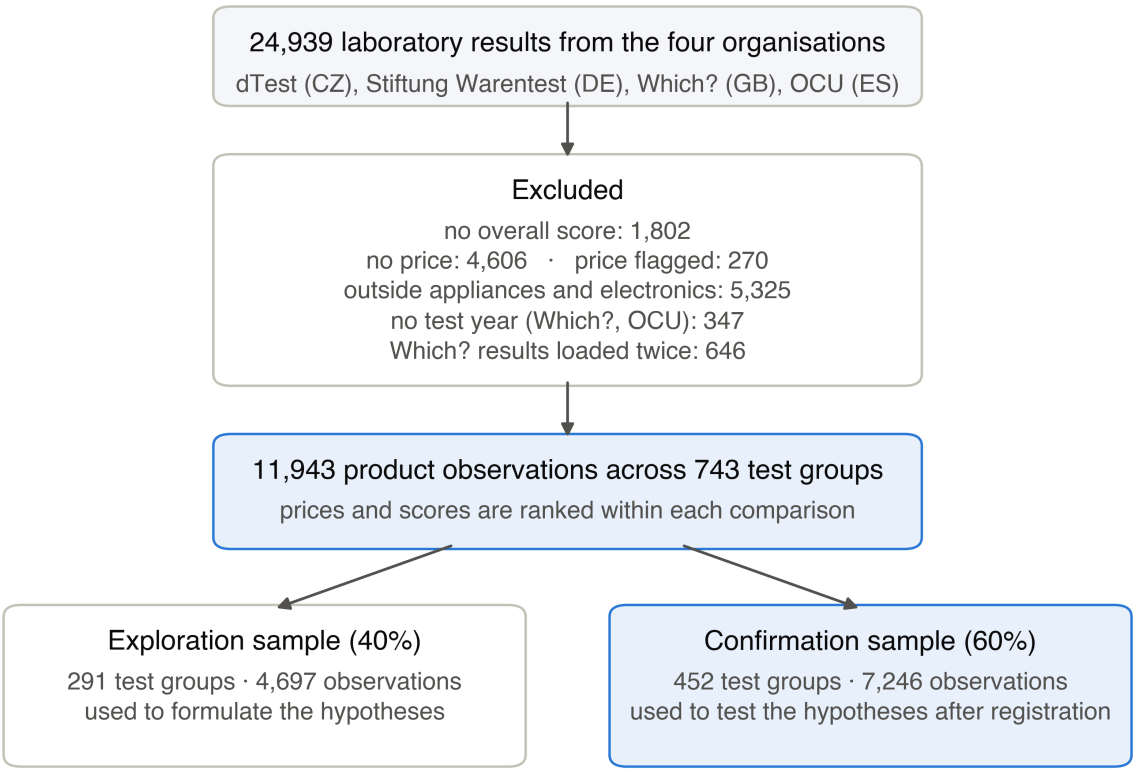


Fig. 1 *From the corpus to the analysis frame: counts after each exclusion step, and the split into an exploration sample, where the hypotheses were formed, and a confirmation sample, opened only after the analysis plan was registered, where they were tested*

