Journal of Retailing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Retailing

journal homepage: www.elsevier.com/locate/retail

From recommendation to revenue: Store-level evidence on the dynamic and spatial effects of the Net Promoter Score (NPS)

Chris Dimos [a,*], Jana Hunt [b], Geoff Pugh [c], Andrew Stephenson [d]

[a] University of Bath, Centre for Governance, Regulation and Industrial Strategy, Claverton Down, Bath, BA2 7AY, UK
[b] EUROPEUM Institute for European Policy, Staroměstské náměstí 4/1,110 00, Praha 1, Czech Republic
[c] Staffordshire University Business School, Leek Road, Stoke-on-Trent, ST4 2DF, UK
[d] Dove Nest Group (Non-Executive Director), Millness Lane, Crooklands, Milnthorpe, LA7 7NS, UK

ARTICLE INFO

Edited by: Dr Katrijn Gielens

Keywords:
Net Promoter Score
Sales
Word of mouth
Dynamic effects
Spatial spillovers
Retail

ABSTRACT

We investigate the effects of customers' propensity to recommend a brand measured by the Net Promoter Score (NPS) on the sales of a publicly listed UK retail firm, which is dominant in the upholstered furniture market in which relatively expensive durable products have a long replacement cycle. From reasoning about word-of-mouth transmission, we theorize that store-level NPS effects are likely to (i) be nonlinear over time and (ii) depend not only on store's own NPS but also on the firm's neighboring stores' NPS. We investigate the relationship between the NPS and sales over a 48-month period, matching every UK store's monthly sales with its monthly NPS. We advance an econometric approach to allow for the identification of a non-linear time path. We find that the NPS, operationalized by surveying established customers six months after purchase, predicts future sales. Moreover, NPS sales effects unfold over many months along an "inverted U" time path: for own-store effects, from eight months after purchase, peaking at 10 months; and for neighboring-store effects, from six months after purchase, peaking at 11 months. Quantitatively, a sustained increase in the NPS across the company gives rise to substantial financial returns.

1. Introduction

The broad academic context of the present study is the "shift from transactional to relationship marketing" and associated use of customer feedback metrics (CFMs) to monitor and improve firm performance (Agag et al., 2023: 1). These include the Customer Effort Score (backward looking), Customer Satisfaction such as the "Top 2 Box" metric (oriented to the present), and the future oriented Net Promoter Score (NPS) (Agag et al., 2023). The present study is located within the literature on CFMs but makes its specific contribution to the narrower literature on the NPS, which is one of the most widely used CFMs (Baehre, 2024), perhaps *the* most widely used (Agag et al., 2023).

The NPS connects with firm performance by a theoretical pathway that highlights customer advocacy and word-of-mouth communication as promoting firms within social networks (de Haan et al., 2015; Lacohee et al., 2024). Recent studies of the NPS literature also emphasize that understanding the economic consequences of customer recommendation metrics remains a central challenge for both marketing theory and practice, particularly given the continued prominence of NPS among managers, the mixed

* Corresponding author.
E-mail address: c.dimos@bath.ac.uk (C. Dimos).

https://doi.org/10.1016/j.jretai.2026.09.004

0022-4359/© 2026 The Authors. Published by Elsevier Inc. on behalf of New York University. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

evidence regarding its relationship with business performance (Baehre et al., 2022; Baehre, 2024; Lacohee et al., 2024), and the importance of industry and national context (Agag et al., 2023; Lacohee et al., 2024).

We contribute to understanding the role of customer advocacy in reliably predicting future business performance. We consider one metric of customer advocacy, the NPS which is simple to calculate and widely used. The widespread adoption of this metric across industries worldwide reflects the persuasiveness of Reichheld's (2003: 54) claim that the NPS can predict sales growth and is thus a suitable metric to inform investment decisions, remuneration policy, and corporate culture (Bendle et al., 2019; Stephenson, 2020; Baehre et al., 2022). Indeed, knowing whether the NPS can reliably predict future sales matters, as it can determine whether the NPS should be treated as a strategic indicator influencing managerial decisions on investments in employee training, customer experience initiatives, and incentive systems tied to NPS performance (e.g., Reichheld, 2003; Keiningham et al., 2007).

The NPS is used by more than two-thirds of the Fortune 1000 (Reichheld et al., 2021) and is often included in company annual reports (Brandt, 2007; Abel, 2017). Nonetheless, despite the widespread use of NPS in corporate practice, academic evaluations of its predictive power remain mixed. Some studies found positive NPS sales effects (Reichheld, 2003; Pingitore et al., 2007; van Doorn et al., 2013; Baehre et al., 2022) whereas others found no relationship (Morgan & Rego, 2006; Keiningham et al., 2007, 2008). Beyond being inconclusive, the extant literature also exhibits methodological shortcomings that question the validity of existing evidence on the predictive capacity of the NPS. Indeed, Baehre et al. (2022: 67) concur with Bendle et al. (2019) that the literature suffers from methodological weaknesses which do not always allow for the identification of causal NPS effects. The paucity of empirical studies, lack of consensus among them, and methodological shortcomings may explain some skepticism among academic researchers towards the use of the NPS as predictive of business performance (Agag et al., 2024).

We advance the existing literature on the NPS-Sales relationship in four ways. First, we shift the level of analysis from between-brand analysis to within-brand analysis by examining the NPS-sales relationship at the store level, i.e., the actual unit at which the NPS is determined and sales are generated. This perspective equips managers with evidence relevant not only for firm-level strategy but also for store-level decision making. Existing studies explore the NPS-sales relationship between brands or industries, using aggregated, company-level data.

Second, we explore cross-store spillover effects by examining how the NPS of neighboring stores influences focal-store sales, whereby customer advocacy in one store affects sales in nearby stores within the same company. This has implications for store strategy and location decisions. Because the literature has only focused on investigating the NPS-sales relationship using aggregated company-level data, it precludes, by definition, the investigation of such spillover effects. Failing to account for spillover effects may overestimate the effects of customer advocacy of the focal unit by modeling sales as a function of the focal unit's NPS alone while neglecting influences from the wider company structure.

Third, we investigate the temporal path through which NPS effects materialize, i.e., whether sales effects emerge with lags, persist over extended periods, and/or follow non-linear paths. Understanding the time horizon over which customer advocacy translates into sales informs managerial planning while underscoring the importance of incorporating time dynamics into scholarly analyses of customer advocacy and business performance. Prior research has not adequately examined the temporal structure of the NPS–sales relationship; existing studies limit themselves to contemporaneous or short-horizon associations. This constitutes a severe limitation as NPS effects on sales may require time to materialize, especially for products with long inter-purchase cycles.

Fourth, unlike a good part of the literature that only provides associations between the NPS and sales, we provide robust evidence of the predictive capability of NPS for sales by matching store-level data with an econometric approach designed to accommodate multiple lagged values of the NPS without being compromised by multicollinearity. By comparing results across successively more sophisticated statistical and econometric approaches and conducting robustness checks, we demonstrate the substantial influence of modeling choices on estimated NPS effects, with important implications for researchers undertaking quantitative analysis in the field.

We aim to answer the following research questions:

1. Can customer advocacy, as captured by NPS, predict future sales performance?
2. Do these effects operate only locally (i.e., does the store's NPS impact only its own sales), or do they extend beyond the store where the recommendation originates (i.e., does a store's NPS impact the sales of the company's other stores)?
3. What is the temporal structure of the NPS-sales relationship (i.e., is it contemporaneous or prolonged over time; and, if prolonged, does it unfold in a linear or non-linear fashion)?
4. How robust are observed effects to alternative modeling choices?

We make our contributions by investigating the NPS-sales relationship over a period of up to 48 months and *within* the same company. We use longitudinal NPS and sales data from individual stores belonging to DFS, a major UK retail firm – during our sample

period, listed in the FTSE 250 – with nearly 100 stores nationally together accounting for over two-thirds of the UK living room furniture market during our sample period (DFS, 2018). Therefore, the present study is the first to respond to the call of Keiningham et al. (2007: 44) for 'a longer-term, longitudinal study' to investigate whether 'changes in satisfaction/loyalty metrics are important predictors of relative changes in revenue *within* companies'.

The study is organized as follows. In Section 2, we provide a brief literature review and critique of studies exploring the NPS-sales relationship. In Section 3 we present our institutional setting. Section 4 discusses the conceptual framework underpinning the study while Section 5 discusses the data and samples used. Section 6 outlines our modeling and estimation strategy. In Section 7, we present, interpret and check the robustness of the findings. Finally, in Section 8 we discuss their implications for researchers and managers.[1]

2. Literature review

The extant literature does not provide consistent evidence of a positive relationship between NPS and sales, with some studies reporting no association. Table 1 summarizes studies reporting quantitative estimates of the relationship between the NPS and sales. The seven studies detailed in Table 1 (i.e., excluding the present study) reveal mixed findings: three report mainly no relationship and four report positive correlations or marginal effects from regression analysis (Table 1; Columns 12 and 15).

The lack of consensus may reflect differences in context including country settings, industries examined, brands, product types (corresponding to varying product life cycles and differing degrees of repeat purchasing) and sample periods, as well as differences in methodological choices, including data type (i.e., longitudinal versus cross-sectional), data frequency (e.g., quarterly versus annual), time-series depth, and econometric estimation techniques.

2.1. Measurement and operationalization of NPS

The original version of the NPS is based on Reichheld (2003) and it is derived from one question: 'How likely is it that you would recommend [company X] to a friend or colleague?'. Reponses are recorded on a scale from 0 to 10 and categorized into three groups: customers giving a score of 0 to 6 are classified as detractors; 7 and 8 as passives; and 9 and 10 as promoters. The NPS is calculated by subtracting the proportion of detractors from the proportion of promoters, and thus ranges from −1 to +1 (where −1 is 100% detractors and +1 is 100% promoters). However, there is much variation in the literature in operationalizing and measuring the NPS (Table 1, Column 5).

In terms of operationalization, some studies derive the NPS from the original question used in Reichheld (2003). Other studies use different questions to gauge the propensity to recommend. For example, Keiningham et al. (2007, 2008) use 'business relation' in the place of 'colleague' whereas Morgan and Rego (2006) use two questions to derive the NPS: 'Have you discussed your experiences with [brand or company x] with anyone?'; and 'Have you formally or informally complained about your experiences with [brand or company x]?'. This variation in questions may be related to different findings. DFS remains close to Reichheld's original question but adds "family" to "friends" and "colleagues" to better capture the social network of the customer (i.e., 'How likely are you to recommend DFS to your friends, family and colleagues?').

With respect to measurement, some studies deviate from Reichheld's original 11-point NPS scale and use alternative 10-point scales by excluding the "0″ value (Keiningham et al., 2007, 2008).[2] These studies tend to report weaker or null results, raising the possibility that differences in measurement weaken construct validity. In our study, the NPS is derived from the original 11-point scale and method of calculation.[3] Therefore, our construct's validity is not subject to issues arising from the omission of the "0″ value while the inclusion of "family" in the NPS question more fully captures the social network of the customer.

2.2. Contextual variation

The studies examining the relationship between the NPS and sales exhibit substantial variation in terms of geographical context, ranging from multi-country datasets to single-country analyses (Table 1, Column 1). The earliest study by Fred Reichheld (2003) relies on data from three countries, namely the United States, the United Kingdom, and South Korea, covering several industries and brands. Subsequent studies tend to rely on single-country datasets, particularly from the United States. Country heterogeneity even within the

[1] To keep this article to a reasonable length, much detailed explanation is provided in 10 appendices. We anticipate that different groups of appendices will be of interest to different groups of readers. Appendices A, B, C and D present and discuss theory and empirical evidence that: (i) customer word-of-mouth is the transmission mechanism between the NPS and future sales; (ii) motivates our hypothesized spillover effects between the NPS at one store and sales at neighboring stores; and (iii) motivates our hypothesized non-linear time path for NPS sales effects. Appendices E, G, and H report in full all the auxiliary econometric and statistical results that informed the modelling approach explained and implemented in this article. Appendices F and J enable replication of our econometric approach by providing: (i) detailed explanation of how our variables are constructed and our model specified; and (ii) details of our data and syntax files. Appendix I reports the estimates conducted as robustness checks on our main findings.

[2] However, the original 11-point scale (0–10) is generally preferable, as it allows respondents to explicitly indicate that they would not recommend the product or service at all. Removing the "0" option may constrain respondents' ability to express the most negative evaluation and could lead them either to select a higher scale point that does not fully reflect their views or to avoid responding altogether.

[3] Studies using the original NPS question as Reichheld (2003), or close variants, all use the original method for calculating the NPS. The one exception is Morgan and Rego (2006) whose different question entails a different calculation (Table 1, Column 5).

ARTICLE IN PRESS

Table 1
Studies reporting quantitative estimates of the relationship between the NPS and sales.

