Profiting from the Decoy Effect: A Case Study of the Online Diamond Marketplace

Chunhua Wu and Koray Cosguner*

Update: October 2018
First Version: November 2014

Abstract

The decoy effect (DE), first introduced by Huber et al. (1982), has been robustly documented across dozens of product categories and choice settings using lab experiments. However, in the literature, the DE has never been verified in a real marketplace. In this paper, we empirically test and quantify the DE in a major online diamond marketplace. We develop a diamond-level proportional hazard framework by jointly modeling market-level decoy–dominant detection probabilities and the boost in sales upon detection of dominants. Results suggest that decoy–dominant detection probabilities are low (10%-29%) in the diamond marketplace; however, upon detection, the DE increases dominant diamonds' sales hazards significantly (2.3-4.4 times). To understand the DE's managerial significance, we quantify its profit impact and find that it contributes 21.4% of the diamond retailer's profit. Finally, we explore various strategies that might help the retailer to further increase profitability. We find that the retailer's profit can increase up to 5.4% via effective utilization of the DE.

Keywords: Decoy Effect, Attraction Effect, Asymmetric Dominance Effect, Context Dependent Choice, Proportional Hazard Model, Diamond Pricing

*Chunhua Wu (chunhua.wu@sauder.ubc.ca) is Assistant Professor of Marketing and Behavioral Science at the University of British Columbia. Koray Cosguner (kcosgun@iu.edu) is Assistant Professor of Marketing at Indiana University. The authors thank seminar participants at the University of British Columbia, University of Alberta, University of Colorado at Boulder, Indiana University, University of Pittsburgh, Dartmouth College, the 2014 Marketing Science Conference, the 2017 UW & UBC Annual Marketing Conference, and participants from the University of Texas at Dallas, Case Western University, and University of Southern California at the 2017 Summer AMA Conference in San Francisco for their valuable feedback.

1 Introduction

The decoy effect (Huber et al., 1982), also called attraction or asymmetric dominance effect, refers to the phenomenon of consumers having different preferences for existing choice alternatives with and without dominated (i.e., decoy) options in their choice sets. By design, these decoys are inferior to some, but not to all, existing choice options. When such decoys exist, all else equal, dominant options' choice likelihoods and choice shares get greater compared to cases when the decoys are not present. Since its introduction, the decoy effect (DE henceforth) has become one of the most popular and frequently cited context effects in the consumer behavior literature, and it has been thoroughly examined across dozens of product categories and choice domains using lab experiments (see, for example, Huber et al., 1982; Huber and Puto, 1983; Wedell, 1991; Lehmann and Pan, 1994; Royle et al., 1999).

Despite its high popularity, the DE's practical validity has been severely challenged recently by a series of unsuccessful replication attempts that shed light on the limits and boundaries of the effect. Frederick et al. (2014) showed that the DE can be observed only in very stylized settings, such as the presentation of two products with two numerically depicted attributes. Yang and Lynn (2014) provided additional support to these findings and questioned whether the DE has any practical significance, or if it is just an experimental artifact. The lack of documentation on the practice of the DE in today's marketplace was also noted by Huber et al. (2014); and this has further put the practical validity and significance of the DE into question. In this paper, in response to these recent studies, we provide strong empirical evidence that not only validates the DE in a real product marketplace, but shows the managerial significance of the effect through quantifying its substantive profitability impact.

Even though it has been more than three decades since the DE was introduced, to the best of our knowledge there has been no empirical study similar to ours that tests and quantifies the DE with real world practices. To empirically test and quantify the DE in a real marketplace, multiple aspects (challenges) need to be considered (resolved). First, to study the DE in a real product market, a researcher needs to *calibrate decoy–dominant relationships* among existing product alternatives. Since products typically have horizontal attributes—such as brand, taste, content, package size, and even package design—and consumers have heterogeneous preferences among these attributes, decoys to some consumers may not be decoys to others. Because of this, in most product markets, strict decoy–dominant relationships may not exist, let alone permit calibration. Second, the decoy–dominant relationships should be *perceived by consumers.* Unlike in lab experiments where alternatives with only two or three attributes are presented, choice scenarios in real life are far more complex: In a typical market, products have a much larger number of attributes, e.g., brand, size, design, color, weight, packaging, content, taste and price, to name a few. Thus, it is much harder for consumers to detect decoy–dominant relationships even if decoys/dominants exist, so that, as noted in Huber et al. (1982), "the effect may be lessened." Along this line, Simonson (2014) emphasized the importance of consumers discovering or detecting decoy–dominant relationships among existing choice alternatives in validating the DE, and called for a systematic study separating decoy–dominant detection from the DE (i.e., sales boost in dominants upon their decoys being detected). Similarly, Huber et al. (2014) emphasized the lack of decoy–dominant detection as one of the mitigating factors of the DE. Therefore, in real product markets, the detection of decoy–dominant relationships becomes a critical pre-condition for the DE to have an impact on alternatives' (namely decoys, dominants, or neither) sales likelihoods. For this reason, the choice decision must be salient enough and require some cognitive processing so that consumers' preferences can be constructed rather than revealed (Huber et al., 2014). For example, for trivial decisions, consumers may just make their choices without paying much attention to the alternatives; consequently, they may not be able to detect existing decoys/dominants. Similarly, for repeat-purchase products, added decoys may not significantly impact the choices of consumers who have already developed clear preferences towards existing alternatives over time. For example, it is hard to expect a consumer to switch to a less preferred alternative from a long-term loyal product just because a decoy to this less preferred option is introduced. Third, it is quite possible that decoy pricing strategies may not generate a positive profit impact for a firm, which limits the existence of the *decoy pricing practice* in the real world.[1]

Due to the above-mentioned challenges, to empirically test and quantify the DE, we need data

[1]We use the term decoy pricing to refer to the practice of introducing decoy–dominant relationships through pricing, i.e., charging higher prices for products with the same or inferior physical characteristics.

from a product category 1) with a reasonably small number of vertical product attributes that are uniformly preferred by consumers; 2) that is important to consumers but not repeatedly purchased; and 3) that has the decoy pricing practice. The online diamond marketplace is a highly appropriate case for this purpose: Diamonds are commodity-type and unbranded products with quality clearly defined on a few vertical attributes such as color, cut and clarity; diamond purchases are important but not repeated lifetime decisions for consumers; and, finally, we frequently observe decoy pricing patterns in the online diamond marketplace.

We use diamond price and sales data from a major online jewelry retailer in the U.S. to empirically test the DE's existence in the field, quantify its magnitude, and show its real significance regarding the profitability. Modeling the impact of decoy diamonds on sales of their dominants in the marketplace requires us to separate the market-level decoy–dominant detection from the sales boost once diamonds are detected as dominants. This is especially important in a real market context because, with numerous alternatives available, consumers may search and construct consideration sets with a limited number of alternatives that have no or few decoy–dominant relationships even though such relationships exist within a broader set of options in the marketplace. To achieve that, we develop a diamond-level proportional hazard framework incorporating two critical components: *market-level decoy/dominant detection probability* and *dominant boost hazard* upon dominant detection. The *detection probability* component captures the market-level probabilities of detecting decoys/dominants, and the *dominant boost hazard* component captures the boost in sales hazard once a diamond is detected as a dominant. Thus, in our setting, upon the dominant detection, the *dominant boost hazard* component serves as the test of the existence of the DE.

In the estimation, we use a diamond's characteristics and price, daily demand fluctuations, and competition from other similar diamonds to control for the differences in the daily sale hazards among diamonds. To capture potential consumer heterogeneity in response to decoy pricing, we classify the diamonds into three segments based on price (low: $2K-$5K; medium: $5K-$10K; and high: $10K-$20K) and estimate separate parameters for detection probability and dominant boost hazard. Results on daily sales hazard suggest that it is relatively easier to sell diamonds with smaller carats and moderate price, cut, color and clarity levels. Results also help us quantify how the relevant decoy–dominant structure impacts the detection probability and dominant boost hazard. The market-level detection probability for a decoy (or dominant) diamond is, in general, quite low, ranging from 10% to 29% based on the diamond's price segment. The low decoy–dominant detection probabilities show that it is a requirement to explicitly model the decoy–dominant detection probabilities in the online marketplace when quantifying the DE. As opposed to low detection probabilities, we find a significant sales boost for dominant diamonds: Conditional on detection, the sales hazard becomes 2.3 to 4.4 times larger for a dominant diamond. This finding offers strong real-life evidence of the practical validity of the DE in response to recent studies questioning this aspect, including Frederick et al. (2014) and Yang and Lynn (2014). Across the three segments, the low price segment has the highest detection probability but lowest dominant boost effect, while the high price segment demonstrates the opposite.

We further quantify the profit impact of the DE, i.e., the managerial significance of the DE. The direct measurement of the DE in terms of profit comes from our model estimates of the dominant boost hazard component. We compare differences in the expected profits when we turn the dominant boost parameters on and off. Overall, we find that the DE contributes 21.4% of the retailer's gross profit; and the profit contribution is quite even across different diamond price segments. This finding shows that the DE is not only real but quite substantive managerially. Finally, through our policy studies, we investigate how decoys and dominants can be used/priced in more effective ways to further increase the retailer's gross profit. We manipulate three strategies the retailer can implement: 1) changing the number of listed decoys/dominants (called as frequency strategy); 2) changing the price dispersion for diamonds with the same attributes (called as range strategy); and 3) changing the baseline decoy–dominant detection probabilities (called as awareness strategy). Our simulation studies indicate that the retailer could gain additional profits from all three strategies. The potential gain is the largest from the awareness strategy: At the optimal consumer-awareness levels, the retailer might acquire an additional 5.4% gross profit compared to the current setting.

Our study contributes to the literature on the DE by empirically separating the market-level decoy–dominant detection from the DE boost of decoys on dominants. More importantly, for the first time in the literature, we 1) validate the existence of the DE in a real marketplace; 2) quantify its magnitude across different segments; and 3) show its substantive profit impact. This paper thus attenuates the recent concerns (Frederick et al., 2014; Yang and Lynn, 2014) about the practical validity of this classical context effect beyond traditional lab settings. Finally, through our policy studies, we advise the retailer how to acquire higher profits from the DE via easy-to-implement strategies.

2 Literature Review

This paper contributes to two streams of literature: the general consumer behavior literature on context dependent choices (in particular, the decoy effect) and the empirical consumer choice modeling literature in marketing and economics.

Standard rational choice models in economics and marketing are built upon the revealed preference assumption, which implicitly assumes two principles: the principle of regularity (Luce, 1977) and the principle of independence of irrelevant alternatives (IIA)(Luce, 1959). In contrast, consumer behavior researchers adopted the notion of constructed preference (Bettman et al., 1998) and extensively documented context effects in consumer choices (Tversky, 1972; Simonson, 1989). The DE (Huber et al., 1982), which is a classic example of such context effects, violates both IIA and regularity principles. The DE has been examined across dozens of product categories and choice domains (see, for example, Huber et al., 1982; Huber and Puto, 1983; Wedell, 1991; Lehmann and Pan, 1994; Royle et al., 1999). Further, the literature features investigations of cognitive processes and mechanisms underlying as well as moderating the DE and related context effects (see, for example, Ratneshwar et al., 1987; Heath et al., 1995; Müller et al., 2014; Guo and Wang, 2016). In this domain, Khan et al. (2011) studied the influence of how choice construal on context effects, finding that high construal as opposed to low increases the size of the DE. In a recent paper, Morewedge et al. (2018) demonstrated that when comparisons of alternatives for choice makers require social comparisons, the context effects get stronger. Guo and Wang (2016) studied underlying causes of context effects. Specifically, they found that the response time can mediate the compromise effect but the context information can not. Although the DE was introduced more than three decades ago, empirical testing and quantification of the effect in a real marketplace has not been achieved yet. Our paper fills this important gap in the literature by validating the DE in the field.

The limits and boundaries of the DE have been debated by multiple studies.Frederick et al. (2014) stated that the DE can only be observed in very stylized lab settings with 2×2 numerical depictions of the products (two products with two attributes, with a decoy to one product added to the choice set later). Through 38 replication attempts, their study showed that when the product attributes are depicted with perceptual representations and verbal descriptions (rather than numerical), the DE weakens, dies, or gives way to the repulsion effect (i.e., the decoy option decreases the share of the dominant option). Through 91 replication attempts,Yang and Lynn (2014) also showed that replicating the DE is very difficult with verbal and pictorial depictions of product attributes. They went further and questioned the usefulness of marketing academics by accusing the studies of placing insufficient emphasis on the practical significance of the examined behavioral concepts. With the current research, we respond to concerns of Frederick et al. (2014) and Yang and Lynn (2014) by providing strong empirical evidence of the existence of the DE in a real product marketplace.

