# Profiting from the Decoy Effect:
# A Case Study of an Online Diamond Retailer

Chunhua Wu and Koray Cosguner*

**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, it has never been verified in a real product market in the literature. In this paper, we empirically test and quantify the DE in the diamond sales of a leading online jewelry retailer. 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 (11%–25%) in the diamond market; however, upon detection, the DE increases dominant diamonds' sale hazards significantly (1.8–3.2 times). In terms of the managerial significance, we find that the DE substantially increases the diamond retailer's gross profit by 14.3%. We further conduct simulation studies to understand the DE's profit impact under various dominance scenarios.

**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 research idea originated from the first author's engagement ring purchase experience. The authors thank the review team and 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, 5th Empirical & Theoretical (ET) Symposium, and the 2017 UW & UBC Annual Marketing Conference for their valuable feedback. The authors also thank Bridd Xu at Zbird.com for providing industry insights.

# 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 get larger 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 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 product markets was 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 market but also illustrates its managerial significance through quantifying the substantial profit impact.

Even though it has been almost four decades since the DE was introduced, to the best of our knowledge, there has been no empirical study that tests and quantifies the DE with field data. To achieve this, one must consider and resolve a few key challenges. First, a researcher needs to *calibrate* decoy–dominant relationships among product alternatives. Since products typically have horizontal attributes—such as brand, taste, size, and packaging—and consumers have heterogeneous preferences for them, decoys to some consumers may not be decoys to others. Therefore, in most product markets, strict decoy–dominant relationships may not exist, let alone permit calibration. Second, consumers should be able to *detect* the decoy–dominant relationships. 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 consumer product category, alternatives have a much larger number of attributes, e.g., brand, size, design, color, weight, packaging, taste, and price, to name a few. Thus, it is much harder for consumers to fully evaluate the trade-offs and detect existing decoy–dominant relationships, so that, as noted in Huber et al. (1982), "the effect may be lessened," and the lack of detection becomes one important mitigating factor of the DE (Huber et al., 2014). For consumers to detect decoy–dominant relationships, the choice decision must be salient and require enough cognitive processing that consumers' preferences can be constructed rather than already 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 impact the choices of consumers who have already developed clear preferences for existing alternatives. From a technical perspective, Simonson (2014) further called for a systematic study separating decoy–dominant detection from the DE (i.e., sales boost in dominants upon detection). Lastly, it is quite possible that decoy pricing strategies, i.e., introducing decoy–dominant relationships by charging higher prices for the same or inferior quality products, may not generate positive profit impacts for firms, which limits the existence of the *decoy pricing practice* in the real world.

Due to the above-mentioned challenges, to empirically test and quantify the DE, we need data from a product category 1) with a reasonably small number of vertical product attributes; 2) that is important to consumers but not repeatedly purchased; and 3) that has the decoy pricing practice. The online diamond market is a highly appropriate case for this purpose: diamonds are commodity-type products with quality clearly defined on a few vertical attributes such as carat, color, cut, and clarity (4Cs); diamond purchases are important but not repeated lifetime decisions; and, finally, we frequently observe decoy pricing patterns in the online diamond market.

We use diamond pricing and sales data from a major U.S. online jewelry retailer to empirically test the DE's existence, quantify its magnitude, and show its significance for firm profitability. We validate that a diamond's value is predominantly determined by its most important vertical attributes, i.e., the 4Cs. Yet we also observe significant price variation in the market for diamonds with the same characteristics, and we construct dyadic decoy–dominant relationships among diamond pairs based on their 4Cs and prices for our analysis. Data patterns show that dominants have significantly larger sale probabilities than diamonds that are neither decoys nor dominants, while the opposite is true for decoys. The effect for dominants, interestingly, doubles that for decoys (29% vs. 14%). Sales share regressions reveal that increasing the proportion of dominants in the market would extract a dis-proportionally larger sales share from other diamonds. These data patterns are consistent with the predictions of the DE but could also be explained *qualitatively* by alternative mechanisms such as the reference price effect, and consumer search. We develop formal statistical tests to show that, *quantitatively*, there are strong statistical supports favoring the DE over the reference price explanation and that the observed price variation are consistent with the DE coexisting with consumer search but cannot be explained solely by consumer search.

Given the data providing evidence of the DE, we formally develop a proportional hazard framework in our empirical analysis. Modeling the impact of decoy diamonds on sales of their dominants requires us to separate the market-level decoy–dominant detection from the sales boost once diamonds are detected as dominants. To achieve that, we incorporate two critical components into our proposed hazard framework: *market-level decoy–dominant detection probability* and *dominant boost hazard* upon dominant detection. Thus, in our setting, upon the dominant detection, the *dominant boost hazard* component is used to test the existence and measure the magnitude of the DE.

In the estimation, we use a diamond's characteristics, daily demand factors, competition from other similar diamonds, and the observed decoy–dominant structure to control for the differences in the sale hazards of diamonds. To capture potential consumer heterogeneity in response to decoy pricing, we divide the diamonds into three price segments (low: \$2K–\$5K; medium: \$5K–\$10K; and high: \$10K–\$20K) and estimate segment-specific detection probabilities and sale boosts upon dominant detection. Further, to capture the unobserved heterogeneity and correlations in diamond sales, we use a diamond-level random effect specification. Results suggest that the market-level detection probability for a decoy (or dominant) diamond is quite low, ranging from 11% in the high-price segment to 25% in the low-price segment. The low decoy–dominant detection probabilities confirm that modeling the detection probabilities explicitly is required to quantify the DE accurately.

As opposed to low detection probabilities, upon detection, we find that a dominant diamond's sale hazard gets 2.7, 1.8, and 3.2 times larger in the low-, medium-, and high-price segments, respectively. This finding validates the DE in the field in response to recent studies questioning this aspect, including Frederick et al. (2014) and Yang and Lynn (2014). Next, we quantify the profit impact of the DE and find that the DE improves the retailer's gross profit by 14.3%. This finding shows that the DE is not only real but also highly substantive managerially. Finally, through additional simulation studies, we show that the profit impact of the DE gets larger as the number of decoys increases; as the price variation increases; and as the market-level detection probability remains similar, decreases, and increases in the low-, medium-, and high-price segments, respectively.

To reiterate our contribution, we advance 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 market; 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.

## 2 Literature Review

This paper contributes to two streams of literature: the general consumer behavior literature on context-dependent choices (in particular, the DE) 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 they 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 regularity and IIA 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 moderating the DE and related context effects (see, for example, Ratneshwar et al., 1987; Heath et al., 1995; Khan et al., 2011; Müller et al., 2014; Guo and Wang, 2016; Morewedge et al., 2018). In this domain, Khan et al. (2011) studied the influence of choice construal on context effects and found that high construal as opposed to low increases the size of the DE. 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 and found that the response time can mediate the compromise effect, but the context information cannot. Despite an ample amount of research devoted to the DE, empirical test and quantification of the effect in a real product market has not been achieved yet. Our paper fills this important gap by validating the DE in the field.

Despite its wide acceptance, the limits and boundaries of the DE have been debated by multiple studies recently. Frederick et al. (2014) stated that the DE could 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. 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. 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 market.

In response to Frederick et al. (2014) and Yang and Lynn (2014), 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, they may not be able to make their decisions based on existing decoy–dominant configurations, especially if such configurations are difficult to detect. He thus called for a systematic study on the drivers of decoy–dominant detection. In addition, Huber et al. (2014) recognized the lack of practice of the DE in today's product markets, noting that it is difficult to observe the DE in a real product market since the detection of decoys is typically very hard for consumers due to numerous alternatives with many attributes. With this research, we respond to the call of Simonson (2014) by explicitly modeling decoy–dominant detection in the studied online diamond market to empirically quantify the DE.

Our study is also closely related to consumer choice models in 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. A few empirical and analytical methods have been developed to incorporate the 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 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 decoy 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 that explicitly accounts for the DE using pricing and sales data.

Since the studied online diamond retailer offers a large number of diamonds daily, as discussed earlier we separate the detection of decoys/dominants and the boost in sales upon dominant detection. Accordingly, our decoy/dominant detection component serves the role of a market-level consideration model. Due to that, our research is also related to the empirical literature that separately considers consumers' consideration sets and choices. Existing studies in that domain use consumer-level data to model consumers' consideration and choice decisions together. Some studies use purchase data only (see, for example, Siddarth et al., 1995; Chiang et al., 1998; Van Nierop et al., 2010), while others use purchase along with search-related data (see, for example, De los Santos et al., 2012; Honka, 2014). Unlike those studies, we don't observe either purchase or search behaviors at the individual level. Hence, we derive a diamond-level hazard model from individual primitives, including consideration sets and choice. Consequently, our hazard specification provides a framework to separately estimate the market-level decoy/dominant detection from the DE by solely utilizing the aggregate product pricing and sales data.

In the following sections, we first describe the online diamond market, our dataset, and how we calibrate the decoy–dominant relationships, then we provide data evidence on the existence of the DE in this market. We develop our model framework in Section 5, and we present the estimation results and illustrate the DE's managerial implications in the subsequent two sections. Finally, we conclude with a discussion of the current study's limitations and offer directions for future research.

# 3 Data

## 3.1 Online Diamond Market

Several U.S. retailers emerged in the online market for diamonds and jewelry products in the past two decades. We use panel data from a major retailer in this market. In fiscal year 2015, the retailer reported net sales of $480 million. According to industry reports, it has around 50% market share of the U.S. online diamond market, with sales approximately three times greater than its closest competitor. These figures clearly indicate that the retailer is the leading player in the market.

The retailer sells a variety of jewelry products to end consumers, such as unbranded loose diamonds, gemstones, engagement and wedding rings,[1] bracelets, necklaces, and earrings. Loose diamonds account for the core part of 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

[1] Buying a diamond ring from the retailer requires a consumer to choose his/her loose diamond first and then a ring setting. Consumers pay the total price, and 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.

drop-shipping business model, i.e., the retailer, in most cases, does not physically carry inventories of loose diamonds listed on its website. Instead, it purchases diamonds from corresponding suppliers when consumers place their orders with 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 more than a hundred thousand diamonds every day.

In this setting, suppliers list their diamonds on the retailer's website and establish wholesale prices. The retailer then adds a fixed percentage markup to the wholesale prices. Per the retailer's annual report, the markup is fixed at around 18–20% for all diamonds. Because the retailer chooses a fixed margin over wholesale prices, the decoy pricing structure ultimately comes from the suppliers.[2] Nevertheless, consumers are expected to respond to the decoy–dominant structure regardless of whether the retailer or the suppliers create it. That being the case, as we demonstrate later, the dominance structure still affects the retailer's sales and ultimately its profitability to a large extent.

Since more than a hundred thousand diamonds are listed each day, the retailer provides filtering and sorting tools that help consumers to search for diamonds. On the website, a consumer can filter diamonds based on a desired range of diamond characteristics (such as price, carat, clarity, etc.). The website then returns a list of all the diamonds that fall into the filtering criteria on a single page in the default ascending price order. This one-page structure requires the consumer to scroll down to check all diamonds filtered. The web page displays each diamond in a row with its carat, cut, color, clarity (4Cs henceforth), and price information. The consumer needs to further click into a diamond's details page to check other less-significant characteristics such as symmetry and polish. The website also allows the consumer to sort the diamonds based on any one of the 4Cs or price. Given abundant diamonds from the retailer, the consumer would still face a long list of diamonds (hundreds to thousands) even after a few rounds of filtering and sorting. The list typically includes many decoys and dominants that are not easy to detect without laborious investigation. In addition, because consumers are presented with a list of diamonds according to their set criteria, this search process resembles the non-sequential search that is consistent with our derivation of the

[2] We note that the supplier price variation may exist due to multiple reasons. First, due to 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 observed price variation in our data.

diamond-level hazard model from individual primitives (see Web Appendix B for the details).