Study (Country)	Context			Data Characteristics							Methodology			Findings
1.	2 Number of industries / brands	3. Level of aggregation / analysis	4. Short product cycle / repeat sales OR long product cycle / new customers (WoM)	5. Reichheld NPS question (yes/no) / 11-point or 10-point scale / calculation method (yes/no)	6. Data type (upon which results are based)	7. Frequency	8. Sample period	9. Number of time periods	10. #obs available for estimation	11. #obs based on primary data / # individual responses	12. Results reported as correlations OR marginal effects	13. Controlling for unobserved cross-sectional influences	14. Non-linear effects	15. NPS - sales relationship
Reichheld (2003) (US, UK, South Korea)	7 industries (computers, life insurance, Korean auto insurance, airlines, Internet service providers, UK supermarkets, and car rental) – 14 brands	Brand level	Short cycle / Repeat Sales	Yes 11-point Yes	Pooled	Quarterly	1999–2003 (different periods for different industries)	4	Not reported	10,000–15,000	correlations	No	No	Positive
Morgan and Rego (2006) (US)	Fewer than 38 industries – 80 brands	Brand level	Short cycle / Repeat Sales	No Not applicable No	Longitudinal	Annual	1994–2000	7	569	Not reported	marginal effects	No	No	No relationship
Keiningham et al. (2007) (US, Norway) Keiningham et al. (2008) (US, Norway)	5 industries (Banking, gasoline stations, home furnishing retailer, security systems,	Brand level	Short cycle / Repeat Sales	No (10-point Likert scale) Yes	Pooled (pooling cross-sections)	Annual	2000–2005 (depending on industry)	Ranging between 2 and 5 (depending on industry)	17	Not reported (smaller than 15,500)	correlations / partial correlations	No	No	No relationship (positive only for banking)

(*continued on next page*)

ARTICLE IN PRESS

Table 1 (*continued*)

Study (Country)	Context			Data Characteristics							Methodology			Findings
	transportation) – 17 brands													
Pingitore et al. (2007) (US)	4 industries (auto insurance, full-service investment, airline, rental car) – 64 brands	Brand level	Short cycle / Repeat Sales	Yes 11-point Yes	Cross-sectional	N/A	2005/06	1	64	26,944	correlations	No	No	Positive
van Doorn et al. (2013) (Netherlands)	4 industries (banking, insurance, utilities, telecom) – 46 brands	Brand level	Short cycle / Repeat Sales	Yes 11-point Yes	Cross-sectional	N/A	2008	1	46	11,967	marginal effects	No	No	Positive
Baehre et al. (2022) (US)	1 industry (sportswear) – 7 brands	Brand level	Short cycle / Repeat Sales	Yes 11-point Yes	Longitudinal (balanced)	Quarterly	2013–2017	19	133	38,644	marginal effects	Yes	No	Positive
Present Study (UK)	1 industry (upholstered furniture) – 1 brand – 96 stores	Store level	Long cycle / new customers (via WoM)	No (family added) 11-point Yes	Longitudinal (unbalanced & balanced)	Monthly	August 2011- July 2015	48	1654 (unbalanced) 456 (balanced)	44,585	marginal effects	Yes	Yes	Positive

same study raises questions about the identified NPS-sales relationships in the literature, because differences in consumer behavior, market structures, and competitive environments across countries may affect how recommendation intentions translate into sales. The present study contributes to the literature by examining the NPS–sales relationship using data from the United Kingdom, thereby providing evidence from a different national context. Also, single-country studies keep the national context invariant, thereby enabling better identification of the underlying relationship.

Early work (Reichheld, 2003; Morgan & Rego, 2006; Keiningham et al., 2007) pools brands across multiple industries, implicitly assuming a universally homogeneous NPS–sales relationship (Table 1, Column 2). However, industries differ substantially in purchase frequency, customer involvement, and competition intensity. Evidence from Keiningham et al. (2007), who find a positive relationship only in banking, suggests that industry heterogeneity may mask effects in pooled samples. In contrast, the only study that focuses on a single industry (Baehre et al., 2022) finds a positive relationship, consistent with the view that the NPS–sales link is context-dependent. In our study, we also focus on one industry to eliminate the risk of identifying a relationship that is driven by cross-industry heterogeneity.

Besides different industries, all existing empirical studies have also exclusively examined the NPS-sales relationship between multiple brands (ranging from 7 to 80 brands) (Table 1, Columns 2 and 3). All studies used brand-level data where customer feedback and sales were aggregated across brands corresponding to one respective data point for NPS and sales for each brand. Similar to drawing data from different industries, investigating the NPS-sales relationship across multiple brands is subject to potentially significant cross-brand heterogeneity that may obscure the NPS-sales relationship. Furthermore, the use of brand-level information limits our understanding of whether the relationship holds at lower levels where economic activity actually occurs, e.g., in retail, the physical store. To address these limitations, we focus on a single industry and a single brand, using store-level data that allows us to capture finer-grained variation in both NPS and sales. This is the first study to examine the NPS-sales relationship at the store level.

Another critical dimension is whether sales are driven primarily by repeat purchases (short product cycles) or by new customer acquisition via word-of-mouth (long product cycles) (Table 1, Column 4). All prior studies focus on industries with relatively short purchase cycles, where repeat purchasing is common. In such contexts, NPS may overlap conceptually with other loyalty or satisfaction measures, making its incremental predictive power more difficult to detect. The present study is distinct in examining a long product cycle context, where repeat purchases are rare and sales effects are more plausibly transmitted through word-of-mouth. This difference in the underlying demand mechanism may explain why some studies find weak or no relationships while others find positive effects.

2.3. Data and methodological design

Studies substantially differ in their data and methodology (Table 1, Columns 5–14). We show that divergent findings in the literature can be associated with heterogeneity in research design.

Data characteristics (Columns 5–11). In investigating the NPS-sales relationship, existing studies use cross-sectional, pooled, or longitudinal data (Table 1, Column 6). Several studies rely on cross-sectional designs (Pingitore et al., 2007; van Doorn et al., 2013), which can identify contemporaneous associations but cannot establish whether changes in NPS precede changes in sales. Pooled cross-sectional approaches (Reichheld, 2003; Keiningham et al., 2007) similarly conflate cross-sectional and temporal variation. In contrast, longitudinal studies (Morgan & Rego, 2006; Baehre et al., 2022) are better suited to assessing predictive relationships. Notably, positive findings are more common in longitudinal designs, particularly when they exploit richer time variation. To focus on within-store variation over time, we use longitudinal samples which are among the largest in this literature (Table 1, Column 10).

Studies differ in data frequency and the time horizon of their samples. Prior longitudinal studies typically use annual or quarterly data with relatively few time periods (Table 1, Columns 7–9), which may be insufficient to detect delayed or distributed sales effects. If NPS influences sales only gradually, lower-frequency data may fail to capture the relationship, contributing to null findings in earlier work (as reported by Morgan and Rego (2006)). We use monthly data over four years, enabling the identification of lagged and cumulative effects over time.

Econometric specification and unobserved heterogeneity. Studies differ in whether results are reported as simple correlations or as marginal effects from multivariate models (Table 1, Column 12). Correlation-based studies (Reichheld, 2003; Pingitore et al., 2007) generally report positive relationships, while more rigorous regression-based analyses often yield weaker or mixed results. This pattern suggests that failure to control for confounding factors may inflate estimated relationships. At the same time, only Baehre et al. (2022) explicitly control for unobserved cross-sectional influences using longitudinal methods (Table 1, Column 13). Studies that do not account for between-group (in our case, between-store) heterogeneity may suffer from omitted-variable bias, leading either to spurious positive findings or to attenuation of true effects, depending on the direction of bias. Accordingly, we use a fixed-effects estimation strategy to control for potential confounding factors and unobserved heterogeneity.

Non-linear effects: Finally, none of the earlier studies allow for non-linear time paths of NPS effects on sales (Table 1, Column 14). If the impact of customer advocacy unfolds gradually and non-linearly, linear specifications may mischaracterize or underestimate the true relationship. The present study is the only one to explicitly model non-linear dynamics, which may explain why it identifies positive effects that are not detected in previous studies.

Estimation Strategy: Baehre et al. (2022) address methodological shortcomings in the previous literature relating to unobserved heterogeneity and unmodelled dynamics by estimating dynamic panel models. However, Baehre et al. (2022) specify a dynamic random effects (RE) model, which is known to cause biased and inconsistent estimates (Pesaran & Smith, 1995). Consequently, we do not adopt the Baehre et al. (2022) econometric strategy but, instead, develop a new approach that is designed to model dynamic effects while being statistically valid.

While previous research has advanced our understanding of NPS as a leading indicator of firm performance, it has offered mixed

evidence on whether the NPS predicts sales – reflecting variation in measurement, context, and empirical strategy. Important questions remain about not only *whether,* but also *when* and *how* it predicts sales. In answering these questions, this study contributes a temporally explicit and store-level analysis by advancing a robust dynamic panel methodology to: (i) separate direct (local) and indirect (area-level) impacts; and (ii) test for non-linear time paths.

3. Institutional setting

DFS is a suitable setting to explore the NPS-sales relationship not only from the perspective of a dominant retail firm in a specific product market (Sections 1; and 8.4 for sales revenue and profit) but also from the more general perspective of a product category characterized by high prices and corresponding low purchase frequency and substantial customer deliberation. These features reinforce the relevance of DFS during our sample period for studying NPS sales effects where these arise from word-of-mouth recommendations (see Section 4). First, DFS manufactured, marketed and retailed products under a single brand and, across the UK, led all sectors of the upholstered furniture market by both product and customer type (Section 1 above; and Stephenson, 2020). The single-brand feature is consistent with the NPS as a measure of brand or company loyalty, while all-sector domination widens the scope for effective word-of-mouth recommendations across both potential customers and products. Second, substantial customer deliberation is likely to positively affect both generation and impact of word-of-mouth recommendations (Section 4).

The DFS operating model enforced homogeneity among its UK stores, which – by minimizing between-store variation – is most useful for our analysis (Section 6.2). DFS in the period studied was a heavily centralized operation with local stores and regions not being involved in marketing, promotions or even store layouts. All marketing was centrally delivered through national contracts; stores had no scope for local agreements or activities. Marketing was delivered through national TV and media advertising, which was balanced to deliver maximum results by a media buying agency. As an example, should a TV region over index on the "credit fan" customer type, the advertising would be adjusted accordingly. The only activity that might have been perceived as "local" was the sponsorship of local charities and sports teams; however, this was also centrally controlled with no involvement from stores. Store layouts were managed by the head office space planning team who had detailed plans for each location and ensured that the relevant ranges were appropriately placed. For example, should the "Zinc" sofa be the "hero" product in that month's advertising then it would be in the central front section of all stores. This was monitored closely and there was no latitude for local management to change the store layout. In sum, the store teams had two functions only, to sell, and to deliver customer service, both within strictly defined models. They had no responsibility for purchasing, merchandising, promoting, advertising or managing stock, all of which were controlled by central teams with their own management structures.

At DFS, the NPS was introduced as the fundamental metric for creating and then sustaining a "customer-centered" culture. To this end, DFS adopted the NPS as part of a company-wide NPS "system", whereby employee remuneration, recognition and performance management at all levels were tied to the NPS rather than to sales directly (Stephenson, 2020 analyses in fine detail this corporate change process).[4] Consequently, maximizing response rates and ensuring that survey responses were representative of the customer base were corporate priorities for two separate reasons: (i) to ensure the validity of NPS and other survey findings informing Board-level strategy; and (ii) at least as important, to ensure the legitimacy among employees of corporate and personal outcomes determined by NPS responses. These sampling issues are discussed at length by Stephenson (2020: 123–32); the remainder of this Section draws on this account.

In this study, we analyze the sales effects of NPS responses from the Established Customer (EC) Survey, which is collected six months after purchase. Whereas the temporal distance between purchase and experience of disposable products (e.g., some fashion or electronics categories) is relatively short, DFS durable products with long replacement cycles entail a substantially longer period before the product can be delivered (typically two months after purchase) and experienced. Accordingly, compared to either of the other two DFS surveys – respectively implemented immediately post-purchase and post-delivery – only the EC responses reflect prolonged experience of the product that can inform credible recommendations leading to subsequent sales (Section 4.3 and Appendices D and I, Part 2). After six months, customers have had sufficient time for impressions formed by purchase, after-sales service and delivery to fade while gaining experience of the product – e.g., its comfort and build quality – and coming to a settled conclusion about the brand.

In our sample period, the Established Customer survey was sent to all customers and responded to by circa 5 per cent. Previously, DFS had undertaken significant work to maximize response rates. For example, for EC surveys, communications were timed to coincide with likely product usage, which increased engagement (DFS Internal Documents: A1, 2011). In addition, DFS rejected monetary or prize-draw incentives, because trials showed that these not only increased response rates but also inflated NPS scores (DFS Internal Documents: A37, 2011). Instead, charitable donations were used as a non-personal incentive, demonstrably increasing response rates without distorting scores (DFS Internal Documents: A1, 2011).