Simonson (2014) underlined the importance of recognizing the set formation, i.e., subjects being aware of decoy–dominant relationships, in being able to replicate the DE. He argued that consumers' choices require them to make multiple trade-off contrasts simultaneously. As a result, consumers may not discover or make their decisions based on existing decoy–dominant configurations, especially if such configurations are difficult to detect. Further, Simonson (2014) called for a systematic study on the drivers of decoy–dominant detection. Huber et al. (2014) recognized the lack of practice of the DE in today's marketplace, noting that it is very difficult to observe the DE in a real market place since detection of decoys is typically very hard for consumers due to numerous alternatives with many attributes. With this research, we respond to the call that Simonson (2014) made by explicitly modeling decoy–dominant detection in the studied online diamond marketplace to empirically test and quantify the DE's magnitude.

Our study is closely related to the studies modeling consumer choice in the economics and marketing literature. Classic multinomial logit and probit models are built upon the revealed preference assumption, thus they cannot directly account for context effects. Given the extensive documentation on the prevalence of context effects in the behavioral literature, a few empirical and analytical methods have been developed to incorporate these context effects into the choice models. Tversky (1972) formulated his well-recognized Elimination-By-Aspects (EBA) model to account for the similarity effect. Kamakura and Srivastava (1984) modified the standard multinomial probit model in order to account for the similarity effect by modifying the error structure through incorporating similarity-based error correlations. Kivetz et al. (2004) proposed a choice model that can account for the compromise effect. Orhun (2009) developed an analytical choice model to study the DE and compromise effects under the loss-aversion assumption. Rooderkerk et al. (2011) proposed an empirical choice model that can incorporate decoy, compromise and similarity effects all together. They used choice-based conjoint data to estimate their proposed model and showed that ignoring context effects significantly biases the choice model's predictions. Our paper adopts a different approach by developing a proportional hazard model framework that explicitly accounts for the DE by using data from a real product marketplace. Our proposed framework allows researchers to quantify the DE even without the consumer-level search data by using the aggregate product sales data.

In the following sections, we first describe the online diamond marketplace, our dataset, and how we calibrate the decoy–dominant relationships. Second, we provide data evidence on the existence of the DE in this online diamond marketplace. Third, we discuss our model framework. Fourth, we present our estimation results. Fifth, we discuss the DE's managerial implications. We conclude with a discussion of the current study's potential limitations and directions for future research.

3 Data

3.1 Online Diamond Marketplace

Several U.S. retailers emerged in the online market for diamonds and fine jewelry products in the past two decades. We use a panel data set from one major retailer in this market. The retailer sells a variety of jewelry products to end consumers such as unbranded loose diamonds, gemstones, wedding rings[2], bracelets, necklaces, and earrings. Loose diamonds account for the core part of

[2]Buying a diamond ring from the retailer requires a consumer to choose his/her loose diamond first and then a ring setting. The retailer provides a limited number of standard ring settings. Consumers pay the total price and

the retailer's business in terms of revenue contribution. According to its annual report, the retailer works with dozens of diamond suppliers worldwide under an "exclusivity" agreement, which requires suppliers to sell their diamonds only through the retailer's online channel, and not through their own or other competing online and offline channels. For listed diamonds, the identities of the suppliers are not revealed on the retailer's website so that consumers cannot differentiate diamonds based on the suppliers. Instead, consumers recognize only the retailer name as the diamond brand.

To operate in a cost-efficient manner, the retailer uses a drop-shipping business model, i.e., the retailer, in most cases, does not physically carry inventories of loose diamonds listed on its website, instead purchasing diamonds from corresponding suppliers when consumers place their orders from the retailer. Unlike traditional brick-and-mortar stores, where only a limited number of diamonds are available, this drop-shipping model allows the retailer to list tens of thousands of diamonds every day. The retailer's website is designed to allow potential buyers to effectively search loose diamonds based on their physical characteristics such as carat, clarity, color, cut (4Cs henceforth), and price. After such a search, a consumer's consideration set typically contains multiple decoy diamonds, i.e., diamonds with the same 4Cs (the same grade henceforth) or even inferior 4Cs, but with higher prices than their counterparts (dominants henceforth). As stated earlier, we call this frequently observed pricing behavior the decoy pricing strategy.

In this setting, suppliers list their diamonds on the retailer's website and establish the wholesale prices. The retailer then adds a fixed percentage markup to the wholesale prices. As per the retailer's annual report, the markup is fixed at around 18-20% for all diamonds. Thus, the decoy pricing structure actually comes from the suppliers instead of the retailer strategically manipulating it.[3] Nevertheless, consumers are expected to respond to an existing decoy–dominant structure regardless of whether it is created by the retailer or suppliers.That being the case, as we demonstrate later, the decoy–dominant structure still affects the retailer's diamond sales and ultimately its profitability

the diamond ring is then assembled by the retailer. Typically, the loose diamond accounts for more than 90% of the total price paid by consumers.

[3]We note that the supplier price variation may exist due to multiple reasons. First, since there are consumer search costs, the observed prices can be the outcome of a mixed-strategy price equilibrium on the supplier side. Second, suppliers might have different costs, resulting in different pricing functions. Third, suppliers may change prices at different times. Understanding the source of the price variation is beyond the scope of the current study. Instead, we focus on quantifying the DE given the existing decoy–dominant structure in the marketplace, i.e., we take the existing price variations as given and model the demand-side responses to such variations by incorporating the DE.

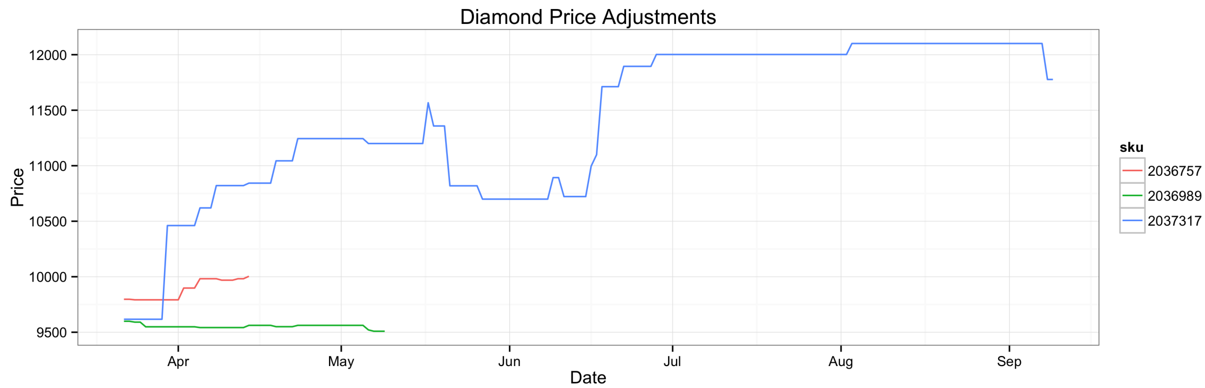

Figure 1: Diamond Price Patterns Over Time

to a large extent. Furthermore, strategic pricing on top of suppliers' wholesale prices remains an option for the retailer. We explore this possibility in our policy analysis to demonstrate how the retailer can further profit from the DE.

3.2 Data Description

We construct a panel data set of diamond prices and sales from this online retailer. We collect our daily data from the retailer's website through a web crawler for the period from February 2011 to September 2011. For each diamond listed during our sample period, we observe the diamond's inherent physical characteristics (its 4Cs and some other attributes such as symmetry and polish) and daily prices until the diamond is sold. In the data, diamond prices typically change overtime: On average, each diamond's price changes once every 21 days, conditional upon it being unsold. Figure 1 provides an example of price dynamics among three 1.0 carat diamonds from the day of introduction in the market till each is sold. As shown in the figure, the diamond price can go up-and-down, and each diamond may have its unique price patterns over time. We infer that a diamond is sold through the retailer's website on the last day it is listed as available, based on the unique SKU number. We believe this is a reasonable approach because, as discussed earlier, the suppliers are under an exclusive channel agreement with the retailer so that the diamond sale would not have happened through other channels. On average, it takes about 50 days to sell a diamond on the online platform.

In our analysis, we focus specifically on round-shaped diamonds with prices ranging between \$2K and \$20K. These diamonds are the most popular ones among those listed and sold.[4] Diamonds in different price ranges might be more attractive to different segments of potential buyers with different budgetary constraints. To account for some of the potential heterogeneity in the DE across different consumer segments, we further divide the diamonds into low- (\$2K-\$5K), medium- (\$5K-\$10K) and high-price (\$10K-\$20K) segments based on their first-day market prices.

Before calibrating the decoy–dominant structure, we first examine what determines diamond prices. For this purpose, we run an ordinary least squares regression with (log- of) daily diamond prices as our dependent variable and the diamonds' physical characteristics as independent variables to uncover the secret diamond-pricing formula. To control for potential demand variations across different days, we also add day fixed effects to the regression model. We report the regression results in Table 1. The adjusted R-squared for the model with 4Cs, along with day fixed effects, is as high as 96.67%. Individual regressions for each day result in adjusted R-squared from 94.92% to 96.57%. The results provide strong evidence that the 4Cs are the predominant factors in determining diamond prices.

To further check the robustness, we ran several regressions by incorporating other diamond attributes, such as symmetry and polish, into our main regression model. Overall, the R-squared measure does not improve. Moreover, the effects of these additional variables on prices are mostly insignificant, and their estimates are notably smaller in magnitude compared to the estimates of the 4Cs. For example, the implied price difference contributed by symmetry and polish turns out to be less than 0.5%. Thus, we have strong supporting figures to conclude that the quality of a diamond can be measured very precisely by simply looking at its 4Cs. In other words, a diamond can be characterized as a combination of five characteristics: 4Cs and price. This is, indeed, quite consistent with industry reports on diamond valuations and with articles educating consumers on purchasing diamonds. Even though the variation in diamonds' physical attributes explains a large portion of the price variation, we still observe significant within-grade (same 4Cs) and within-day

[4]Diamond shape can be considered as a horizontal attribute. Thus, including only round-shaped diamonds would not affect our decoy–dominant constructions, i.e., no decoy–dominant relationships exist across diamond shapes. Further, in the data, the prominent diamond shape is round, accounting for 74% of listing and 78% of sales. Last but not least, when buying a diamond, most consumers commit to a particular shape before choosing among other attributes.

price variation[5]. This variation is essential to characterize the dyadic decoy–dominant relationships among every diamond-pair, as discussed next.

Table 1: Price Regression: ln(price) on 4Cs and day fixed effects

Variable		**Estimate**	**Std. Err.**
Carat		1.768	0.0004
Cut	Poor	0.000	
	Good	0.056**	0.0009
	Very Good	0.114**	0.0009
	Ideal	0.180**	0.0009
	Signature Ideal	0.231**	0.0013
Color (Low to High)	J	0.000	
	I	0.141**	0.0004
	H	0.268**	0.0004
	G	0.361**	0.0003
	F	0.455**	0.0003
	E	0.503**	0.0004
	D	0.583**	0.0004
Clarity (Low to High)	SI2	0.000	
	SI1	0.124**	0.0003
	VS2	0.274**	0.0003
	VS1	0.371**	0.0003
	VVS2	0.455**	0.0003
	VVS1	0.544**	0.0003
	IF	0.619**	0.0004
	FL	0.762**	0.0004
	Daily Dummies	included	
Adj. R-squared		96.67%	
Adj. R-squared w/o daily dummies		95.27%	
Adj. R-squared w/ daily separate regressions		94.92%–96.57%	

Note: Estimates with ** are significant at the 0.05 level.

3.3 Dominance Construction

By definition, a diamond B is a decoy to another diamond A when B is inferior to A in at least one attribute, but has no superior attribute. In our specific setting, we define a diamond as a decoy under two conditions: 1) In terms of 4Cs, B is inferior in at least one attribute to A and has no attribute superior to A but has the same or a higher price than A; and 2) B has the same 4Cs as

[5]To illustrate the amount of within-grade and within-day price variation, we calculate the ratios of price standard deviation to mean price at each day-grade combination. The average (over grades and days) of these ratios turns out to be 0.1, indicating a sufficiently large within-grade and within-day price variation.