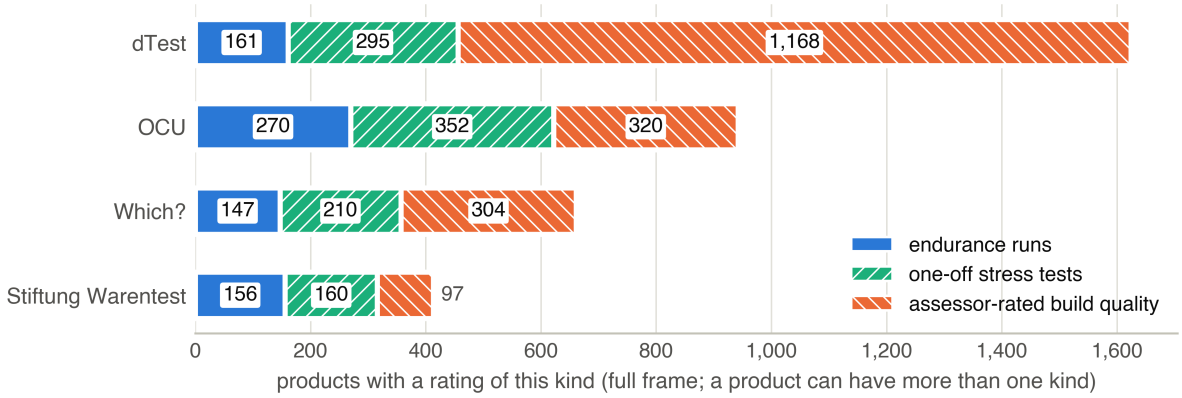


Fig. 2 *What the organisations call "durability": products in the full frame with each kind of rating. Counts only; no outcome is involved, so both samples are shown*

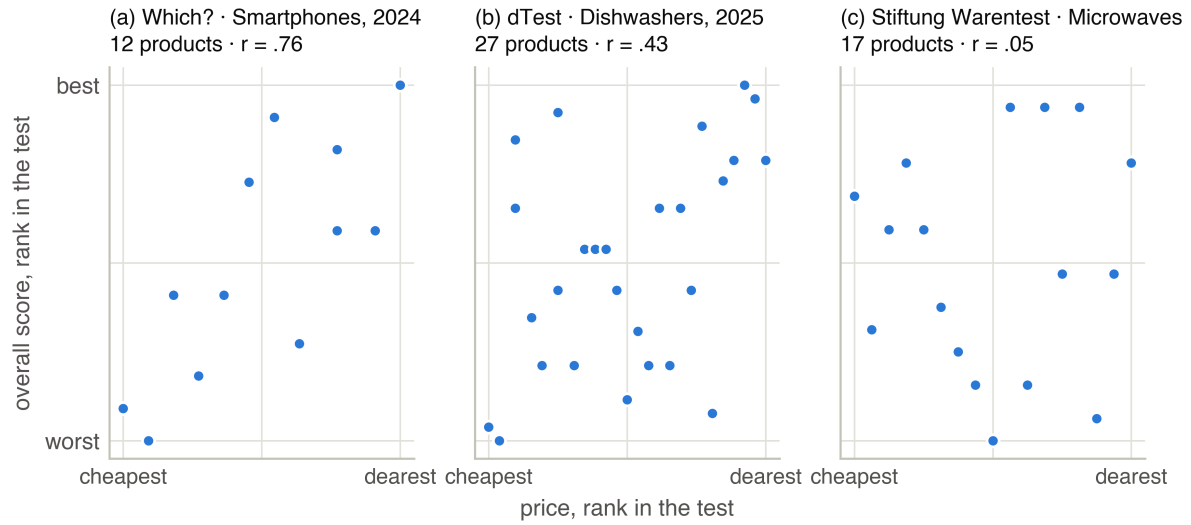


Fig. 3 *How the comparison works. Each dot is a product; both axes are ranks inside the test. The tests were chosen by a fixed rule from exploration-sample tests with 12–30 products: (a) from Which?, the one whose within-test correlation is closest to .75; (b) from dTest, the one closest to the pooled estimate; (c) from Stiftung Warentest, the one closest to .10*