Following a trial, email was retained as the primary delivery mechanism, because almost all customers provided an email address and email enabled the cost-effective collection of large volumes of data (DFS Internal Documents: A37, 2011). To protect representativeness, DFS introduced internal controls to prevent salespersons "gaming" the system; for example, by selectively capturing email addresses from satisfied customers or from friends etc. Basic validation checks were strengthened, including prohibiting store or self-addresses and restricting multiple surveys to the same email (DFS Internal Documents: A39, 2011). In addition, DFS implemented

[4] With two exceptions, this applied to all employees from the CEO down. The exceptions were the responsible Director and the subordinate team member running the NPS surveys whose remuneration was linked to the accuracy of stores' NPS returns according to external triangulation (described below).

parallel survey mechanisms to ensure that all customers were surveyed. To include customers who did not provide email addresses, these received a one-question NPS survey by text message, while a small residual group was randomly sampled for telephone follow-up. These additional responses were not included in headline NPS results but were used diagnostically to detect systematic bias between customers with and without email contact details (DFS Internal Documents: A29, A35, 2011). This approach explicitly prioritized data integrity – e.g., balance across stores and administrative areas – over cost. Moreover, because every survey was attached to an order number, survey responses were assessed for representativeness by comparing respondents and the overall customer base with respect to (i) the products purchased, (ii) the methods of payment across finance or non-finance agreements, and (iii) trading patterns by day. As responses closely mirrored operational data, the survey results were accepted as a meaningful representation of the estate (DFS Internal Documents: A37, 2011). In sum, prior to our sample period DFS had implemented internal procedures to increase response rates substantially, ensure data validity, and deliver survey responses representative of the customer base.

In addition, EC survey responses were continuously compared with an independent third-party brand-tracking NPS survey of the whole upholstered furniture sector. To this end, DFS engaged a well-respected third-party marketing agency to complete whole of market NPS studies at regular intervals. These included double-blind surveys, using their independent and comprehensive database of UK consumers, which included both purchasers and non-purchasers. Comparisons restricted to DFS customers revealed that these third-party surveys consistently delivered NPS results within ±1 percentage point of the DFS Established Customer surveys. Accordingly, during our sample period, external triangulation consistently checked the results from the internal procedures to ensure that the NPS responses from the EC survey were representative of the customer base across the entire store estate. Whilst there were mechanisms in place that could have adjusted the data should systematic over- or under-representation of a store, product or payment method occur, during our sample period these were never needed.

4. Conceptual framework

We start with a simple model of word-of-mouth recommendation as the transmission mechanism between the NPS and subsequent sales (Section 4.1). We then extend this model to yield predictions regarding (i) direct and indirect NPS effects (4.2) and (ii) the time path of NPS effects (4.3).

4.1. Linking NPS to sales: core mechanism

The NPS is intended to capture customers' propensity to recommend a brand or retailer to others. The central hypothesis underpinning its use as a managerial tool is that higher NPS should generate greater customer acquisition via positive word-of-mouth (WOM). The core mechanism involves the following sequential logic.

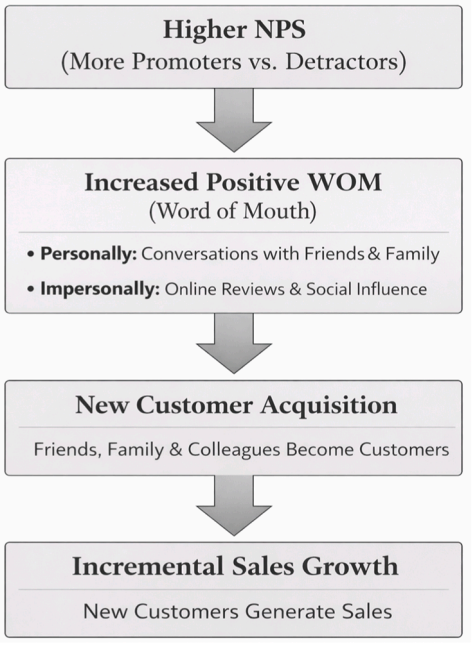


Fig. 1. NPS and sales: the core mechanism.
Source: authors.

1. A higher NPS reflects a greater share of promoters relative to detractors.
2. Promoters engage in more positive WOM, either personally (e.g., conversations with friends/family/colleagues) or impersonally (e.g., online reviews, social influence).[5]
3. Positive WOM, in turn, increases the likelihood that family members, friends and colleagues will consider purchasing from the retailer, visit one of the company's stores, and become new customers.
4. Over time, these new customers generate incremental sales.[6]

This mechanism (set out in Fig. 1) is especially relevant in product categories with: (i) low-frequency purchases (where repeat purchasing is rare within the time window);[7] (ii) high involvement (where trust and recommendation matter); and (iii) considerable social discussion and pre-purchase consideration.

4.2. Direct vs. indirect effects

We distinguish between: (i) direct (own-store) effects, where a store's NPS influences its own future sales through local WOM; and (ii) indirect (area-level) effects, where the sales at one store are affected by NPS scores at nearby stores.

According to Meyners et al. (2017: 10), distance matters less the stronger the social ties under consideration. Theoretically, this statement can be rooted in social capital theory and helps to motivate our emphasis on area effects. The DFS version of the NPS question refers to "family" and "friends" as well as to "colleagues", which together give the social context for particularly effective word-of-mouth recommendations, because according to social capital theory kinship and friendship are social ties of the "strong ties" type, which is the type characterized by high trust (Granovetter, 1973) and thus the least likely to be attenuated by distance.

Although 'insights into how geographic proximity directly affects the strength, and therefore the success, of social influence remain sparse' (Meyners et al., 2017: 2), there is evidence that the distances over which social influences (including via face-to-face word-of-mouth recommendations) can be effective do not exceed the distances over which we estimate inter-store NPS sales effects. Meyners et al. (2017: 6) give the following summary of the empirical regularity arising from research into the distance dependence of the strength of social ties:

> Multiple studies have shown that the likelihood of the formation and maintenance of social ties between individuals can be expressed as an inverse logarithmic function of the geographic distance between them … this relationship holds true for large distances (such as those between cities or regions) … Interestingly, the emergence of new communication technologies whose financial costs are independent of geographic distance has not altered this relationship.

According to this function, the relative strength of social ties attenuates very sharply as distances increase to around five miles but relatively gently as we approach the 100-mile limit of interest to our study :[8] in round terms, the strength of social ties weakens by 50 per cent as the distance between individuals increases from 10 to 100 miles.[9]

The relative effects on a focal store's sales of the focal-store's own NPS and the NPS of neighboring stores in the area will depend on the respective numbers of new customers at the focal store consequent upon recommendations made by customers of the focal store

[5] There is evidence that word-of-mouth remains important in the presence of on-line recommendation, both in general and in the particular context of the present investigation. In general, according to Danziger (2017): 'The Word of Mouth Marketing Association quantifies this claim, finding that two-thirds of WOM's impact is from offline communications, and only one-third from online'. What's more, it finds that 'an offline WOM impression drives at least 5 times more sales than a paid advertising impression, and much more (as much as 200 times more) for higher-consideration categories'. Danziger (2017) also reports the following, which suggests that word-of-mouth is important in the context of the present investigation: 'In every survey Unity Marketing conducts with retailers on their marketing and advertising methods, word of mouth is hands down the most important, being both most used and most effective. For example, in a recent survey among furniture retailers, over 80% of furniture retailers named word-of-mouth their most powerful marketing tool.'

[6] We follow Baehre et al. (2022: 69) in highlighting the importance of word-of-mouth recommendation as the link between the NPS and sales. In our data, we have no direct observations on word-of-mouth recommendations. However, DFS survey data yields two findings consistent with positive linkage between recommendations and subsequent sales. First, we find a strong statistical association between intended recommendation (responses to the NPS question) and actual recommendation (binary responses from established customers to the question: 'Have you actually recommended DFS to friends, family and colleagues, since you placed your order or received delivery of your product?') (see online Appendix B for the detailed evidence). And, second, 13 per cent of respondents to the DFS post-purchase surveys covering 2014, 2015, and 2016 indicated that their visit to a DFS store was the result of a recommendation.

[7] Reflecting customer loyalty, the NPS can drive sales by repeat purchases from existing customers and/or new customers via unremunerated person-to-person word-of mouth recommendations from trusted individuals who are DFS "promoters". In the case of sofas, two considerations support the assumption that NPS effects rarely unfold via the repeat purchase channel: (i) the long replacement cycle (seven years, as noted in the Introduction, above); and (ii) the high cost of a sofa. Moreover, this assumption is supported by DFS survey data. From over 19,000 Existing Customer surveys in the first three waves of our sample (the financial years 2011-12, 2012-13, and 2013-14), only 59 repeat purchases (0.31% of the total) are recorded. (See online Appendix A for detailed analysis of repeat purchases during our sample period.) We conclude that our findings are not to any substantial extent influenced by repeat sales. Although we have no direct observational data on word-of-mouth transmission, ruling out repeat purchase as a substantial influence points strongly towards word-of-mouth as the main channel linking the NPS and future sales.

[8] For our data, within each of the 11 DFS administrative areas, no two of the 24 stores in our estimation sample are separated by >100 miles (indeed only one is separated by >70 miles from the other stores in its region).

[9] On-line Appendix C discusses this inverse log relationship in more detail.

and by customers of the other stores. From a theoretical perspective, we cannot predict the relative strength of the focal-store and area effects, because the area effects are expected to be: (i) stronger than the focal-store effects because of scale effects – the number of stores in each area; and (ii) potentially weaker because of distance effects – the potentially attenuating influence of distance. As these two forces pull in opposite directions, their joint outcome is theoretically indeterminate. However, in this study, our focus is on estimating the relative strength of focal-store and area effects rather than on the two forces driving the area effects.

In our empirical analysis, besides direct NPS effects on sales (NPS measured only from each store's own customers), we also allow for indirect NPS effects on sales with considerable geographic reach, whereby the NPS of each DFS store influences not only its own sales but also the sales of neighboring DFS stores (NPS measured as the mean NPS across all other stores – i.e., net of the focal store's own NPS – in the focal store's DFS UK administrative area, irrespective of whether these stores are homogeneous or heterogeneous).

4.3. Temporal dynamics and inverted-U shape

Our second extension of the WOM transmission mechanism is to investigate the time path over which NPS could potentially influence sales. We theorize that both store-level and area-level effects unfold over a prolonged period and, possibly, in a nonlinear fashion: first from the perspective of recommendees; and second from the perspective of recommenders.

From the perspective of recommendees. We assume that those influenced by recommendations will take varying amounts of time – according to their varying circumstances (e.g., the urgency of the purchase and competing claims on their time) – to research a highly differentiated product range, to make a purchase decision, and then to make a purchase. Accordingly, the simplest explanation of why NPS sales effects are likely to unfold over a non-linear time path – i.e., according to an "inverted-U" pattern – is that some recommendees act relatively quickly (reflecting favorable circumstances), a larger group take more time (average circumstances), and some act relatively slowly (unfavorable circumstances). Accordingly, from the time of recommendation, we conjecture that the sale effects are initially relatively small but then rise before eventually diminishing in successive months. Such a temporal pattern is consistent with established models of consumer adoption, in which purchase rates of consumer durables are distributed non-uniformly over time, typically rising to a peak before declining (Bass, 1969).

From the perspective of recommenders. Established customers require periods, each potentially taking many weeks or months, (i) to arrive at a settled and credible opinion regarding their purchase, and (ii) to host or otherwise communicate with family and friends to provide actual word-of-mouth recommendations to diffuse to prospective customers. Accordingly, the pathway from intended to actual recommendation and purchase, linking the NPS and subsequent sales, is likely to be protracted. To motivate, in addition, the possibility of a non-linear time path of NPS sales effects from the perspective of established customer recommenders, we posit that the NPS sales effect is determined by the total strength of recommendation which, in turn, is the product of the quality (i.e., the average credibility) of the recommendations and the quantity of recommendations. We consider these in turn.

The quality effect. Concepts from Martin and Lueg (2013) suggest that word-of-mouth recommendations may become (i) an effective influence on sales with a lag and (ii) increasingly effective over time up to a maximum. "Source experience" (familiarity of the customer with the product) is a necessary condition for "source trustworthiness" and "source expertise" (competence to make the recommendation) to inform credible and, hence, effective recommendations. (On-line Appendix D, Part 1, discusses these concepts together with their relevance to the products retailed by DFS and the unfolding of NPS sales effects over a prolonged period.) Moreover, because "source experience" and corresponding credibility is gained not all at once, but accumulates over time, word-of-mouth recommendations may initially become an increasingly effective influence on sales as the time elapsed from purchase increases. However, it is reasonable to assume that the quality of recommendations cannot increase without limit but, over time, increases at a decreasing rate (as increments to recommender source experience and source expertise diminish with familiarity) up to some maximum (i.e., recommendations reach their maximum quality in some particular month and then maintain the same quality in subsequent months). The corollary of these quality effects – if considered in isolation – is that NPS sales effects increase at a declining rate to reach a maximum level that persists in subsequent periods. Moreover, as the recommender's network approaches saturation with recommendations so a state of "message fatigue" may tend to reduce the quality and corresponding effectiveness of recommendations.