A, but is priced higher[6]. Under these two definitions, for any two diamonds on a particular day, we define the relationship between them as follows: A dominates B ($A \succ B$), B dominates A ($B \succ A$), and no dominance ($A \sim B$). Notice that, under the strict definition, two diamonds with the same attributes but different prices must have a strict dominance relationship. However, in real purchase situations, consumers may not care much about (or even notice) the difference if the price gap is not large enough. Thus, we use a conservative approach in our analysis: We define a dominance relationship for two diamonds with the same 4Cs only if the price difference between the two is larger than 5%.[7] Our conservative 5% rule also helps us avoid the potential problem of defining a false dominance relationship when the dominated diamond is, indeed, superior in other non-critical attributes such as symmetry and polish. As mentioned before, in a full regression model (with polish and symmetry), the price premium contributed by each of these attributes is smaller than 0.5%. Thus, we believe the 5% rule for two diamonds with the same 4Cs is conservative enough.

In our data sample, every pairwise decoy–dominant relationship between all listed diamonds is constructed for each day (due to within diamond price variation over time). For a particular diamond j, for each day t, we calculate the number of diamonds that are decoys (N_{jt}^{Decoy}) and dominants ($N_{jt}^{Dominant}$) to that diamond. The median number of decoys and dominants that a diamond has is 7 in the data sample, while the distribution is right skewed. 86% of the decoys/dominants are defined based on the first part of our definition in which diamonds differ in at least one of the 4Cs. As N_{jt}^{Decoy} gets larger, we expect that it gets easier for consumers to discover diamond j and its decoys at the same time, resulting in an increase in the size of the consumer segment detecting it as a dominant. Furthermore, when diamond j is detected as a dominant to more decoys, it would become more attractive, i.e., the sales hazard is expected to increase more. With similar reasoning, as $N_{jt}^{Dominant}$ gets larger, the size of the consumer segment detecting diamond j as a decoy is expected to grow. As a result, the sales hazard for diamond j will decrease given that consumers would not purchase it once they detect it as a decoy. Thus, N_{jt}^{Decoy} and $N_{jt}^{Dominant}$ are

[6]We do not consider the condition in which B is superior in 4Cs, but the high price premium compared to diamond A does not justify the difference in superior attributes (i.e., the near-dominance relationship as discussed in Huber et al. (1982)). There are two reasons for this. First, this is not a strict definition of decoy. Second, and more importantly, it might be significantly more difficult for consumers to detect such decoy–dominant relationships since diamonds are not directly comparable.

[7]If two diamonds have different 4Cs and have a dominance relationship, we use the strict definition, i.e., the inferior diamond with the same or higher price is labeled as a decoy. For diamonds with same 4Cs, we also conducted our analysis under the 1% and 10% rules, and the results turned out to be qualitatively very similar to the 5% rule.

the two key measures in our model.

Under our dominance definition, a diamond j can belong to one of the four mutually exclusive groups on day t: 1) neither decoy nor dominant: $N_{jt}^{Decoy} = 0$ and $N_{jt}^{Dominant} = 0$; 2) decoy only : $N_{jt}^{Decoy} = 0$ and $N_{jt}^{Dominant} > 0$; 3) dominant only: $N_{jt}^{Decoy}>0$ and $N_{jt}^{Dominant}=0$; and 4) both decoy and dominant: $N_{jt}^{Decoy} > 0$ and $N_{jt}^{Dominant} > 0$. The middle column of Table 2 shows the count of diamond-day observations for each type. Due to the significant within-grade price variations, we observe that a majority of the diamonds fall into the *both decoy and dominant* type. However, even in the smallest group, i.e., *neither decoy nor dominant*, we have a sufficient number of observations (27,077) to allow the identification of our model, as we discuss later.

Table 2: Summary Statistics Across Diamond Types

Diamond Type	**Diamond-Day Observations**	**Daily Percentage Sales**
Neither decoy nor dominant	27,077	1.96%
Decoy only	404,283	1.68%
Dominant only	340,456	2.53%
Both decoy and dominant	1,945,009	2.08%
Total	2,716,825	2.07%

3.4 Data Evidence of the DE

In this subsection, we discuss some observed data patterns that are suggestive of the DE. First, we provide statistics on how the likelihood of a diamond's sale is affected by which of the four dominance types the diamond belongs to. The last column of Table 2 summarizes the percentage of diamonds sold in the total diamond-day observations across different diamond types. Each cell is calculated as follows: For example, there are in total 404,283 diamond-day observations for *decoy only* diamonds, out of which 6,792 were sold. Therefore, the average sale probability in this cell becomes 1.68%. Table 2 shows that *decoy only* diamonds have the lowest average sale probabilities (1.68%), while the opposite is true for dominant only diamonds (2.53%). The pattern is consistent across the three diamond price segments. Qualitatively, such patterns could be explained by consumer search processes: A *decoy only* diamond is less likely to sell compared to a *neither decoy nor dominant* type since there exists a segment of consumers in the market who are able to detect this diamond as a decoy and therefore would not purchase it. Yet the sale probability is not zero because consumers have search costs and there exists another segment that fails to detect this diamond as a decoy and may end up purchasing it. Thus, the significant difference in sales probabilities between *decoy only* and *neither decoy nor dominant* type diamonds highlights the importance of controlling for market-level decoy detection in the model. Under random consumer search, *dominant only* diamonds should have similar probabilities of being included in consumers' consideration sets as do the diamonds in the *neither decoy nor dominant* type, and slightly greater sale probabilities because of their relative quality and/or price advantages. The fact that they have on average a significant 29% larger sale probability (2.53% vs. 1.96%) could be suggestive of effects beyond consumer search only. If it is solely search-effect, the retailer could acquire a significant profit gain by slightly decreasing prices of diamonds in the *neither decoy nor dominant* type to make them *dominant only*, thereby increasing their demand by 29%. Thus, we attribute this 29% demand boost to the potential DE: Once dominants are detected by some consumers in the marketplace, their purchase probabilities increase significantly.

Second, we develop a formal statistical method to test whether the existing price dispersion can be solely explained by consumer search. We show the details of this test in Web Appendix A. The intuition is that under pure consumer search without the DE, a supplier sets prices to maximize the expected profit of each individual diamond, and thus identical diamonds with different prices are expected to generate the same level of profit for the supplier in equilibrium (Burdett and Judd, 1983). However, if there is DE together with consumer search, a supplier needs to consider the price optimization beyond each individual diamond because decoys would serve as *loss-leaders* that help generate higher expected profits from their dominants. Consequently, one would expect the profit contribution of dominants to be higher than that of decoys.

Under the hypothesis of no DE (i.e., there is only search effect), we can utilize the observed price and sales information to recover the cost of each diamond j on day t (c_{jt}) from the corresponding supply-side pricing optimality conditions. Recovered costs should be approximately the same for diamonds with identical 4Cs, and should be exactly the same for the same diamond over time. However, if there exists DE along with consumer search — since the demand for higher-priced decoys is more price elastic — recovered costs will increase with their relative price levels[8]. Test of the pure search effect versus additional DE thus becomes the same as testing whether recovered

[8]On day t, we define the relative price of diamond j (i.e., rp_{jt}) as the percentage price difference between the price of diamond j and the average price of all diamonds that belong to diamond j's grade (i.e., same 4Cs).

costs are increasing with relative prices. We conduct two statistical tests: 1) using cross-diamond price variations (labeled as Test I); and 2) using within-diamond price variations over time (labeled as Test II). Results are provided in Table A1 in Web Appendix A. Results consistently reject the null hypothesis of a pure consumer search explanation across all three diamond price segments. All the estimated coefficients of relative price levels (rp_{jt}) are positive and significant, and thus are directionally consistent with the hypothesis supporting the existence of DE together with consumer search in the observed data. Further, the coefficient estimates are larger in the higher price segments, suggesting that the potential DE would be stronger for more expensive diamonds.

Third, we investigate how the overall dominance structure might impact the extent to which consumer search and the DE affect diamond sales likelihoods. We run two linear regression models to check these relationships. In the results reported in Table 3, we use the daily percentage sales share of decoys (dominants) as the dependent variable of the first (second) regression model. We use price-segment dummies and percentage of *decoy only* and *dominant only* diamonds as independent variables in both regression models. Results show that the percentage of *decoy only* diamonds significantly increases the decoy' sales share; while the percentage of *dominant only* diamonds significantly increases the dominants' sales share but decreases the decoy' sales share. Specifically, as the share of the *decoy (dominant) only* diamonds increases by 10%, the predicted share of decoys (dominants) from the overall sales increases by 9% (19.7%). The decoy sales share regression is consistent with the consumer search story: Having relatively more *decoy only* diamonds compared to dominants would reduce the likelihood of consumers discovering these decoys along with their dominants. In other words, as there are more *decoy only* diamonds, the size of the market segment detecting these diamonds as decoys decreases, i.e., their sales probability increases. Regarding the dominant sales share regression, the elasticity of percentage of dominants on sales share is significantly larger than 1 (1.97), suggesting that including additional *dominant only* diamonds would extract a disproportionally larger sales share from other diamonds, which is strong supporting evidence of the DE.

Table 3: Diamond Sales Share Regression

Variable	Decoy Sales Share		Dominant Sales Share	
	Estimate	Std. Err.	Estimate	Std. Err.
Intercept	0.023	0.026	−0.048	0.040
Medium Segment(5K-10K)	0.006	0.016	−0.082**	0.025
High Segment(10K-20K)	−0.008	0.021	−0.086**	0.032
% Decoys	0.899**	0.198	0.161	0.301
% Dominants	−0.312**	0.155	1.972**	0.236
Adj. R-squared	0.118		0.117	

Note: Estimates with ** are significant at the 0.05 level.

To sum up, our simple data analyses yield the following results: 1) Diamond prices can be characterized by 4Cs with high precision; 2) there is still sufficient price dispersion for diamonds with identical physical attributes; 3) the likelihood of a diamond purchase depends on whether it is a decoy and/or a dominant; and 4) there is suggestive evidence of the DE beyond consumer search in the online diamond marketplace. Therefore, a careful investigation of the decoy phenomenon through a deliberately developed empirical model is required. We achieve this in the next section.

4 Model

4.1 A Proportional Hazard Model of Diamond Sales Embedding the DE

We develop a diamond-level proportional hazard framework to model the daily sale likelihoods of diamonds. In general, a specific diamond's daily sale likelihood could be determined by a few factors: 1) its physical attributes and price; 2) time-variant diamond needs of consumers; and 3) competition from other comparable diamonds to this focal diamond. We utilize these three sets of information in the daily-diamond sale hazard component of our proportional hazard framework. However, since decoy pricing is widely observed in our market setting, and such pricing can potentially affect the diamond sales, we must also incorporate the DE into our hazard specification. As discussed earlier, due to the existence of large numbers of decoys/dominants, a focal diamond's demand should depend on: 4) the size of the market segment that is able to detect whether the diamond is a decoy and/or dominant; and if the diamond is a dominant, 5) the level of the demand boost upon the diamond's detection as as a dominant. We capture these two components, i.e., market-level decoy–dominant detection and dominant boost hazard, in the decoy–dominant hazard component of our proportional hazard framework.

Separating the market-level decoy–dominant detection and the boost in sales upon dominant detection is a challenging task since we do not observe any individual consumer level behaviors. To achieve our objective, we derive our diamond-level proportional hazard model from fundamental consumer-level primitives including *the consumer arrival process*, *product search*, *consideration set formation*, and *conditional choice probabilities with the embedded DE*. Please see Web Appendix B for the details of this derivation, along with how we embed the DE in consumers' conditional choice probabilities and how our proposed specification serves as a test for the DE at the aggregate level.