## 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 and daily prices until the diamond is sold. In the data, diamond prices typically change over time: 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 the end of the observation period.

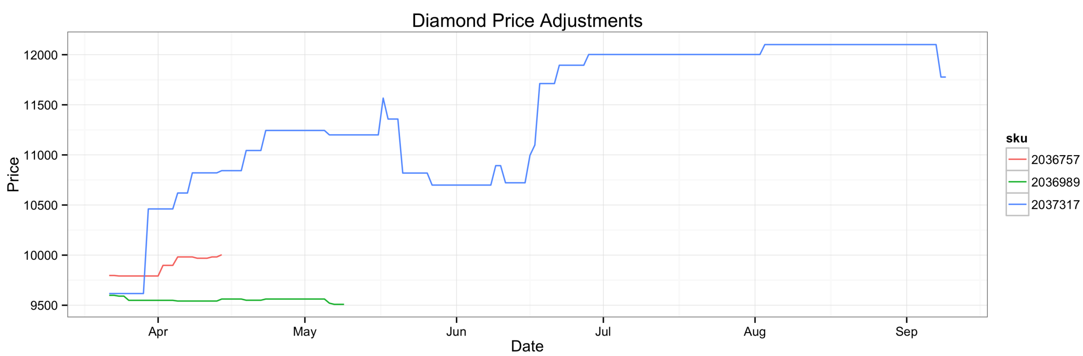


Figure 1: Diamond Price Patterns Over Time

As seen in Figure 1, the diamond prices can go up and down, and each diamond may have a unique price pattern 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 its unique SKU number.[3] On average, it takes about 50 days to sell a diamond.

In our analysis, we focus specifically on round-shaped diamonds with prices ranging between $2K and $20K. Round-shaped diamonds are the most popular ones among those listed (74%) and

[3]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. Accordingly, until a diamond is sold, its supplier is expected to keep listing the diamond on the retailer's website. That said, we acknowledge the sale time of a diamond to be inferred based on its removal day as a limitation of the current research.

sold (78%).[4] Diamonds in different price ranges might be more attractive to different segments of buyers with various budget levels. To account for the potential heterogeneity in the DE across 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. We run a linear 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 variation across periods, we also add day fixed effects to the regression. We report the regression results in Table 1. The adjusted R-squared measure for the model with 4Cs, along with day fixed effects, is as high as 96.67%. Individual regressions for each day yield adjusted R-squared measures ranging between 94.92% and 96.57%. The results provide evidence that 4Cs are the predominant attributes 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 regression model. Overall, the R-squared measure does not improve. Moreover, these additional variables have mostly insignificant estimates that are notably smaller in magnitude compared to the 4C estimates. For example, the implied price difference contributed by symmetry and polish turns out to be less than 0.5%. Thus, we have strong statistical evidence to conclude that 4Cs of a diamond can represent its quality very precisely. Accordingly, in our analysis, we label each unique 4C combination as a *grade*.

Characterizing a diamond as a combination of its 4Cs and price 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 and within-day price variation. Specifically, in the data, the average of the ratios of price standard deviation to the mean price at each day-grade combination is 0.1, indicating a sufficiently large within-grade and within-day price variation. This variation is essential to characterize the dyadic decoy–dominant relationships among every diamond-pair, as discussed next.

[4] Diamond shape can be considered as a horizontal attribute. Thus, including only round-shaped diamonds would not affect our decoy–dominant constructions since no decoy–dominant relationships exist across diamond shapes. Further, we believe most consumers commit to a particular shape before choosing among other attributes.

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

| **Variable** | | **Estimate** | **S.E.** |
|---|---|---|---|
| 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 $^{*}$ and $^{**}$ are significant at the 0.10 and 0.05 levels, respectively.

### 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 $A$ but is priced at least 5% higher.[5] Under the two decoy definitions, for any two diamonds on a

[5] Under the strict definition, two same-grade diamonds with different prices must have a dominance relationship. However, in real purchase situations, consumers may not care much about (or even notice) small price differences. Thus, we use a conservative approach, defining a dominance relationship only if the price difference between the two is larger than 5%. This 5% rule 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. Our regressions show that the price premium contributed by these non-critical attributes is less than 0.5%. As a robustness check, we replicate our analysis under the 1% and 10% rules and obtain similar results.

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 ($A \sim B$).

In our data sample, every pairwise decoy–dominant relationship between all listed diamonds is constructed for each day.[6] For a particular diamond $j$, for each day $t$, we create two measures: 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, while the distribution is right-skewed. Since the number of diamonds varies across different grades significantly, there exist large variation in the number of decoys/dominants across grades. Hence, we normalize the two measures by dividing them by the number of diamonds in the grade on the same day, and label the grade-level percentage measures of decoys and dominants as $R_{jt}^{Decoy}$ and $R_{jt}^{Dominant}$, respectively.

# 4 Data Evidence of the DE

In this section, we discuss some data patterns and reduced-form analyses that are suggestive of the DE. We show that important data patterns cannot be solely explained by alternative mechanisms such as the reference price effect or consumer search without considering the DE.

## 4.1 Dominance Types and Diamond Sales

Based on the dominance construction, on day $t$, diamond $j$ can belong to one of the four mutually exclusive groups: 1) neither decoy nor dominant: $R_{jt}^{Decoy} = 0$ and $R_{jt}^{Dominant} = 0$; 2) decoy only: $R_{jt}^{Decoy} = 0$ and $R_{jt}^{Dominant} > 0$; 3) dominant only: $R_{jt}^{Decoy} > 0$ and $R_{jt}^{Dominant} = 0$; and 4) both decoy and dominant: $R_{jt}^{Decoy} > 0$ and $R_{jt}^{Dominant} > 0$. The middle column of Table 2 shows the count of diamond-day observations for each diamond type. Due to the significant within-grade price variation, we observe that a majority of the diamonds fall into the *both decoy and dominant* type.

[6] We would like to acknowledge that omitting diamonds from competing retailers may create biases in the estimated size of the DE if the majority of consumers search and compare diamonds from multiple retailers. Given that the studied retailer is the dominant player in the market with about 50% market share, we expect such practices to be less common. Even if consumers search across competitors, it is laborious for them to compare diamonds from multiple retailers altogether since each retailer provides its search/filtering tools, i.e., pooling diamonds from different retailers is not trivial. As discussed earlier, even with the focal retailer's filtering tools, the lists of diamonds consumers face tend to be very large. In other words, it is very difficult for consumers to detect the decoy–dominant relationships from a single retailer—a fact also confirmed by our estimated decoy–dominant detection probabilities—let alone comparing across sites. Finally, each retailer's brand may be perceived as a horizontal attribute, making decoy–dominant calibrations invalid across retailers. That being said, if consumer search data across multiple retailers become available, modeling the DE beyond a single retailer can be an interesting future research direction.

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.

We summarize the percentage of diamonds sold in the total diamond-day observations across different diamond types in the last column of Table 2. Each cell is calculated by dividing the number of diamonds sold in each diamond type by that type's total diamond-day observations. For example, there are 404,283 *decoy only* diamond-day observations, out of which 6,792 were sold, yielding the average sale probability of 1.68%. Table 2 shows that *decoy only* diamonds have the lowest average sale probability (1.68%), while the opposite is true for *dominant only* diamonds (2.53%). The *both* type diamonds have a slightly higher average sale probability than the *neither* type.

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% |

In addition, we explore how the overall dominance structure impacts the sales shares of different diamond types using two linear regressions. We use the daily sales share of decoys (in percentage) and the sales share of dominants as the dependent variables and include price-segment dummies and percentage of *decoy only* and percentage of *dominant only* diamonds in each segment as independent variables. Results reported in Table 3 show that the percentage of *decoy only* diamonds significantly increases the decoys' sales share, while the percentage of *dominant only* diamonds significantly increases the sales share of dominants but reduces that of decoys.

Table 3: Diamond Sales Share Regression

| **Variable** | **Decoy Sales Share** | | **Dominant Sales Share** | |
|---|---|---|---|---|
| | **Estimate** | **S.E.** | **Estimate** | **S.E.** |
| 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 * and ** are significant at the 0.10 and 0.05 levels, respectively. The average value of % Decoys and % Dominants are 0.14 and 0.13, respectively.

Qualitatively, the patterns presented in Tables (2) and (3) could be explained by reference price effect or consumer search. Dominants are in general more preferred by consumers because their prices are relatively lower than comparable diamonds, whereas the opposite is the case for decoys (reference price effect). Beyond the relative price disadvantage, a *decoy only* diamond may also be detected as a decoy by a segment of consumers who search extensively, and it would never be purchased upon detection, leading to a further reduction in the sale probability (consumer search). Notably, the average sale probability of a decoy is still much larger than zero; thus, it is critical to control for decoy detection in our analysis. The sales share regression results in Table (3) are also directionally consistent with consumer search because having relatively more *decoy only* diamonds in the market would reduce the chances of consumers discovering these decoys along with their dominants. In other words, as there is a larger share of *decoy only* diamonds, the size of the market segment detecting them as decoys is expected to decrease, leading the sales share of decoys to increase. Similarly, as the percentage of *dominant only* diamonds increases, consumers are more likely to include them in their search process, leading to an increase in the sales share of dominants.

Quantitatively, however, the magnitude of differences in the average sale probabilities across the diamond types and the magnitude of the regression estimates could be suggestive of additional effects beyond reference price and consumer search. The average sale probability difference between a *dominant only* and a *neither* type diamond (0.57%) is twice the difference between a *neither type* and a *decoy only* diamond (0.28%). Since buying decoys and dominants can be seen (by consumers) as monetary losses and gains, respectively, based on the prospect theory (Tversky, 1972), reference price effect would predict the exact *opposite* pattern, i.e., the former probability difference is expected to be smaller than the latter. Consumer search, on the other hand, could explain the significant reduction in the average sale probability for decoys (due to the segment detecting them never purchasing them), but it could hardly explain the significant boost in the sales of dominants, especially when the effect is twice that of decoys. In addition, we would expect the sales share of dominants to increase proportionally with the increase in the percentage of dominants based on random consumer search, yet the estimated elasticity is as high as 1.97, significantly larger than 1.0, suggesting that including additional *dominant only* diamonds would extract a disproportionally larger sales share from other diamonds. All this evidence suggests that there is a significant boost effect in the sale probabilities for dominants, which is not completely explained by reference price

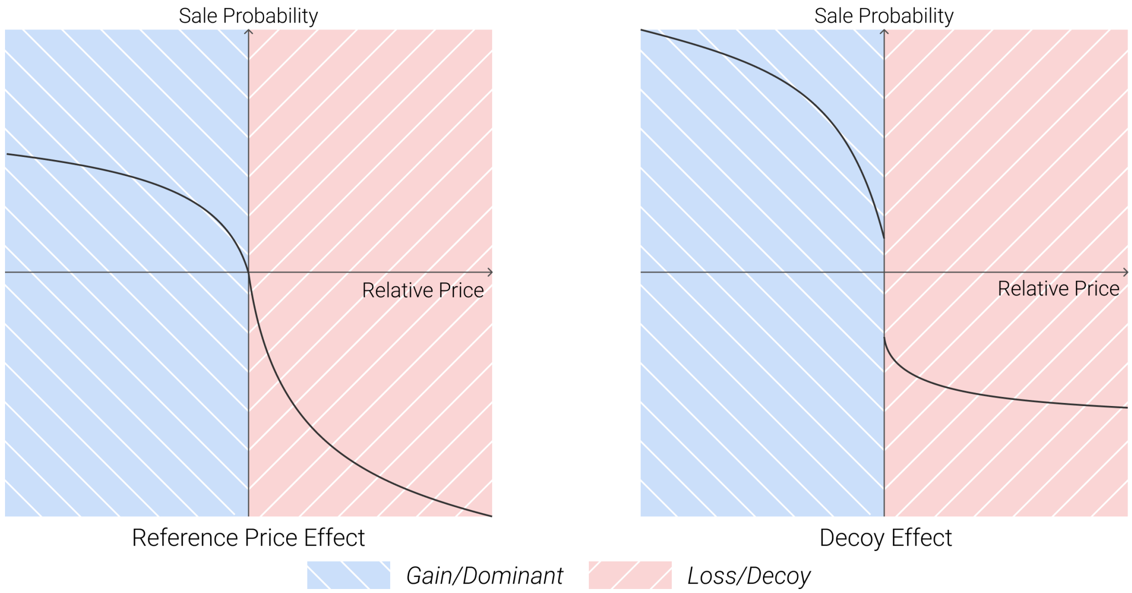


Figure 2: Sale Probability Function According to Different Theories

effect or consumer search, yet is very consistent with the predictions from the DE.