The quantity effect. According to social capital theory, high-trust networks involving family, friends and colleagues sufficiently close to be invited to customers' homes, or to be in regular communication, are characterized by "strong ties" and are thus necessarily restricted in size. The corollary is that each additional recommendation reduces the remaining number of potential recommendations within a given social network. Consequently, over time the number of new recommendations must diminish.[10] In addition, the quantity effect may be reinforced by the novelty effect, whereby over time the purchase loses novelty, becomes less of a talking point, and thus generates fewer recommendations.

The quality effect rises to a peak and then at best plateaus, although it may actually decline with "message fatigue". Rising to a culminating plateau or constant value can be represented by the rising portion of a quadratic function. Conversely, the quantity effect – reinforced by the novelty effect – continuously declines and can be represented by a linear function. The product of these two effects yields a function representing the strength of WOM recommendation through time. (A more formal presentation of these functions and their combination is presented in Appendix D, Part 2.) In the early periods, the rising quality effect may outweigh the falling quantity effect thereby yielding a rising strength function; conversely, if from the beginning the quality effect is outweighed by the quantity

[10] Here, we assume the necessity of first-hand experience for a credible recommendation, so that – to fix ideas – Customer A's first-order recommendation can influence Household B with whom Customer A has a close relationship, but that the chain of recommendation does not extend to second-order recommendations from Household B to further Households known to Household B but not Customer A.

Table 2
Data reduction and sample sizes.

UK Stores	Mean Months	Store-month observations available for estimation *	Description and use
91	38.62	3514	**Unbalanced dataset:** intermediate correlations and regressions (no lagged values)
Reductions (1): Estimating with 12 lags of the independent variables (substantial loss of observations is expected for two reasons: (i) 12 months per store are mechanical due to the lags; and (ii) additional losses reflect missing NPS_EC and calendar gaps that break the 13-month window required at each month).			
91	18.47	1681	**Unbalanced dataset:** intermediate two-way fixed effects regression
Reductions (2): 27 observations lost by calculating the net mean area variables			
91	18.18	1654	**Final unbalanced dataset**: used to estimate Eq. E1, reported in Appendix E (observations per store available for estimation vary from 2 to 31; mean = 18.18)
Reductions (3): Constructing a balanced dataset with sufficient time-series depth, the maximum number of observations, and statistical characteristics not significantly different from those of the unbalanced dataset entailed the following reductions: (i) 63 stores that because of missing observations reduced the sample size (by reducing the overall time-series depth by more than their contribution to cross-section width); (ii) two stores in regions with insufficient stores to calculate the area net mean average NPS, (iii) one month from the inclusion of the first lag of the dependent variable in Eq. 2, and (iv) the two stores with the highest average sales.†			
24	19	456	**Final balanced dataset:** used to estimate Eq. 1 (Section 6.5)

Notes:.
* After taking into account observations lost due to estimating with forward and lagged values (see Section 6).
† The first three reductions are dictated by, respectively, the need for (i) a balanced dataset with the maximum possible number of observations, (ii) sufficient stores in each area to define net mean area variables, and (iii) a dynamic model. These reductions gave a balanced dataset with 26 stores, which is used to support our robustness checks (reported in Section 7.1). The final reduction, from 26 to 24 stores, was undertaken to obtain a balanced dataset that, uniformly and unambiguously, is statistically homogeneous with the full unbalanced dataset (as demonstrated in Section 6.4 and Appendix G).

Table 3
Modeling strategy.

Models	Contemporaneous relationship between Sales and NPS	Time path specified by deep lags (contemporaneous plus 12 lags of store and area NPS)	Time path specified by deep lags (contemporaneous plus 12 lags of store and area NPS)	Time path specified by deep lags subject to the Almon transformation (contemporaneous plus forward values plus lags of store and area NPS)
Estimator(s)	Correlation coefficient; OLS regression; One-way Fixed Effects	Two-way FE with period (monthly) dummies	Bias-corrected LSDV (see Section 6.2) with period (monthly) dummies	Bias-corrected LSDV (see Section 6.2) with period (monthly) dummies
Static or dynamic?	Static	Static	Dynamic (model includes the lagged dependent variable)	Dynamic (model includes the lagged dependent variable)
Corresponding Dataset (see Table 2)	Unbalanced: all 91 stores; maximum number of observations (3514)	Unbalanced: all 91 stores; reduced number of observations (1681)	Unbalanced: all 91 stores; further reduction in number of observations (1654)	Balanced: 24 stores (456 observations)
Valid estimates?	No	No	No	Yes
• **Why?**	No time path (by definition)	Dynamic misspecification (no lagged dependent variable); Severe Multicollinearity	Severe multicollinearity	Almon transformation addresses multicollinearity
Potential selection bias arising from omitting stores?	No	No	No	Yes
Potential selection bias addressed?	Not applicable	Not applicable	Not applicable	Yes
• **How?**				Construction of the balanced dataset to mimic the statistical characteristics of the unbalanced datasets with all 91 stores
Evidence	Appendix E, Part 1	Appendix E, Part 1	Appendix E, Part 2	Section 7

Source: Authors.

effect then the strength function will decline continuously. The resulting strength function thus cannot be predicted a priori, because it may either rise from the outset (if at first the increasing quality effect prevails) or decline from the outset (if at first the quantity effect prevails).

We therefore expect word-of-mouth recommendations to impact sales over a protracted period after purchase and, most likely, along a non-linear time path. From the recommendee perspective, we present a simple explanation as to why NPS sales are most likely to unfold over a non-linear time path, while from the recommender perspective we conclude that the time path is theoretically indeterminate, allowing the possibility of a non-linear time path. The corollary is that we must specify our econometric model to allow for NPS sales effects to unfold along a linear or nonlinear time path according to the data.

To summarize the theoretical expectations arising from our conceptual framework, we do not expect NPS sales effects to be only an own-store effect or to arise immediately. Instead, we expect stores' sales to be influenced both directly by their own NPS and indirectly by the NPS of neighboring stores and that, via word-of-mouth transmission, both effects will be positive, lagged, and – most likely – unfold along a non-linear time path (following an inverted-U shape). The relative size and duration of direct own-store and indirect (area) NPS effects will be determined by the competing strengths of scale and distance effects. Accordingly, our empirical strategy is designed to (i) test our theoretical predictions on both own-store and area effects and their evolution over time, and (ii) quantify the magnitude and identify the time path of these effects.

5. Data & sample construction

The data are monthly and contain information from 96 UK stores over a period of four years (August 2011 - July 2015). However, five stores opened too late during the sample period to report sufficient observations for estimation. We therefore use monthly data from 91 DFS stores with 44,585 individual Established Customer survey responses, which were completed after six months from purchase. We used these responses to calculate the NPS for each store and in each month: this is a store-level measure derived by subtracting the proportion of detractors from the proportion of promoters for each store. DFS defines 11 areas in the UK, which we use to define area-level NPS variables enabling us to estimate the NPS effects of other DFS stores in the area on each store's sales. DFS also provided complete monthly sales data for every store during the same four-year period.

Our analysis reveals a fundamental trade-off: while simpler models (all subject to econometric issues producing invalid estimates) allow the use of more observations from the full 91-store sample, the most appropriate estimation method for our data imposes restrictions that reduce the usable sample. This reduction raises the potential for selection bias. Nevertheless, we promote the most appropriate, 'data-hungry' approach, as it is the only method capable of producing valid estimates while simultaneously accounting for (i) sales persistence, (ii) a prolonged time path of NPS effects on sales, and (iii) multicollinearity stemming from modeling this time path (Section 6.3). To address concerns that selection bias may result from estimating on a sample including fewer stores, we test both our sales and NPS variables to demonstrate that they do not differ statistically between the reduced (24-store) and full (91-store) samples with respect to means, variance, skewness, and kurtosis, and, in addition, stationarity (i.e., the variables in both samples are untrended). This evidence of homogeneity between samples is explained in Section 6.4 with test results reported in detail in Appendix G, and is consistent with: (i) institutional evidence of the homogeneity of DFS stores across the UK (Section 3); (ii) minimal between-store variation in the NPS across both full and reduced samples (Section 6.2); and (iii) evidence of locational homogeneity across DFS stores (Section 6.4 and Appendix G).

However, because the sample reduction is substantial, we begin our investigation with the 91-store sample, using three increasingly improved model specifications, each corresponding to an incrementally reduced set of store-month observations. The different sample sizes used corresponding to these three 91-store modeling approaches, together with the sample corresponding to our preferred modeling approach, are presented in Table 2. The following Section describes in more detail the modeling approaches used.

6. Modeling and estimation strategy

This Section sets out the modeling approaches we employ in relation to different samples, our preferred model, and why selection bias is not influencing our results.

6.1. Modeling approaches

We briefly set out four modeling approaches – three estimated using data from 91 stores and one (our preferred) from 24 stores. These explain how our preferred approach maintains the benefit of a well-specified model (valid estimates) without suffering the cost of reduced sample size resulting from increasingly "data-hungry" models (potential selection bias). We report the results from our preferred approach in Section 7 and, for comparison, the results from the other approaches in Appendices.

Modeling Approach 1. We model the contemporaneous relationship between sales and NPS by three approaches: (i) calculating the correlation coefficient; (ii) OLS regression of sales on NPS; and (iii) static one-way fixed effects regression of sales on NPS, which controls for store-specific influences on sales. These approaches allow us to use the maximum number of usable observations (3514 from 91 stores) but do not allow for the identification of a time-path of the NPS effects or account for sales persistence. (The corresponding results are presented in Appendix E, Part 1.)

Modeling Approach 2. To model the time path of NPS sales effects we employ a static two-way FE model including the contemporaneous and 12 lagged values of both the direct own-store NPS effects and the indirect area NPS effects. This model uses fewer observations (1681 from 91 stores) due to specifying lagged values. However, this approach is impaired by dynamic misspecification

(i.e., not accounting for sales persistence) and by severe multicollinearity due to the many lagged values used (see Appendix E, Part 1).

Modeling Approach 3. Compared to the previous modeling approach, this approach additionally accounts for sales persistence by including one lagged value of the dependent variable. This dynamic model uses fewer observations (1654 from 91 stores) due to the specification of the lagged dependent variable. These estimates are qualitatively consistent with our hypothesized non-linear time path of NPS sales effects. However, consistent with statistical theory, the signs and sizes of these estimates reveal the presence of severe multicollinearity, which is the corollary of estimating with a deep lag structure. (Appendix E, Part 2, reports these estimates and demonstrates how multicollinearity affects their size and sign.)

Modeling Approach 4. In addition to the previous modeling approach, this approach addresses the issue of multicollinearity through the use of the Almon transformation (Almon, 1965). This approach, which is explained in Section 6.2 and Appendix F, allows us to obtain valid estimates. However, this approach requires the use of a balanced dataset (i.e., each store has a complete set of sales and NPS values for the same set of months) and thus substantially reduces the number of observations and stores used for estimation (456 observations from 24 stores). To address the possibility of selection bias (i.e., such that the 24 stores in our balanced dataset may not be representative of the 91 stores in the full unbalanced dataset), the balanced dataset is constructed to mimic the statistical characteristics of the full unbalanced dataset (Table 2). This is our preferred modeling approach.

Table 3 provides an overview of the four modeling approaches while highlighting the trade-off between sample size and valid estimation.

6.2. Dynamic modeling: the bias-corrected least squares dummy variable (LSDV) estimator

We use *dynamic* fixed effects (FE) panel estimation to account for otherwise unmodeled dynamics. This approach provides a solution to accommodating the joint occurrence of otherwise unobserved heterogeneity at the level of individual stores (fixed effects) and persistence in stores' sales over time (dynamics).

Our FE approach estimates the within-store effects while controlling for (conditional on) between-store differences. Identification of the effect of independent variables is best achieved when these have a high ratio of within-group to between-group variation. This is a characteristic of both our unbalanced (91 stores) and balanced samples (24 stores). The NPS of individual stores typically varies much more in relation to their own scores over time than in relation to the NPS of other stores: in our unbalanced and balanced samples, respectively, for our variable of interest (NPS_EC) the within variation (SD= 0.422; 0.390) is around four times as large as the between variation (SD= 0.097; 0.103). Accordingly, we judge there is sufficient within variation in the data for valid FE estimation.

However, FE estimation yields biased and inconsistent estimates from dynamic panel models estimated on datasets with "finite" time series depth (Nickell, 1981) as in our case. To address this issue, we turn to the bias corrected least-squares dummy variable (LSDV) estimator with bootstrapped standard errors. Bias-corrected LSDV estimation, which is an alternative nomenclature for FE estimation, is suitable for both unbalanced and balanced data (Bruno, 2005).[11] Therefore, we use the LSDV estimator for both our 91-store unbalanced sample model (Appendix E, Part 2) and our 24-store balanced sample model (Section 7).

Unlike other estimators including the difference- and system-GMM dynamic estimators (Arellano & Bond, 1991; Blundell & Bond, 1998) which are designed for "wide-N" and "shallow-T" data, the LSDV estimator is a "Large-N" and "Large-T" dynamic estimator which better fits our data. As the time dimension increases, the difference- and system-GMM dynamic estimators suffer from increased proliferation of potential instruments making specification choices increasingly arbitrary which leads to unstable estimates (Roodman, 2009). Therefore, the LSDV estimator is our preferred estimator.[12]

6.3. The Almon approach and the problem of selection bias

To estimate our panel model while addressing the consequences of multicollinearity, we integrate two strands of time-series econometrics.