Based on the derivation in Web Appendix B, we use the following to denote the hazard that diamond j will be sold at day t:

$$h_j(t) = \psi_j(X_j, p_{jt}, Z_t, W_{jt}|\theta)\phi_j(N_{jt}, rp_{jt}|\gamma), \tag{1}$$

where $\psi_j(.)$ is the daily-diamond sale hazard and $\phi_j(.)$ is the decoy–dominant hazard, i.e., the key component of our model. Table 4 provides the definitions of the variables used. We model the daily-diamond sale hazard $\psi_j(.)$ as an exponential function of 1) diamond j's characteristics: its price segment–dummy coded low-, medium-, or high-price; physical attributes—dummy coded 4Cs (X_j); and, price and price square (p_{jt}, p_{jt}^2); 2) daily demand proxies capturing the daily diamond needs of consumers in the marketplace–Google search trends and weekday dummies (Z_t); and 3) competition from comparable diamonds–number of diamonds with the same and in the surrounding 4Cs (W_{jt}). Specifically, with $\theta = \{\alpha_0, \alpha_Z, \alpha_X, \alpha_W, \beta_1, \beta_2\}$, we have

$$\psi_j(X_j, p_{jt}, Z_t, W_{jt}|\theta) = \exp(\alpha_0 + Z_t\alpha_Z + X_j\alpha_X + W_{jt}\alpha_W + p_{jt}\beta_1 + p_{jt}^2\beta_2). \tag{2}$$

We define a grade as a unique combination of 4Cs and use $\overline{p_{jt}}$ to represent the average price for all diamonds for the grade that diamond j belongs to. We denote rp_{jt} as the relative price of diamond j compared to prices of other diamonds in diamond j's grade. The relative price measurement rp_{jt} is defined as $rp_{jt} = (p_{jt} - \overline{p_{jt}})/\overline{p_{jt}}$. As seen in Equation (1), we model the decoy–dominant hazard $\phi_j(.)$ as a function of 1) number of decoys and dominants diamond j has ($N_{jt} = [N_{jt}^{Decoy}, N_{jt}^{Dominant}]$), and 2) the relative price measurement rp_{jt} . The variables we choose to model the decoy–dominant hazard are consistent with the work of Huber et al. (1982) that examines two factors impacting the salience of the DE: 1) the *frequency effect*, which measures the number of attributes that the dominant alternative is dominating; and 2) the *range effect*, which measures the degree of dominance within an attribute. In our context, because the model is built upon the sales hazard of each diamond, we do not directly measure the number of attributes that form strict dominance relationships. Instead, we aggregate the number of alternatives that a diamond is dominating/dominated by (N_{jt}^{Decoy} and $N_{jt}^{Dominant}$). Therefore, this first measurement could be viewed as a proxy to the *frequency effect*. The relative price measure (rp_{jt}), on the other hand, can be viewed as a close proxy for the *range effect*.

Table 4: List of Variables Used in Model Estimation

	Variable Name	**Description**
Z_t	Google search	Daily Google search trends index of diamond-related keywords
	Weekday dummies	Dummy variables of weekdays
X_j	Diamond characteristics	Dummy coded 4Cs of each diamond
W_{jt}	Daily competitiveness	Log of number of diamonds of the same grade, and of the neighboring grades, Residual from the price regression, % of price change from last period
p_{jt}	Daily price	Daily prices of each diamond (in 1000)
rp_{jt}	Relative price index	Price of each diamond relative to the average price with same 4Cs
N_{jt}	Decoys	Number of diamonds that are dominated by diamond j
	Dominants	Number of diamonds dominating diamond j

Central to our empirical test is the modeling of the decoy–dominant hazard component $\phi_j(.)$, in which we incorporate two critical elements: 1) the *market-level decoy and dominant detection probabilities* as denoted by $Pr_{jt}^{Decoy}(.)$ and $Pr_{jt}^{Dominant}(.)$, respectively; and 2) the *dominant boost hazard* upon dominant detection, denoted as $Q_{jt}^{D}(.)$. We model market-level decoy and dominant detection probabilities as follows:

$$\begin{aligned} Pr_{jt}^{Dominant}(N_{jt}, rp_{jt}) &= I(N_{jt}^{Decoy} > 0)\frac{\exp(V_{jt}^{Dominant})}{1+\exp(V_{jt}^{Dominant})} \\ Pr_{jt}^{Decoy}(N_{jt}, rp_{jt}) &= I(N_{jt}^{Dominant} > 0)\frac{\exp(V_{jt}^{Decoy})}{1+\exp(V_{jt}^{Decoy})} \end{aligned}, \tag{3}$$

where $I(.)$ is the indicator function and $V_{jt}^{Dominant}$ and V_{jt}^{Decoy} are specified as:

$$\begin{aligned} V_{jt}^{Dominant} &= \gamma_0^{Dominant} + \gamma_1^{Dominant} \ln(N_{jt}^{Decoy}) + \gamma_2^{Dominant} I(rp_{jt} < 0)(-rp_{jt}) \\ V_{jt}^{Decoy} &= \gamma_0^{Decoy} + \gamma_1^{Decoy} \ln(N_{jt}^{Dominant}) + \gamma_2^{Decoy} I(rp_{jt} < 0)(rp_{jt}) \end{aligned} . \tag{4}$$

The intercept terms ($\gamma_0^{Dominant}$ and γ_0^{Decoy}) are modeled at each of the diamond price segments since decoy–dominant detection probabilities defined in Equation (3) might differ across various market segments with different consumer budgetary levels. The other γ s capture how both the number of decoys/dominants and the relative price would impact the likelihood of decoy–dominant detection probabilities.

Next, we define the dominant boost hazard $Q_{jt}^D(.)$ as the following:

$$Q_{jt}^D(N_{jt}, rp_{jt}) = \exp\left[\gamma_0^{Boost} + \gamma_1^{Boost} \ln(N_{jt}^{Decoy}) + \gamma_2^{Boost} I(rp_{jt} < 0)(-rp_{jt})\right]. \tag{5}$$

Similarly, we model the intercept term (γ_0^{Boost}) at each of the diamond price segment since the dominant boost effects might differ across market segments. γ_1^{Boost} captures the frequency on the DE and γ_2^{Boost} captures the range effect of the DE.

Given market-level detection probabilities and dominant boost hazard definitions, we operationalize the decoy–dominant hazard as follows:

$$\phi_j(.) = \begin{cases} 1, & \text{if } j \text{ is } \textit{Neither} \\ (1 - Pr_{jt}^{Decoy}), & \text{if } j \text{ is } \textit{Decoy Only} \\ (1 - Pr_{jt}^{Dominant}) + Pr_{jt}^{Dominant} Q_{jt}^D, & \text{if } j \text{ is } \textit{Dominant Only} \\ (1 - Pr_{jt}^{Decoy})(1 - Pr_{jt}^{Dominant}) + (1 - Pr_{jt}^{Decoy}) Pr_{jt}^{Dominant} Q_{jt}^D, & \text{if } j \text{ is } \textit{Both} \end{cases} . \tag{6}$$

Before discussing Equation (6) in detail, we would like to note that we make an implicit assumption in the derivation of the decoy–dominant hazard such that once consumers detect a specific diamond to be a decoy, they would never purchase it, as they can always choose the dominant one. This assumption is consistent with the existing literature. For example, Huber et al. (1982) verified that fully-informed subjects would seldom make "mistakes" of choosing decoys in lab experiments.

Equation (6) shows how the decoy–dominant hazard depends on diamond j's type: If the diamond is *neither decoy nor dominant* type, the DE has no impact on the sale hazard of the diamond, i.e., the proportional decoy–dominant hazard is normalized to one. If the diamond is *decoy only* type, it is considered only by the consumer segment that fails to detect it as a decoy under our assumption. Further, the DE does not play a role in the sale hazard of the diamond given no detection, resulting in the overall decoy–dominant hazard being the size of this segment, i.e., $\phi_j(.) = 1 - Pr_{jt}^{Decoy}(.)$. For a diamond belonging to the *dominant only* type, there exist two market segments: the segment that fails to detect the diamond as a dominant, and the segment that is able to detect it. The DE does not have any impact on the former segment (i.e., $Q_{jt}^{D}(.) = 1$), while we expect there is a demand boost effect (i.e., $Q_{jt}^{D}(.) > 1$) for the latter. The overall sale hazard thus becomes the expression in the third line of Equation (6). Finally, if the diamond is *both decoy and dominant* type, the diamond is only considered for purchase by the market segment that fails to detect it as a decoy, with the size of the segment being $1 - Pr_{jt}^{Decoy}(.)$. Similar to the *decoy only* case, the remaining consumer segment will never purchase it, i.e., $Q_{jt}^{D}(.) = 0$. Within the segment that fails to detect the diamond as a decoy, we can further divide them into two sub-segments: the sub-segment that fails to detect the diamond as a dominant and the one that detects it. Similar to the *dominant only* case, the former sub-segment with size $1 - Pr_{jt}^{Dominant}(.)$ will not be impacted by DE, i.e., $Q_{jt}^{D}(.) = 1$, while sale hazard from the other sub-segment will be boosted by $Q_{jt}^{D}(.) > 1$. Combining all the scenarios, we have the derivation as shown in the last line of the above equation.

Denote the total number of days since diamond j is on market to the end of our observation period as T_j, and the day diamond j is sold since its introduction as T_j^s. Given J diamonds in the data set, the total likelihood we use for estimation becomes the following:

$$L = \Pi_{j=1}^{J} \left\{ \left[I(T_j^s \leq T_j)(1 - e^{-h_j(T_j^s)}) \Pi_{t=1}^{T_j^s - 1} e^{-h_j(t)} \right] \times \left[I(T_j^s > T_j) \Pi_{t=1}^{T_j} e^{-h_j(t)} \right] \right\}. \tag{7}$$

We now discuss a few properties of our model. First, the unit of our analysis is each diamond, which is different from classic choice models defined by brand or product shares. Second, in terms of how to model the DE conditional on dominant detection, we choose to use a scalar function $Q_{jt}^{D}(.)$. When $Q_{jt}^{D}(.) > 1$, our specification becomes consistent with the DE theory, i.e., upon detection, there is a sales boost for dominant diamonds. In other words, under our framework, testing the existence of the DE becomes the same as testing whether $Q^D_{jt}(.) > 1$. Details of this test are provided in Web Appendix B. Third, it is quite possible that consumers are heterogeneous, so we allow our daily-diamond hazard $\psi_j(.)$ and the decoy–dominant hazard $\phi_j(.)$ to differ across different diamond price segments to capture such consumer heterogeneity.

4.2 Model Identification

We now discuss the intuition of our identification strategy. For illustrative purposes, let us think about the simpler case where each diamond's attributes (i.e., the type, prices, number of decoys, and dominants over time) do not change over days. Our identification strategy relies on the fact that it takes different number of days to sell different types of diamonds (*neither decoy nor dominant, decoy only, dominant only*, or *both decoy and dominant*). First, based on our normalization in Equation (6), the sales hazard for *neither* type is the baseline daily-diamond sale hazard. Thus, we use the variation of the time it takes to sell this type of diamonds with different price segments, 4Cs, price, and under different daily demand proxies and competition characteristics (i.e., with different X_j, p_{jt}, Z_t and W_{jt}) to identify the parameters in the daily-diamond hazard component (i.e., θ). Second, conditional on the identification of parameters in the daily-diamond sale hazard, we identify the parameters related to the detection probabilities and dominant boost hazards. Since decoy diamonds could only be purchased by the market segment that fails to detect them as decoys (see the second line of Equation (6)), we use the variation of time it takes to sell *decoy only* diamonds with different numbers of dominants, $N^{Dominant}_{jt}$, and relative prices, rp_{jt}, to identify the parameters of the market-level decoy detection probabilities (i.e., the parameters in the second line of Equation (4)-γ^{Decoy}). Third, as seen in the third and fourth lines of Equation (6), it is not possible to separately identify the parameters of the market-level dominant detection probabilities (parameters in the first line of Equation (4)-$\gamma^{Dominant}$) and of the dominant boost hazard (parameters in Equation (5)-γ^{Boost}) since $Pr^{Dominant}_{jt}$ and Q^D_{jt} are always bundled together in the form of $(1 - Pr^{Dominant}_{jt}) + Pr^{Dominant}_{jt} Q^D_{jt}$. To separately identify $\gamma^{Dominant}$ from γ^{Boost}, we need to make an additional assumption: that is, $\gamma^{Decoy} = \gamma^{Dominant}$, i.e, all else equal, the market-level probability of detecting a diamond with n decoys as a dominant is identical to the probability of detecting a diamond with n dominants as a decoy. Since the decoy–dominant relationships are calibrated at the diamond-pair level, the probability of discovering one diamond dominating another is the flip side of discovering that one is dominated by the other. Thus, parameters quantifying the decoy and dominant detection probabilities in the marketplace should be the same.[9] Based on this symmetric market-level detection (of decoys and dominants) assumption, and conditional on the dominant detection parameters ($\gamma^{Dominant}$) being identified, we use the variation of the time to sell *dominant only* and *both decoy and dominant* type diamonds (with different numbers of decoys, N_{jt}^{Decoy}, and relative prices, rp_{jt}) to identify the parameters of the *dominant boost hazard* (i.e., γ^{Boost} in Equation (5)). In our empirical setting, the diamond prices change over days and subsequently the decoy–dominant structure also changes at the market level. This level of data variation further empowers the identification of our model parameters.