We further develop formal statistical tests to show evidence consistent with the DE instead of reference price effect and beyond consumer search explanations in the following subsections.

## 4.2 Decoy Effect vs. Reference Price Effect

The marketing literature has extensively examined the importance of relative price comparisons in consumers' purchase decisions (Monroe, 1973). The reference prices could be formed based on contextual factors such as the distribution of market prices (Biswas and Blair, 1991; Rajendran and Tellis, 1994), or temporal anchoring stimuli such as a product's past prices (Lynch Jr et al., 1991; Rajendran and Tellis, 1994). In our particular context, given that diamonds are not repeat-purchase products and there are thousands of diamonds available each day, it is more likely that consumers form reference prices based on contextual factors rather than temporal anchors (Kalyanaram and Winer, 1995). When consumers use average grade-level price as the reference, the prospect theory (Kahneman, 1979) predicts that 1) consumers respond positively to dominants (in the monetary gain domain) and negatively to decoys (in the loss domain); and 2) the response to decoys would be stronger than to dominants (of the same size) due to loss aversion.

Even though the reference price effect (RPE henceforth) and decoy effect may look similar at first glance, we would like to emphasize a few notable differences that we illustrate in Figure 2. First, the RPE gradually vanishes as the monetary gains or losses approach zero. In other words, the RPE predicts the sale probability function to be continuous as the price of a diamond moves from the gain to the loss domain (i.e., around the reference price point). To the contrary, under the decoy effect, we expect a discrete jump in the sale probability function around the reference point—because, as a non-decoy diamond turns into a decoy one, a market segment emerges with zero probability of purchasing it upon detection, resulting in a downward jump in the sale probability function around the decoy threshold price. The opposite becomes the case as a non-dominant diamond turns into a dominant one: a market segment with significantly large purchase probability emerges, and this segment induces the sale probability to jump upward around the dominant threshold price. Thus, the sale probability function is expected to be discontinuous around the reference price point with the decoy effect. Further, the magnitude of the discrete jump (around the reference price) depends on the market-level detection probabilities for decoys; and on both the market-level detection probabilities and the strength of the DE (i.e., sales boost) for dominants.

Second, the predicted relative effects across the domains (loss to gain/decoy to dominant) are different under the RPE and the DE. The RPE predicts the slope of the sale probability function to be steeper for decoys than for dominants. We would argue the opposite is more likely to be the case under the DE explanation, because further reducing the prices of dominants would not only increase the size of the consumer segment who detects them as dominants but also elevate the attractiveness of these dominants (i.e., the sales boost level) upon detection. However, as prices of decoys increase, it gets easier for consumers to detect them, i.e., the size of the market segment that fails to detect them decreases. Note that only this shrinking segment would respond to the increasing prices (or monetary losses) of decoys since the other segment would eliminate them upon detection irrespective of the loss amount's size. Therefore, with the DE, the sale probability function is expected to have a steeper slope in the gain domain, and a flatter slope in the loss domain.

We estimate three diamond-level logistic sales regressions to test whether our data support the DE theory rather than the RPE. The results of these three specifications are reported in Table 4. In the regressions, we control for the following variables: a diamond's 4Cs, daily demand effects such as weekday and holiday dummies, the Google search indexes for diamond-related keywords, and the number of diamonds in the same grade. We define the reference price as the average price of diamonds belonging to the same grade. Central to our test, we include separate intercepts and separate price slopes (*PGain* and *PLoss* as percentage differences relative to the reference price) for diamonds in the *Gain* and *Loss* domains in the first regression. We replace the definition of *Gain* and *Loss* with *Dominant* and *Decoy*, and the relative prices with interactions with the *Dominant* and *Decoy* dummies in the second and third regressions. The difference between the last two specifications is that the definition of *Dominant* and *Decoy* is based on whether a diamond has any decoys or dominants in specification (I), while in specification (II) we use a stricter definition of whether a diamond is a *decoy only* or *dominant only* type. The regression specifications follow the tradition in empirical tests of the RPE (Hardie et al., 1993; Bell and Lattin, 2000).

Table 4: Diamond Sales Response Function with Logistic Regressions

| **Variable** | **Gain/Loss** | | **Decoy/Dominant(I)** | | **Decoy/Dominant(II)** | |
|---|---|---|---|---|---|---|
| | **Estimate** | **S.E.** | **Estimate** | **S.E.** | **Estimate** | **S.E.** |
| Controls | included | | included | | included | |
| Intercept—*Gain* | −0.028 | 0.019 | | | | |
| Intercept—*Loss* | −0.006 | 0.019 | | | | |
| Intercept—*Dominant* | | | 0.152** | 0.014 | 0.140** | 0.018 |
| Intercept—*Decoy* | | | −0.186** | 0.014 | −0.205** | 0.019 |
| Slope—*PGain* | −2.444** | 0.145 | | | | |
| Slope—*PLoss* | −1.733** | 0.156 | | | | |
| Slope—*PGain* × *Dominant* | | | −1.491** | 0.131 | −1.831** | 0.192 |
| Slope—*PLoss* × *Decoy* | | | −1.059** | 0.137 | −0.567** | 0.211 |
| Slope Difference (Loss−Gain) | 0.686** | 0.215 | 0.432** | 0.218 | 1.264** | 0.292 |
| AIC | 525,333 | | 525,016 | | 525,191 | |

*Note:* Estimates with * and ** are significant at the 0.10 and 0.05 levels, respectively.

The results show a very consistent pattern with the price coefficient more negative in the gain/dominant domain than in the loss/decoy domain across the three regressions. The differences between the slopes (loss minus gain) are all significantly positive at the 0.05 level, as shown in the table. In addition, intercepts for dominants are significantly positive while those for decoys are significantly negative in the last two regressions, indicating the existence of a discrete jump around the reference price point. This is consistent with our earlier discussion about the role of the DE. Lastly, the models with the DE outperform the RPE-only model based on the AICs. These results are all robust to alternative model specifications, such as separating the effects across the three price segments and including quadratic terms for relative prices. Overall, our test results show strong statistical support favoring the DE over the RPE, as illustrated in Figure 2.

### 4.3 Decoy Effect vs. Consumer Search

As Simonson (2014) emphasized, consumers' detection of decoy–dominant relationships is a precondition for the DE. Given the extremely large number of alternatives in the online diamond market, consumers rely on search to finalize their consideration sets. Thus, consumer search is an inherent component in our test for the DE (see our Model section). In this subsection, we develop a formal statistical test to check whether the data patterns can be solely explained by consumer search, or if there is evidence of the DE in addition to consumer search. Our test relies on the observed market price dispersion and sales information. We provide the details of this test in Web Appendix A. The intuition behind the test 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 along with consumer search, a supplier needs to consider the price optimization beyond each individual diamond because decoys would serve as "loss leaders" helping the supplier get higher expected profits from the 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 prices 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 exactly the same for the same diamond over time. However, if there exists DE along with consumer search, because of the positive profit externality from decoys to dominants, the recovered costs from the optimality conditions will be higher for decoys and lower for dominants, as compared to their true costs. This essentially leads to a positive correlation between the diamonds' recovered costs and their relative price levels. The test of the pure search effect versus the search effect along with the DE thus becomes the same as testing whether recovered costs are increasing with relative prices. Accordingly, we conduct two statistical tests: 1) using cross-diamond price variation (labeled as Test I); and 2) using within-diamond price variation over time (labeled as Test II). Results (see Table A1 in Web Appendix A) 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 the DE together with consumer search in our data.

To sum up, our data analyses yield the following results: 1) A diamond's sale probability largely depends on whether it is a decoy and/or a dominant; 2) It is critical to account for decoy-dominant detection in validating the DE; and 3) There is suggestive evidence of the DE beyond alternative mechanisms of the RPE and consumer search. Hence, empirical investigation of the decoy phenomenon requires a deliberately developed model, which we introduce in the next section.

# 5 Model

We develop a diamond-level proportional hazard model to study daily diamond sales. We control for the effect of diamond characteristics and market demand and supply factors in a baseline *daily diamond sale hazard* component. We further capture the role of the DE on diamond sales with our *dominance hazard* component. Separating the DE from other factors that affect diamond sales is a challenging task since consumer-level search and purchase behaviors are unobserved in our setting. To achieve the objective, we derive our diamond-level proportional hazard from underlying consumer primitives including *consumer arrival process*, *search*, *consideration set formation*, and *conditional choice probabilities with the embedded DE*. Please see Web Appendix B for the details of our derivation including how the DE is embedded into consumers' conditional choice probabilities, and how the proposed hazard specification is used to quantify the DE at the market level. Following this derivation, we use the following to denote the hazard that diamond $j$ will be sold on day $t$:

$$h_j(t) = \psi_j(X_{jt};\beta) \cdot \phi_j(D_{jt};\gamma), \tag{1}$$

where $\psi_j(\cdot)$ and $\phi_j(\cdot)$ are the daily diamond sale hazard and the dominance hazard components, respectively. We next discuss each component with the choice and rationale of the corresponding variables. We present an overview of the variables in Table 5.

## 5.1 Daily Diamond Sale Hazard

The daily diamond sale hazard component ($\psi_j(\cdot)$) captures the baseline daily diamond sale likelihoods without the DE consideration. In general, a diamond's daily sale likelihood depends on a few essential factors. First, since diamond buyers with various budgetary constraints have different

Table 5: List of Variables Used in Model Estimation

| | Variable Name | Description |
|---|---|---|
| **Daily Diamond Sale Hazard ($X_{jt}$)** | | |
| $K_j$ | Diamond segment | Segment dummies, low (2K–5K), medium (5K–10K), high (10K–20K) |
| $Z_t$ | Google search indexes | Daily Google search trends index of diamond-related keywords |
| | Weekday dummies | Dummy variables of weekdays |
| | Holiday dummies | Dummy variables for Valentine's Day and federal holidays |
| $H_j$ | Diamond characteristics | Dummy coded cut, color and clarity of diamond $j$, |
| | | log- of carat of diamond $j$, |
| | | indicator of carat 1.0 for diamond $j$ |
| $p_{jt}$ | Diamond price | Daily price of diamond $j$ (in 1,000) |
| $W_{jt}$ | Daily competitiveness | Log of number of diamonds of the same grade |
| **Dominance Hazard ($D_{jt}$)** | | |
| $rp_{jt}$ | Relative price index | % price difference between diamond $j$ and its grade-level average |
| $R_{jt}$ | Percentage decoys | Number of diamond $j$'s decoys / number of diamonds in $j$'s grade |
| | Percentage dominants | Number of diamond $j$'s dominants / number of diamonds in $j$'s grade |
| $sp_{jt}$ | Price standard deviation | Standard deviation of prices in diamond $j$'s grade on day $t$ |

preferences for the 4Cs, diamonds at different price levels and grades are likely to have different sale likelihoods. Second, the sale likelihood of a diamond is expected to decrease as the number of similar diamonds listed increases. Third, because consumers may have different purchase intentions on different days, the sale likelihoods of diamonds may also change over time. For example, consumers may have different purchase likelihoods on different days of a week, or special occasions and holidays. Accordingly, we model the daily diamond sale hazard $\psi_j(\cdot)$ as an exponential function of 1) diamond $j$'s price segment (dummy coded low-, medium-, or high-price, labeled as $K_j$); 2) dummy-coded cut, color, and clarity, log- of carat[7] (labeled as $H_j$); 3) its price (labeled as $p_{jt}$); 4) log- of the number of diamonds with the same 4Cs (labeled as $W_{jt}$); 5) Google search trends to capture consumer interests, weekday and holiday dummies (labeled as $Z_t$).

