Online Appendix for

The Impact of Online Product Reviews on Product Returns

A Variable Summaries

Variable Names	Range	**Summary statistics** (min, median, mean, max) of the continuous variables % of units with 1, for the dummy variables
		Browsed and then purchased
Purchased	{0,1}	61954 purchased out of the 1319561 browsed instance
return	{0,1}	9846 returned
Review variables based on purchased products (time-variant)		
nratings	{0,1,…,∞}	0, 0.04, 0.22, 9.94
av_rating	1—5	1, 4.35, 4.24, 5
sdev_rating	≥0	0, 0.51, 0.53, 2.83
helpful_reviews	0—1	0, 0.41, 0.43, 1
from top reviewers	0—1	0, 0.11, 0.22, 1
length	{0,1,…,∞}	1, 54.75, 49.64, 391
Context (time-variant)		
Holiday	{0,1}	20.5%
in-store	{0,1}	26.5%
Price	> 0	0, 59, 74, 286
Consumer activity prior to purchase (time-variant)		
repeated browses	{0,1,…,∞}	1, 1, 1.52, 19
# products browsed	{0,1,…,∞}	1, 25, 36.5, 188
# keyword searches	{0,1,…,∞}	0, 0, 1.1, 22
Recency	{-∞, …, -2,-1}	-685, -23, -58, -1
Marketing variables (time-variant)		
Emails	{0,1,…,∞}	0, 9, 9.5, 63
Promo	{0,1}	26.5%
Product Information (time-invariant)		
Category	Accessories Clothes Home & Furniture Misc	20.9% 63.9% 14.4% 0.8%
Consumer variables (time-invariant)		
frequency (per year)		0, 29, 45.9, 403.8
Age > 36	{0,1}	47.8%
Male	{0,1}	8.3%
Distance	≥0	0, 7.93, 35.56, 5676
online_shopper	{0,1}	60.8%

Table 1 Variable summary statistics based only on the products browsed by the consumers, i.e., the data used in the analysis in this paper.

Variables	**Variance Inflation Factor**
Number of reviews	1.12
Valence of reviews	1.16
Dispersion of reviews	1.56
Helpful reviews	1.49
From top reviewers	1.45
Length of reviews	2.27
Time trend	1.27
In-store purchase	1.46
Discount	1.10
Holiday	1.01
Total returns	1.10
Repeated browses	1.07
Products browsed	1.17
Keyword searches	1.09
Recency	1.11
Promo	1.24

Table 2 Variance Inflation Factors

B A Model of Product Returns

In this appendix, we develop an analytical model that examines how return probabilities are affected by the information that the consumer obtains about the product by reading reviews about the product online.

We assume that customers are rational, but risk-averse, even though our modeling specification is general enough to include risk-neutral customers as a special case. Our analysis examines the impact of two constructs of interest on return probabilities: (1) precision of product quality information, and (2) uncertainty of product fit.

Modeling purchases. Before we can model the probability of returns, we first need to develop a model of purchases in settings where customers can return products. Our model is a generalization of a model proposed by (Anderson et al. 2009).[1] We consider a customer i who is contemplating whether to purchase an item j. Once the customer purchases, she has the option to keep or return the item. We assume an exponential utility function $U(z)=(1-e^{-\alpha z})/\alpha$, $\alpha \geq 0$ that allows us to model risk-averse customers and also converges to risk-neutral behavior ($U(z)=z$) as α goes to zero. The customer's ex-ante beliefs are given by:

$$\begin{aligned} z_{ij|keep} &= y_{ij}+\theta_{ij}+\varepsilon_{ij} \\ z_{ij|return} &= -r_{ij} \end{aligned}$$

which corresponds to:

$$\begin{aligned} U_{ij|keep} &= \left(1-\exp\left(-\alpha(y_{ij}+\theta_{ij}+\varepsilon_{ij})\right)\right)/\alpha \\ U_{ij|return} &= \left(1-\exp(\alpha r_{ij})\right)/\alpha \end{aligned} \tag{1}$$

In the above expressions, y_{ij} denotes the customer's beliefs about product quality, θ_{ij} represents the

[1] Anderson et al. (2009) only consider risk-neutral consumers; our model allows for risk aversion and treats risk-neutrality as a limiting case.

customer's beliefs about the fit of product j to her taste, and $r_{ij} \geq 0$ are the costs of returning the product. Term ε_{ij} is a standard, normally distributed, econometric error term that is known to the customer before purchase, but remains unknown to the econometrician. Its precision is normalized to one.[2]

We assume that product quality is unknown to the customer before purchase. Customers have normally distributed prior beliefs about product quality with mean μ_{ij} and precision τ_{ij}. During the path-to-purchase, each customer obtains a signal of quality x_{ij} whose precision we denote by ρ_{ij}. In the context of this paper this signal abstracts all the information that the consumer obtains about the product by browsing product pages, looking at product reviews online, examining the product at a retail store, etc. We assume that signal x_{ij} follows a normal distribution that is centered at the product's true quality. Using elementary probability theory it can then be shown that the consumer's posterior beliefs y_{ij} about product quality are normally distributed with mean $(\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij})/(\tau_{ij} + \rho_{ij})$ and precision $\tau_{ij} + \rho_{ij}$.

The fit term θ_{ij} models the fact that almost all items have subjective dimensions whose appeal differs from one customer to another. Each customer has normally distributed prior beliefs about product fit with mean λ_{ij} and precision ξ_{ij}. During the path-to-purchase, each customer obtains a signal of fit ϕ_{ij} whose precision we denote by ω_{ij}. This signal abstracts all the information that the consumer obtains about the product by reading the text of reviews about the product online, examining the product, etc. We

[2] The above specification does not explicitly include the disutility of product cost. We assume that it is factored into y_{ij}.

assume that signal y_{ij} follows a normal distribution. Using elementary probability theory it can then be shown that the consumer's posterior beliefs θ_{ij} about the fit of product to her individual taste are normally distributed with mean $\left(\xi_{ij}\lambda_{ij}+\omega_{ij}\phi_{ij}\right)/\left(\xi_{ij}+\omega_{ij}\right)$ and precision $\xi_{ij}+\omega_{ij}$.