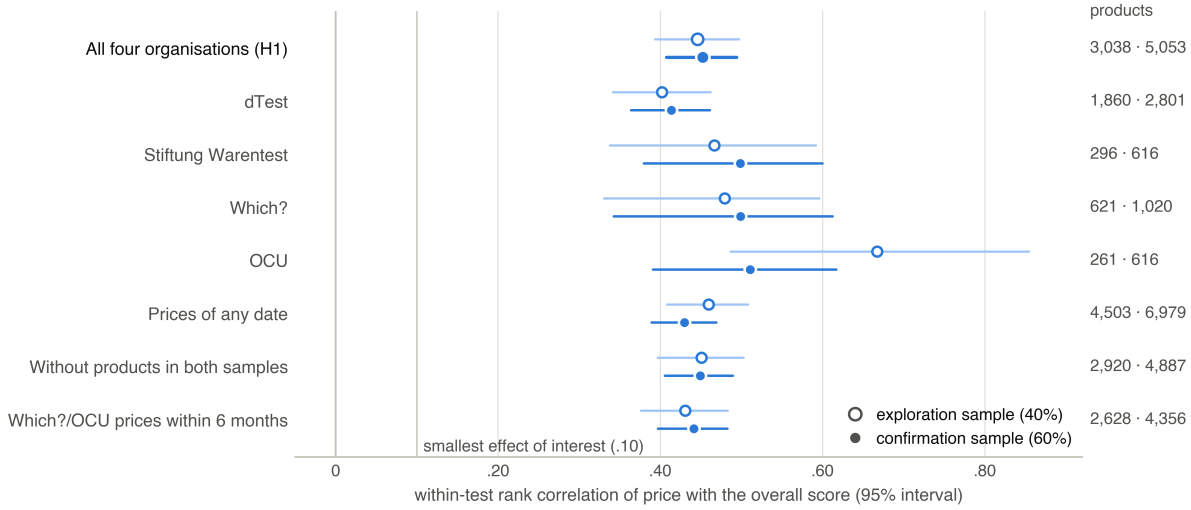


Fig. 4 *Price and the overall score inside a test (H1). Open markers: exploration sample; filled markers: confirmation sample. Dots are within-test rank correlations and lines 95% bootstrap intervals; the vertical line at .10 marks the smallest effect of interest. Right: products in the exploration and confirmation samples*

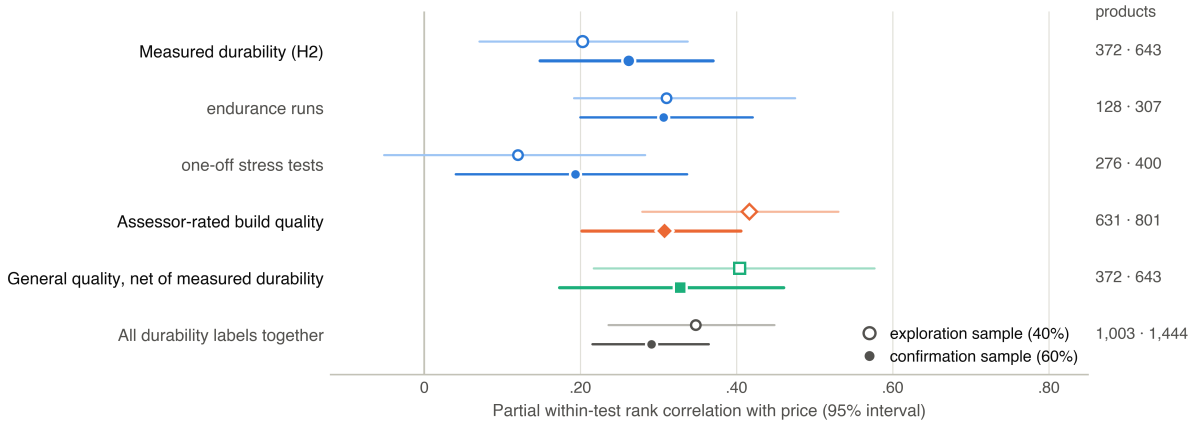


Fig. 5 *Partial within-test rank correlations with price and 95% intervals. Open markers show exploration estimates and filled markers show confirmation estimates. The general-quality row adjusts for measured durability. All other rows adjust for general quality. Counts at right show the number of products in each sample*