With $X_{jt} = \{K_j, H_j, p_{jt}, W_{jt}, Z_t\}$, and $\beta = \{\beta_K, \beta_H, \beta_p, \beta_W, \beta_Z\}$, we define the daily diamond sale hazard as,

$$\psi_j(X_{jt}; \beta) = \exp(b_j + K_j\beta_K + H_j\beta_H + p_{jt}\beta_p + W_{jt}\beta_W + Z_t\beta_Z). \tag{2}$$

We use a random effect specification, i.e., $\exp(b_j) \sim \Gamma(1, \sigma^2)$, to allow the intercept term to be diamond-specific.[8] The random coefficient specification helps in capturing unobserved correlations

[7]Since 1.0-carat diamonds are more preferred in our data, in addition to controlling for carat as a continuous variable, we use an indicator variable that takes the value of 1 if diamond $j$'s carat is equal to 1.0.

[8]We would like to note that modeling the unobserved heterogeneity beyond the intercept term is not possible in our setting due to lack of consumer-level panel data. That being said, ignoring price response heterogeneity may

in sale likelihoods among diamonds. This is important in our setting since if there exist demand shocks affecting two diamonds at the same time, without such specification, the effect of one diamond on the other one's demand would likely be inferred with biases. We adopt the Gamma distribution assumption following Lancaster (1979), where the random effect can be analytically integrated out.

## 5.2 Dominance Hazard

We capture the DE by our *dominance hazard* component $\phi_j(\cdot)$. Per our derivation in Web Appendix B, a diamond's sale likelihood further depends on 1) the size of the market segment detecting it to be a decoy/dominant; and 2) the sale boost upon it is detected as a dominant. We separate these two parts as the *market-level decoy–dominant detection* and the *dominant boost hazard*.

### 5.2.1 Market-Level Decoy–Dominant Detection

The likelihood of detecting decoys and dominants in the diamond market may depend on various context-related factors. First of all, it should naturally depend on the decoy–dominant structure in the market. As consumers typically sample diamonds based on their desired grades using the filtering and sorting tools, a diamond with a larger percentage of decoys/dominants in its grade would have a higher chance of being included together with its decoys/dominants in the consideration sets. Second, the relative price of a diamond in its grade matters because a diamond is more likely to stand out either on the top or the bottom of the returned lists under the default price-sorting design when its absolute relative price differences get larger. Third, diamond purchases are unique, first-time experiences for most consumers who have limited knowledge regarding diamond pricing. Thus, learning about the prices becomes an inherent part of their decision process. A greater variation in the prices of comparable diamonds indicates more opportunities to save, which will likely motivate consumers to spend more time on the retailer's site searching and comparing alternatives. Consequently, they would be more likely to detect existing decoy–dominant relations. As such, we expect the grade-level price dispersion to be another moderator of dominance detection.

cause overestimation of the DE's magnitude if there is selection in the market where more price-sensitive consumers enter into the market only when the market prices are low (see Bell and Lattin (2000) for the overestimation of the magnitude of loss aversion in the absence of controlling price response heterogeneity). This type of selection might be less an issue in our particular setting because a diamond is a one-time purchase product category in which consumers are less likely to have well-formed reference prices from past experiences. In addition, the average price within each grade stays almost constant week-over-week, making it less likely for the retailer to attract a substantially different consumer segment (i.e., more price-sensitive) on a weekly basis.

Based on the above rationales, we model the market-level decoy–dominant detection part of our dominance hazard as a function of the following variables: 1) percentages of decoys and dominants that diamond $j$ has ($R_{jt}^{Decoy}$ and $R_{jt}^{Dominant}$), 2) the relative price measurement ($rp_{jt}$), and 3) the price standard deviation in diamond $j$'s grade ($sp_{jt}$).

We denote the *market-level decoy and dominant detection probabilities* as $Pr_{jt}^{Decoy}(\cdot)$ and $Pr_{jt}^{Dominant}(\cdot)$, respectively. Given $D_{jt} = \{K_j, R_{jt}^{Decoy}, R_{jt}^{Dominant}, rp_{jt}, sp_{jt}\}$, we model these two terms as the following:

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

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

$$\begin{aligned} V_{jt}^{Dominant} &= K_j\gamma_0^{Dominant} + \gamma_1^{Dominant}\ln(R_{jt}^{Decoy}) + \gamma_2^{Dominant} I(rp_{jt} < 0)(-rp_{jt}) + \gamma_3^{Dominant} sp_{jt} \\ V_{jt}^{Decoy} &= K_j\gamma_0^{Decoy} + \gamma_1^{Decoy}\ln(R_{jt}^{Dominant}) + \gamma_2^{Decoy} I(rp_{jt} > 0)(rp_{jt}) + \gamma_3^{Decoy} sp_{jt} \end{aligned}. \tag{4}$$

The intercept terms ($\gamma_0^{Dominant}$ and $\gamma_0^{Decoy}$) 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 $\gamma$s capture how the percentage measures of decoys/dominants, the relative price, and standard deviation of the grade-level prices would impact the decoy–dominant detection probabilities.

### 5.2.2 Dominant Boost Hazard

Upon a diamond being detected as a dominant, the size of the boost in its sale likelihood, i.e., the DE, may also depend on a few important context-related factors. First, all else equal, dominants with more decoys are likely to become more *attractive* (DE is also called attraction effect) compared to diamonds with fewer decoys, especially when consumers engage in multiple comparisons. Second, based on the prospect theory (Kahneman, 1979), a consumer's purchase decision depends on whether the price paid is perceived as fair with respect to her reference price point. Since diamonds are not repeat-purchase products, a consumer is expected to form her reference price based on the prices of available diamonds rather than the past price histories (Mazumdar et al., 2005). Thus, upon detection of a dominant, we expect its sale likelihood to increase as the size of the price gain relative to comparable diamonds increases. Lastly, following our discussion on consumer learning in dominance detection, we expect consumers to spend less time as the variation of the grade-level prices decreases. Accordingly, we expect that with a smaller price dispersion, consumers become more likely to settle down with their detected dominants rather than continuing the search for other diamonds. As such, we expect the sale boost of a detected dominant to be more substantial as the within-grade price variation decreases. Given these rationales, we include the same set of variables as in the dominance detection component. We label $D_{jt} = \{K_j, R_{jt}^{Decoy}, R_{jt}^{Dominant}, rp_{jt}, sp_{jt}\}$, and define the dominant boost hazard $Q_{jt}$ as follows:

$$Q_{jt}(D_{jt};\gamma) = \exp\left[K_j\gamma_0^{Boost} + \gamma_1^{Boost}\ln(R_{jt}^{Decoy}) + \gamma_2^{Boost}I(rp_{jt} < 0)(-rp_{jt}) + \gamma_3^{Boost}sp_{jt}\right]. \quad (5)$$

Similar to the intercept term of the market-level decoy–dominant detection probabilities, we model the intercept term ($\gamma_0^{Boost}$) at each of the diamond price segments since the dominant boost sizes might differ across market segments. The other $\gamma$s capture how the percentage measures of decoys/dominants, the relative price, and standard deviation of the grade-level prices would impact the sales boost upon dominant detection.

Given market-level detection probabilities and dominant boost hazard specifications, we operationalize the dominance hazard as follows:

$$\phi_j(t|\cdot) = \begin{cases} 1 & \text{if } j \text{ is } \textit{Neither} \\ (1 - Pr_{jt}^{Decoy}) & \text{if } j \text{ is } \textit{Decoy Only} \\ (1 - Pr_{jt}^{Dominant}) + Pr_{jt}^{Dominant}Q_{jt} & \text{if } j \text{ is } \textit{Dominant Only} \\ (1 - Pr_{jt}^{Decoy})\left[(1 - Pr_{jt}^{Dominant}) + Pr_{jt}^{Dominant}Q_{jt}\right] & \text{if } j \text{ is } \textit{Both} \end{cases}. \quad (6)$$

We would like to note that we make an implicit assumption in the derivation of the dominance hazard in Equation (6). We assume that once a consumer detects a specific diamond to be a decoy, she would never purchase it, as she 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. We argue that it is highly unlikely for a consumer to buy a detected decoy diamond given that diamonds are high ticket item products and the monetary cost of doing so is significant.

Equation (6) shows how the dominance hazard depends on a diamond's type: if it is *neither decoy nor dominant*, the DE has no impact on the sale hazard of the diamond, i.e., the dominance 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 dominance hazard being the size of this segment, i.e., $\phi_j(\cdot) = 1 - Pr_{jt}^{Decoy}(\cdot)$. For a *dominant only* type diamond, there exist two market segments: the segment that fails to detect the diamond as a dominant, and the segment that is able to. The DE does not have any impact on the former segment (i.e., $Q_{jt}(\cdot) = 1$), while we expect a boost in sale hazard (i.e., $Q_{jt}(\cdot) > 1$) for the latter. The overall sale hazard thus becomes the expression in the third line of Equation (6). Finally, if the diamond is a *both decoy and dominant* type, it is considered for purchase only by the consumer segment that fails to detect it as a decoy, with the size of the segment being $1 - Pr_{jt}^{Decoy}(\cdot)$. Similar to the *decoy only* case, the remaining consumer segment never purchases it, i.e., $Q_{jt}(\cdot) = 0$. The segment that fails to detect the diamond as a decoy can be divided into two sub-segments: the sub-segment that fails to detect the diamond as a dominant and the one that is able to do so. Similar to the *dominant only* case, the former sub-segment with size $1 - Pr_{jt}^{Dominant}(\cdot)$ will not be impacted by DE, i.e., $Q_{jt}(\cdot) = 1$, while the sale hazard from the other sub-segment will be boosted by $Q_{jt}(\cdot) > 1$. Combining all the scenarios, for the dominance hazard, we have the expression in the last line of Equation (6).

## 5.3 Model Estimation

Based on the model components outlined above, we can further arrange the terms and derive the following log-hazard representation (see Web Appendix C for details),

$$\begin{aligned}\ln h_j(t) = \quad & \underbrace{b_j + X_{jt}\beta}_{\text{daily diamond sale hazard}} \quad \underbrace{-I(Decoy)\ln(1 + e^{D_{jt}^{Decoy}\gamma^{Decoy}})}_{\text{dominance hazard of a decoy}} \\ & \underbrace{+I(Dominant)\Big[\ln(1 + e^{D_{jt}^{Dominant}(\gamma^{Dominant}+\gamma^{Boost})}) - \ln(1 + e^{D_{jt}^{Dominant}\gamma^{Dominant}})\Big]}_{\text{dominance hazard of a dominant}},\end{aligned} \tag{7}$$

where $D_{jt}^{Decoy} = \{K_j, \ln R_{jt}^{Dominant}, I(rp_{jt} > 0)rp_{jt}, sp_{jt}\}$ and $D_{jt}^{Dominant} = \{K_j, \ln R_{jt}^{Decoys}, I(rp_{jt} < 0)(-rp_{jt}), sp_{jt}\}$. The effects of being a decoy and being a dominant on the sale hazards are clearly outlined in the last two terms of this representation.