Note that the precision of the fit variable depends on both on the customer and the product. Certain products have inherently more subjective qualities than other. Similarly, some customers are more uncertain about their own taste or needs.

Finally, we assume that return costs r_{ij} are known to the customer before purchase.

According to the above assumptions, from the perspective of a consumer contemplating a purchase, $z_{ij|keep}=y_{ij}+\theta_{ij}+\varepsilon_{ij}$ is the sum of two independently distributed normal random variables[3] and is, thus, also normally distributed with mean $m_{ij}=\left(\tau_{ij}\mu_{ij}+\rho_{ij}x_{ij}\right)/\left(\tau_{ij}+\rho_{ij}\right)+\left(\xi_{ij}\lambda_{ij}+\omega_{ij}\phi_{ij}\right)/\left(\xi_{ij}+\omega_{ij}\right)+\varepsilon_{ij}$ and variance $s_{ij}^2=1/(\tau_{ij}+\rho_{ij})+1/(\xi_{ij}+\omega_{ij})$.

The customer's expected utility from purchasing an item can be written as:

$$U_{ij}=E[U_{ij|keep}\,|\,keep]Pr(keep)+E[U_{ij|return}]Pr(return) \tag{2}$$

Customers will return the product if and only if $z_{ij|keep}<z_{ij|return}=-r_{ij}$. Customer i's expected probability of return is thus simply:

---

[3] Recall that ε_{ij} is known to the consumer and is, therefore, not a random variable from her perspective.

$$Pr(return_{ij}) = \Phi\left(\frac{-m_{ij} - r_{ij}}{s_{ij}}\right) = 1 - \Phi\left(\frac{m_{ij} + r_{ij}}{s_{ij}}\right) \tag{3}$$

with corresponding expected utility $E[U_{ij|return}] = (1 - \exp(\alpha r_{ij}))/\alpha$. Conversely, customers will keep the product if the normally distributed $z_{ij|keep} \geq -r_{ij}$. Therefore, they expect that:

$$E[U_{ij|keep} \mid keep]Pr(keep) = \int_{-r_{ij}}^{\infty} U(z)\frac{1}{s_{ij}}\phi\left(\frac{z - m_{ij}}{s_{ij}}\right)dz$$

$$= \alpha^{-1}\left[\Phi\left(\frac{m_{ij} + r_{ij}}{s_{ij}}\right) - \left(e^{-\alpha m_{ij} + \alpha^2 \frac{s_{ij}^2}{2}}\right)\Phi\left(\frac{m_{ij} + r_{ij}}{s_{ij}} - \alpha s_{ij}\right)\right]$$

Substituting into (2) yields:

$$U_{ij}(m_{ij}, r_{ij}, s_{ij}) = \alpha^{-1}\left[1 - e^{\alpha r_{ij}}\left[1 - \Phi\left(\frac{m_{ij} + r_{ij}}{s_{ij}}\right)\right] - \left(e^{-\alpha m_{ij} + \alpha^2 \frac{s_{ij}^2}{2}}\right)\Phi\left(\frac{m_{ij} + r_{ij}}{s_{ij}} - \alpha s_{ij}\right)\right]$$

Therefore:

$$Pr(order) = Pr(U_{ij}(m_{ij}, r_{ij}, s_{ij}) > 0)$$

We define H as a function of $z = m_{ij}$ as follows:

$$H(z, r_{ij}, s_{ij}) = 1 - e^{\alpha r_{ij}}\left[1 - \Phi\left(\frac{z + r_{ij}}{s_{ij}}\right)\right] - \left(e^{-\alpha z + \alpha^2 \frac{s_{ij}^2}{2}}\right)\Phi\left(\frac{z + r_{ij}}{s_{ij}} - \alpha s_{ij}\right)$$

and observe that function H has a negative lower bound ($1 - e^{\alpha r_{ij}}$), a positive upper bound (1) and increases monotonically with z . This implies that H has a unique root $\zeta(r_{ij}, s_{ij})$, such that

$H(\zeta(r_{ij}, s_{ij}), r_{ij}, s_{ij}) = 0$. Note that $\zeta(r_{ij}, s_{ij})$ is the value of m_{ij} that makes the customer indifferent between purchasing and not purchasing. A purchase occurs iff $m_{ij} = (\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij})/(\tau_{ij} + \rho_{ij}) + (\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij})/(\xi_{ij} + \omega_{ij}) + \varepsilon_{ij} > \zeta(r_{ij}, s_{ij})$. Therefore:

$$Pr(purchase) = Pr\left(m_{ij} - \zeta(r_{ij}, s_{ij}) > 0\right) = \Phi\left(\frac{\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij}}{\tau_{ij} + \rho_{ij}} + \frac{\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij}}{\xi_{ij} + \omega_{ij}} - \zeta(r_{ij}, s_{ij})\right) \tag{4}$$

Modeling returns. We are now ready to model the probability of product returns. Equation (3) gives the probability of return for a given ε_{ij} (known to the customer before purchase, unobservable by the econometrician). This is the ex-ante probability of return of a given customer i at a given purchase instance. The ex-post probability of product returns, as observed by the econometrician, is the expected probability of returns, conditional on having made a purchase, for all possible values of ε_{ij}:

$$R_{ij} = Pr(return \mid purchase) = \frac{Pr(return \wedge purchase)}{Pr(purchase)}$$

Denote $\underline{\varepsilon}(r_{ij}, x_{ij}, \rho_{ij}, \omega_{ij}) = \zeta(r_{ij}, s_{ij}) - (\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij})/(\tau_{ij} + \rho_{ij}) - (\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij})/(\xi_{ij} + \omega_{ij})$. Recall, from (4), that purchases happen iff $\varepsilon_{ij} > \underline{\varepsilon}(r_{ij}, x_{ij}, \rho_{ij}, \omega_{ij})$ and it is

$$\begin{aligned}Pr(purchase) &= \Phi\left(\frac{\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij}}{\tau_{ij} + \rho_{ij}} + \frac{\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij}}{\xi_{ij} + \omega_{ij}} - \zeta(r_{ij}, s_{ij})\right) \\ &= \Phi\left(-\underline{\varepsilon}(r_{ij}, x_{ij}, \rho_{ij}, \omega_{ij})\right) = \int_{\underline{\varepsilon}(r_{ij}, x_{ij}, \rho_{ij}, \omega_{ij})}^{\infty} \phi(\varepsilon)\,d\varepsilon\end{aligned}$$