1. As reflected in the results reported in Appendix E, Part 2 (see especially Figure E1) valid point estimates cannot be obtained from models including many "lead" and/or "lag" values of one or more independent variables, because estimation is greatly impaired by multicollinearity (Almon, 1965: 179; Gujarati, 1988: 512). The solution introduced by Almon (1965) was to reduce multiple values – current together with leads and/or lags – of an independent variable to polynomial functions (which can be modeled as variables). We implement the Almon approach by transforming large numbers of leading and/or lagged values of the NPS variables (i.e., both own-store and mean-area NPS) into three functions denoted $Z0$, $Z1$, and $Z2$. Finally, we use the estimates for $Z0$, $Z1$, and $Z2$ in

[11] We implement the LSDV estimation by using the Stata user-written program *xtlsdvc* (Bruno, 2005: 473).

[12] Bruno (2005) provides both citations and additional evidence – Monte Carlo testing for N=20, which is not much different from the N=24 of our balanced sample, and for N=40 – that, for models limited in the number of cross-section units (N) and strictly exogenous regressors, the bias-corrected LSDV estimator shows better performance than either static LSDV or GMM estimation with respect to both (i) bias in the estimated effects of the lagged dependent variable and of the independent variables and (ii) Root Mean Square Error. Finally, approaches to time-series modeling designed to identify long-run cointegrating relationships and corresponding error-correction mechanisms are not suitable for modeling relationships between variables that are both stationary (indicated by the extensive unit root testing reported and apparent in the graphical evidence presented in Appendix E, Part 3) and of short span (four years only).

post-estimation calculations to recover the effect of each individual leading, current, and lagged NPS variable. Appendix F explains in detail how we calculate these functions and how we recover the NPS effects.

2. Appealing to well-established principles in time-series econometrics (Spanos, 1986: Chapters 23 and 24 – in particular, 601–2; Hendry, 1995: 339), the estimated coefficient on the lagged dependent variable ($\widehat{\lambda}$) measures the persistence of sales. Accordingly, $1\text{-}\widehat{\lambda}$ measures the rate of adjustment of current sales to changes in sales in the previous month, and $1/(1\text{-}\widehat{\lambda})$ can be interpreted as a persistence factor by which the estimated effects of $Z0_t$, $Z1_t$ *and* $Z2_t$ can be multiplied to obtain their long-run effects. We apply this multiplier to each recovered individual lagged, current, and leading NPS effect to derive their corresponding *long-run sales effects*. Next, we sum the derived long-run coefficients on NPS to derive the estimated *total long-run NPS effect*, which reflects both (i) the direct short-run effects of all the lagged, current or leading NPS variables and (ii) their indirect effects induced by persistence over time. (Appendix F explains in detail how we integrate the Almon procedure into our dynamic model with one lag of the dependent variable.)

In adopting the Almon approach to modeling distributed lags, we must consider that we are estimating a panel model and that our 91-store dataset has missing values. Although bias-corrected LSDV estimation is designed for use with both unbalanced and balanced datasets, the Almon approach to overcoming multicollinearity requires a balanced dataset. If we create our *Z* variables from data with missing values then they would be defined differently across our observations making them inappropriate to use.[13] In turn, missing values with uneven temporal spacing would be a source of unstable estimation. Accordingly, we must use a "balanced" dataset for valid estimation.

6.4. *Investigating potential selection bias*

Sample reduction may introduce selection bias, potentially limiting the representativeness of estimates. However, several pieces of evidence mitigate this concern. First, institutional evidence indicates that DFS actively promotes store homogeneity (Section 3). Second, consistent with this, statistical evidence shows limited between-store variation in NPS (Section 6.2). We extend this by demonstrating that there are no significant differences between the statistical characteristics of the full 91-store unbalanced sample and the reduced 24-store balanced sample, including across DFS administrative areas.

For valid statistical inference, we compare the 24-store sample to its complement, i.e., the full 91-store sample excluding the 24-store sample.[14] This procedure is applied to both SALES and the Established Customer NPS (NPS_EC) to test the null of equality between our two samples of, respectively, their means, variances and measures of skewness and kurtosis. In each case, the statistical characteristics of our variables of interest cannot be compared using "off-the-peg" tests but instead require tests crafted to take account of the complex characteristics of our data: (i) heteroskedasticity across stores (e.g., larger stores might have more variable monthly sales than smaller stores); (ii) within-store serial correlation (NPS / sales persistence over months); (iii) dependence between the sales of stores in the same area (e.g., because of common shocks such as weather events); and (iv) dependence between areas (because of macro influences originating from either the company – such as national promotions – or the national economy). In addition, in each of our tests we control for area-month shocks and store fixed effects that would otherwise be absorbed into the residuals of the respective model-based testing procedures as sources of omitted variable bias (hence, invalidating estimates and inference). To calculate standard errors and p-values, we use the Wild Cluster Bootstrap (WCB) approach.

Our results suggest that across the full unbalanced sample and the balanced subsample there is no significant difference between the respective means of SALES and NPS_EC. We also investigate potential heterogeneity by location in these mean comparisons and find that the means for both variables are not significantly different across locations, which is consistent with a high degree of homogeneity across DFS stores. Appendix G (Parts 1 and 2) sets out in detail our testing approach and reports the corresponding results. We additionally test the null of equal variance (volatility) across the samples and we fail to reject the equal variance null across the unbalanced full sample and the balanced subsample for both SALES and NPS_EC. The finding of no statistically different variances holds when we apply the same tests of variance equality to each area (see Appendix G, Parts 1 and 2). Finally, we test the null of equality between our full sample and our subsample for the higher moments of skewness and kurtosis. We found no statistically significant differences across the two samples (see Appendix G, Part 1).

Furthermore, in both our 91-store unbalanced sample and our balanced 24-store sample, a variety of first-generation (assuming cross-section independence) and second-generation (allowing cross-section dependence) panel unit root tests rejected the unit root null for both variables of interest. Hence, there is no evidence that our variables are trended (for visual confirmation, see also the plots of SALES and NPS_EC for each store in Appendix E, Part 3).

In sum, for both variables, the subsample and complement are statistically indistinguishable in mean, variance, skewness, and kurtosis (under dependence-robust procedures). This rules out outcome-based selection on level, dispersion, asymmetry, and tail behavior — i.e., the channels through which economically meaningful differences in store performance or customer satisfaction would

[13] For example, consider two stores: Store 1 with complete NPS data (i.e., one value for each month); and Store 2 with just one missing value (say, for Month 15). In this case, for Store 1, in Month 20, Z0t is calculated (correctly) from 13 values (from t=24, the fourth month forward, to t=12, the eighth month back), X 24 + X 23 + X 22 +···+X 15 + X 14 + X 13 + X12, whereas for Store 2, Z0t is calculated (incorrectly) from 12 values, X24 + X23 + X22 +···+X16 + X14 + X13 + X12 (i.e. omitting X15).

[14] Because the subsample is nested within the full sample, valid inference requires comparing it to its complement. This comparison enables valid statistical inference regarding whether the subsample differs from the overall sample.

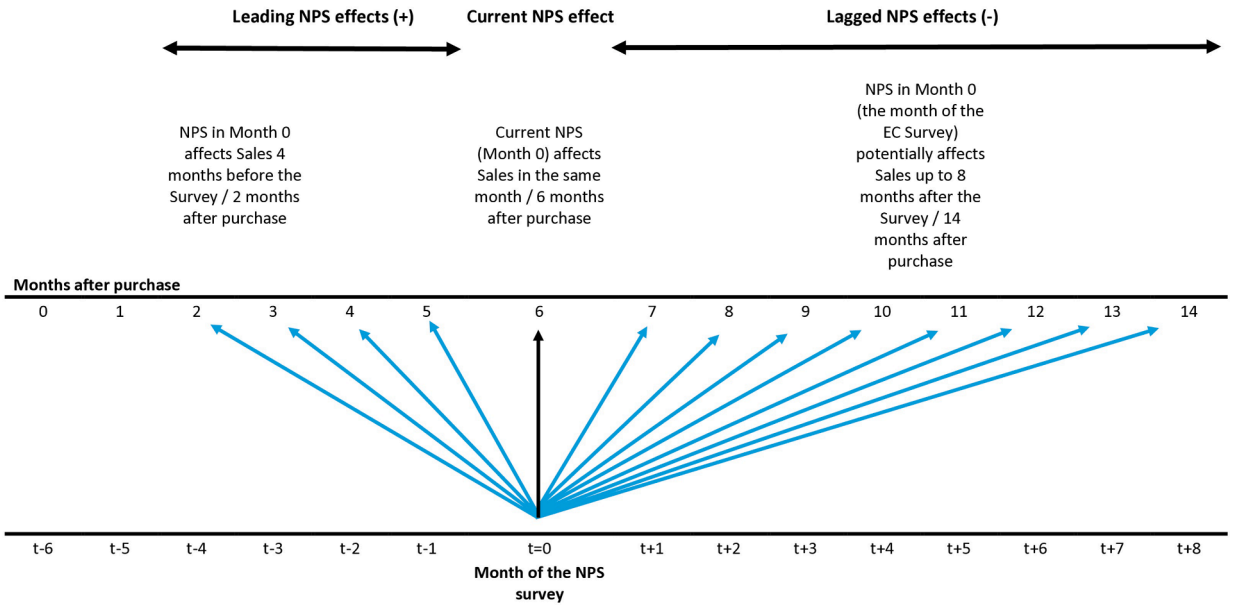


Fig. 2. NPS effects on sales over time (from 4 months before to 8 months after the EC Survey – i.e., from 2 to 14 months post purchase). Source: authors.

be expected to surface.

In the absence of many covariates, which are not available to us, no finite set of tests can *prove* representativeness. Yet, the alignment of all four moments for both variables of interest makes it unlikely that the SALES–NPS_EC relationship differs materially between the groups. Because both samples have very similar marginal distributions in both NPS_EC and SALES — not just in means, but also in dispersion and tail behavior — it becomes much less plausible that one sample has a fundamentally different functional relationship $Y = f(X)$ – where Y (sales) is a function of X (NPS) – than the other.[15] In general, when the inputs and outputs are distributed in the same way across groups, it is harder to construct a story in which the mapping from X to Y would be systematically different. In addition, when we investigate representativeness with respect to variation across DFS administrative areas – the only covariate available in our dataset – tests of both mean equality and variance equality across our samples – implemented for both SALES and NPS_EC – revealed overwhelming homogeneity (see Appendix G, Part 2).

DFS culture and policy tends towards store homogeneity (Section 3), thereby minimizing the unobservable influences under store control that potentially might differ between the samples (i.e., internal unobservables). However, we acknowledge the theoretical possibility of factors outside the control of the company (such as local competition) that could – however unlikely – systematically differ between the subsample and complement (i.e., external unobservables) and so (i) leave the unconditional distributions of SALES and NPS_EC unchanged yet (ii) shift the *conditional* response of SALES to NPS_EC. Nonetheless, discussion with a senior DFS Board member during the sample period suggested no candidates for one or more external unobservables that could have been both important and exerting differential influence on the full and reduced samples. We conclude that the available evidence suggests that our balanced sample is, in all likelihood, representative of DFS UK stores in the sample period.

6.5. The model

Consistent with our theoretical framework, we model both direct (own-store) and indirect (area) effects of NPS on sales, allowing for nonlinear dynamics over time. We specify horizons of up to 14 months after purchase for direct effects and 18 months for indirect effects. In the absence of theoretical guidance and by considering that too few months cannot identify a non-linear time path and too many wastes degrees of freedom, we determine these time horizons by drawing on estimates from the non-preferred specifications (Appendix E, Parts 1 and 2).

Each store's NPS is derived for each month from the EC survey completed by customers six months after purchase. We illustrate the potential effects of a store's NPS measured in the month of the EC survey with reference to the direct (own-store) sales effects. Fig. 2 provides a graphical depiction of potential NPS effects on sales, beginning from the month of purchase (Month 0 on the top scale) and continuing through the month of the EC Survey (Month 6) to the months following the EC Survey (Months 7–14). This allows the NPS to reflect settled judgements that may have already informed recommendations *in advance* of the EC Survey as well as having the

[15] A marginal distribution is the full, one-variable distribution of a random variable for one group at a time, *ignoring* (i.e., "marginalizing over") other variables.

potential to inform recommendations in months *following* the EC Survey. We therefore investigate the potential effects of each month's NPS on sales via:

- the "Current NPS sales effect" – i.e., in the same month as the EC Survey, which is denoted 0 at the bottom of Fig. 2;
- "Forward (i.e., leading) NPS sales effects" in each of the four months preceding the EC Survey (i.e., sales effects from the *future* NPS); and
- "Backward (i.e., lagged) NPS effects" in each of the eight months following the EC Survey (i.e., sales effects from the *past* NPS).[16]

Because delivery typically takes place a couple of months after purchase, customers immediately post purchase lack the "source expertise" and "source experience" required for credible – hence, effective – word-of-mouth recommendation (Section 4.3). Accordingly, we do not expect a discernible NPS sales effect in either the month of purchase or the first post-purchase month.