Denote the total number of days since diamond $j$ (in total $J$ diamonds) is on the market to the end of our observation period as $T_j$, and the day diamond $j$ is sold since its introduction as $T_j^s$. Following the derivation in Lancaster (1979) for the random effect hazard model, the total likelihood we use for estimation becomes the following:

$$L = \prod_{j=1}^{J} \left\{ \left[ I\left(T_j^s \leq T_j\right) \left(S_j(T_j^s - 1) - S_j(T_j^s)\right) \right] \times \left[ I\left(T_j^s > T_j\right) S_j(T_j^s) \right] \right\}, \tag{8}$$

where $S_j(t) = \left[1 + \sigma^2 \sum_{\tau=1}^{t} \bar{h}_j(\tau)\right]^{-\sigma^{-2}}$ is the survivor function and $\bar{h}_j(\tau)$ is the mean value of the hazard $h_j(\tau)$ where $b_j = 0$. Notice that $\lim_{\sigma^2 \to 0} S_j(t) = \prod_{\tau=1}^{t} e^{-\bar{h}_j(\tau)}$, i.e., it reduces to the case without random effects. We estimate the model by using the maximum likelihood approach.

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. Second, in terms of how to model the DE conditional on dominant detection, we choose to use a scalar function $Q_{jt}(\cdot)$. When $Q_{jt}(\cdot) > 1$, our specification becomes consistent with the DE theory, i.e., upon detection, there is a boost in sales for dominant diamonds. In other words, under our framework, testing the existence of the DE becomes the same as testing whether $Q_{jt}(\cdot) > 1$. 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 sale hazard $\psi_j(\cdot)$ and the dominance hazard $\phi_j(\cdot)$ to differ across different diamond price segments.

## 5.4 Model Identification

Our identification strategy relies on the fact that it takes different numbers of days to sell different types of diamonds (*neither decoy nor dominant, decoy only, dominant only*, or *both decoy and dominant*); or equivalently, the sale hazards vary across different diamond types. Accordingly, we use different parts of the data to identify different components of our specification. First, based on the normalization in the first line of Equation (6), the sale hazard for the *neither* type equals to the daily diamond sale hazard. Thus, we use the portion of the data regarding the sales of *neither* type diamonds with different $K_j$, $H_j$, $W_{jt}$, and $Z_t$ in $X_{jt}$ to identify the parameters $\beta$ in the daily diamond sale hazard component. Second, conditional on the identification of $\beta$, we identify the parameters related to the detection probabilities and dominant boost hazard. 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 portion of the data regarding the sales of *decoy only* diamonds with different $R_{jt}^{Dominant}$, $rp_{jt}$, and $sp_{jt}$ in $D_{jt}$ to identify the parameters of the market-level decoy detection probabilities, i.e., $\gamma^{Decoy}$ (also see the second component of Equation (7)). 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 ($\gamma^{Dominant}$) and of the dominant boost hazard ($\gamma^{Boost}$) since $Pr_{jt}^{Dominant}$ and $Q_{jt}$ are always bundled together in the form of $(1-Pr_{jt}^{Dominant})+Pr_{jt}^{Dominant}Q_{jt}$. To separately identify $\gamma^{Dominant}$ from $\gamma^{Boost}$, we make the following assumption: $\gamma^{Decoy} = \gamma^{Dominant}$, 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 in the same grade 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, it is reasonable to assume parameters quantifying the decoy and dominant detection probabilities in the diamond market to be the same.[9]

Based on this symmetric market-level detection assumption, and conditional on the dominant detection parameters ($\gamma^{Dominant}$) being identified, we use portion of the data regarding the sales of *dominant only* and *both decoy and dominant* type diamonds with different $D_{jt}$ to identify the parameters of the *dominant boost hazard*, $\gamma^{Boost}$. In our empirical setting, the diamond prices change over days and subsequently the decoy–dominant structure also changes daily. This data variation empowers the identification of our model parameters ($\beta$, $\gamma^{Decoy}$, and $\gamma^{Boost}$). Finally, variation in the time it takes to sell diamonds with similar $X_{jt}$ and $D_{jt}$ enables identification of the unobserved heterogeneity parameter (i.e., $\sigma^2$).

In addition, the identification of the parameters relies on our exclusion restrictions. Specifically, we observe two types of variables in the data–the ones that are directly observable by consumers (such as diamond physical attributes and daily demand factors, i.e., $X_{jt}$), and the ones that require extensive consumer search and comparisons (i.e., the context-related variables, $D_{jt}$). The Daily Diamond Sale Hazard component captures the baseline sale hazard of a diamond that does not depend on deliberate diamond comparisons. Hence, we exclude context-related variables ($D_{jt}$) in the modeling of this component. In contrast, the Dominance Hazard component captures the

[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.

critical impact of potential context effect under study. Accordingly, we exclude the variables that are directly observable to consumers without extensive search ($X_{jt}$).[10]

Finally, given that the derived log-hazard specification is highly non-linear, a valid concern is whether various components in the function can be identified accurately. We confirm that our statistical test of dominance is valid through Monte Carlo simulations (see details in Web Appendix D). We simulate diamond sales by using the proposed model specification under various detection and dominant boost levels. Estimation with these simulated datasets yields that we can correctly recover the assumed parameters with high precision. Further, we show that the dominance hazard component is not identifiable without the existence of the dominance effect in the simulated data, assuring that the quantified DE is not an artifact of the specific non-linear functional form.

# 6 Results

## 6.1 Main Estimation Results

We report our estimation results in Table 6. Results suggest that the daily diamond sale hazard increases with the diamond's carat size. In addition, 1.0-carat diamonds are significantly easier to sell. As expected, as a diamond's price increases, its sale hazard decreases. Regarding the cut, color, and clarity attributes, we observe an inverse U-shape relationship, i.e., the daily diamond sale hazard is the largest for diamonds with moderate cut, color and clarity levels.[11] Estimates of daily demand proxies suggest that Google search indexes for diamond-related keywords are significant proxies for the sales. The daily diamond sale hazard increases significantly when the search indexes on the keywords of "engagement ring" and the studied retailer's name are high, whereas it decreases when the search intensity is high on the keywords of "diamond," "diamond ring," and the competitor's name. The daily diamond sale hazard also differs significantly across weekdays, with Thursday

[10]It might be possible that some context-related variables such as the relative price and the price variation could directly affect choice instead of the DE as we operationalize. Similarly, it is possible that diamond physical attributes could also moderate the extent of the DE. We conduct robustness checks that a) include relative price and price variation in the baseline component and b) move diamond 4C attributes to the dominance hazard component. Our main model specification in the paper generally outperforms these alternative specifications based on the AIC. More importantly, the quantified detection probabilities and DE remain quantitatively similar as compared to our main model specification. Details are available upon request.

[11]As previously mentioned, our unit of analysis is each individual diamond; thus the basic sale hazard would be determined by both consumer demand for and 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.

being the best day for diamond sales, while Saturday and Sunday are the worst days. The results also suggest that diamonds are easier to sell during holidays. Regarding the competition from other diamonds, we find that, intuitively, as the number of same-grade diamonds increases, the daily sale hazard decreases. Lastly, the estimated variance of the random effect is reasonably large, i.e., the standard deviation is around 27% of the intercepts, highlighting the importance of controlling for unobserved correlations in the sale likelihoods among diamonds.

Table 6: Model Estimates

| **Variable** | **Estimate** | **S.E.** |
|---|---|---|
| **Daily Diamond Sale Hazard** | | |
| Low-Price Segment (2K–5K) | $-3.349^{**}$ | 0.114 |
| Medium-Price Segment (5K–10K) | $-3.169^{**}$ | 0.121 |
| High-Price Segment (10K–20K) | $-3.312^{**}$ | 0.131 |
| ln(Carat) | $0.462^{**}$ | 0.079 |
| Is 1.0 Carat | $0.542^{**}$ | 0.018 |
| Price (in 1000) | $-0.057^{**}$ | 0.006 |
| Cut: Poor | 0.000 | |
| Cut: Good | $0.325^{**}$ | 0.096 |
| Cut: Very Good | $0.835^{**}$ | 0.095 |
| Cut: Ideal | $1.227^{**}$ | 0.094 |
| Cut: Signature Ideal | $0.431^{**}$ | 0.127 |
| Color: J | 0.000 | |
| Color: I | $0.075^{**}$ | 0.029 |
| Color: H | $0.228^{**}$ | 0.030 |
| Color: G | $0.248^{**}$ | 0.031 |
| Color: F | $0.299^{**}$ | 0.032 |
| Color: E | $0.065^{*}$ | 0.035 |
| Color: D | 0.046 | 0.039 |
| Clarity: SI2 | 0.000 | |
| Clarity: SI1 | $0.166^{**}$ | 0.023 |
| Clarity: VS2 | $0.310^{**}$ | 0.025 |
| Clarity: VS1 | $0.248^{**}$ | 0.027 |
| Clarity: VVS2 | $0.078^{**}$ | 0.031 |
| Clarity: VVS1 | $-0.112^{**}$ | 0.036 |
| Clarity: IF | $-0.564^{**}$ | 0.044 |
| Clarity: FL | $-0.300$ | 0.415 |
| Google Search: “diamond” | $-0.205^{*}$ | 0.110 |
| Google Search: “diamond ring” | $-0.527^{**}$ | 0.065 |
| Google Search: “wedding ring” | $-0.003$ | 0.057 |
| Google Search: “engagement ring” | $0.079^{**}$ | 0.031 |
| Google Search: retailer’s name | $0.183^{**}$ | 0.020 |
| Google Search: competitor’s name | $-0.446^{**}$ | 0.064 |
| Weekday: Monday | 0.000 | |
| Weekday: Tuesday | $-0.026^{*}$ | 0.014 |
| Weekday: Wednesday | $-0.053^{**}$ | 0.014 |
| Weekday: Thursday | $0.052^{**}$ | 0.014 |

Continued on next page

Table 6: Model Estimates

| Variable | Estimate | S.E. |
|---|---|---|
| Weekday: Friday | $-0.186^{**}$ | 0.015 |
| Weekday: Saturday | $-1.737^{**}$ | 0.026 |
| Weekday: Sunday | $-1.158^{**}$ | 0.021 |
| Is Holiday | $0.033^{**}$ | 0.013 |
| ln(# Diamonds of the Same Grade) | $-0.026^{**}$ | 0.011 |
| $\sigma^2$ (Variance of the Random Effect) | $0.787^{**}$ | 0.012 |
| **Dominance Hazard—Market-Level Detection Probability** | | |
| Low-Price Segment (2K–5K) | $-1.341^{**}$ | 0.114 |
| Medium-Price Segment (5K–10K) | $-1.850^{**}$ | 0.190 |
| High-Price Segment (10K–20K) | $-2.430^{**}$ | 0.388 |
| $\ln(R_{jt}^{Dominant})$ | $0.241^{**}$ | 0.041 |
| $I(rp_{jt} > 0)(rp_{jt})$ | $1.816^{**}$ | 0.734 |
| $sp_{jt}$ | $0.265^{*}$ | 0.151 |
| **Dominance Hazard—Dominant Boost Hazard** | | |
| Low-Price Segment (2K–5K) | $0.965^{**}$ | 0.090 |
| Medium-Price Segment (5K–10K) | $0.552^{**}$ | 0.155 |
| High-Price Segment (10K–20K) | $1.072^{**}$ | 0.305 |
| $\ln(R_{jt}^{Decoy})$ | $0.072^{**}$ | 0.027 |
| $I(rp_{jt} < 0)(-rp_{jt})$ | $1.227^{**}$ | 0.496 |
| $sp_{jt}$ | 0.087 | 0.082 |
| Log-likelihood | $-257{,}666$ | |
| AIC | 515,432 | |

*Note:* Estimates with ** and * are significant at the 0.05 and 0.10 levels, respectively.