The ex-ante consumer beliefs regarding the probability of a return are given by (3). If we assume that these beliefs are correct, we obtain:

$$R_{ij} = \frac{\int_{\underline{\varepsilon}(r_{ij},x_{ij},\rho_{ij},\omega_{ij})}^{\infty} \left[1-\Phi\left(\frac{m_{ij}+r_{ij}}{s_{ij}}\right)\right]\phi(\varepsilon)d\varepsilon}{\int_{\underline{\varepsilon}(r_{ij},x_{ij},\rho_{ij},\omega_{ij})}^{\infty} \phi(\varepsilon)d\varepsilon} = 1 - \frac{\int_{\underline{\varepsilon}(r_{ij},x_{ij},\rho_{ij},\omega_{ij})}^{\infty} \Phi\left(\frac{m_{ij}+r_{ij}}{s_{ij}}\right)\phi(\varepsilon)d\varepsilon}{\int_{\underline{\varepsilon}(r_{ij},x_{ij},\rho_{ij},\omega_{ij})}^{\infty} \phi(\varepsilon)d\varepsilon} \tag{5}$$

where $m_{ij} = \left(\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij}\right)/\left(\tau_{ij} + \rho_{ij}\right) + \left(\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij}\right)/\left(\xi_{ij} + \omega_{ij}\right) + \varepsilon$.

Comparative statics yield the following result:

Proposition 1. *If the probability of product returns is given by equation (5) and return costs are sufficiently high, each of the following reduces the probability of product returns:*

i) Lower prior uncertainty about product quality.

ii) A more precise quality signal.

iii) Lower prior uncertainty about product fit.

iv) A more precise fit signal.

v) A higher valence of the quality signal.

Consequences of Incorrect Signal Interpretation. The preceding analysis assumes that customers correctly interpret the precision and valence of the various signals that are available to them. In practice, this is not always the case. Claims about a product made on a website might be overblown. Reviews might be biased (Hu et al. 2009, Mayzlin et al. 2012). Consumers do not always perceive these irregularities before purchase (Hu et al. 2012). The following result can be shown:

Proposition 2. *Relative to the case where customers correctly interpret the information available to them:*

i) *If they erroneously assign higher valence to a quality signal, return probabilities increase.*

ii) *When return costs are sufficiently high and customers sufficiently risk averse, if they erroneously assign higher precision to a quality or fit signal, return probabilities increase.*

Proposition 2 shows that if signal attributes that are commonly associated with higher precision lead to higher returns then this constitutes evidence of misleading information (or of wrong interpretation of this information by consumers).

Effect of product uncertainty on the purchase of substitutes. So far we have considered the purchase of only one product at a time. However, consumers when uncertain about the products available at an online store might purchase more than one substitute products with the intention of returning all but one that meets their need. Extending our model to such a scenario we arrive at the following result:

Proposition 3. *In environments with quality and fit uncertainty, unit demand consumers may find it optimal to purchase $n > 1$ substitute products with the intention of keeping the best one and returning the rest. The optimal number of substitute products to purchase is monotonically increasing with the amount of uncertainty that exists in the system.*

B.1 Analytical Proofs

Lemma 1: $\frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}}$ *is increasing with* α *and is less than or equal to* $\frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}}$

Proof: Implicit differentiation of $H(\zeta(r_{ij}, s_{ij}), r_{ij}, s_{ij}) = 0$ with respect to s_{ij} gives:

$$\frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}} = \frac{\alpha s_{ij} \Phi\left(\frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}} - \alpha s_{ij}\right) - \phi\left(\frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}} - \alpha s_{ij}\right)}{\Phi\left(\frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}} - \alpha s_{ij}\right)}$$

From the properties of the normal distribution it is easy to see that: (1) $\frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}}$ is monotonically increasing with α, (2) $\lim_{\alpha \to 0} \frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}} = -\phi\left(\frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}}\right) / \Phi\left(\frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}}\right) < 0$ and, (3) $\lim_{\alpha \to \infty} \frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}} = \frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}} > 0$. A sufficient condition for the latter inequality is $\zeta(r_{ij}, s_{ij}) \geq 0$.

To see why this is true, observe that, for sufficiently large α , it is:

$$H(0, r_{ij}, s_{ij}) = 1 - e^{\alpha r_{ij}}\left[1 - \Phi\left(\frac{r_{ij}}{s_{ij}}\right)\right] - \left(e^{\alpha^2 \frac{s_{ij}^2}{2}}\right)\Phi\left(\frac{r_{ij}}{s_{ij}} - \alpha s_{ij}\right) < 0$$

which, together with the monotonicity of *H*, implies that, for large α , the root $\zeta(r_{ij}, s_{ij})$ that satisfies $H(\zeta(r_{ij}, s_{ij}), r_{ij}, s_{ij}) = 0$ is also positive.

Lemma 2*: The threshold:*

$$\underline{\varepsilon}(r_{ij}, x_{ij}, \rho_{ij}, \omega_{ij}) = \zeta(r_{ij}, s_{ij}) - \frac{\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij}}{\tau_{ij} + \rho_{ij}} - \frac{\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij}}{\xi_{ij} + \omega_{ij}}$$

above which consumers are willing to purchase, decreases when:

i) *the valence x_{ij} of the quality signal observed by the consumer increases*

If return costs are sufficiently large and consumers sufficiently risk-averse, the above threshold also decreases when:

ii) *the precision ρ_{ij} of the quality signal observed by the consumer increases*

iii) *the precision ω_{ij}* of the *fit signal observed by the consumer increasing.*