5 Results

5.1 Main Estimation Results

We report our estimation results in Table 5. Results suggest that the daily-diamond sale hazard varies across the diamond price segments: Diamond-day level sales hazard is highest for the medium-price segment while it is lowest for the high-price segment. The hazard also decreases with the diamond's carat size. For the cut, color, and clarity attributes, we observe an inverse U-shape relationship, i.e., the daily-diamond sales hazard is the largest for diamonds with moderate attributes. As mentioned earlier, our unit analysis is each individual diamond; thus the basic sales hazard would be determined by both potential consumer demand and the supply level of diamonds. Therefore, an inverse U-shape relationship does not imply that given the same price, a consumer does not prefer diamonds with better physical attributes. Similarly, price also has an inverse U-shape effect. The competition related control variables such as the (log-) number of diamonds in the same and neighboring grades have no significant effect on the daily-diamond sale hazard.

Estimates of daily demand proxies suggest that Google search indexes for diamond-related keywords are significant proxies for the daily demand needs of consumers. The daily-diamond sale hazard increases significantly when the Google search indexes on the keywords of "diamond," "en-

[9] Our data limit us from testing whether they are empirically equal. Web browsing information from individual consumers would potentially help construct measurements to test this. Due to the stringent data requirement, we leave this exercise for future research.

gagement ring," and the specific name of the retailer are high. However, if the search intensity is high on the competitor's name, diamonds' sales hazards decrease. Another interesting finding is that when more people are searching for the keywords "diamond ring" and "wedding ring," the sale hazards decrease. One reason might be that the retailer is not able to get a premium position in the search results for those keywords, thus; losing some potential consumers to its competitors. The daily-diamond sale hazard also differs significantly across weekdays, with Monday and Thursday being the best days for diamond sales, while Saturday and Sunday are the worst days. One explanation might be that people use weekends to do research on diamonds, possibly go to brick-and-mortar stores and educate themselves about how to pick up the right diamond, and then place their orders online during weekdays.

Table 5: Model Estimates

Variable	**Estimate**	**Std. Err.**
Daily Diamond Sale Hazard		
Component D_j		
Low-Price Segment (2K-5K)	-2.756^{**}	0.092
Medium-Price Segment (5K-10K)	-2.722^{**}	0.100
High-Price Segment (10K-20K)	-2.832^{**}	0.126
Component X_j:		
ln(Carat)	-0.402^{**}	0.098
Cut: Poor	0.000	
Cut: Good	0.144^{*}	0.069
Cut: Very Good	0.439^{**}	0.068
Cut: Ideal	0.649^{**}	0.069
Cut: Signature Ideal	0.115	0.096
Color: J	0.000	
Color: I	-0.019	0.022
Color: H	0.062^{**}	0.024
Color: G	0.041	0.027
Color: F	0.035	0.030
Color: E	-0.156^{**}	0.033
Color: D	-0.192^{**}	0.037
Clarity: SI2	0.000	
Clarity: SI1	0.029	0.017
Clarity: VS2	0.073^{**}	0.022
Clarity: VS1	-0.004	0.026
Clarity: VVS2	-0.145^{**}	0.030
Clarity: VVS1	-0.316^{**}	0.035
Clarity: IF	-0.628^{**}	0.004
Clarity: FL	-0.620^{*}	0.291
Component p_{jt}:		
Price (in 1000)	0.068^{**}	0.016
Price Squared	-0.004^{**}	0.000
Component Z_t:		
Google Search: "diamond"	0.217^{*}	0.107
Google Search: "diamond ring"	-0.712^{**}	0.064
Google Search: "wedding ring"	-0.047	0.057
Google Search: "engagement ring"	0.093^{**}	0.030
Google Search: retailer's name	0.172^{**}	0.020
Google Search: competitor's name	-0.635^{**}	0.063
Weekday Dummy: Monday	0.000	
Weekday Dummy: Tuesday	-0.072^{**}	0.014
Weekday Dummy: Wednesday	-0.118^{**}	0.014
Weekday Dummy: Thursday	-0.032^{**}	0.014
Weekday Dummy: Friday	-0.264^{**}	0.015
Weekday Dummy: Saturday	-1.828^{**}	0.026
Weekday Dummy: Sunday	-1.169^{**}	0.021
Component W_{jt}:		
ln(# Diamonds of the Same Grade)	-0.003	0.011

Continued on next page

Table 5: Model Estimates

Variable	**Estimate**	**Std. Err.**
ln(# Diamond of Neighboring Grades)	0.001	0.006
Market-Level Detection Probability		
Low-Price Segment (2K-5K)	-1.519^{**}	0.166
Medium-Price Segment (5K-10K)	-2.327^{**}	0.142
High-Price Segment (10K-20K)	-2.688^{**}	0.832
$\ln(N_{jt}^{Decoys}+1)$	0.158^{**}	0.032
$I(rp_{jt}<0)(-rp_{it})$	4.344^{**}	0.924
Dominance Boost Hazard		
Low-Price Segment (2K-5K)	0.651^{**}	0.103
Medium-Price Segment (5K-10K)	1.110^{**}	0.117
High-Price Segment (10K-20K)	1.338^{**}	0.693
$\ln(N_{jt}^{Decoys}+1)$	0.066^{**}	0.022
$I(rp_{jt}<0)(-rp_{it})$	-0.001	0.432
Log-likelihood	$-257{,}824.7$	
BIC	516.390	

Note: Estimates with ** are significant at the 0.05 level.

We now discuss the estimation results regarding the *market-level decoy–dominant detection probabilities* and the *dominant boost hazard*, which are the most critical components of our model for addressing the paper's central research questions. First, our results suggest that the base market-level detection probability of a decoy (or dominant) diamond is significantly higher for diamonds in the low-price segment compared to those in the medium- and high-price segments; however, there is no statistically significant difference between the latter two segments. One potential explanation might be that consumers of the low-price ($2K-$5K) segment are usually on tight budgets and are more likely to spend more time searching for better prices. Thus, they are more likely to have larger consideration sets, and as a result they are more likely to detect existing decoy–dominant relationships. The positive significant estimate of (log- of) number of decoys/dominants (0.158) shows that when a diamond has more decoys/dominants, it is relatively easier for the market to detect that diamond as a dominant/decoy. The positive significant estimate of (the absolute value of) the relative price index (4.344) shows that the further a decoy (dominant) is priced from the average grade level price, the higher is the probability of decoy (dominant) detection in the marketplace. We next calculate the market-level decoy–dominant detection probabilities by using our model estimates. A detailed summary table about the distribution across the three diamond price segments is presented in the upper panel of Table 6. Interestingly, we find that for decoy and dominant diamonds, the market-level detection probabilities are quite low: 0.29 for the low-, 0.14 for the medium-, and 0.10 for the high-price segments, respectively. These findings show that our real-life scenario with a large number of diamonds defined on 4Cs and price with many decoys and dominants greatly contrasts with usual lab settings, in which participants are typically aware of the decoy–dominant relationships by the construction of the experimental design. Along this line, Simonson (2014) stated that consumers' detection of decoy–dominant relationships among the existing choice alternatives is a vital precondition for decoys to increase their dominants' sale likelihoods and called for a systematic study that accounts for consumers' decoy–dominant detection likelihoods. Responding to Simonson (2014)'s call, our study contributes to the literature by developing a framework to guide marketing researchers about how to quantify decoy–dominant detection probabilities in real product markets without any consumer-level search information.

Table 6: The Distribution of Detection Probability and Dominant Boost Hazard

	Min	**Q1**	**Median**	**Mean**	**Q3**	**Max**
Market-Level Detection Probability						
2K-5K	0.20	0.23	0.27	0.29	0.33	0.88
5K-10K	0.10	0.11	0.13	0.14	0.16	0.83
10K-20K	0.07	0.08	0.09	0.10	0.11	0.51
Dominant Boost Hazard						
2K-5K	2.00	2.16	2.29	2.30	2.44	3.16
5K-10K	3.18	3.32	3.51	3.52	3.69	4.55
10K-20K	3.99	4.09	4.33	4.38	4.57	5.78

Second, the intercept estimates of our dominant boost hazard component are all positive, confirming that, upon dominant detection, the sale hazard would be significantly boosted. This provides direct evidence of the existence of the DE in a real marketplace. Interestingly, this demand boost effect is lower for the low-price segment (0.651) than for the medium- (1.110) and high-price (1.338) segments. The reason may be, as mentioned previously, that consumers with limited budgets search more intensively for better prices and, therefore, are less responsive to savings gained from a single dominant once they detect it–i.e., they may be more likely to continue to search. The parameter estimate for the (log- of) number of decoys is positive and significant (0.066), indicating that having more decoys would further increase the dominant diamonds' sales hazards. The relative price index, on the other hand, turns out to be insignificant. We also calculate the dominant boost effect in proportional terms based on the model estimates. We report the distribution of these effects in the lower panel of Table 6. On average, conditional on detection of a diamond as a dominant, its sale hazard increases by 130% for the low-, 250% for the medium-, and 340% for the high-price segments. This finding is in line with Heath et al. (1995), who showed that the DE is stronger for higher quality products (that correspond to the higher priced diamonds in our case). It is also consistent with Mourali et al. (2007), who showed that promotion focus (less budget constrained buyers in our case) would increase the DE.

In summary, our estimation results suggest that it is difficult for consumers to detect the decoy–dominant relationships in the online diamond marketplace, especially among diamonds in the medium- and high-price segments. For this reason, it is critical to model the decoy–dominant detection process in real marketplaces before modeling the sales impact of decoys on dominants. On the other hand, even though the market-level decoy–dominant detection probabilities are low, once an alternative is detected as a dominant, its sales hazard increases quite significantly, especially in the medium- and high-price segments. With this finding, we not only provide strong field evidence about the existence of the DE, we also respond to Frederick et al. (2014) and Yang and Lynn (2014), who questioned the practical validity and usefulness of the DE.

5.2 Model Comparison

As discussed earlier, capturing the market-level decoy–dominant detection is critical in testing the DE in a real product market such as ours. Thus, to obtain accurate inferences, we carefully calibrate our model by separating the market-level detection from the dominant boost hazard. To test the importance of this separation, we estimate a benchmark proportional hazard model without such separation by using the same set of variables from our proposed framework.

Table 7 reports the results of the decoy–dominant hazards for both decoy and dominant diamonds[10]. Without the separation of the market-level detection and the dominant boost hazard, the estimates have a compound effect on diamonds' sales hazards. In other words, one may get significantly biased results using such estimates to quantify the DE.

[10] We use the same label for dominant diamonds and call the boost in their sales the dominant boost hazard in Table 7. Meanwhile, for decoy diamonds, we use the label of decoy shrinkage hazard to denote the decrease in their sales. Further, the daily-diamond hazard parameters are skipped in this table to save space. They are available upon request.

Table 7: Benchmark Model Estimates

Variable	**Estimate**	**Std. Err.**
Decoy Shrinkage Hazard		
Low-Price Segment (2K-5K)	-0.231^{**}	0.022
Medium-Price Segment (5K-10K)	-0.046^{**}	0.022
High-Price Segment (10K-20K)	-0.085^{**}	0.025
$ln(N_{jt}^{Dominants} + 1)$	-0.032^{**}	0.007
$I(rp_{jt} > 0)(rp_{jt})$	-1.035^{**}	0.154
Dominant Boost Hazard		
Low-Price Segment (2K-5K)	0.162^{**}	0.024
Medium-Price Segment (5K-10K)	0.095^{**}	0.024
High-Price Segment (10K-20K)	0.052^{**}	0.025
$ln(N_{jt}^{Decoys} + 1)$	0.070^{**}	0.006
$I(rp_{jt} < 0)(-rp_{jt})$	0.979^{**}	0.162
Log-likelihood	−257,948.7	

Note: Estimates with ** are significant at the 0.05 level.

As seen in Table 7, first of all, we find that the log-likelihood (as well as BIC) for this benchmark model is much worse than our proposed model, suggesting that explicitly separating the market–level decoy–dominant detection from the dominant boost hazard better explains data variations better. Second, as expected, results show that being decoys (dominants) would have a negative (positive) impact on the sales hazard. However, compared to the dominant boost hazard estimates from Table 5, the estimates from Table 7 turn out to be much smaller in magnitude. The average decoy shrinkage and dominant boost hazards across different price segments are reported in Table 8. Results suggest that, on average, a decoy (dominant) diamond is 25% less (47% more) likely to be sold. Results also suggest that dominant boost hazards (from this benchmark model) are much smaller (ranging from 1.35 to 1.59 times) compared to ones from the proposed model (2.30 to 4.38 times). In addition, the dominant boost hazard is largest for the low-price segment and smallest for the high-price segment, which is directionally opposite of the finding from our proposed model. Thus, the model comparison demonstrates that not controlling for the market-level detection not only biases the overall magnitude of the DE but also yields directionally wrong inferences regarding the DE's magnitude across different diamond price segments. This finding brings additional support to Simonson (2014)'s statement about the importance of controlling the decoy–dominant detection in recovering the true impact of the DE.