Conversely, there are two possible reasons why the four forward ("lead") values of our independent variables may significantly influence future sales. First, the time required for word-of-mouth effects to translate into purchases (Section 4) may fall within the four months preceding the EC survey. Second, lead effects may capture unobserved, time-varying store-specific factors that are not fully controlled for by store and period fixed effects (Eq. (1)). Accordingly, the individual and, more importantly, joint significance of the first four leads provides a test of both (i) the economic significance of word-of-mouth recommendation in the period between delivery and the EC survey and (ii) the influence of store-specific time-varying unobservables. Rejection of the null that these lead effects are zero indicates the presence of one or both mechanisms; failure to reject suggests that neither is present.

We estimate the effects of NPS on sales by estimating the model specified in Eq. (1):

$$\begin{aligned} Sales_{s,\ t} = {} & \widehat{\lambda}_1 Sales_{s,\ t-1} + \widehat{\gamma}_0 Z0_{s,\ t} + \widehat{\gamma}_1 Z1_{s,\ t} + \widehat{\gamma}_2 Z2_{s,\ t} + \widehat{\phi}_0 Z0_net\ area\ mean_{s,\ t} + \widehat{\phi}_1 Z1_net\ area\ mean_{s,\ t} \\ & + \widehat{\phi}_2 Z2_net\ area\ mean_{s,\ t} + \sum \widehat{Store_DV}_s + \sum \widehat{Month_DV}_t + \varepsilon_{s,t} \end{aligned} \tag{1}$$

where *s* indexes the individual stores, *t* indexes the months included in the balanced dataset, and ($\varepsilon_{s,t}$) is the usual regression error.

Sales$_{t-1}$. The first lag of the dependent variable (*t-1*) proved to be statistically significant at the five per cent level but the second did not.

Z0, Z1 and Z2. These Almon variables are designed to capture the time path of *direct* (own-store) NPS sales effects – whereby recommendations from customers at a store lead to future sales at the same store.

Z0_net area mean, Z1_net area mean, and *Z2_net area mean*. These variables are designed to capture the time path of *indirect* NPS sales effects – whereby recommendations from customers at one store lead to future sales at neighboring stores. Each store belongs to one of 11 company-defined administrative areas in the UK. Hence, for each store we controlled for the net area monthly average NPS – i.e., area monthly averages net of the store for which the mean NPS is calculated (to avoid double counting).

Store_DV. Store fixed effects (i.e., the store dummy variables) control for all time-invariant characteristics of the store; for example, location-specific effects, the sociodemographic composition of the local population and/or customers, influences from the regional/sub-regional economy etc.

Month_DV. Time period dummies (i.e., the month dummy variables) absorb shocks common to all stores in each month — e.g., seasonal influences, macro cycles, national marketing, or price inflation[17] — so unobservable effects of this type cannot confound the NPS-sales link. In addition, the inclusion of month dummies is the most flexible way to control for unobserved trend effects, should there be any.

From estimation of Eq. (1) we derive: 13 coefficients estimating the direct long-run marginal sales effects of the four lead values, the current value, and the eight lag values of stores' own NPS ($\beta_0{}^{LR} \ldots \beta_{12}{}^{LR}$); and 17 coefficients estimating the indirect long-run marginal sales effects of the four lead values, the current value, and the 12 lag values of the area NPS ($\beta_{13}{}^{LR} \ldots \beta_{29}{}^{LR}$).

7. Results and interpretation

The results of estimating the model in Eq. 1 are reported in Table 4a. The derived estimates of the long-run effects of each variable of interest are reported in Table 4b. (Store and month dummies have a control function and are not reported or discussed individually.)

The estimated coefficient on lagged sales ($\widehat{\lambda}$) is in the valid range of $0 < \widehat{\lambda} < 1$ and gives a persistence factor ($1/[1-\lambda]$) of 1.25, thereby precluding a random walk or explosive growth in sales. At the same time, it suggests that the long-run effects will be only a

[16] Leading (or forward) values of the *independent* variable(s) influence past values of the *dependent* variable, while lagged values of the *independent* variable(s) influence future values of the *dependent* variable. To fix ideas with an example, NPS values derived from the EC Survey in June are allowed to have (i) a contemporaneous sales effect, (ii) leading sales effects in February, March, April and May (in relation to which, June NPS values are, respectively, +4, ..., +1 month[s] in the future), and (iii) lagged sales effects in July, August, September, October, November, December and in January and February in the following year (in relation to which, June NPS values are, respectively, −1, ..., −8 month[s] in the past).

[17] Our initial inclination was to deflate sales according to monthly changes in the price level. However, we were persuaded otherwise by two main considerations: (i) the lack of a deflator sufficiently precisely defined to apply to our particular firm; and (ii) advice from a DFS senior board member that company prices were not influenced by inflation during our sample period, as the then low rates of inflation were of much less concern to the company than the desirability of maintaining its price points. For these reasons, we control for the potential effects of inflation alongside (and indistinguishable from) other systematic period influences (for a recent example of this practice, see Gashi et al., 2020).

Table 4a
Established Customer NPS sales effects – estimates from Bias-corrected LSDVC dynamic regression (bootstrapped SEs).
Dependent variable: Sales.
Number of observations: 480 (456 used in estimation, due to the first lag of the dependent variable).

	Coefficient	Z statistic *	P>|z| (p value)
Lagged dependent variable: Persistence			
L1.SALES	0.203	4.09	0.000
Own-store NPS effects: Almon functions (≡ *Z0, Z1*, and *Z2* in Eq. 1). Each of these is a function of four lead values, the current value, and the eight lag values of stores' own NPS.			
Z0_12_NPSEC_bal	−3656.95	−0.45	0.653
Z1_12_NPSEC_bal	3095.44	1.10	0.270
Z2_12_NPSEC_bal	−183.480	−0.84	0.399
Net area NPS effects: Almon functions (≡ *Z0_net area mean, Z1_net area mean*, and *Z2_net area mean* in Eq. 1). Each of these is a function the four lead values, the current value, and 12 lag values of the Net Area Mean NPS.			
Z0_16_NPSEC_Mean_bal	2211.31	0.22	0.827
Z1_16_NPSEC_Mean_bal	4271.19	1.64	0.102
Z2_16_NPSEC_Mean_bal	−241.28	−1.57	0.115

19 monthly dummy variables (September 2013-March 2015); except for seven, all statistically significant at the five per cent level as well as jointly significant, $p < 0.0000$); August 2013 is the omitted - baseline - category.
24 Store fixed effects.
*Computed from bootstrapped cluster-robust standard errors (clustered on store) (2000 replications – there are no noteworthy differences from the SEs computed using either 400 or 100 replications).

little larger than the short-run effects with the former arriving very quickly, within two to three months.[18]

First, we report our joint test, outlined in Section 6.5, of the economic significance of word-of-mouth recommendation in the four months between delivery and the EC survey and of the statistical influence of store-specific time-varying unobservables. The estimates of all four forward lead effects of both the own-store and the area NPS are individually (Table 4b) and jointly non-significant (own-store – $p = 0.950$; area – $p = 0.346$). These non-rejections indicate no significant presence of either the potential economic effect or potentially confounding statistical effects from store-specific time-varying unobservables. This diagnostic supports the validity of our Table 4a and 4b estimates.

Table 4b reports statistically significant estimates suggesting that a store's own NPS has a *direct* positive and economically meaningful effect on the store's own sales over a five-month period (from eight to 12 months after purchase). This effect begins in the second month after the EC Survey, increases during the third month after the survey, reaches a peak in the fourth month after the survey, and then fades during the fifth and sixth months after the survey. Fig. 3, left-hand panel, shows the sales effects resulting from a sustained unit increase in the stores' own NPS for an average store.

Because NPS is bounded by ±1 (where 1 is 100% promoters and −1 is 100% detractors), a unit increase is an increase of the NPS from e.g. 0 to 1, i.e. 0 to 100%. However, econometrically estimated effects measure the effects of marginal, i.e. small changes. Moreover, for most businesses, feasible sustained increases in NPS are likely to be small. Accordingly, on grounds of both econometric validity and business practicality, we interpret the effects of an increase in the NPS of 0.01 (i.e., an increase of 1 percentage point). Hence, for each of the long-run effects reported in Table 4b, the corresponding effect on sales in pounds sterling (£) is given by multiplying the estimated coefficient by 0.01.

In round terms, a *sustained* one percentage point improvement in NPS delivers on average – holding other influences constant – an increase in sales of £104 (i.e. $0.01 \times \beta_6^{LR}$) plus a further £113 (i.e. $0.01 \times \beta_7{}^{LR}$), £117, £117 and £112, respectively, over the five-month period. Both before and immediately after this period, the effects are small and are not estimated with sufficient precision to warrant inclusion in the calculation.[19] Nonetheless, a statistically significant total increase in sales of £564 (the sum of the unrounded values; $p = 0.024$) in every subsequent month – so long as the initial increase in NPS is sustained – represents an economically substantial effect from a small increase in NPS. (The sum across all the estimated store effects is statistically significant only at the 10 per cent level (£936: $p = 0.094$.) If a similar increase were to have been achieved by all 96 of the company's UK stores and sustained for a year, then the annual increase in sales would have amounted to around £650,000 (rounded: £564 × 12 × 96 = £649,728).

Especially noteworthy is the *indirect* sales effect on an individual store of the average NPS of the other stores in its administrative area. Here, a one percentage point improvement, ceteris paribus, were it to be sustained, would give rise to statistically significant effects in the month of the EC survey and in each of the next 10 months: respectively, in round terms, £193 (i.e. $0.01 \times \beta_{17}{}^{LR}$), £220 (i.e. $0.01 \times \beta_{18}{}^{LR}$) … and so on, yielding a statistically significant total of £ 2580 ($p = 0.006$) accumulating in a quadratic pattern with its

[18] If a sales increase of £1 took place in the previous month, then according to the estimated coefficient on the lagged dependent variable, other factors held constant, current sales (t=0) would increase by £0.20 (rounded); in the next month (t=1), the induced sales effect would be £0.20×0.20 = £0.04; in the month after that (t=2), £0.20×0.20×0.20 = £0.008; … and so on. By the third month (t=3), the cumulative induced sales effect is 0.248, only slightly less than the full effect of 0.25, which in formal mathematical terms is reached after an infinite number of periods. This suggests that the long-run effects occur sufficiently quickly to be of practical rather than purely mathematical significance.

[19] Moreover, outside of the range of the significant effects, negative estimates are to be expected – as Almon (1965): 184 noted – 'since the polynomial cannot, of course, continue at zero'.

Table 4b
Derived long-run effects of the Established Customer NPS on Sales (balanced sample: N = 24; T = 19) *.

Month relative to purchase	Month of the NPS derived from the EC Survey *relative to* the month of the sales effect (leads/lags indicated by + and – respectively)	Coefficient derived (post estimation) from estimating Eq. 1	Estimated long-run sales effects			Month relative to purchase	Month of the NPS derived from the EC Survey *relative to* the month of the sales effect (leads/lags indicated by + and – respectively)	Coefficient derived (post estimation) from estimating Eq. 1	Estimated long-run sales effects		
Stores' own NPS						**Area net average NPS**					
			Estimated Coefficient	Z statistic	p value				Estimated Coefficient	Z statistic	p value
+2	+4	β_0^{LR}	−4588	−0.45	0.653	+2	+4	β_{13}^{LR}	2774	0.22	0.827
+3	+3	β_1^{LR}	−935	−0.12	0.903	+3	+3	β_{14}^{LR}	7830	0.72	0.474
+4	+2	β_2^{LR}	2258.	0.38	0.704	+4	+2	β_{15}^{LR}	12280	1.25	0.212
+5	+1	β_3^{LR}	4991	0.97	0.334	+5	+1	β_{16}^{LR}	16125	1.72	0.085
+6	0	β_4^{LR}	7263	1.44	0.151	**+6**	**0**	β_{17}^{LR}	19365	**2.08**	**0.038**
+7	−1	β_5^{LR}	9074	1.74	0.082	**+7**	**−1**	β_{18}^{LR}	21999	**2.32**	**0.020**
+8	**−2**	β_6^{LR}	10426	**1.95**	**0.051**	**+8**	**−2**	β_{19}^{LR}	**24028**	**2.47**	**0.013**
+9	**−3**	β_7^{LR}	11317	**2.13**	**0.033**	**+9**	**−3**	β_{20}^{LR}	25451	**2.58**	**0.010**
+10	**−4**	β_8^{LR}	11747	**2.25**	**0.024**	**+10**	**−4**	β_{21}^{LR}	26269	**2.66**	**0.008**
+11	**−5**	β_9^{LR}	11718	**2.24**	**0.025**	**+11**	**−5**	β_{22}^{LR}	26482	**2.72**	**0.006**
+12	**−6**	β_{10}^{LR}	11227	**1.96**	**0.050**	**+12**	**−6**	β_{23}^{LR}	26089	**2.77**	**0.006**
+13	−7	β_{11}^{LR}	10277	1.45	0.146	**+13**	**−7**	β_{24}^{LR}	25091	**2.80**	**0.005**
+14	−8	β_{12}^{LR}	8866	0.96	0.339	**+14**	**−8**	β_{25}^{LR}	23487	**2.76**	**0.006**
						+15	**−9**	β_{26}^{LR}	21278	**2.60**	**0.009**
						+16	**−10**	β_{27}^{LR}	18464	**2.23**	**0.026**
						+17	−11	β_{28}^{LR}	15044	1.66	0.096
						+18	−12	β_{29}^{LR}	11019	1.04	0.297

*Bold indicates statistical significance at the 5% level (or more precise).