Proof: Partial differentiation of $\underline{\varepsilon}(r_{ij}, x_{ij}, \rho_{ij}, \omega_{ij})$ yields the following:

i) The impact of higher quality signal valence x_{ij} is straightforward:

$$\frac{\partial \underline{\varepsilon}}{\partial x_{ij}} = -\frac{\rho_{ij}}{\tau_{ij} + \rho_{ij}} < 0$$

ii) The impact of higher quality signal precision ρ_{ij} depends on the degree of customer risk aversion. It is:

$$\frac{\partial \underline{\varepsilon}}{\partial \rho_{ij}} = \frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}} \frac{\partial s_{ij}}{\partial \rho_{ij}} - \frac{\tau_{ij}(x_{ij} - \mu_{ij})}{(\tau_{ij} + \rho_{ij})^2}$$

where $\frac{\partial s_{ij}}{\partial \rho_{ij}} = -\frac{1}{2 s_{ij}(\tau_{ij} + \rho_{ij})^2}$ and, by Lemma 1, $\frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}} < 0$ for small α and $\frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}} \approx \frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}} > 0$ for large α. Therefore, for small α:

$$\frac{\partial \underline{\varepsilon}}{\partial \rho_{ij}} = \left| \frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}} \right| \frac{1}{2 s_{ij}(\tau_{ij} + \rho_{ij})^2} - \frac{\tau_{ij}(x_{ij} - \mu_{ij})}{(\tau_{ij} + \rho_{ij})^2}$$

The above expression is positive, except when the signal valence x_{ij} is sufficiently higher than the prior beliefs μ_{ij}. This means that, if customers are not very risk averse, if they receive a signal that is higher than their prior expectations, the more precise the signal, the lower the threshold $\underline{\varepsilon}$ above which they are willing to purchase. On the other hand, if they receive a signal that is lower than their prior expectations, the more precise the signal, the higher the threshold $\underline{\varepsilon}$ above which they are willing to purchase. For large α:

$$\frac{\partial \underline{\varepsilon}}{\partial \rho_{ij}} \approx -\frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{2 s_{ij}^2 (\tau_{ij} + \rho_{ij})^2} - \frac{\tau_{ij}(x_{ij} - \mu_{ij})}{(\tau_{ij} + \rho_{ij})^2}$$

is negative for sufficiently large return costs r_{ij}.

iii) Observing the equivalence of ρ_{ij} and ω_{ij} in expressions $\frac{\tau_{ij}\mu_{ij} + \rho_{ij} x_{ij}}{\tau_{ij} + \rho_{ij}}$ and $\frac{\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij}}{\xi_{ij} + \omega_{ij}}$, as well as in $s_{ij}^2 = 1/(\tau_{ij} + \rho_{ij}) + 1/(\xi_{ij} + \omega_{ij})$, we conclude that the impact of increasing the precision ω_{ij} of the fit signal is identical to that of increasing the precision ρ_{ij} of the quality signal. The proof is identical to that of part ii) by substituting the corresponding symbols.

Proof of Proposition 1

We rewrite (5) as

$$R_{ij} = 1 - \frac{\int_{\underline{\varepsilon}(z)}^{\infty} \Phi\left(w(z,\varepsilon)\right)\phi(\varepsilon)d\varepsilon}{\int_{\underline{\varepsilon}(z)}^{\infty} \phi(\varepsilon)d\varepsilon} \quad (6)$$

where Z represents the vector of variables $(\tau_{ij}, \rho_{ij}, \xi_{ij}, \omega_{ij})$ and

$$w(z,\varepsilon) = \frac{m_{ij} + r_{ij}}{s_{ij}} = \frac{\frac{\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij}}{\tau_{ij} + \rho_{ij}} + \frac{\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij}}{\xi_{ij} + \omega_{ij}} + r_{ij} + \varepsilon}{s_{ij}}$$

$$\underline{\varepsilon}(r_{ij}, z) = \zeta(r_{ij}, s_{ij}) - \frac{\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij}}{\tau_{ij} + \rho_{ij}} - \frac{\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij}}{\xi_{ij} + \omega_{ij}}$$

Differentiating with respect to any of the variables contained in vector Z gives:

$$\frac{\partial R_{ij}}{\partial z}=-\frac{\left[\int_{\underline{\varepsilon}(z)}^{\infty}\left(\frac{\partial w(z,\varepsilon)}{\partial z}\right)\phi(w(z,\varepsilon))\phi(\varepsilon)d\varepsilon-\left(\frac{\partial\underline{\varepsilon}(z)}{\partial z}\right)\Phi(w(z,\underline{\varepsilon}(z)))\phi(\underline{\varepsilon}(z))\right]\left(\int_{\underline{\varepsilon}(z)}^{\infty}\phi(\varepsilon)d\varepsilon\right)+\left(\frac{\partial\underline{\varepsilon}(z)}{\partial z}\right)\left(\int_{\underline{\varepsilon}(z)}^{\infty}\Phi(w(z,\varepsilon))\phi(\varepsilon)d\varepsilon\right)\phi(\underline{\varepsilon}(z))}{\left(\int_{\underline{\varepsilon}(z)}^{\infty}\phi(\varepsilon)d\varepsilon\right)^{2}} \tag{7}$$

$$=-\frac{\left(\int_{\underline{\varepsilon}(z)}^{\infty}\left(\frac{\partial w(z,\varepsilon)}{\partial z}\right)\phi(w(z,\varepsilon))\phi(\varepsilon)d\varepsilon\right)\left(\int_{\underline{\varepsilon}(z)}^{\infty}\phi(\varepsilon)d\varepsilon\right)+\left(\frac{\partial\underline{\varepsilon}(z)}{\partial z}\right)\phi(\underline{\varepsilon}(z))\left[\left(\int_{\underline{\varepsilon}(z)}^{\infty}\Phi(w(z,\varepsilon))\phi(\varepsilon)d\varepsilon\right)-\Phi(w(z,\underline{\varepsilon}(z)))\left(\int_{\underline{\varepsilon}(z)}^{\infty}\phi(\varepsilon)d\varepsilon\right)\right]}{\left(\int_{\underline{\varepsilon}(z)}^{\infty}\phi(\varepsilon)d\varepsilon\right)^{2}}$$