Next, we discuss the managerial implications of our study. These findings shed some new light on the DE's practical significance and show that it is not simply an experimental artifact.

Table 8: Decoy/Dominant Hazards from the Benchmark Model

Diamond Type	**2K-5K**	**5K-10K**	**10K-20K**	**Overall**
Decoy Shrinkage Hazard	0.66	0.82	0.81	0.75
Dominant Boost Hazard	1.59	1.42	1.35	1.47

6 Managerial Significance

To demonstrate the substantive implications of our model, we use the estimated parameters and run policy simulations. First, we quantify the overall profit impact that is contributed by the DE. Second, we explore opportunities for the retailer to improve its profitability through three easy-to-implement strategies by that include changing 1) the number of dominants/decoys; 2) the degree of price dispersion; and 3) the baseline decoy–dominant detection probabilities.

6.1 Profit Impact of the DE

In this subsection, we investigate the retailer's profit gain or loss from carrying decoy (and dominant) diamonds. To check the profit impact of the DE, we calculate the retailer's profit under our proposed model estimates and compare it with a scenario with no DE; i.e., the parameters of the dominant boost hazard (γ^{Boost}) are turned off. The retailer's expected profit for a given diamond j at time t is calculated as:

$$\pi_{jt} = \Pr_j(t|.) \times (p_{jt} - w_{jt}), \tag{8}$$

where p_{jt} is the price, and w_{jt} is the wholesale price, which can be easily calculated by subtracting out the retailer's mark-up of 18% (given in the retailer's annual report) from the observed daily retail prices. $\Pr_j(t|.) = 1 - \exp(-h_j(t|.))$ is the discrete time hazard, or the probability that diamond j would be sold in day t, conditional on not being sold until that day. The DE on profit comes from the differences in the sale probability represented by $\Pr_j(t|.)$ with and without the dominant boost hazard component.

The results are presented in Table 9. Without the DE, on average, each diamond would contribute \$20.49 in gross profit; whereas with the DE the contribution would become \$26.07. In other words, the DE contributes 21.40% of the retailer's gross profit. The contribution is quite similar across the three diamond price segments. Based on the financial information of the retailer, this percentage increase would translate into $15.4 million per year in absolute terms. This result shows that even though decoy–dominant detection probabilities are quite low in the online diamond marketplace, the DE still has a very significant profit impact due to the significant boost in sales likelihoods upon the dominant detection. Indeed, this profit impact is what really matters the most from the substantive point of view. This result further mitigates the concerns of Frederick et al. (2014) and Yang and Lynn (2014) about the practical significance of the DE.

Table 9: The Impact of the DE on Retailer's Gross Profit

Effect	**2K-5K**	**5K-10K**	**10K-20K**	**Overall**
Avg Daily Revenue Per Diamond W/O the DE	8.39	25.56	39.36	20.49
Avg Daily Revenue Per Diamond	11.00	32.60	49.37	26.07
% Revenue from the DE	21.54%	23.20%	20.80%	21.40%

6.2 Further Profitability from the DE

After showing the significant profit impact of the DE, we next investigate how the retailer can further improve its profitability by effectively utilizing the DE. We look at three strategies the retailer could potentially adopt. The first strategy is to list more dominants or decoys. The second strategy is to manipulate the price dispersion levels while keeping the current dominance structure unchanged. Borrowing the terminology from the literature, we label the first strategy the *frequency strategy* and the second strategy the *range strategy.* A third strategy the retailer could use is to change the baseline decoy–dominant detection probabilities in the marketplace. If the retailer wants to increase detection probabilities, it might simply recommend decoys/dominants to consumers. On the other hand, by making the search more difficult through modifying its website design, the retailer might decrease detection probabilities. We call this third strategy the *awareness strategy.*

For the *frequency strategy*, we compare the profitability gains or losses when we add a few *decoy only* or *dominant only* diamonds. Results are presented in Table 10. Overall, the retailer can gain some additional profit when it adds *dominant only* diamonds. For example, when the retailer adds one *dominant only* diamond for each of the diamonds that has at least one dominant in the current pricing schedule, it can gain an additional 0.61% profit, which translates into an additional $440,000 yearly profit. Adding a second dominant diamond further increases the retailer's profit by 0.51%($368,000). Further, as seen in Table 10, the profit gains are quite similar across different diamond price segments. However, unlike adding *dominant only* diamonds, adding *decoy only* diamonds reduces the retailer's profitability.

Table 10: The Frequency Effect of Dominance Structure on the Retailer's Gross Profit

Dominance Structure	**2K-5K**	**5K-10K**	**10K-20K**	**Overall**
Add 1 decoy	−0.36%	−0.23%	−0.17%	−0.23%
Add 1 dominant	0.51%	0.63%	0.64%	0.61%
Add 1 decoy and 1 dominant	0.16%	0.40%	0.47%	0.38%
Add 2 decoys	−0.65%	−0.42%	−0.31%	−0.42%
Add 2 dominants	0.96%	1.16%	1.17%	1.12%
Add 2 decoys and 2 dominants	0.30%	0.74%	0.86%	0.70%

Regarding the *range strategy*, we change the price dispersion levels for diamonds within the same grades. To achieve this, we simply enlarge or reduce the relative price measure for each diamond by a factor. For example, think about a diamond that is priced at $11,000, with a calculated mean grade price of $10,000. We change the price of this diamond to $10,500 (dispersion factor 0.5), $10,800 (dispersion factor 0.8), $11,200 (dispersion factor 1.2), and $11,500 (dispersion factor 1.5) in our simulation studies[11]. Table 11 reports the results. Overall, the retailer could make additional profits by reducing the price dispersion compared to the current pricing schedule in the low- and medium-price segments. This is because, for those diamonds, the market-level detection probability is relatively large, but the dominant boost hazard is relatively small. On the other hand, the current pricing scheme for diamonds in the high-price segment seems to be already optimal. By changing the price dispersion by a factor of 0.5, the retailer could gain an additional 0.38% profit, which translates into an additional $274,000 yearly profit.

Table 11: The Range Effect of Dominance Structure on the Retailer's Gross Profit

Dispersion Factor	**2K-5K**	**5K-10K**	**10K-20K**	**Overall**
0.5	1.75%	1.28%	−0.83%	0.38%
0.8	0.84%	0.60%	−0.23%	0.26%
1.2	−1.04%	−0.72%	0.08%	−0.41%
1.5	−2.96%	−2.05%	−0.11%	−1.32%

The third (i.e., the *awareness*) strategy is to change the probability that consumers' in the marketplace would detect the existing decoy–dominant structure. We achieve this by changing the estimated intercepts for the three price segments (see Table 6 – Market-Level Detection Probabilities) and exploring the optimal values for these intercepts. In this exercise, we first generate a

[11]Note that the mean price level is preserved in this exercise.

sequence of values between -4.0 and 4.0 with the increments of 0.1; we then use each value in this sequence as the intercepts, and compute the corresponding expected profits. We compare across the trial combinations and choose the optimal intercept levels that maximize the expected profit. Table 12 reports the mean market-level detection probabilities under current awareness levels and under optimal awareness levels. We also report the additional profit impact when we move towards the optimal awareness levels. Interestingly, we find that the awareness levels in the medium- and high-price segments are lower than optimal levels; while for the low-price segment, the retailer should strongly discourage consumers from detecting the decoy–dominant relationships by making the consumer search for diamonds harder. For the medium-price (high-price) segment, though, the retailer should increase the detection level from 14% to 29% (10% to 34%). The retailer might achieve this through personalized product recommendations. If the retailer could manage the awareness levels optimally, the overall profit impact becomes highly significant: With a 5.4% increase in the gross profit, the retailer could gain an additional $3.9 million net profit annually.

Table 12: The Effect of the Awareness Strategy on the Retailer's Gross Profit

Awareness Level	**2K-5K**	**5K-10K**	**10K-20K**	**Overall**
Current Mean Level	28.95%	13.97%	9.90%	20.66%
Optimal Mean Level	6.88%	28.50%	34.11%	18.77%
Profit Increase	7.53%	1.32%	7.10%	5.36%

In summary, our simulation studies show the possibility for the retailer to further increase profitability by effectively utilizing the DE. We find that the additional profit increase might be limited when the retailer manipulates the number of dominants and decoys (*frequency strategy*) and the price dispersion (*range strategy*). However, if the retailer can manage the consumers' awareness levels optimally (*awareness strategy*), there is significant room to increase its profit.

7 Conclusions

In this research, we empirically validate the DE by using a unique panel data from a large online jewelry retailer. We first estimate a proportional hazard model with embedded market-level decoy–dominant detection probabilities and the sales boost upon dominant detection. Our proposed model is derived from individual consumer primitives. We model daily sale likelihoods of diamonds as a function of their 4Cs, prices, daily demand fluctuations, competitive effects from other diamonds, and the observed decoy–dominant structure. We find that, in general, the market-level probability of detecting a diamond as a decoy (dominant) is quite low, especially for the medium- and high-price diamond segments, but this detection probability increases significantly as the number of dominants (decoys) that the diamond has increases. More importantly, we find that once a diamond is detected as a dominant, its sale hazard increases quite significantly (2.3 to 4.4 times). Thus, we empirically validate the existence of the DE in a real product market and show that the effect is not merely an experimental artifact as argued in Yang and Lynn (2014). Model comparisons reveal that not controlling for decoy–dominant detection probabilities yields biased and directionally inconsistent results regarding the magnitude of the DE.

In addition, we contribute to the substantive issue of measuring the overall profit impact, i.e., true managerial significance, of the DE. We quantify the overall profit impact of the DE using model estimates and find that it contributes about 21.4% of the retailer's gross profit. Next, we explore various strategies that the retailer can adopt to improve its profitability through further utilizing the DE. We test the implications of three strategies: listing more decoys/dominants – frequency strategy; changing the grade-level price dispersion – range strategy; and changing the baseline market-level decoy–dominant detection probabilities (for example, through recommending diamonds to consumers, or making the consumer's search easier (or harder) by changing the website design) – awareness strategy. From our simulation studies, we find that the awareness strategy turns out to be the most effective among the three strategies and brings an additional 5.4% gross profit to the retailer.

Our study is the first empirical attempt to quantify the widely documented DE in the consumer behavior literature by using real world data. It is exciting to apply the well-developed context dependent choice theory to real-life data and empirically quantify the managerial implications. Several directions might be pursued to extend the understanding of this topic in future research. One direction is to jointly model demand for and supply of diamonds. We are less concerned with the diamond suppliers' optimal pricing decisions in our application since our focus is the DE on the demand side. Modeling the suppliers' pricing decisions under the DE might be an avenue for future studies if such supplier–level information is observed. A second direction is to model the DE with other context effects, such as compromise (Simonson, 1989) and similarity (Tversky, 1972) effects. We face significant identification challenges in jointly estimating multiple context effects in our setting due to the aggregate nature of our data. Future research could potentially address this issue when consumer–level search data is available. Finally, a third direction is to model the competition between our focal retailer and the other online diamond retailers because consumers might search for diamonds from different online retailers. When data of consumers from the multiple online retail outlets are available, this direction might be another interesting direction to pursue for future research.

References

Bettman, James R, Mary Frances Luce, John W Payne. 1998. Constructive consumer choice processes. *Journal of consumer research* **25**(3) 187–217.

Burdett, Kenneth, Kenneth L Judd. 1983. Equilibrium price dispersion. *Econometrica: Journal of the Econometric Society* 955–969.

De los Santos, Babur, Ali Hortaçsu, Matthijs R Wildenbeest. 2012. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review* **102**(6) 2955–80.

Frederick, Shane, Leonard Lee, Ernest Baskin. 2014. The limits of attraction. *Journal of Marketing Research* **51**(4) 487–507.

Guo, Liang, Wenbo Wang. 2016. Compromise effects explained by contextual deliberation reveal the rationality in preference construction. *Working paper, Chinese University of Hong Kong* .

Heath, T B, S Chatterjee, K R France. 1995. Mental accounting and changes in price: the frame dependence of reference dependence. *Journal of consumer research* **22** 90–97.

Hong, Han, Matthew Shum. 2006. Using price distributions to estimate search costs. *The RAND Journal of Economics* **37**(2) 257–275.

Honka, Elisabeth, Pradeep Chintagunta. 2016. Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science* **36**(1) 21–42.

Huber, J, J W Payne, C Puto. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research* **9**(1) 90–98.

Huber, J, C Puto. 1983. Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of consumer research* **10** 31–44.