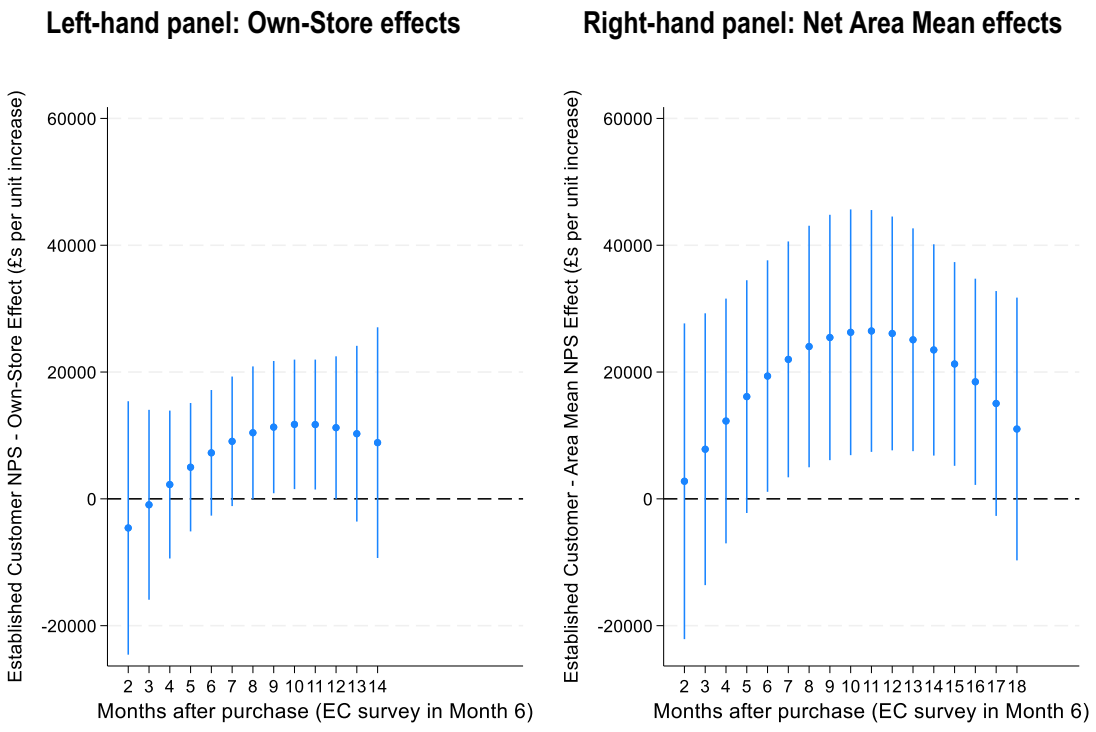


Fig. 3. The sales effects of a sustained one unit increase in the established customer NPS (£s) (95% confidence intervals)*.
* As explained in the text, the sales effect in £s of a marginal – and more realistic – one percentage point increase (0.01) is obtained by multiplying each estimate by 0.01. Point estimates (shown by the dots) with confidence intervals (shown by the bars) that overlap the dotted line at zero are not statistically significant at the conventional 5% level. Produced using Stata's "matrix input" command and the user-written program coefplot.

turning point five months after the EC Survey. Projected across 96 stores for a year this would have amounted to an annual sales increase of almost £3 million (£2580.02 × 12 × 96 = £2,972,181), which is just over 4.5 times the store-level increase. This is consistent with the potentially larger number of word-of-mouth recommendations from the customers of several neighboring stores compared to the smaller number of word-of-mouth recommendations from its own customers.

By aggregating the direct and indirect sales effects, we conclude that if every store could have achieved a sustained increase in its NPS of one percentage point then, across 96 stores, the additional annual sales over the long run would have been in the region of £3.5 million (£649,728 + £2,972,181 = £3,621,909) which, in the Financial Year 2016, would have been an addition to Gross Sales of 0.36 per cent (DFS, 2016: 3).

Fig. 3 displays graphically the non-linear (quadratic) sales effects of both the store's own NPS (left-hand panel) and the area NPS (right-hand panel) as they unfold over time, at first strengthening and then declining. Each dot depicts the respective (long-run) "Estimated Coefficient" reported in Table 4b, while the vertical bars depict the associated 95 per cent confidence intervals (such that the shorter the bar the more precise the estimate). Estimates with bars overlapping the zero reference line indicate an estimate that cannot be statistically distinguished from zero at the 5% level; for example, in the right-hand panel of Table 4b, the estimate for five months after purchase is not significant at the 5% level but is significant at the 10% level.

7.1. Robustness checks

We subject our findings to two robustness checks.

1. To check whether the indirect sales effects were uniquely associated with the 100-mile maximum distance between DFS stores in the same DFS administrative area, we redefined the net area effect as the mean NPS of all other stores within 70 miles in the same DFS area.
2. We change the dependent variable to estimate the effect on percentage changes in sales (instead of sales measured in levels) of absolute changes in the NPS (as in Baehre et al., 2022).

We report these robustness checks in detail in Appendix I (Part 1).

In both cases, both the direct and indirect NPS sales effects are qualitatively the same as reported for our preferred model – unfolding along a similarly non-linear time path – and quantitatively similar. Robustness Check 2 uses a balanced dataset with 26

stores – i.e. including the two stores omitted to obtain our balanced dataset of 24 stores, which uniformly reproduces the statistical characteristics of the full unbalanced dataset. Accordingly, Robustness Check 2 demonstrates that our findings are not greatly affected by small changes in the estimation sample and its statistical characteristics. In addition, this Robustness Check demonstrates that (i) our findings are not an artefact of model specification and (ii) the Baehre et al. (2022) preferred specification does not have the exclusive claim to validity that they suggest (a point we mention in Section 8.5, Implications for Future Research).

In addition, we estimate the model specified in Eq. 1 above on data from the Post Delivery Survey, which focusses on customers' experiences with delivery. Strictly, this is not a robustness check, because – as we explain in Appendix I (Part 2) – Post Delivery responses to the NPS question are not best suited to explore the NPS-Sales relationship. Nonetheless, two related conclusions for future research are suggested by comparison of the NPS sales effects from the EC dataset and the Post Delivery (PD) dataset (Appendix I, Part 2): (i) as expected, the touch point at which the NPS is measured can matter greatly (post-delivery, only the store-level effects are statistically significant and these decline linearly from the time of the survey, reflecting the one-off experience of delivery); and (ii) the near linear pattern of own-store NPS effects in the PD data – suggesting that the delivery experience steadily loses salience in relation to experience of the product – demonstrates that while our methodology is designed to reveal a quadratic pattern of estimates should such a pattern be present in the data, it does not impose such a pattern should such a pattern not be present in the data.

8. Discussion and conclusions

In this Section, we: (i) address the research questions set out in the Introduction; (ii) discuss word-of-mouth transmission as the conceptual link between the NPS and sales; and (iii) highlight our methodological contribution. In addition, we discuss further implications of our findings for (iv) managerial engagement with the NPS and (v) future research into NPS sales effects.

8.1. Key findings: addressing the research questions

We investigate a single large retail company using store-level data at monthly frequency over a period of four years. In line with our research questions in the Introduction, we provide evidence that:

(1) the NPS can predict future changes in sales;
(2) NPS sales effects extend beyond the store where the recommendation originates (i.e., a store's NPS impacts the sales of the company's neighboring stores) with these effects being more than four times larger than the own store's NPS impact, and
(3) both own-store and area NPS sales effects (i) take place over a period of many months (for the direct own-store's sales from eight to twelve months after purchase, and for the indirect area effects from six to sixteen months after purchase) and (ii) unfold over a non-linear (inverted-U) time path (for the direct own-store's sales peaking ten months after purchase, and for the indirect area effects peaking eleven months after purchase).

The evidence for these three findings is summarized and visually displayed by Fig. 3.

(4) The validity of these findings was confirmed by two robustness checks: (i) estimating the same model with a modified area variable; and (ii) estimating a different model specification on a slightly larger balanced dataset (26 stores rather than 24).

There are disadvantages as well as advantages to investigating a single case. The disadvantage is that it may compromise external validity (generalizability). For example, because the company accounts for two-thirds of the market of interest, any relationship between NPS and sales that we uncover might not reflect the same relationship for smaller companies, even those in the same industry. However, while we as researchers cannot generalize our findings from our quantitative study of one firm, readers can decide for themselves the extent to which – if any – the characteristics and results of this case study may provide usable insights in other contexts. For example, similarly long replacement cycles together with a high product price supported transfer of the NPS system of measurement and monitoring developed at DFS to a large retail company in the auto sector (Stephenson, 2020).

Finally, we demonstrated that a modeling strategy producing valid estimates in the presence of (i) sales persistence, (ii) a prolonged time path of NPS sales effects, and (iii) inherent multicollinearity, may require substantial reductions in the number of observations available for estimation (the modeling requirements of the present study and the associated reductions in the estimation sample are summarized in Tables 2 and 3). Consequently, we had to consider potential selection bias arising from the consequent reduction in the number of stores in the estimation sample from 91 to 24. To this end, we first note the institutional evidence that DFS policy is to make its stores homogeneous (Section 3). Second, consistent with the institutional evidence, is statistical evidence that the NPS displays relatively minimal between-store variation (Section 6.2). Third, further statistical evidence suggests homogeneity of DFS stores across the 11 DFS administrative areas (Section 6.4 and Appendix G, Part 2). And, fourth, we provide comprehensive evidence (Section 6.4 and Appendix G) that there are no significant differences between the statistical characteristics of the 91-store unbalanced sample and the reduced 24-store balanced sample.

8.2. Conceptual implications: word-of-mouth as the mechanism linking the NPS and sales to new customers

The present study makes a conceptual contribution by focusing on the word-of-mouth influence on sales to *new* customers (Section 4). Here, we relate this transmission channel to an issue highlighted by Bendle et al. (2019: 223); i.e., 'whether the relationship between NPS and (revenue) growth is causal or purely correlational seems to be under discussed'. In the case of DFS, our empirical analysis provides direct evidence that NPS changes can predict sales changes. However, the case for this being a causal relationship

depends on the credibility of customer word-of-mouth as the mechanism linking the NPS and subsequent sales. To help make this case, we consider three arguments, each directly or indirectly suggesting that word of mouth is the mechanism linking the NPS and subsequent sales.

1. Because the long replacement cycle of upholstered furniture (c. 7 years) goes beyond the span of our data (four years), we infer that, in this study, it is the acquisition of new customers that will drive sales rather than customer retention. This inference is confirmed by the very low rate of repeat purchase in our data (Sections 2.2 and 4.1 and Appendix A). This is consistent with the role of word-of mouth recommendation.
2. We provide direct evidence from customer survey responses that customers' self-reported propensity to recommend, measured by the NPS, is the antecedent of both (i) actual word-of-mouth recommendations (i.e., that there is a strong association between intended and actual recommendations) and (ii) the subsequent acquisition of new customers (i.e., at least some actual recommendations are converted into effective recommendations revealed in sales) (Section 4.1 and Appendix B).
3. Finally, we theorize about word-of-mouth influence to support hypotheses that inform our empirical model and are supported by our findings:
   a. We draw upon social capital theory and evidence on the transmission of word-of-mouth influence over distance to theorize cross-store spillovers, such that a store's NPS will affect both its own future sales and the future sales of neighboring stores (Section 4.2).
   b. We theorize (Section 4.3) that word-of-mouth recommendations will give rise to NPS sales effects along an extended and nonlinear time path.

The consistency between these two hypotheses and our empirical findings (Fig. 3) provides additional support for word-of-mouth recommendations as a mediating link between the NPS and future sales. Although the actual word-of-mouth recommendations are unobserved, their existence can be inferred from their effects.

Together, these arguments suggest a plausible if not yet definitive case for a causal relationship between the NPS and subsequent sales that, in the context of the present study, is brought about via the mechanism of customers' word-of-mouth recommendations. This observation has the virtue of refocusing attention from the NPS and its computation to the underlying corporate culture and corresponding policies and behaviors that drive scores, and thus on how DFS acts on NPS information – e.g., via investment policy, remuneration policy, and employee awards – to increase customer recommendations and subsequent sales. (Stephenson, 2020 analyses in detail how DFS uses the NPS to shape its corporate culture and policies.)