Because $w(z,\varepsilon)$ is monotonically increasing with ε and $\Phi(w)$ is monotonically increasing with w, it is $\left(\int_{\underline{\varepsilon}(z)}^{\infty}\Phi(w(z,\varepsilon))\phi(\varepsilon)d\varepsilon\right)-\Phi(w(z,\underline{\varepsilon}(z)))\left(\int_{\underline{\varepsilon}(z)}^{\infty}\phi(\varepsilon)d\varepsilon\right)>0$. This and (7) then give:

$$\begin{aligned}&\frac{\partial w(z,\varepsilon)}{\partial z}>0 \text{ and } \frac{\partial\underline{\varepsilon}(z)}{\partial z}>0 \text{ imply } \frac{\partial R_{ij}}{\partial z}<0\\&\frac{\partial w(z,\varepsilon)}{\partial z}<0 \text{ and } \frac{\partial\underline{\varepsilon}(z)}{\partial z}<0 \text{ imply } \frac{\partial R_{ij}}{\partial z}>0\end{aligned} \tag{8}$$

Partial differentiation of $w(z,\varepsilon)$ with respect to each variable of vector z gives:

$$\begin{aligned}\frac{\partial w}{\partial\tau_{ij}}&=\frac{1}{s_{ij}(\tau_{ij}+\rho_{ij})^{2}}\left(\rho_{ij}(\mu_{ij}-x_{ij})+\frac{m_{ij}+r_{ij}}{2s_{ij}^{2}}\right)>0\\\frac{\partial w}{\partial\rho_{ij}}&=\frac{1}{s_{ij}(\tau_{ij}+\rho_{ij})^{2}}\left(\tau_{ij}(x_{ij}-\mu_{ij})+\frac{m_{ij}+r_{ij}}{2s_{ij}^{2}}\right)>0\\\frac{\partial w}{\partial\xi_{ij}}&=\frac{1}{s_{ij}(\xi_{ij}+\omega_{ij})^{2}}\left(\omega_{ij}(\lambda_{ij}-\phi_{ij})+\frac{m_{ij}+r_{ij}}{2s_{ij}^{2}}\right)>0\\\frac{\partial w}{\partial\omega_{ij}}&=\frac{1}{s_{ij}(\xi_{ij}+\omega_{ij})^{2}}\left(\xi_{ij}(\phi_{ij}-\lambda_{ij})+\frac{m_{ij}+r_{ij}}{2s_{ij}^{2}}\right)>0\end{aligned} \tag{9}$$

where positivity requires that $r_{ij} > -2s_{ij}^2 \max\begin{bmatrix} \rho_{ij}(\mu_{ij}-x_{ij}) \\ \tau_{ij}(x_{ij}-\mu_{ij}) \\ \omega_{ij}(\lambda_{ij}-\phi_{ij}) \\ \xi_{ij}(\phi_{ij}-\lambda_{ij}) \end{bmatrix} - m_{ij}$, i.e., that return costs are sufficiently high.

From (8) and (9) we can infer that $\frac{\partial R_{ij}}{\partial \rho_{ij}} < 0$ and $\frac{\partial R_{ij}}{\partial \omega_{ij}} < 0$ for the parameter regions where it is $\frac{\partial \underline{\varepsilon}(z)}{\partial \rho_{ij}} > 0$ and $\frac{\partial \underline{\varepsilon}(z)}{\partial \omega_{ij}} > 0$ respectively.

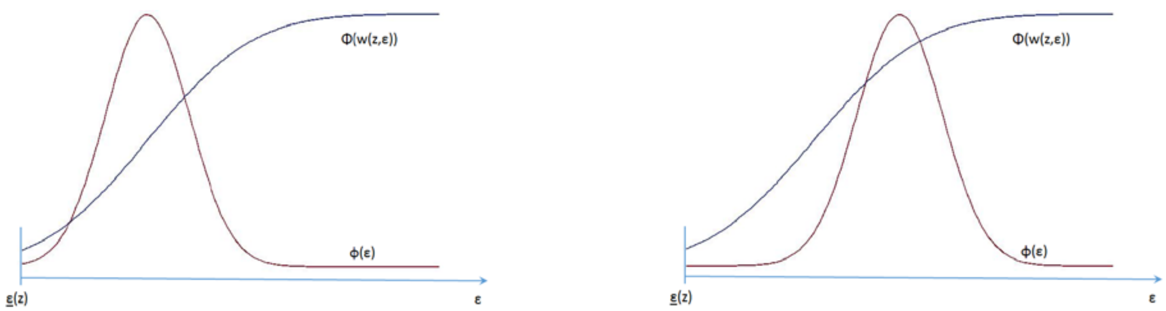
Since it is $\frac{\partial w}{\partial x_{ij}} = \frac{\rho_{ij}}{s_{ij}(\tau_{ij}+\rho_{ij})} > 0$ and (from Lemma 2) $\frac{\partial \underline{\varepsilon}(z)}{\partial x_{ij}} < 0$ the above analysis does not provide firm conclusions about the impact of signal valence x_{ij} on returns. However, we observe that (6) can be rewritten as:

$$R_{ij} = 1 - \frac{\int_0^\infty \Phi\left(w(z,\underline{\varepsilon}(z)+\varepsilon)\right)\phi(\underline{\varepsilon}(z)+\varepsilon)d\varepsilon}{\int_0^\infty \phi(\underline{\varepsilon}(z)+\varepsilon)d\varepsilon} \qquad (10)$$

where

$$w(z,\underline{\varepsilon}(z)+\varepsilon) = \frac{r_{ij}+\zeta(r_{ij},s_{ij})+\varepsilon}{s_{ij}}$$

is independent of x_{ij}. This implies that increases in x_{ij} do not affect the values of $\Phi(\cdot)$ that get integrated in equation (10) but, because $\frac{\partial \underline{\varepsilon}(z)}{\partial x_{ij}} < 0$, simply shift the probability density that these values are multiplied by to compute the expected value, to the right. We assume that μ_{ij} and x_{ij} are sufficiently high, so that $\underline{\varepsilon}(z) \leq 0$. Recall that the normal density function is monotonically increasing for negative values of ε and attains its maximum at $\varepsilon = 0$, and $\Phi(w(z, \underline{\varepsilon}(z) + \varepsilon))$ is also monotonically increasing with ε. Thus, shifting the distribution to the right implies that the center of mass of the distribution is multiplied with higher values of $\Phi(w(z, \underline{\varepsilon}(z) + \varepsilon))$. This is graphically illustrated in Figure A.1. But this, in turn implies that, as χ_{ij} grows and $\underline{\varepsilon}(z)$ declines, the expected value $\frac{\int_0^\infty \Phi\left(w(z, \underline{\varepsilon}(z) + \varepsilon)\right)\phi(\underline{\varepsilon}(z) + \varepsilon)d\varepsilon}{\int_0^\infty \phi(\underline{\varepsilon}(z) + \varepsilon)d\varepsilon}$ grows and R_{ij} declines.