Huber, Joel, John W Payne, Christopher P Puto. 2014. Let's be honest about the attraction effect. *Journal of Marketing Research* **51**(4) 520–525.

Kamakura, Wagner A, Rajendra K Srivastava. 1984. Predicting choice shares under conditions of brand interdependence. *Journal of Marketing Research* 420–434.

Khan, Uzma, Meng Zhu, Ajay Kalra. 2011. When Trade-Offs Matter: The Effect of Choice Construal on Context Effects. *Journal of Marketing Research* **48**(1) 62–71.

Kivetz, Ran, Oded Netzer, V Srinivasan. 2004. Alternative models for capturing the compromise effect. *Journal of marketing research* **41**(3) 237–257.

Lehmann, Donald R, Yigang Pan. 1994. Context Effects, New Brand Entry, and Consideration Sets. *Journal of Marketing Research* **31**(3) 364.

Luce, R Duncan. 1959. *Individual Choice Behavior*. John Wiley & Sons, Inc, New York.

Luce, R Duncan. 1977. The choice axiom after twenty years. *Journal of Mathematical Psychology* **15** 215–233.

Morewedge, Carey K, Meng Zhu, Eva C Buechel, Andrea C Morales. 2018. Hedonic contrast effects are larger when comparisons are social. *Journal of Consumer Research* .

Mourali, M, U Böckenholt, M Laroche. 2007. Compromise and attraction effects under prevention and promotion motivations. *Journal of consumer research* **34** 234–247.

Müller, Holger, Victor Schliwa, Sebastian Lehmann. 2014. Prize decoys at work—new experimental evidence for asymmetric dominance effects in choices on prizes in competitions. *International Journal of Research in Marketing* **31**(4) 457–460.

Orhun, A Yesim. 2009. Optimal product line design when consumers exhibit choice set-dependent preferences. *Marketing Science* **28**(5) 868–886.

Ratneshwar, S, A D Shocker, D W Stewart. 1987. Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of consumer research* **13** 520–533.

Rooderkerk, Robert P, Harald J Van Heerde, Tammo HA Bijmolt. 2011. Incorporating context effects into a choice model. *Journal of Marketing Research* **48**(4) 767–780.

Royle, Jone R, David J. O'Connor, Gareth M. Reynolds, Paul A. Bottomley. 1999. The robustness of the asymmetrically dominated effect: Buying frames, phantom alternative, and instore purchases. *Psychology and Marketing* **16**(3) 225–243.

Simonson, I. 1989. Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research* **16** 158–174.

Simonson, Itamar. 2014. Vices and virtues of misguided replications: The case of asymmetric dominance. *Journal of Marketing Research* **51**(4) 514–519.

Tversky, A. 1972. Elimination by aspects: A theory of choice. *Psychological review* **79** 281–299.

Wedell, Douglas H. 1991. Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **17**(4) 767–778.

Yang, Sybil, Michael Lynn. 2014. More evidence challenging the robustness and usefulness of the attraction effect. *Journal of Marketing Research* **51**(4) 508–513.

Zhang, Xing, Tat Y Chan, Ying Xie. 2017. Price search and periodic price discounts. *Management Science* **64**(2) 495–510.

Web Appendices

A A Statistical Test for the Source of the Observed Price Variation

In this web appendix, we develop a statistical test to understand whether the observed price variations in the data can be solely explained by consumer search, or if they are suggestive of both consumer search and the DE.

Burdett and Judd (1983) proved that when consumers search for price, even for homogeneous products, price variations can arise in equilibrium. The intuition is that because of consumer search cost, consumers may not discover all options; and as a result, high-priced options may be purchased by some consumers (i.e., these options may still have positive sale likelihoods). Further, in the mixed-strategy pricing equilibrium, each observed price point generates the same expected profit. Hong and Shum (2006) used this idea of mixed-strategy pricing equilibrium to recover consumer search-cost distribution purely from observed price variations for textbooks–a typical homogeneous good. We borrow this idea and use it to conduct our statistical test. For diamonds with identical 4Cs (i.e., the same grade), denoted by the set J, we observe two types of price variations: 1) across diamonds in the set J on day t, and 2) within the same diamond (identified by SKU) over time. We use p_{jt} to denote the price of diamond j on day t. We use $Ph_j(t|p_{jt})$ to denote the sales response function conditional on price. Because the retailer has a fixed margin ($1 - r = 18\%$) in our market setting, the wholesale price of suppliers becomes $w_{jt} = r \times p_{jt}$. We finally denote the marginal cost for suppliers as c_{jt}.

If there is no DE in the sales response function $Ph_j(t|p_{jt})$ (i.e., the observed price variation is driven solely by consumer search), for a supplier, setting prices either through maximizing the total expected profit from a set of diamonds, or through maximizing the expected profit for each individual diamond (in the corresponding set) will yield the same set of prices and ultimately the same total profit. However, when the DE exists along with consumer search, this no longer holds since decoys boost their dominants' demand, i.e., decoys bring positive externality to the profitability of their dominants. Therefore, we predict the expected profit from dominants to be higher than that of decoys. In other words, if the DE exists together with consumer search, each price point in the support of the observed price distribution would no longer yield the same expected profit, i.e., dominants will have higher expected profits compared to decoys. The supplier's expected profit from diamond j on day t is:

$$\pi_{jt} = (r \times p_{jt} - c_{jt}) Ph_j(t|p_{jt}). \tag{A.1}$$

Under the assumption of no DE, assume that suppliers have priced optimally, in which case, the following first order condition should hold: $\frac{\partial \pi_{jt}}{\partial p_{jt}} = 0$. Note that this optimality condition holds for any price observed in the marketplace, because under mixed pricing equilibrium, each price point generates the (same) optimal profit. From this optimality condition, we can invert the following marginal cost:

$$c_{jt} = r \times p_{jt}(1 + \frac{1}{\eta_{jt}}), \tag{A.2}$$

where $\eta_{jt} = [\frac{\partial Ph_j(t|p_{jt})}{\partial p_{jt}}][\frac{p_{jt}}{Ph_j(t|p_{jt})}]$ is the price elasticity at price p_{jt}.

However, when there is DE, this relationship no longer holds, because decoys serve as "loss-leaders" and generate less profit than their dominants. Consequently, for high-priced decoys, true price elasticities will be larger in absolute values (more elastic) compared to their low-priced dominants. Therefore, using Equation (A.2) leads us to a relationship where the calculated cost c_{jt} increases with the observed price p_{jt} for identical diamonds. We use this idea in our proposed statistical tests. In the first test (labeled as Test I), we assume that the suppliers' marginal costs of diamonds with identical attributes are the same, i.e, $c_{jt} = c, \forall j \in J, \forall t$. It might be a reasonable assumption in this particular industry, because diamonds are supplied globally by a few dominant manufacturers. Further, this assumption was also used by both Burdett and Judd (1983) and Hong and Shum (2006). To conduct our test, we proceed with the following steps:

1. Use a proportional hazard model to fit the sales response function $Ph_j(t|p_{jt})$ with polynomials of p_{jt} (we use linear, quadratic, and cubic forms), diamond characteristics, days on market, and day fixed effects to control for focal and over-time demand effects.

2. Invert the implied cost $\widehat{c_{jt}}$ using Equation (A.2) for each observed price point under our null hypothesis that observed price variation is driven by consumer search only.

3. Regress $\widehat{c_{jt}}$ over the relative price index $rp_{jt} = (p_{jt} - \overline{p_{Jt}})/\overline{p_{Jt}}$ ($\overline{p_{Jt}}$ is the average price of diamond j on day t, and other control variables such as diamond characteristics and day fixed effects.)

If the estimated coefficient for rp_{jt} is insignificant, then the test favors the null hypothesis that price dispersion could be explained based on consumer search alone; if the coefficient is positive and significant, we would have the statistical support to reject the null hypothesis, and the results would be consistent with the price variations being driven by the DE along with consumer search. In our second test (labeled as Test II), we relax the cost assumption ($c_{jt} = c, \forall j \in J, \forall t$) from Test I. Instead, we impose the following assumption: For the same diamond j, the cost for the supplier would be the same over time, i.e., $c_{jt} = c, \forall t$. In other words, diamonds in the same grade might have different costs, but this cost is time-invariant. We use the within-diamond over time price variation to test our hypothesis. The test follows the same steps as in Test I, except that in Step 3, we run the regressions using diamond-level (i.e., SKU-leve) fixed effects as controls and test whether the coefficient for rp_{jt} is significant. We repeat Test I and II for each of the three diamond price segments (low-, medium-, and high-price) and report the results in Table A1. Both Test I and II reject the null hypothesis (i.e., the price variation arises solely from consumer search) and support the DE as coexisting with consumer search. Furthermore, consistent with our proposed model predictions (see Estimation Results section in the manuscript), the effects are stronger (i.e., coefficients for rp_{jt} are larger) for medium-, and high-price segments.

Table A1: Test of the Source of the Observed Price Variation

Variable	**2K-5K**	**5K-10K**	**10K-20K**
Test I			
Controls		4Cs, day fixed effects, rp_{jt}	
rp_{jt}	1.314**(0.001)	3.163**(0.004)	6.073**(0.007)
Adj. R-squared	0.960	0.965	0.965
Test II			
Controls		diamond fixed effects rp_{jt}	
rp_{jt}	2.349**(0.024)	2.965**(0.051)	4.440**(0.143)
Adj. R-squared	0.960	0.963	0.946

Note: Estimates with ** are significant at the 0.05 level. Dependent variable—estimated cost—is in 1000 dollars.

B Derivation of Diamond-Level Sales Hazard from Consumer Primitives

In this web appendix, we derive our diamond-level proportional hazard model (Equation 1 in the manuscript) from consumer primitives including *consumer arrival process*, *search*, *consideration set formation* and *conditional choice probabilities*. We further discuss 1) how we embed the DE in consumers' conditional choice probabilities, and 2) how our specification serves as a test for the DE.

B.1 Individual Primitives and Continuous Time Diamond Hazard

We assume that potential diamond consumers arrive at random times to the retailer's website. In each specific time τ (can be a millisecond), we assume that at most one consumer would be making a diamond purchase decision. A representative consumer i, arriving at time τ_i, searches the retailer's website to form her consideration set and then decides whether to purchase one of the diamonds from that set that maximizes her utility or not to purchase any diamonds. In terms of search, we assume independence across diamonds—that is, the probability that a particular diamond is included is independent of any other diamonds being included or not.[1] We denote the consumer i's consideration set as M_i, and the super set containing all the possible consideration sets as $\mathbb{M}$.

We define the conditional choice probability of consumer i choosing a particular diamond j from her consideration set , given j has not been sold before τ_i, as $s_i(j|M_i)$ (to be discussed in the following subsection in detail). The expected sales probability of diamond j at time τ_i is thus the sum of the choice probabilities over all possible consideration sets:

$$\omega_j(\tau_i) = \sum_{M_i \in \mathbb{M}} Pr(M_i) \times s_i(j|M_i). \tag{B.1}$$

Note that this consumer and diamond specific choice probability is equal to the sales hazard of diamond j at time τ_i. This is the case because it is the conditional probability that diamond j will be sold at τ_i conditional on it not being sold until τ_i, and the consumer i is the only consumer deciding whether j would be purchased at this particular time.

[1] Note that this assumption holds under simultaneous but not sequential search. Researchers have documented empirical evidence in support of both sequential (Zhang et al., 2017) and simultaneous search in the literature (De los Santos et al., 2012; Honka and Chintagunta, 2016).