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 the highest for diamonds in the low-price segment and the lowest in the high-price segment. One potential explanation is that consumers of the low-price (\$2K–\$5K) segment are usually on tight budgets and are more motivated to spend extra time searching for better prices, leading to larger consideration sets. As a result, they are more likely to detect existing decoy–dominant relationships. The positive significant estimate of (log- of) grade-level percentage of decoys/dominants (0.241) shows that when a larger percentage of diamonds are decoys in a diamond's grade, it gets relatively easier for the market to detect that diamond as a dominant. The positive significant estimate of the relative price index (1.816) shows that the further a decoy/dominant diamond is priced from the average grade-level price, the higher is the probability of decoy/dominant detection in the diamond market. In addition, results suggest that as the price variation increases in a grade, the market-level detection probabilities of decoys/dominants marginally increase. This finding suggests that as the within-grade price variation gets larger, consumers search more to identify decoys (and dominants). By using our model estimates, we next calculate the market-level decoy–dominant detection probabilities. Interestingly, we find that these probabilities are quite low: 0.25 for the low-, 0.17 for the medium-, and 0.11 for the high-price segments. 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 almost always aware of the decoy–dominant relationships.

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 and conclusive evidence of the existence of the DE in a real product market. The demand boost effect is lower for the medium-price segment (0.552) compared to the low- (0.965) and high-price (1.072) segments. The parameter estimate for the (log- of) percentage of decoys is positive and significant (0.072), indicating that having more decoys would increase the dominant diamonds' sale hazards. The relative price measure is also positive and significant (1.227) suggesting that as a dominant is discounted further away from the grade-level average price, its sale likelihood is boosted more. The estimate of the grade-level price dispersion turns out to be insignificant. We next calculate the average dominant boost effect in proportional terms using the estimates. We find that, conditional on a diamond being detected as a dominant, its sale hazard increases by 2.7 times for the low-, 1.8 times for the medium-, and 3.2 times for the high-price segments, suggesting that—due to the DE—there is a substantial boost in the sale hazards of dominant diamonds.

In summary, our estimation results show that in general, consumers have a low chance of detecting the decoy–dominant relationships in the online diamond market, especially among diamonds in the high-price segment. Thus, it is critical to model the decoy–dominant detection process in real product markets in order to correctly quantify the sales impacts 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 sale hazard increases quite significantly, especially in the low- and high-price segments. With this finding, we not only provide strong field evidence about the existence of the DE, but also respond to Frederick et al. (2014) and Yang and Lynn (2014), who questioned the practical validity and usefulness of the DE.

### 6.2 Model Comparison and Robustness

In addition to our main model estimation, we conduct comparisons with two simpler modeling approaches and check the robustness of the results with three alternative specifications. We show that one would obtain worse data fits and biased inferences of the DE under these alternative approaches. We also confirm that the estimates are very robust under alternative specifications with other ways of heterogeneity control and different specifications of the reference price.

One important modeling contribution of this paper is the separation of the DE from decoy–dominant detection. To test the importance of this separation, in our first model comparison, we estimate a benchmark proportional hazard model by including the same set of controls directly in the baseline hazard. Specifically, the log-hazard specification in Equation (7) becomes the following:

$$\ln(h_j(t)) = b_j + X_{jt}\beta + D_{jt}^{Decoy}\gamma^{Decoy} + D_{jt}^{Dominant}\gamma^{Dominant}. \tag{9}$$

The second column of Table 7 (column "No Detection") reports the estimates of the dominance variables for decoys (decoy shrinkage hazard) and dominants (dominant boost hazard). This simple modeling approach is outperformed by our proposed model based on the AIC, suggesting that explicitly separating the market–level decoy–dominant detection from the dominant boost hazard better explains data variation. More importantly, the alternative model's estimates have a compound effect of detection and boost on diamonds' sale hazards, causing inaccurate inferences in understanding the magnitude of the DE if one directly uses these estimates. Results suggest that dominant boost hazards from this benchmark model are much smaller (ranging from 1.1 to 1.4 times) compared to ones from the proposed model (1.8 to 3.2 times). Thus, not explicitly controlling for the market-level detection causes a significant under-estimation of the DE's magnitude.

One potential explanation for the observed sales patterns is heterogeneous competitive effects. For example, there could exist two diamond segments with different competition intensities where diamonds in the first segment are minimally affected by competition while those in the second are highly affected. Further, the likelihood of being in the second segment depends on a diamond's relative price in the grade. Accordingly, if a dominant diamond is priced significantly lower than other similar diamonds, the competitive effect plays a role implying a boost in the diamond's sale likelihood, and *vice-versa*.[12] To test this alternative explanation, we use a simple 2-segment latent-class model with diamond price segment dummies and absolute relative price differences determining the competitive segment membership, and a competition hazard (conditional on being in the competitive segment) that is modeled by the (log- of) number of competing diamonds in the same grade. This specification yields a much higher AIC compared to our proposed specification (516,718 vs. 515,432), strongly supporting the DE over the heterogeneous competitive effects explanation.

Table 7: Model Comparisons and Robustness Checks

| **Variable** | **No Detection** | **Homogeneous** | **Reference Price** | **Fixed Effect** |
|---|---|---|---|---|
| Diamond Random Effects | included | | included | included |
| Day Fixed Effects | | | | included |
| $PGain$:$I(rp_{jt} < 0)rp_{jt}$ | | | $-1.251^{**}(0.209)$ | |
| $PLoss$:$I(rp_{jt} > 0)rp_{jt}$ | | | $-0.599^{**}(0.222)$ | |
| **Market-Level Detection Probability/Decoy Shrinkage Hazard** | | | | |
| Low-Price Segment | $-0.245^{**}(0.025)$ | $-1.548^{**}(0.110)$ | $-1.376^{**}(0.120)$ | $-1.578^{**}(0.145)$ |
| Medium-Price Segment | $-0.090^{**}(0.029)$ | $-2.048^{**}(0.208)$ | $-1.993^{**}(0.217)$ | $-1.742^{**}(0.156)$ |
| High-Price Segment | $-0.095^{**}(0.037)$ | $-3.205^{**}(0.413)$ | $-2.388^{**}(0.433)$ | $-2.176^{**}(0.265)$ |
| $\ln(R_{jt}^{Dominant})$ | $-0.044^{**}(0.008)$ | $0.205^{**}(0.047)$ | $0.243^{**}(0.041)$ | $0.293^{**}(0.045)$ |
| $I(rp_{jt} > 0)(rp_{jt})$ | $-0.621^{**}(0.198)$ | 1.164 (0.836) | | 0.752 (0.829) |
| $sp_{jt}$ | −0.046 (0.029) | $0.587^{**}(0.141)$ | 0.261 (0.167) | $0.304^{**}(0.140)$ |
| **Dominant Boost Hazard** | | | | |
| Low-Price Segment | $0.314^{**}(0.027)$ | $0.720^{**}(0.089)$ | $0.989^{**}(0.092)$ | $1.118^{**}(0.114)$ |
| Medium-Price Segment | $0.050^{*}(0.029)$ | $0.415^{**}(0.168)$ | $0.493^{**}(0.173)$ | $0.416^{**}(0.130)$ |
| High-Price Segment | $0.064^{*}(0.033)$ | $1.446^{**}(0.334)$ | $0.941^{**}(0.321)$ | $0.722^{**}(0.202)$ |
| $\ln(R_{jt}^{Decoy})$ | $0.076^{**}(0.007)$ | −0.040 (0.027) | $0.078^{**}(0.027)$ | 0.027 (0.030) |
| $I(rp_{jt} < 0)(-rp_{jt})$ | $1.237^{**}(0.220)$ | $0.910^{*}(0.508)$ | | $1.768^{**}(0.560)$ |
| $sp_{jt}$ | $0.088^{**}(0.028)$ | −0.010 (0.085) | 0.095 (0.089) | $0.107^{**}(0.007)$ |
| N. parameters | 50 | 49 | 50 | 247 |
| Log-likelihood | $-257,679$ | $-262,156$ | $-257,669$ | $-254,071$ |
| AIC | $515,458$ | $524,410$ | $515,438$ | $508,636$ |

*Note:* Numbers reported are mean estimates and standard errors. Estimates with ** and * are significant at the 0.05 and 0.10 levels, respectively. Other controls (diamond 4Cs, competition, etc.) are included in the models but not reported here to save space.

We also conduct a series of robustness checks with alternative model specifications and report the results in Table 7. The first specification omits the random effect in the hazard specification (column "Homogeneous"), the second adopts a RPE specification with the relative price variables entering into the daily diamond sale hazard component instead of our dominance hazard specification (column "Reference Price"), and the last specification includes day fixed effects instead of

[12]We thank the Associate Editor for pointing out this. Due to the aggregate nature of our data, it is not possible to identify the segment-specific log-hazards and the segment sizes simultaneously. Thus, for identification purposes, we normalize the hazard for diamonds in the first segment to one, which represents no competitive effect.

Google search indexes and weekday dummies as the control for demand effects across days (column "Fixed Effect"). Across these three specifications, we find consistent and robust results: most of the estimates have the same direction and similar magnitude as in our main model. One notable difference is that the relative magnitude of the DE is reversed for the low- and high- price segments under the homogeneous model. We believe our random effect specification captures more unobserved heterogeneity across diamonds and thus should be more accurate, with the log-likelihood also much improved. The change in moving the relative price variables into the diamond daily hazard component seems to make little impact on the model estimates. In addition, the estimates again show that the sales function is more responsive in the gain domain than in the loss domain, consistent with the findings in our data pattern explorations. Finally, including day fixed effects could improve the model fit but it comes at the cost of almost 200 additional parameters. Importantly, the coefficient estimates for the dominance hazard component remain almost the same. Thus, we prefer our main model specification for being parsimonious.

# 7 Managerial Significance

We now explore the managerial implications of our study. We first quantify the DE's overall profit impact to the retailer using model estimates. We also explore through simulation studies how sensitive the DE's profit impact is with respect to changes in different aspects of dominance structure. These exercises shed new light on the DE's practical significance showing that it is not simply an experimental artifact.

To quantify the DE's profit impact, we start with the retailer's profit for a given diamond $j$ at time $t$ that can be calculated as:

$$\pi_{jt} = \Pr_j(t|\cdot) \times (p_{jt} - w_{jt}), \tag{10}$$

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% (provided in retailer's annual report) from the observed daily retail prices. $\Pr_j(t|\cdot) = 1 - \exp(-h_j(t|\cdot))$ is the discrete-time hazard, or the probability that diamond $j$ would be sold on day $t$, conditional on not being sold until that day. The DE's impact on profit can be quantified by calculating the differences in the sale probability $\Pr_j(t|\cdot)$ with and without setting the dominant boost hazard component $Q_{jt}(\cdot)$ to 1.

We present the results in Table 8. Without the DE, on average, each diamond contributes \$22.89 in gross profit each day; whereas with the DE, the contribution becomes \$26.16. The DE increases the retailer's overall gross profit by 14.3%, or equivalently, the DE contributes a 12.5% share of the retailer's gross profit. The profit increase due to the DE is the largest in the low-price segment (25.4%), whereas the opposite is the case in the medium-price segment (8.6%). Based on the financial information of the retailer, this percentage increase would translate into an annual profit increase of \$9 million. This result shows that even though decoy–dominant detection probabilities are low in the online diamond market, the DE still has a substantial profit impact due to the significant boost in sale likelihoods upon dominance detection. Indeed, this profit impact is what matters the most from the substantive point of view.