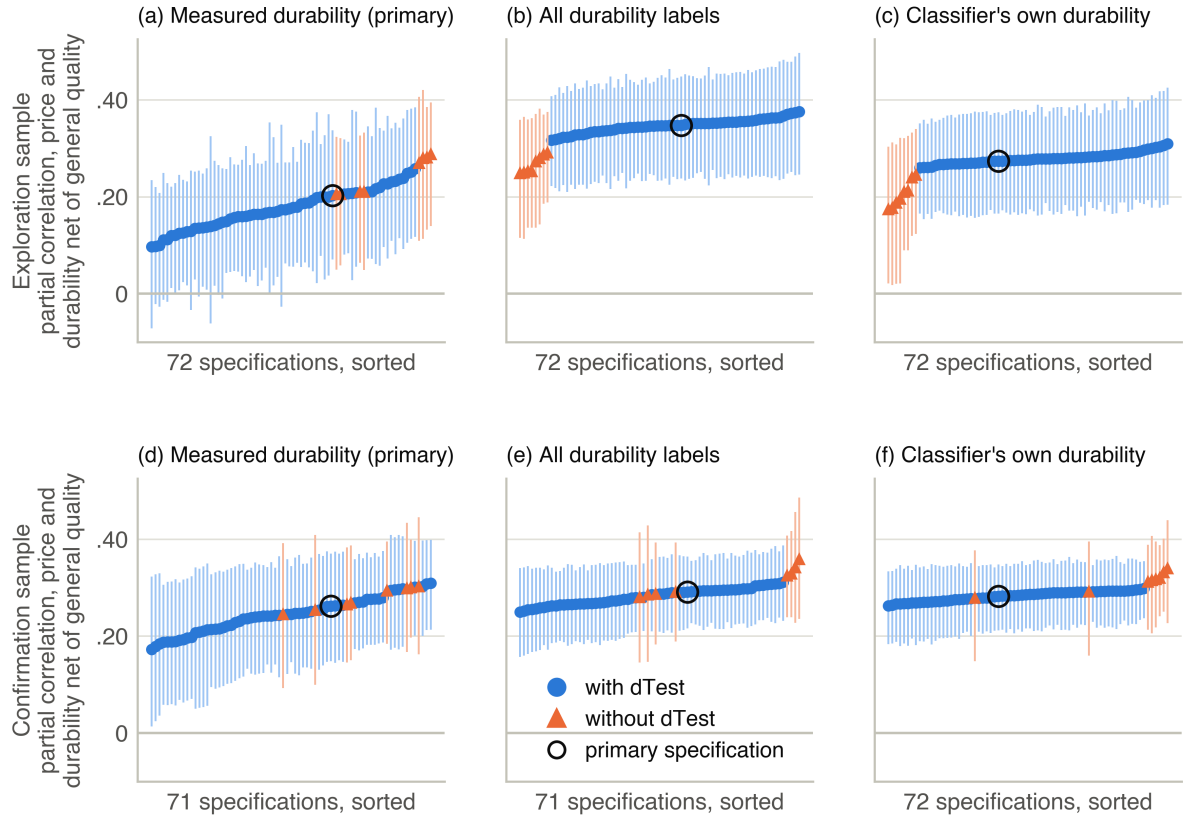


Fig. 6 *Specification curve for H2 in the exploration sample (a–c) and the confirmation sample (d–f). Each point is one distinct specification for a durability definition, sorted by estimate, with its 95% interval; triangles leave dTest out; the ring marks the primary specification in (a) and (d) and the same analytic choices in the other panels. (a, d) Measured durability; (b, e) all durability labels; (c, f) the classifier's own durability*