B.2 Embedding the DE into the Continuous Time Diamond Hazard

For a specific diamond j, we can classify diamonds into three mutually exclusive and collectively exhaustive types based on their relationships to j. The set of C_j^o contains diamonds that are neither dominants nor decoys to j; the set $C_j^{Dominants}$ contains all diamonds that are dominating j; and finally, the set C_j^{Decoys} represent the collection of j's decoys. Denote as M_{ij}^o the set of diamonds from C_j^o that are in consumer i's consideration set, and similarly $M_{ij}^{Dominants}$ and M_{ij}^{Decoys} the sets of diamonds (in i's consideration set) that are from j's dominants set ($C_j^{Dominants}$) and decoys set (C_j^{Decoys}). Denote $\mathbb{M}_j^o$, $\mathbb{M}_j^{Dominants}$, and $\mathbb{M}_j^{Decoys}$ as the super sets of M_{ij}^o, $M_{ij}^{Dominants}$ and M_{ij}^{Decoys}, respectively. We now can express consumer i's consideration set M_i as a combination of j, M_{ij}^o, $M_{ij}^{Dominants}$ and M_{ij}^{Decoys}, and define the choice probability $s_i(j|M_i)$ accordingly:[2]

$$s_i(j|M_i) = \begin{cases} 0, & \text{if } j \notin M_i \\ s_i(j|j \cup M_{ij}^o), & \text{if } j \in M_i \ \& \ M_{ij}^{Dominants} = \varnothing \ \& \ M_{ij}^{Decoys} = \varnothing \\ 0, & \text{if } j \in M_i \ \& \ M_{ij}^{Dominants} \neq \varnothing \\ s_i(j|j \cup M_{ij}^o \cup M_{ij}^{Decoys}), & \text{if } j \in M_i \ \& \ M_{ij}^{Decoys} \neq \varnothing \ \& \ M_{ij}^{Dominants} = \varnothing. \end{cases} \tag{B.2}$$

In the above equation, j has a choice probability of zero in the first case simply because it is not in the consideration set. The choice probability is also zero in the third case, because we assume that consumers are rational, and once a dominant is included in the choice set, the inferior decoy diamond j will never be purchased. Also, notice that in the last case, the consumer will only choose an option from the subset, not from M_{ij}^{Decoys}, because diamonds in M_{ij}^{Decoys} are inferior to option j and a rational consumer would not buy those decoys. In other words, the effective choice set becomes the same as the second case. If there is no DE, $s_i(j|j \cup M_{ij}^o) = s_i(j|j \cup M_{ij}^o \cup M_{ij}^{Decoys})$; however, if there is DE, we would expect the presence of diamonds from M_{ij}^{Decoys} to increase the attractiveness and thus the choice probability of diamond j. To capture that demand boost due to the DE, we denote the relationship between the two conditional choice probabilities as follows.

$$s_i(j|j \cup M_{ij}^o \cup M_{ij}^{Decoys}) = s_i(j|j \cup M_{ij}^o) \times q_i(j, M_{ij}^o, M_{ij}^{Decoys}), \tag{B.3}$$

where $q_i(j, M_{ij}^o, M_{ij}^{Decoys})$ is a scalar that affects the choice probability of option j, and is a function

[2] The functional form of $s_i(j|M_i)$ could be very general. Since we do not have individual-level data, we do not specify its functional form here. A natural choice could be the multinomial logit model.

of dominance relationships in the choice set. Note that, $q_i(j, M^o_{ij}, M^{Decoys}_{ij}) = 1$ if there is no DE; $q_i(j, M^o_{ij}, M^{Decoys}_{ij}) > 1$ if there is DE. Given Equations (B.2) and (B.3), we next aggregate the choice probability over the possible consideration sets to derive the continuous time diamond hazard from Equation (B.1) as follows:

$$\begin{aligned}
\omega_j(\tau_i) = & \sum_{M_i \in \mathbb{M}} Pr(M_i) \times s_i(j|M_i) \\
= & \sum_{M^o_{ij} \in \mathbb{M}^o_j} Pr(M_i = j \cup M^o_{ij}) s_i(j|j \cup M^o_{ij}) + \\
& \sum_{M^o_{ij} \in \mathbb{M}^o_j} \sum_{M^{Decoys}_{ij} \in \mathbb{M}^{Decoys}_j} Pr(M_i = j \cup M^o_{ij} \cup M^{Decoys}_{ij}) s_i(j|j \cup M^o_{ij} \cup M^{Decoys}_{ij}) \\
= & \sum_{M^o_{ij} \in \mathbb{M}^o_j} Pr(M_i = j \cup M^o_{ij}) s_i(j|j \cup M^o_{ij}) + \\
& \sum_{M^o_{ij} \in \mathbb{M}^o_j} \sum_{M^{Decoys}_{ij} \in \mathbb{M}^{Decoys}_j} Pr(M_i = j \cup M^o_{ij} \cup M^{Decoys}_{ij}) s_i(j|j \cup M^o_{ij}) q_i(j, M^o_{ij}, M^{Decoys}_{ij}) \\
= & Pr(j \in M_i) Pr(M^{Decoys}_{ij} = \varnothing) Pr(M^{Dominants}_{ij} = \varnothing) \sum_{M^o_{ij} \in \mathbb{M}^o_j} Pr(M^o_{ij} \in M_i) s_i(j|j \cup M^o_{ij}) + \\
& Pr(j \in M_i) Pr(M^{Dominants}_{ij} = \varnothing) \sum_{M^o_{ij} \in \mathbb{M}^o_j} Pr(M^o_{ij} \in M_i) s_i(j|j \cup M^o_{ij}) \times \\
& \sum_{M^{Decoys}_{ij} \in \mathbb{M}^{Decoys}_j, M^{Decoys}_{ij} \neq \varnothing} Pr(M^{Decoys}_{ij} \in M_i) q_i(j, M^o_{ij}, M^{Decoys}_{ij})
\end{aligned} \quad . \tag{B.4}$$

In the above derivation, we first use Equation (B.2) (i.e., two non-zero choice probabilities on the second and fourth lines) to obtain the second line of Equation (B.4). Next, we use Equation B.3 to move from the second to the third line of Equation (B.4). We next use the independence assumption to move from the third to the fourth line of Equation (B.4). Next, we define the following:

$$s_i(j|j, \mathbb{M}^o_j) = \sum_{M^o_{ij} \in \mathbb{M}^o_j} Pr(M^o_{ij} \in M_i) s_i(j|j \cup M^o_{ij}). \tag{B.5}$$

Given $s_i(j|j, \mathbb{M}^o_j)$ from Equation (B.5), we then define the following:

$$q_i(j, \mathbb{M}^o_j, \mathbb{M}^{Decoys}_j) = \sum_{M^o_{ij} \in \mathbb{M}^o_j} Pr(M^o_{ij} \in M_i) s_i(j|j \cup M^o_{ij}) q_i(j, M^o_{ij}, \mathbb{M}^{Decoys}_j) / s_i(j|j, \mathbb{M}^o_j), \tag{B.6}$$

where $q_i(j|M^o_{ij}, \mathbb{M}^{Decoys}_j) = \sum_{M^{Decoys}_{ij} \in \mathbb{M}^{Decoys}_j, M^{Decoys}_{ij} \neq \varnothing} Pr(M^{Decoys}_{ij} \in M_i) q_i(j, M^o_{ij}, M^{Decoys}_{ij})$ $/Pr(M^{Decoys}_{ij} \neq \varnothing)$. Plugging $s_i(j|j, \mathbb{M}^o_j)$ and $q_i(j, \mathbb{M}^o_j, \mathbb{M}^{Decoys}_j)$ from Equation (B.5) and (B.6) into (B.4), and rearranging the terms, we can simplify Equation (B.4) as follows:

$$
\begin{aligned}
\omega_j(\tau_i) = &\sum_{M_i \in \mathbb{M}} Pr(M_i) \times s_i(j|M_i) \\
=&Pr(j \in M_i) Pr(M_{ij}^{Decoys} = \varnothing) Pr(M_{ij}^{Dominants} = \varnothing) s_i(j|j, \mathbb{M}_j^o) + \\
&Pr(j \in M_i) Pr(M_{ij}^{Decoys} \neq \varnothing) Pr(M_{ij}^{Dominants} = \varnothing) s_i(j|j, \mathbb{M}_j^o) q_i(j, \mathbb{M}_j^o, \mathbb{M}_j^{Decoys}) \\
=&Pr(j \in M_i) s_i(j|j, \mathbb{M}_j^o) Pr(M_{ij}^{Dominants} = \varnothing) \times \\
&\left[Pr(M_{ij}^{Decoys} = \varnothing) + Pr(M_{ij}^{Decoys} \neq \varnothing) q_i(j, \mathbb{M}_j^o, \mathbb{M}_j^{Decoys}) \right].
\end{aligned}
\tag{B.7}
$$

Note that $q_i(j, M_j^o, \mathbb{M}_j^{Decoys})$ is the average of $q_i(j, M_j^o, M_j^{Decoys})$ weighted by the probability of having M_{ij}^{Decoys}; and $q_i(j|\mathbb{M}_{ij}^o, \mathbb{M}_j^{Decoys})$ is the average of $q_i(j|M_{ij}^o, \mathbb{M}_j^{Decoys})$ weighted by the probability of having M_{ij}^o. Thus, the final $q_i(j, M_{ij}^o, \mathbb{M}_j^{Decoys})$ is the average of all the $q_i(j|.)$ weighted by the probability of all the possible combinations of M_{ij}^{Decoys} and M_{ij}^o.

Equation (B.7) tells us that the aggregate sales hazard in the continuous time can be decomposed to the sum of two components: 1) the probability that no decoy-dominant relationships are included in the consideration set times the aggregate choice probability over all no dominants–no decoys sets (i.e., the baseline choice probability); and 2) the probability that a diamond's decoys but not dominants are included in the consideration set times the baseline choice probability multiplied by an additional aggregate term q that depends on the decoy-dominant structure. From the DE theory, we know that for each specific consideration set, $q_i(j, M_{ij}^o, M_j^{Decoys}) \geq 1$, i.e., it cannot be the case that adding dominated options to the choice set would reduce the choice share of a dominant. When this q function aggregates to the market level, it is a weighted average of all the consideration set-level qs. Therefore, the aggregate $q_i(j, \mathbb{M}_{ij}^o, \mathbb{M}_j^{Decoys}) \geq 1$. As $q_i(j, M_{ij}^o, M_j^{Decoys}) \geq 1$ captures the potential DE at the individual-choice level, testing whether $q_i(j, \mathbb{M}_{ij}^o, \mathbb{M}_j^{Decoys}) \geq 1$ equals testing whether on average the DE exists in individual choices, i.e., whether $Eq_i(j, M_{ij}^o, M_j^{Decoys}) \geq 1$.

B.3 Daily Diamond Sales Hazard

We now derive the aggregate-level sales hazard for diamond j at discrete time, i.e., day t. We assume n_t potential consumers arrive randomly during day t, and each consumer can be represented by consumer i. By definition in survival analysis, we know the survival function for diamond j at the end of day t, thus $S_j(t)$ is defined as:

$$
S_j(t) = e^{-H_j(t)}, \tag{B.8}
$$

where $H_j(t) = \int_0^t \omega_j(\tau_i) d\tau$ is the cumulative hazard function.

We use $Ph_j(t)$ to denote the hazard in the discrete time for diamond j on day t:

$$\begin{aligned} Ph_j(t) &= \frac{S_j(t-1) - S_j(t)}{S_j(t-1)} \\ &= 1 - e^{-(H_j(t) - H_j(t-1))} \\ &= 1 - e^{-h_j(t)}, \end{aligned} \tag{B.9}$$

where $h_j(t) = n_t \omega_j(\tau_i)$, and the corresponding survival function is $S_j(t) = \Pi_{k=1}^t e^{-h_j(k)}$.

Functional Form Specification

Based on our derivation of $\omega_j(\tau_i)$ in Equation (B.7), we now can write $h_j(t)$ in Equation (B.9) as follows:

$$\begin{aligned} h_j(t) =& n_t \omega_j(\tau_i) \\ =& n_t Pr(j \in M_i) s_i(j|j, \mathbb{M}_j^o) Pr(M_{ij}^{Dominants} = \varnothing) \times \\ & \left[Pr(M_{ij}^{Decoys} = \varnothing) + Pr(M_{ij}^{Decoys} \neq \varnothing) q_i(j, \mathbb{M}_j^o, \mathbb{M}_j^{Decoys}) \right]. \end{aligned} \tag{B.10}$$

Note that, in the above equation,$n_t Pr(j \in M_i) s_i(j|j, \mathbb{M}_j^o)$ component represents the daily-diamond hazard for diamond j. In our functional specification, we model $n_t Pr(j \in M_i) s_i(j|j, \mathbb{M}_j^o)$ as the daily-diamond hazard ($\psi_j(.)$) as a function of diamond price segments, 4Cs, price, daily demand proxies and competition from other diamonds. Note also that, in the above equation, $Pr(M_{ij}^{Dominants} = \varnothing)[Pr(M_{ij}^{Decoys} = \varnothing) + Pr(M_{ij}^{Decoys} \neq \varnothing) q_i(j, \mathbb{M}_j^o, \mathbb{M}_j^{Decoys})]$ component represents the probability of diamond dominance detection and the boost in sales upon dominant detection. In our functional specification, we use the DE hazard ($\phi_j(.)$) to control for this component. Therefore, we operationalize the hazard $h_j(t)$ at discrete time in Equation(B.10) as follows:

$$h_j(t) = \psi_j(.)\phi(.). \tag{B.11}$$

Note that Equation (B.11) above is the same as Equation (1) in our manuscript. This completes our derivation of the diamond-specific proportional hazard from the individual consumer primitives.