Table 8: The Impact of the DE on Retailer's Gross Profit

| **Effect** | **2K–5K** | **5K–10K** | **10K–20K** | **Overall** |
|---|---|---|---|---|
| Avg Daily Profit Per Diamond W/O the DE | 8.71 | 30.17 | 43.67 | 22.89 |
| Avg Daily Profit Per Diamond W/ the DE | 10.92 | 32.76 | 49.72 | 26.16 |
| % Profit Increase due to the DE | 25.37% | 8.58% | 13.85% | 14.29% |

We next investigate the sensitivity of the DE's profit impact to different dominance configurations in the diamond market. In the simulation studies, we implement changes in the dominance structure within the range of observed data. In the first simulation study, we check how the DE's profit impact changes when each existing decoy/dominant diamond has one more decoy and/or dominant in the market. Results are presented in the top portion of Table 9. We find that, intuitively, having one more decoy would increase the DE's profit impact by 0.37% due to the effect of decoys in boosting the sales of dominants. Whereas, when one more dominant is added, the DE's profit impact decreases by 0.24% because the decoy detection probability would increase leading to a decrease of sales from the decoys. The DE's profit impact slightly increases (0.14%) if each decoy/dominant diamond has one more decoy and one more dominant.

We examine the effect of price dispersion on the DE's profit impact in the second simulation study. Specifically, we enlarge or shrink the relative prices of 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. By preserving the mean price level, we change the price of this diamond to \$10,500 (dispersion factor

Table 9: Percentage Changes in the DE's Profit Impact as the Dominance Structure Changes

| | 2K–5K | 5K–10K | 10K–20K | Overall |
|---|---|---|---|---|
| **Changing the Number of Decoys/Dominants** | | | | |
| Each Diamond Has 1 More Decoy | 0.43% | 0.25% | 0.39% | 0.37% |
| Each Diamond Has 1 More Dominant | −0.26% | −0.25% | −0.17% | −0.24% |
| Each Diamond Has 1 More Decoy and Dominant | 0.16% | 0.01% | 0.22% | 0.14% |
| **Changing the Price Variation** | | | | |
| Multiplying the Dispersion by 0.5 | −0.52% | −0.28% | −1.08% | −0.60% |
| Multiplying the Dispersion by 0.8 | −0.19% | −0.10% | −0.44% | −0.23% |
| Multiplying the Dispersion by 1.2 | 0.17% | 0.10% | 0.46% | 0.22% |
| Multiplying the Dispersion by 1.5 | 0.37% | 0.22% | 1.17% | 0.53% |
| **Changing the Market-Level Detection Probabilities** | | | | |
| Market-level Detection Probability in the Data | 0.25 | 0.17 | 0.11 | 0.19 |
| Market-level Detection Probability of 0.05 | −3.36% | 1.49% | −2.51% | −1.41% |
| Market-level Detection Probability of 0.10 | −1.71% | 0.99% | −0.25% | 0.01% |
| Market-level Detection Probability of 0.20 | −0.13% | −0.78% | 2.45% | 1.00% |
| Market-level Detection Probability of 0.30 | −0.45% | −4.03% | 2.98% | −0.04% |
| Market-level Detection Probability of 0.40 | −2.28% | −7.65% | 1.66% | −2.40% |

0.5), \$10,800 (dispersion factor 0.8), \$11,200 (dispersion factor 1.2), and \$11,500 (dispersion factor 1.5) in our simulation study. The middle portion of Table 9 reports the results of this simulation study. Results suggest that the DE's profit impact increases with the increase in price dispersion (between the dispersion factor of 0.5 and 1.5). Based on our model estimates, an expanded price dispersion would increase both the detection probabilities of decoys/dominants and the effect of demand boost on dominants. The profit gains from dominants (due to higher detection and sales boost) outperform the losses from decoy sales (due to higher detection), leading to an increase in the retailer's profit in the studied dispersion range.[13]

Given that dominance detection is a critical pre-condition for the DE, we evaluate the profit impact of the DE under different dominance detection levels in our last simulation study. By changing the intercepts in the detection equation, we set the market-level detection probabilities at 5%, 10%, 20%, 30%, and 40%. The detection probabilities are between 25% for the low-price segment and 11% for the high-price segment in the data. Results reported in the bottom portion of Table 9 suggest that the DE's profit impact gets larger when the detection probability is lower in the medium-price segment and higher in the high-price segment, as compared to the current levels. On the other hand, the DE's profit impact is highest around the current detection level in

[13]Additional analyses reveal that as the dispersion levels further increase to a few times, the DE's profit impact levels off, and the retailer's profit start to decrease due to the DE.

the low-price segment. These patterns are driven by the fact that dominance detection has opposite impacts on the profits from decoys versus dominants. Thus, the DE's profit impact varies based on the detection levels and the magnitude of the DE across the diamond price segments.

In summary, our study offers a framework to guide marketing researchers about how to quantify decoy–dominant detection probabilities in real product markets with pricing and sales information. We show that the DE has a substantial impact on the studied retailer's profitability. Further, we find that the retailer could potentially gain additional profits from the DE when 1) there are more decoys in the market, 2) within-grade price variation increases, and 3) dominance detection probabilities stay similar, decrease, and increase in the low-, medium-, and high-price diamond segments, respectively. Having a better understanding of the profit impact could enable the retailer to more effectively utilize the DE in its marketing and operations activities. One interesting observation is that the retailer recently included a new section of comparable alternatives on the details pages of a selected list of diamonds, in which their decoys/dominants are displayed for some of the diamonds. This practice could likely change the chances of consumers detecting decoys and dominants.

# 8 Conclusions

In this research, we empirically validate the DE by using unique panel data from a leading online jewelry retailer. We estimate a proportional hazard model (derived from consumer primitives) with embedded market-level decoy–dominant detection probabilities and the sales boost upon dominant detection (i.e., the DE). We find that, the market-level probability of detecting decoys or dominants is quite low (11%–25%). However, upon a dominant is detected, its sale hazard increases by 1.8 to 3.2 times. Thus, we empirically validate the existence of the DE in a real product market. Our model comparisons reveal that not controlling for decoy–dominant detection yields biased inferences regarding the DE's magnitude. We show that our estimates are robust under various alternative specifications. In addition to validating the DE in the field, we contribute to the substantive issue of measuring the DE's profit impact, i.e., the DE's managerial significance. We quantify the profit impact of the DE using model estimates and find that it improves the retailer's gross profit by 14.3%. We explore how sensitive this profit impact is with respect to changes in the dominance structure, and find that the profit impact could potentially increase with more decoys, a larger price variation, and changes in the market-level decoy–dominant detection probabilities.

Our study is the first empirical attempt to quantify the widely documented DE in the consumer behavior literature. It is exciting to apply the well-developed context-dependent choice theory to real-life data and empirically quantify the managerial implications. Accordingly, we would like to note that the studied online diamond market is a unique setting to test the DE since diamonds have vertical attributes, and decoys and dominants are widely observed. That being said, we believe it is possible to find other market settings that also allow the DE to be tested. For example, it is common that multiple sellers carry different price stickers when selling identical products with the same shipping/return policies and warranties in e-commerce platforms such as Amazon where the more expensive alternatives serve the role of decoys. Similarly, in used-goods platforms such as eBay, it is often possible to observe products with inferior conditions from low-rating sellers to be more expensive than some of the better condition ones from top sellers. These decoy alternatives may boost the sale likelihood of less expensive and/or better condition dominants. Further, to substantiate the effect of such decoys on their dominants, these retail platforms can make decoy–dominant detection easier or more difficult by either changing the order of the product listings or recommending the products of particular sellers. However, such operations may have practical limitations for platform designers due to creating equality concerns and discouraging the participation of smaller sellers. In addition to within-product price variation on retail platforms, some producers price their product bundles the same as their main product to make the bundle more attractive. For example, Dyson offers the V11 Torque Drive Cord-Free Vacuum with extra Extension Hose at the same price as the Vacuum itself. Last, as opposed to these strict decoy examples, retailers may position some products as near decoys. For example, the price difference between iPhone X 512GB and 256GB is the same as the difference between the 256GB and 64GB, making upgrading to 512GB more attractive to some consumers. In the future, quantifying the DE across various product categories might help researchers to better understand the DE's limits and boundaries.

We believe there are multiple dimensions that can be pursued to extend understanding of the DE in future research. One direction is to investigate how consumers learn and respond to decoy–dominant relationships in their search process. We face significant modeling challenges in the current context given the aggregate nature of our data. Future research could potentially address such issues when consumer-level search data are available. Another 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 study if such supplier-level information is observed. A third direction is to consider the DE across multiple retailers since consumers might search products from different retailers. In our setting, since the retailer captures about 50% of the market share in the U.S., this issue was not a major concern. It might be worth pursuing in a different product market when data of multiple retail outlets are available. Finally, as Bell and Lattin (2000) studied the impact of price response heterogeneity in quantifying loss aversion, it is worthwhile to investigate the same factor for the DE. We partially achieve this by allowing the DE to differ across diamond price segments. When consumer-level panel data are available, the heterogeneity in both price responses and the DE could be better formulated in the future.

## References


Bell, David R, James M Lattin. 2000. Looking for loss aversion in scanner panel data: The confounding effect of price response heterogeneity. *Marketing Science* **19**(2) 185–200.

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

Biswas, Abhijit, Edward A Blair. 1991. Contextual effects of reference prices in retail advertisements. *Journal of marketing* **55**(3) 1–12.

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

Chiang, Jeongwen, Siddhartha Chib, Chakravarthi Narasimhan. 1998. Markov chain monte carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics* **89**(1-2) 223–248.

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* .

Hardie, Bruce GS, Eric J Johnson, Peter S Fader. 1993. Modeling loss aversion and reference dependence effects on brand choice. *Marketing science* **12**(4) 378–394.

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. 2014. Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics* **45**(4) 847–884.

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.

Kahneman, Daniel. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* **47** 278.

Kalyanaram, Gurumurthy, Russell S Winer. 1995. Empirical generalizations from reference price research. *Marketing science* **14**(3_supplement) G161–G169.

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.

Lancaster, Tony. 1979. Econometric methods for the duration of unemployment. *Econometrica: Journal of the Econometric Society* 939–956.

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.

Lynch Jr, John G, Dipankar Chakravarti, Anusree Mitra. 1991. Contrast effects in consumer judgments: Changes in mental representations or in the anchoring of rating scales? *Journal of Consumer Research* **18**(3) 284–297.

Mazumdar, Tridib, Sevilimedu P Raj, Indrajit Sinha. 2005. Reference price research: Review and propositions. *Journal of marketing* **69**(4) 84–102.

Monroe, Kent B. 1973. Buyers' subjective perceptions of price. *Journal of marketing research* **10**(1) 70–80.

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* .

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.

Rajendran, Kharan N, Gerard J Tellis. 1994. Contextual and temporal components of reference price. *Journal of Marketing* **58**(1) 22–34.

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.

Siddarth, S, Randolph E Bucklin, Donald G Morrison. 1995. Making the cut: Modeling and analyzing choice set restriction in scanner panel data. *Journal of Marketing Research* **32**(3) 255–266.

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.

Van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel, Philip Hans Franses. 2010. Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research* **47**(1) 63–74.

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. 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 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 sale 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 denote the marginal cost for suppliers as $c_{jt}$.

If there is no DE in the sale 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:

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

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 is no longer true, because decoys serve as "loss leaders" and generate less expected profits than their dominants. Consequently, for high-priced decoys, the recovered costs under the optimality condition without the DE will be higher than the true costs; while the opposite is true for 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 sale response function $Ph_j(t|p_{jt})$ with polynomials of $p_{jt}$ (we use linear, quadratic, and cubic forms), diamond characteristics, and daily demand controls 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 diamonds in $J$ on day $t$) and other control variables such as diamond characteristics, daily

demand effects, and grade average prices.

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 and follow the same steps as in Test I, except that in Step 3, we run the regressions using diamond-level (i.e., SKU-level) 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 tests reject the null hypothesis (i.e., the price variation arises solely from consumer search) and support the DE coexisting with consumer search.