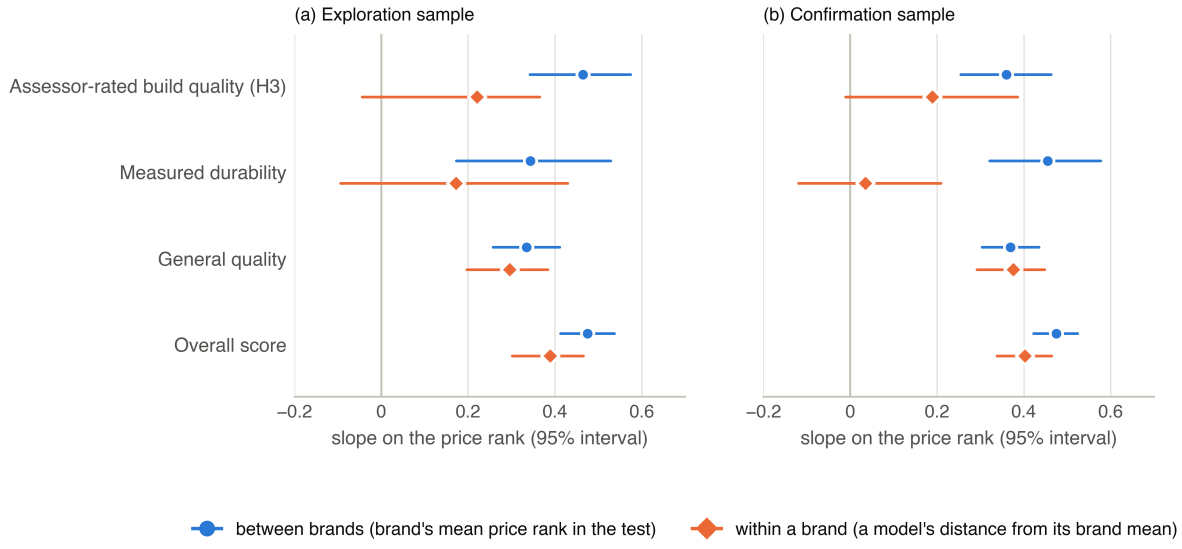


Fig. 7 *Between brands or within a brand (H3), in (a) the exploration sample and (b) the confirmation sample. Slopes of each outcome's rank on the price rank, split into the brand's mean price rank in the test (circles, between brands) and a model's distance from it (diamonds, within a brand), with 95% intervals*

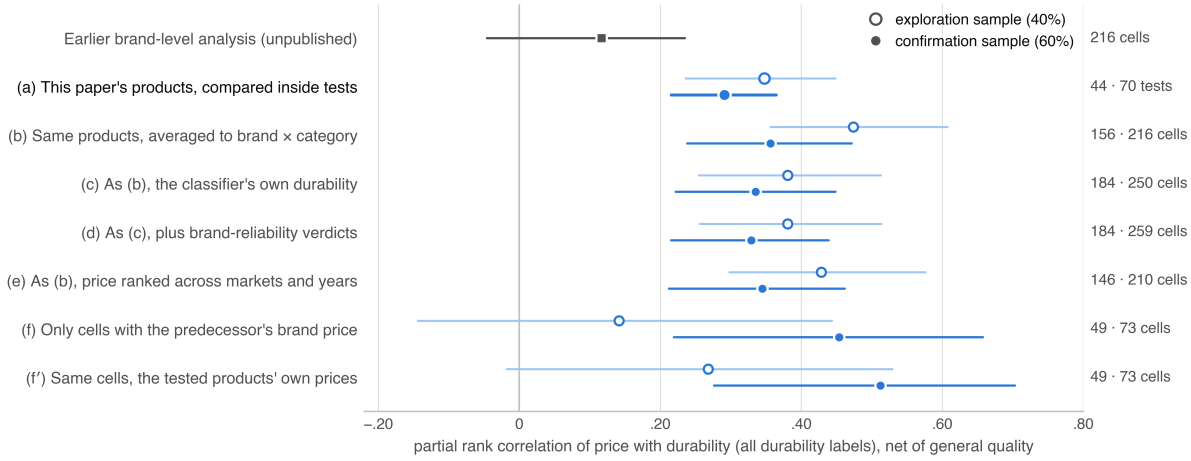


Fig. 8 *From the predecessor's brand-level estimate to this paper's, one change at a time, in the exploration sample (open markers) and the confirmation sample (filled markers; descriptive). Partial correlation of price with durability (all durability labels, as in the predecessor), net of general quality. Square: the earlier brand-level estimate (unpublished analysis; bootstrap clustered on category). Cell intervals resample cells within category. Right: units in the exploration and confirmation samples*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supportingmaterial.docx