(a) Baseline case.

(b) As χ_{ij} grows, threshold $\underline{\varepsilon}(z)$ declines, shifting $\phi(\varepsilon)$ to the right.

Figure A.1: As χ_{ij} grows, the center of mass of distribution $\phi(\varepsilon)$ maps to higher values of $\Phi(w(z, \underline{\varepsilon}(z) + \varepsilon))$. This results in a higher expected value.

Following a somewhat similar line of reasoning we can show that it is $\frac{\partial R_{ij}}{\partial \rho_{ij}} < 0$ and $\frac{\partial R_{ij}}{\partial \omega_{ij}} < 0$

also for the parameter regions where it is $\frac{\partial \underline{\varepsilon}(z)}{\partial \rho_{ij}} < 0$ and $\frac{\partial \underline{\varepsilon}(z)}{\partial \omega_{ij}} < 0$ respectively. The key observation is that

$$\frac{\partial w(z, \underline{\varepsilon}(z))}{\partial s_{ij}} = \frac{1}{s_{ij}^2}\left[\frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}} s_{ij} - (r_{ij} + \zeta(r_{ij}, s_{ij}))\right] \leq 0$$

This is so because, as shown in Lemma 1, $\frac{\partial \zeta(r_{ij}, s_{ij})}{\partial s_{ij}}$ is increasing with α and is less than or equal to $\frac{\zeta(r_{ij}, s_{ij}) + r_{ij}}{s_{ij}}$. The above expression, however, implies that $\frac{\partial w(z, \underline{\varepsilon}(z))}{\partial \rho_{ij}} = \frac{\partial w(z, \underline{\varepsilon}(z))}{\partial s_{ij}} \frac{\partial s_{ij}}{\partial \rho_{ij}} \geq 0$ and, similarly, $\frac{\partial w(z, \underline{\varepsilon}(z))}{\partial \omega_{ij}} \geq 0$. This means that $w(z, \underline{\varepsilon}(z) + \varepsilon)$ is monotonically increasing with ρ_{ij}, ω_{ij} for all values of x_{ij}. Therefore, all values of $\Phi(\cdot)$ that get integrated in equation (10) are increasing with ρ_{ij}, ω_{ij}. In that case, when $\frac{\partial \underline{\varepsilon}(z)}{\partial \rho_{ij}} < 0$ (resp. $\frac{\partial \underline{\varepsilon}(z)}{\partial \omega_{ij}} < 0$) the arguments made in the preceding paragraph can be used to show that as ρ_{ij} (resp. ω_{ij}) grows and $\underline{\varepsilon}(z)$ declines, the expected value $\frac{\int_0^\infty \Phi\left(w(z, \underline{\varepsilon}(z) + \varepsilon)\right)\phi(\underline{\varepsilon}(z) + \varepsilon) d\varepsilon}{\int_0^\infty \phi(\underline{\varepsilon}(z) + \varepsilon) d\varepsilon}$ grows and R_{ij} declines.

The same result can be obtained for the precision τ_{ij} (respectively ξ_{ij}) of prior beliefs by observing the symmetry between τ_{ij} and ρ_{ij} (respectively ξ_{ij} and ω_{ij}) in expression $\frac{\tau_{ij}\mu_{ij} + \rho_{ij}x_{ij}}{\tau_{ij} + \rho_{ij}}$ (respectively $\frac{\xi_{ij}\lambda_{ij} + \omega_{ij}\phi_{ij}}{\xi_{ij} + \omega_{ij}}$).

Proof of Proposition 2

In the proof of Lemma 2 we showed that higher *perceived* valence of product quality reduces the threshold $\underline{\varepsilon}(r_{ij}, x_{ij}, \rho_{ij}, \omega_{ij})$ above which a purchase takes place. Furthermore, if customers are sufficiently risk averse and return costs sufficiently high, we showed that higher *perceived* precision of product quality and fit information has the same effect. Let us now assume that customer beliefs are erroneous. Specifically, let us assume that customer beliefs about valence and precision change, whereas the actual quantities do not. In such a setting, we rewrite equation (6) that gives the observed return probability as follows:

$$R'_{ij} = 1 - \frac{\int_{\underline{\varepsilon}(z)}^{\infty} \Phi(w(z', \varepsilon))\phi(\varepsilon)d\varepsilon}{\int_{\underline{\varepsilon}(z)}^{\infty} \phi(\varepsilon)d\varepsilon}$$

The perceived signal attributes are denoted, as before, by vector z, whereas the actual attributes are now denoted by a different vector z'. The situation we are interested in modeling is one where z (perceptions) change, whereas z' (reality) does not. In that case, the derivative (7) simplifies to:

$$\frac{\partial R'_{ij}}{\partial z} = -\frac{\left(\frac{\partial \underline{\varepsilon}(z)}{\partial z}\right)\phi(\underline{\varepsilon}(z))\left[\left(\int_{\underline{\varepsilon}(z)}^{\infty} \Phi(w(z', \varepsilon))\phi(\varepsilon)d\varepsilon\right) - \Phi(w(z', \underline{\varepsilon}(z)))\left(\int_{\underline{\varepsilon}(z)}^{\infty} \phi(\varepsilon)d\varepsilon\right)\right]}{\left(\int_{\underline{\varepsilon}(z)}^{\infty} \phi(\varepsilon)d\varepsilon\right)^2}$$