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

| **Variable** | **2K–5K** | **5K–10K** | **10K–20K** |
|---|---|---|---|
| *Test I* | | | |
| Controls | | 4Cs, daily demand effects, $\overline{p_{Jt}}$ | |
| $rp_{jt}$ | 2.172**(0.001) | 5.401**(0.002) | 12.721**(0.003) |
| Adj. R-squared | 0.987 | 0.995 | 0.996 |
| *Test II* | | | |
| Controls | | diamond fixed effects | |
| $rp_{jt}$ | 1.178**(0.028) | 1.917**(0.052) | 1.143**(0.167) |
| Adj. R-squared | 0.943 | 0.960 | 0.945 |

*Note:* Estimates with ** and * are significant at the 0.05 and 0.10 levels, respectively. Dependent variable—estimated cost—is in 1000 dollars. In Test II, we randomly sample 500 diamonds in each price segment.

## B Derivation of Diamond-Level Sale 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 $\tau$ (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 $\tau_i$, 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 $\tau_i$, as $s_i(j|M_i)$ (to be discussed in the following subsection in detail). The expected sale probability of diamond $j$ at time $\tau_i$ 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 sale hazard of diamond $j$ at time $\tau_i$. This is the case because it is the conditional probability that diamond $j$ will be sold at $\tau_i$ conditional on it not being sold until $\tau_i$, 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 $C_j^{Dominant}$ contains all diamonds that are dominating $j$; and finally, the set $C_j^{Decoy}$ represents 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 $M_{ij}^{Dominant}$ and $M_{ij}^{Decoy}$ the sets of diamonds (in $i$'s consideration set) that are from $j$'s dominants set ($C_j^{Dominant}$) and decoys set ($C_j^{Decoy}$). Denote $\mathbb{M}_j^o$, $\mathbb{M}_j^{Dominant}$, and $\mathbb{M}_j^{Decoy}$ as the super sets of $M_{ij}^o$, $M_{ij}^{Dominant}$ and $M_{ij}^{Decoy}$, respectively. We now can express consumer $i$'s consideration set $M_i$ as a combination of $j$, $M_{ij}^o$, $M_{ij}^{Dominant}$ and $M_{ij}^{Decoy}$, and define the choice probability $s_i(j|M_i)$ accordingly:[2]

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

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 $M_{ij}^{Decoy}$, because diamonds in $M_{ij}^{Decoy}$ 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, $s_i(j|j \cup M_{ij}^o) = s_i(j|j \cup M_{ij}^o \cup M_{ij}^{Decoy})$; however, if there is DE, we would expect the presence of diamonds from $M_{ij}^{Decoy}$ 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.

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

where $q_i(j, M_{ij}^o, M_{ij}^{Decoy})$ 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 $q_i(j, M_{ij}^o, M_{ij}^{Decoy}) = 1$ if there is no DE; $q_i(j, M_{ij}^o, M_{ij}^{Decoy}) > 1$ 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:

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

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 then use the independence assumption to move from the third to the fourth line of Equation (B.4). Next, we define the following:

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

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

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

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

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

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

Equation (B.7) tells us that the aggregate sale 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, $q_i(j, M_{ij}^o, M_j^{Decoy}) \geq 1$ , 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 $q_i(j, \mathbb{M}_{ij}^o, \mathbb{M}_j^{Decoy}) \geq 1$. As $q_i(j, M_{ij}^o, M_j^{Decoy}) \geq 1$ captures the potential DE at the individual-choice level, testing whether $q_i(j, \mathbb{M}_{ij}^o, \mathbb{M}_j^{Decoy}) \geq 1$ equals testing whether on average the DE exists in individual choices, i.e., whether $\mathrm{E}q_i(j, M_{ij}^o, M_j^{Decoy}) \geq 1$.

### B.3 Daily Diamond Hazard

We now derive the aggregate-level sale 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 $i$[3]. 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:

[3]We use a representative consumer to simplify the derivation of the functional form specification. Our derivation can be extended to include heterogeneous consumer segments where the current model components (i.e., dominance detection and hazard boost) can be interpreted as the weighted average effect from heterogeneous segments.

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

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 $S_j(t) = \Pi_{k=1}^t e^{-h_j(k)}$.

### B.4 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:

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

In the above equation, the $n_t Pr(j \in M_i)s_i(j|j, \mathbb{M}_j^o)$ component represents the daily hazard for diamond $j$. The second component, $Pr(M_{ij}^{Dominant} = \varnothing)[Pr(M_{ij}^{Decoy} = \varnothing) + Pr(M_{ij}^{Decoy} \neq \varnothing)q_i(j, \mathbb{M}_j^o, \mathbb{M}_j^{Decoy})]$, represents the combined effect of the probability of diamond dominance detection[4] and the boost in sales upon dominant detection. Therefore, we operationalize the hazard $h_j(t)$ at discrete time in Equation(B.10) as follows:

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

Equation (B.11) is the same as Equation (1) in our manuscript. This completes our derivation of the diamond-specific sale hazard function based on individual consumer primitives.

[4] In aggregate, the probability of diamond dominance detection can also be interpreted as the size of the consumer segment detecting the dominance relationship.

## C Log-Hazard Derivation

Based on our model specifications in Equations (2) and (6), we can derive the log-hazard specification as:

$$
\begin{aligned}
\ln h_j(t) =& b_j + X_{jt}\beta + I(Decoy)\ln(1 - Pr_{jt}^{Decoy}) + I(Dominant)\ln\left[(1 - Pr_{jt}^{Dominant}) + Pr_{jt}^{Dominant} Q_{jt}\right] \\
=& b_j + X_{jt}\beta + I(Decoy)\ln\frac{1}{1 + e^{D_{jt}^{Decoy}\gamma^{Decoy}}} \\
&+ I(Dominant)\ln\left[\frac{1}{1 + e^{D_{jt}^{Dominant}\gamma^{Dominant}}} + \frac{e^{D_{jt}^{Dominant}\gamma^{Dominant}} e^{D_{jt}^{Dominant}\gamma^{Boost}}}{1 + e^{D_{jt}^{Dominant}\gamma^{Dominant}}}\right] \\
=& \underbrace{b_j + X_{jt}\beta}_{\text{daily diamond sale hazard}} \quad \underbrace{- I(Decoy)\ln(1 + e^{D_{jt}^{Decoy}\gamma^{Decoy}})}_{\text{dominance hazard of a decoy}} \\
&+ \underbrace{I(Dominant)\left[\ln(1 + e^{D_{jt}^{Dominant}(\gamma^{Dominant} + \gamma^{Boost})}) - \ln(1 + e^{D_{jt}^{Dominant}\gamma^{Dominant}})\right]}_{\text{dominance hazard of a dominant}},
\end{aligned}
\tag{C.1}
$$

where $D_{jt}^{Decoy} = \{K_j, \ln R_{jt}^{Dominant}, I(rp_{jt} > 0)rp_{jt}, sp_{jt}\}$, $D_{jt}^{Dominant} = \{K_j, \ln R_{jt}^{Decoys}, I(rp_{jt} < 0)(-rp_{jt}), sp_{jt}\}$ and $I(Decoy)$ and $I(Dominant)$ are indicators for a diamond with dominants and decoys, respectively.

## D Simulation Studies to Illustrate the Model Identification

In our first simulation study, we use the diamonds observed in our data and simulate their sales by using the proposed hazard specification (with the DE). Then, we estimate our proposed model with this simulated sales data. The simulation study yields that we can recover back the assumed parameters with high accuracy. Please see the first two columns of Table A2 below for the results of this simulation study. We also conduct additional simulation studies with various detection-levels and dominant boost-levels and find that we can precisely recover the assumed parameter values irrespective of the assumed size of the detection and dominant boost levels. Results of these additional analyses are available upon request.

Table A2: Results of the Simulation Studies Regarding the Model Identification

| | With Dominance Hazard | | Without Dominance Hazard | |
|---|---|---|---|---|
| **Variable** | **True Value** | **Estimate** | **No Restriction** | **Restrict Boost** |
| **Market-Level Detection Probability** | | | | |
| Low-Price Segment | −3.000 | $-3.105^{**}$(0.200) | −34.758 (*na*) | −12.509 (164.066) |
| Medium-Price Segment | −3.000 | $-2.995^{**}$(0.181) | −23.269 (*na*) | −12.290 (21.587) |
| High-Price Segment | −3.000 | $-3.142^{**}$(0.215) | −21.235 (*na*) | −9.647 (154.853) |
| $\ln(R_{jt}^{Dominant})$ | 0.500 | $0.473^{**}$(0.051) | 5.172 (*na*) | −0.438 (62.541) |
| $I(rp_{jt} > 0)(rp_{jt})$ | 2.000 | $2.489^{**}$(0.771) | 0.059 (*na*) | −0.011 (24.788) |
| $sp_{jt}$ | 0.500 | $0.528^{**}$(0.060) | −10.767 (*na*) | −7.451 (53.051) |
| **Dominant Boost Hazard** | | | | |
| Low-Price Segment | 1.000 | $1.174^{**}$(0.165) | 13.419 (*na*) | 0.000 |
| Medium-Price Segment | 1.000 | $1.037^{**}$(0.152) | −9.239 (*na*) | 0.000 |
| High-Price Segment | 1.000 | $1.150^{**}$(0.180) | −14.611 (*na*) | 0.000 |
| $\ln(R_{jt}^{Decoy})$ | 0.300 | $0.295^{**}$(0.034) | −0.894 (*na*) | 0.000 |
| $I(rp_{jt} < 0)(-rp_{jt})$ | 1.000 | 0.799 (0.545) | −3.705 (*na*) | 0.000 |
| $sp_{jt}$ | 0.600 | $0.591^{**}$(0.038) | 1.572 (*na*) | 0.000 |

*Note:* Numbers reported are mean estimates and standard errors. Estimates with ** and * are significant at the 0.05 and 0.10 levels, respectively. As noted above, separately identifying Decoy Dominant Detection and Dominant Boost components is not possible with no DE in the data. Hence, the standard errors cannot be calculated in the specification with no restriction due to singular numerical hessian. Other controls (diamond 4Cs, competition, etc.) are included in the daily diamond sale hazard but not reported here to save space.

In our second simulation study, we again use the diamonds observed in our data. Unlike the first simulation study, this time, we simulate the sales of diamonds by using only the Daily Diamond Sale Hazard component of our hazard specification (i.e., no DE). In other words, while simulating the sales data, we assume the decoy-dominant detection probabilities of being strictly zero, and the dominance boost of being one. Then, we estimate our proposed hazard model by using the simulated data. This second simulation study suggests that, without the DE in the data, it is not possible to separately identify the detection and dominant boost components because as long as the detection is equal to zero the dominant boost size can take any value. Consequently, even if we can get the point estimates using MLE, the hessian matrix is numerically singular, and we can not compute the standard errors (n.a. in the second column of the following table). The large negative values on the point estimates of the price-segment dummies also show that the detection probabilities are near zero (in the exp transformation), indicating that indeed we would not find significantly non-zero detection probabilities. We further fix the dominant boost component to be one (equivalently parameters are set to zero in the log transformation) and estimate the detection probability parameters. The last column in the following table shows the estimation results. The point estimates are, again, large and negative, which indicates that we would not find numerically non-zero detection probabilities. The standard errors are large in this scenario because the estimated parameters can take a large range of values in the negative domain to make the detection probabilities numerically equal to zero (for example, intercepts of -10 versus -1000 yield practically identical, i.e., zero detection probabilities).

In conclusion, the simulation studies yield that 1) we can recover the detection- and dominant boost-levels precisely as long as there exists DE in the data; 2) without the DE in the data, we cannot identify the dominance boost hazard component of our proposed specification. In other words, the functional form chosen can not cause one to infer the DE if such DE is not present in the data (Simulation 2). However, if the DE exists in the data, the proposed model can recover the size of the effect accurately (Simulation 1).