8.3. Methodological contribution

To estimate NPS sales effects over a prolonged time path, our model was specified with multiple leading and lagged values of the NPS values. Unfortunately, this caused severe multicollinearity, which invalidated point estimates (Section 6.3 and Appendix E, Part 2). Accordingly, our methodological contribution is to combine the long-established Almon lag approach to addressing multicollinearity with more recent innovations in dynamic panel modeling, which enables valid point estimation of both short- and long-run sales effects (Section 6.3). In our case, this approach reveals that NPS sales effects unfold over a prolonged period according to a nonlinear (inverted-U) time path.

This methodological solution to the problem of multicollinearity comes at a cost. Researchers should be aware that the Almon lag approach requires a balanced data set, irrespective of the estimator (Section 6.3). Of course, if researchers work with a complete and balanced "wide-N" and "Long-T" dataset, there is no problem. However, in our case, we had to reduce our full unbalanced dataset to a smaller balanced dataset constructed to replicate the statistical characteristics of the full unbalanced dataset (both overall and by DFS administrative regions). In doing so, we constructed statistical tests to check the equality of means, variance, skewness, and kurtosis, and, in addition, stationarity across the two samples. This ensured that we addressed concerns of potential selection bias (Section 6.4 and Appendix G).

8.4. Managerial implications

Our study offers practical implications for managers by showing that improvements in NPS must be company-wide, sustained, and adequately resourced. Of course, the evidence for an NPS–sales link is most compelling in contexts similar to ours: namely, dominant companies retailing relatively high-priced consumer durables and long interpurchase cycles, where growth depends on attracting new customers from among the colleagues, friends and families of established customers. While the NPS system developed at DFS has been adopted by at least one other major UK retailer (Section 8.1), its relevance will vary across contexts, requiring managers to assess its applicability to their specific circumstances.

First, this study provides a counterpoint to research suggesting no link between customer-based NPS and sales. It indicates that managers using, or considering, NPS systems based on current customers are not necessarily acting against scholarly evidence. This may inform boardroom decisions, particularly in dominant retail firms, where market position may narrow the gap between *customer-based* NPS and broader brand health measures that capture all *potential* customers. Although 'capturing NPS for both current customers and non-customers is now common managerial practice' (Baehre et al., 2022: 71), where a retail company is dominant in its market, managers may consider whether the additional benefit of sampling beyond their established customers justifies the additional cost.

Second, the NPS must be introduced as part of a company-wide, strategic initiative (Bendle et al., 2019; Stephenson, 2020). This is

supported by the finding that the NPSs of nearby stores owned by the company have a very large positive effect on a store's sales revenue.

Third, managers in contexts like those of DFS managers should not expect NPS sales effects to emerge quickly. We find that both own-store and nearby-store effects unfold over several months, following a quadratic pattern, i.e., initially strengthening before declining. Accounting for this time path is essential to identifying NPS effects and estimating their financial impact.

Fourth, generating substantial financial returns from NPS requires sustained improvement, which in turn demands commitment across the organization. When maintained over time, cumulative NPS effects, including both own-store and area spillovers, can be financially substantial. We calculate that a sustained one percentage point increase in NPS across DFS UK stores would yield approximately £3.5 million in additional annual sales (around 0.36% of 2016 revenue of £980.4 million). Considering the firm's profit margin for 2016 (6.58% of total sales) suggests an increase of roughly £238,000 in profit. A three-point increase in NPS could therefore generate profits comparable to the profit of the typical store in our sample period, but without incurring the capital expenditure to open a new store (approximately £1 m, on average, during our sample period; DFS, 2015: 97), making NPS improvement an attractive strategy for management.

Fifth, a sustained company-wide NPS initiative requires significant resources. Firms must invest not only in systems to collect, analyze, and act on customer feedback, but also in robust evaluation. As this study shows, assessing the NPS–sales relationship is highly data-intensive. Reliable estimation requires long time-series data (e.g., several years of monthly store-level observations) and, ideally, a complete ("square") panel dataset. Where data are incomplete, analysts must carefully assess and report potential selection bias arising from reduced samples.

8.5. Implications for future research

Our study reveals several implications for future research. First, our results show that the estimated relationship between NPS and store sales changes markedly as the estimation strategy is refined (Appendix E). Moving from simple correlations to dynamic panel models with Almon lags to address multicollinearity, the estimated effects shift from strongly negative to strongly positive. This supports the view of Baehre et al. (2022) that divergent findings in prior research stem from differences in methodology. However, we also highlight that valid estimation often comes at the cost of substantial sample reduction.

Second, regarding specification of the NPS variable, Baehre et al. (2022) argue that NPS should be measured as changes over time. Our findings do not support this as a general rule: results from our preferred specification, where we measure NPS in levels, are similar to those obtained using their specification. This suggests that alternative specifications of the NPS variable may be equally valid.

Third, while Baehre et al. (2022: 72) conclude that the 'NPS works best in industries … where customers have short interpurchase cycles', we identify significant NPS effects in a setting with a long (approximately seven-year) interpurchase cycle. This indicates that the relationship between NPS and sales is not governed by a short-versus-long interpurchase cycle distinction.

Finally, Baehre et al. (2022) distinguish between two uses of NPS: customer loyalty (based on current customers) and brand health (based on all potential customers), concluding that only the latter has predictive value. At first glance, our customer-based approach appears to challenge this. However, the firm's market dominance and high store homogeneity mean that store-level NPS largely reflects brand-level effects, bringing our measure closer to brand health. As our evidence is based on a single leading company, we do not definitively confirm or reject their conclusion. Instead, we call for further micro-level studies across different market contexts.

Further store-level studies using longitudinal data will contribute to understanding how NPS sales effects vary across different contexts: e.g., short vs long interpurchase cycles; low- vs high-priced consumer products; companies with different degrees of market power; and different industries and countries. The wider the range of contexts studied, the wider the range of managers served with knowledge to inform NPS-related investment decisions. Future research on NPS sales effects would be also strengthened by direct evidence on WoM diffusion of recommendations through social networks. Finally, we recommend that future longitudinal studies allow for the spatial and temporal effects investigated in the present study.

Disclaimer

The authors have no conflicts of interest to report. Company data was provided unconditionally for research purposes. No external funding was either sought or received for this study.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Open AI ChatGPT and the Microsoft Copilot services to generate the graphics in Appendices C and D (Part 2) and to conduct the tests reported in Appendix G. After using these services, the authors reviewed and edited the content as needed and thus take full responsibility for the content of this publication.

Acknowledgements

First, we thank DFS for unconditional access to their customer relations and sale data. Next, we thank participants at the 2018 British Academy of Management Conference for comments on the early findings of this project, and acknowledge a particular debt to Moira Clark and Avi Shankar, who encouraged us to carry out further research. The penultimate draft was read by Fred Reichheld. Finally, comments and suggestions from two anonymous reviewers and an anonymous Associate Editor led to many improvements.

Remaining shortcomings are the sole responsibility of the authors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jretai.2026.09.004.

References

Abel, P. (2017). By the numbers: Net promoter Score (NPS) is not the only number to rely on. Quirk's Insights Media, March 1, 2017, accessed 03-01-2025: By the numbers: Net promoter score (NPS) is not the only number to rely on | articles.

Agag, G., Baseer, A. D., Yasser, S., Majed, A., Hawazen, A., Sherif, E. H., Ahmed, H., & Ziad, H. (2023). Understanding the link between customer feedback metrics and firm performance. *Journal of Retailing and Consumer Services, 73*, Article 103301. https://doi.org/10.1016/j.jretconser.2023.103301. https://www.sciencedirect.com/science/article/pii/S0969698923000486

Agag, G., Durrani, B. A., Abdelmoety, Z. H., Daher, M. M., & Eid, R. (2024). Understanding the link between net promoter score and e-WOM behaviour on social media: The role of national culture. *Journal of Business Research, 170*, Article 114303.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society, 33*(1), 178. https://doi.org/10.2307/1911894

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies, 58*(2), 277. https://doi.org/10.2307/2297968

Baehre, S. (2024). From research to action: Enhancing net promoter score utilization in managerial practice. *International Journal of Market Research, 66*(2–3), 174–181.

Baehre, S., O'Dwyer, M., O'Malley, L., & Lee, N. (2022). The use of Net Promoter Score (NPS) to predict sales growth: Insights from an empirical investigation. *Journal of the Academy of Marketing Science, 50*, 67–84. The use of Net Promoter Score (NPS) to predict sales growth: Insights from an empirical investigation | SpringerLink.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science, 15*(5), 215–227. http://www.jstor.org/stable/2628128.

Bendle, N., Bagga, C., & Nastasoiu, A. (2019). Forging a stronger academic-practitioner partnership – The case of net promoter score (NPS). *Journal of Marketing Theory and Practice, 27*(2), 210–226.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*(1), 115–143.

Brandt, D. R. (2007). For good measure. *Marketing Management, 16*(1), 20–25.

Bruno, G. (2005). Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. *The Stata Journal, 5*(4), 473–500. https://doi.org/10.1177/1536867X0500500401

Danziger, P. (2017). Word-of-mouth marketing: Most used, most misunderstood by retailers. *Forbes Daily*. Dec.11, 2017. Accessed 02-01-2025: Word-Of-Mouth Marketing: Most Used, Most Misunderstood By Retailers.

de Haan, E., Verhoef, P., & Wiesel, T. (2015). The predictive ability of different customer feedback metrics for retention. *International Journal of Research in Marketing, 32*(2), 195–206. https://doi.org/10.1016/j.ijresmar.2015.02.004

DFS (2015). IPO Prospectus. [online] available at http://www.DFScorporate.co.uk/investors/ipo.

DFS (2016). Full Year Results, *2016* (6 October).

DFS (2018). DFS at a Glance. [online] Available at www.DFScorporate.co.uk/who-we-are/DFS-at-a-glance.

DFS Internal Documents: A1 (2011); A29 (2011); A35 (2011), A37 (2011), A39 (2011). Unpublished internal reports for the DFS board.

Gashi, P., Hashi, I., & Pugh, G. (2020). Privatization by auction: Determinants of asset prices in Kosovo. *Eastern European Economics, 58*(4), 327–359. https://doi.org/10.1080/00128775.2020.1754858

Granovetter, M. S. (1973). The Strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380.

Gujarati, D. (1988). *Basic econometrics*. London: McGraw-Hill.

Hendry, D. (1995). *Dynamic econometrics*. Oxford: Oxford University Press.

Keiningham, T., Aksoy, L., & Cooil, B. (2008). A holistic examination of net promoter. *Journal of Database Marketing & Customer Strategy Management, 15*, 79–90. https://doi.org/10.1057/dbm.2008.4

Keiningham, T., Cooil, B., Andreassen, T., & Aksoy, L. (2007). A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing, 71*(3), 39–51. https://doi.org/10.1509/jmkg.71.3.039

Lacohee, H., Souchon, A., Dickenson, P., Krug, L., & Saffre, F. (2024). The Net promoter score interrogated through a services marketing lens: Review and recommendations for service organizations. *International Journal of Market Research, 66*(2–3), 241–260.

Martin, W. C., & Lueg, J. E. (2013). Modeling word-of-mouth usage. *Journal of Business Research, 66*(7), 801–808. https://doi.org/10.1016/j.jbusres.2011.06.004

Meyners, J., Barrot, C., Becker, J., & Goldenberg, J. (2017). The role of mere closeness: How geographic proximity affects social influence. *Journal of Marketing, 81*, 49–66 (September) 49-66. Prepublication version retrieved from ResearchGate (19-02-2025): *The_Role_of_Mere_Closeness_final (1) (1).pdf.

Morgan, N. A., & Rego, L. L. (2006). The value of different customer satisfaction and loyalty metrics in predicting business performance. *Marketing Science, 25*(5), 426–439.

Nickell, S. J. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society, 49*(6), 1417.

Pesaran, M. H., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics, 68*(1), 79–113.

Pingitore, G., Morgan, N. A., Rego, L. L., Gigliotti, A., & Meyers, J. (2007). *The Single-Question Trap. Marketing Research, 19*(2), 8–13.

Reichheld, F. (2003). The one number you need to grow. *Harvard Business Review, 81*(12), 46–54.

Reichheld, F., Darnell, D., & Burns, M. (2021). Net Promoter 3.0. *Harvard Business Review*.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal, StataCorp LLC, 9*(1), 86–136.

Spanos, A. (1986). *Statistical foundations of econometric modeling*. Cambridge: Cambridge University Press.

Stephenson, A. (2020). Using the net promoter system methodology to deliver cultural change in retail organisations: Impacting both the customer and employee experience. Doctoral thesis, Staffordshire University. Using the net promoter system methodology to deliver cultural change in retail organisations: Impacting both the customer and employee experience - STORE - Staffordshire Online Repository (staffs.ac.uk).

Van Doorn, J., Leeflang, P., & Tijs, M. (2013). Satisfaction as a predictor of future performance: A replication. *International Journal of Research in Marketing, 30*(3), 314–318. https://doi.org/10.1016/j.ijresmar.2013.04.002