As discussed in the proof of Proposition 1, it is $\left(\int_{\underline{\varepsilon}(z)}^{\infty} \Phi(w(z', \varepsilon))\phi(\varepsilon)d\varepsilon\right) - \Phi(w(z', \underline{\varepsilon}(z)))\left(\int_{\underline{\varepsilon}(z)}^{\infty} \phi(\varepsilon)d\varepsilon\right) > 0$ and therefore, $\frac{\partial R'_{ij}}{\partial z}$ has the opposite sign of $\frac{\partial \underline{\varepsilon}(z)}{\partial z}$. In Lemma 2 we showed that it is always $\frac{\partial \underline{\varepsilon}(z)}{\partial x_{ij}} < 0$. Furthernore, for sufficiently high return costs and risk averse customers we showed that it is also $\frac{\partial \underline{\varepsilon}(z)}{\partial \rho_{ij}} < 0$ (resp. $\frac{\partial \underline{\varepsilon}(z)}{\partial \omega_{ij}} < 0$). Therefore

$\frac{\partial R'_{ij}}{\partial x_{ij}} > 0$, $\frac{\partial R'_{ij}}{\partial \rho_{ij}} > 0$ and $\frac{\partial R'_{ij}}{\partial \omega_{ij}} > 0$.

Proof of Proposition 3

We base this proof on the same setup as our baseline model. We assume that customers only need one product. In the presence of quality and fit uncertainty, they may decide to purchase $n \geq 1$ substitute products to increase the probability that *at least one* product will be satisfactory. In that case, that we label as $keep_{ij|n}$, we assume that customers will keep the product that offers them the highest utility among the *n* purchased and will return the remaining $n - 1$ products. If no product is satisfactory, customers will return all *n* products. We label the latter case as $return_{ij|n}$.

We model the utilities of substitute items to item *j* as iid normally distributed random variables with mean m_{ij} and variance s_{ij} (see baseline model). Purchasing one additional substitute product increases the expected utility of the max of the bundle. At the same time it also increases the total return costs by the cost of returning one additional product. Customers will only purchase multiple products if the increase in the expected utility of the max is sufficient to offset the increased expected return cosrs. The intuition of the proof is based on the fact that, the higher the uncertainty about quality and fit (i.e. the higher the variance s_{ij}), the higher the element of surprise, which, in this case translates into a higher increase in the expected utility of the best product when one adds one item to the bundle.

Formally, the customer's expected utility from purchasing *n* substitute items (of which she only needs one) can be written as:

$$U_{ij|n} = E\left[U_{keep_{ij|n}}\right]\Pr\left(keep_{ij|n}\right) + E[U_{return_{ij|n}}]\Pr(return_{ij|n}) \quad (1)$$

Customers will return *all* *n* products if and only if *all* items $k = 1, .., n$ satisfy $z_{ij|k} < -r_{ij}$, i.e. iff the maximum of the bundle offers utility that is less than $-r_{ij}$. It is known that the CDF of the maximum of *n* iid normal variables is simply $H(z|m,s,n) = \Phi\left(\frac{z-m}{s}\right)^n$, with corresponding density function $h(z|m,s,n) = \frac{n}{s}\Phi\left(\frac{z-m}{s}\right)^{n-1}\phi\left(\frac{z-m}{s}\right)^n$.

Customer i 's expected probability of returning all products is thus simply:

$$Pr(return_{ij|n}) = H(-r_{ij} \mid m_{ij}, s_{ij}, n) = \Phi\left(\frac{-m_{ij} - r_{ij}}{s_{ij}}\right)^n \tag{2}$$

with corresponding expected utility $E\left[U_{return_{ij|n}}\right] = -nr_{ij}$. Conversely, customers will keep *one* product (the one offering them maximum utility) and will return the remaining *n*-1 products (at cost $(n-1)r_{ij}$) if there is *at least one* product *k* for which the normally distributed $z_{ij|k} \geq -r_{ij}$, i.e. if the maximum utility of the *n* products is higher than the return costs.[4] Therefore, they expect that:

$$E[U_{keep_{ij|n}}]Pr(keep_{ij|n}) = \int_{-r_{ij}}^{\infty} (z - (n-1)r_{ij})h(z \mid m_{ij}, s_{ij}, n)dz \tag{3}$$

Substituting (2) and (3) into (1) and simplifying, we obtain:

$$U_{ij|n} = \int_{-r_{ij}}^{\infty} z \cdot h\left(z \middle| m_{ij}, s_{ij}, n\right)dz + \left(1 - H\left(-r_{ij} \middle| m_{ij}, s_{ij}, n\right)\right) r_{ij} - nr_{ij} \tag{4}$$

Consumers will purchase the number of products *n* that maximizes (4). Our aim is to show that this maximum is monotonically increasing with the variance s_{ij} (which implies that it is monotonically decreasing with the precision of the information signal).

From comparative statics theory, to show this, it suffices to show that: $\frac{\partial^2 U_{ij|n}}{\partial n \partial s_{ij}} \geq 0$. Differentiating (4), we obtain:

$$\frac{\partial^2 U_{ij|n}}{\partial n \partial s_{ij}} = \int_{-r_{ij}}^{\infty} z \cdot \frac{\partial^2 h\left(z \middle| m_{ij}, s_{ij}, n\right)}{\partial n \partial s_{ij}} dz + \frac{\partial^2\left(1 - H\left(-r_{ij} \middle| m_{ij}, s_{ij}, n\right)\right)}{\partial n \partial s_{ij}} r_{ij} \tag{5}$$

Straightforward differentiation gives:

$$\frac{\partial^2 H\left(-r_{ij} \middle| m_{ij}, s_{ij}, n\right)}{\partial n \partial s_{ij}} = \left(\frac{m_{ij} + r_{ij}}{s_{ij}}\right)\left[ln\ \Phi\left(-\frac{m_{ij} + r_{ij}}{s_{ij}}\right) + \frac{1}{n}\right] h\left(-r_{ij} \middle| m_{ij}, s_{ij}, n\right) \tag{6}$$

In the above expression the terms $\left(\frac{m_{ij} + r_{ij}}{s_{ij}}\right)$ and $h\left(-r_{ij} \middle| m_{ij}, s_{ij}, n\right)$ are always positive, whereas the term $ln\ \Phi\left(-\frac{m_{ij} + r_{ij}}{s_{ij}}\right) + \frac{1}{n}$ is negative for all *n*, provided that $m_{ij} + r_{ij}$ is sufficiently large, compared to s_{ij}.[5]

---

[4] We assume for simplicity that the return costs are always lower than the purchase price, so that the unused *n*-1 products, that offer zero additional utility to the consumer, are all returned.
[5] Specifically, the requirement is that $m_{ij} + r_{ij} \geq 0.34 s_{ij}$

Therefore, $\frac{\partial^2 H(-r_{ij}|m_{ij},s_{ij},n)}{\partial n \partial s_{ij}} \leq 0$, which implies that $\frac{\partial^2 \left(1-H(-r_{ij}|m_{ij},s_{ij},n)\right)}{\partial n \partial s_{ij}} \geq 0$.

Now, observe that $1 - H\left(-r_{ij} \middle| m_{ij}, s_{ij}, n\right) = \int_{-r_{ij}}^{\infty} h\left(z \middle| m_{ij}, s_{ij}, n\right) dz$, which implies that $\int_{-r_{ij}}^{\infty} \frac{\partial^2 h(z|m_{ij},s_{ij},n)}{\partial n \partial s_{ij}} dz = \frac{\partial^2 \left(1-H(-r_{ij}|m_{ij},s_{ij},n)\right)}{\partial n \partial s_{ij}} \geq 0$. From the second mean value theorem for integrals, there exists a $\zeta \geq -r_{ij}$ such that:

$$\int_{-r_{ij}}^{\infty} z \cdot \frac{\partial^2 h(z|m_{ij},s_{ij},n)}{\partial n \partial s_{ij}} dz = \zeta \int_{-r_{ij}}^{\infty} \frac{\partial^2 h(z|m_{ij},s_{ij},n)}{\partial n \partial s_{ij}} dz = \zeta \frac{\partial^2 \left(1-H(-r_{ij}|m_{ij},s_{ij},n)\right)}{\partial n \partial s_{ij}} \tag{7}$$

Substituting (7) into (5) we finally find that:

$$\frac{\partial^2 U_{ij|n}}{\partial n \partial s_{ij}} = \frac{\partial^2 \left(1 - H\left(-r_{ij} \middle| m_{ij}, s_{ij}, n\right)\right)}{\partial n \partial s_{ij}} \left(\zeta + r_{ij}\right) \geq 0$$

since it is both $\frac{\partial^2 \left(1-H(-r_{ij}|m_{ij},s_{ij},n)\right)}{\partial n \partial s_{ij}} \geq 0$ and $\zeta \geq -r_{ij}$.

C Additional Robustness Checks

C.1 Effect of volume and helpfulness on return

	(1) Product Fixed Effect		(2) Customer Fixed Effect		(3) Customer & Product Fixed Effects	
Variables	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Number of reviews	-0.2022	(0.0821) *	-0.3217	(0.0705) ***	-0.4334	(0.1288) ***
Number of reviews^2	0.0234	(0.0133) .	0.0424	(0.0118) ***	0.0594	(0.0210) **
Valence of reviews	-0.1231	(0.0296) ***	-0.1069	(0.0279) ***	-0.1599	(0.0479) ***
Dispersion of ratings	-0.0436	(0.0448)	0.0524	(0.0411)	0.0569	(0.0707)
Helpful reviews	-0.3321	(0.0696) ***	-0.1933	(0.0630) **	-0.3836	(0.1099) ***
By top reviewers	0.0761	(0.0879)	-0.0998	(0.0791)	0.1509	(0.1385)
Review length	0.0029	(0.0008) ***	0.0003	(0.0007)	0.0029	(0.0013) *
Time to purchase (in years)	0.0257	(0.0740)	0.1387	(0.0744) .	0.1731	(0.1320)
List Price			0.0063	(0.0004) ***		
Discount	-0.2879	(0.0740) ***	-0.8391	(0.0884) ***	-0.9340	(0.1407) ***
Holiday	-0.0046	(0.0417)	-0.0423	(0.0484)	-0.0211	(0.0735)
Repeated browses	-0.0365	(0.0138) **	-0.0856	(0.0156) ***	-0.1192	(0.0225) ***
Products browsed	0.0076	(0.0005) ***	0.0046	(0.0008) ***	0.0052	(0.0011) ***
Keyword searches	-0.0172	(0.0068) *	-0.0131	(0.0090)	-0.0253	(0.0137) .
Recency	0.0009	(0.0002) ***	-0.0006	(0.0003) .	-0.0011	(0.0005) *
Above	-0.4809	(0.1265) ***	-0.4804	(0.1632) **	-0.6290	(0.2526) *
Promo	0.1612	(0.0376) ***	0.0208	(0.0476)	0.0205	(0.0725)
Offline shopper	-0.4358	(0.0349) ***				
Frequency (per year)	-0.0019	(0.0004) ***				
Age > 36	-0.2196	(0.0345) ***				
Male	-0.5054	(0.0722) ***				
Distance	-0.0004	(0.0001) **				
Accessories (baseline)						
Clothing			0.4143	(0.0473) ***		
Home item or furniture			-0.8276	(0.0843) ***		
Misc			0.3223	(0.1747) .		
Brand A (baseline)						
Brand B			-0.3103	(0.0993) **		
Brand C			-0.1458	(0.0842) .		
Number of consumers	7416		2155		1832	
Number of products	3364		6783		2669	
Number of transactions	32940		24361		16389	

Significance codes are 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 3 Effect of free shipping of online orders.

	(1) Product Fixed Effect		(2) Customer Fixed Effect		(3) Customer & Product Fixed Effects	
Variables	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Number of reviews	-0.2188	(0.0933) *	-0.2745	(0.0707) ***	-0.3499	(0.1257) **
Number of reviews^2	0.0294	(0.0152) .	0.0343	(0.0118) **	0.0433	(0.0210) *
Valence of reviews	-0.0932	(0.0342) **	-0.1230	(0.0285) ***	-0.1808	(0.0493) ***
Dispersion of ratings	-0.0304	(0.0509)	0.0345	(0.0415)	0.0663	(0.0714)
Helpful reviews	-0.1817	(0.0793) *	-0.1080	(0.0638) .	-0.1641	(0.1112)
By top reviewers	0.0035	(0.1005)	-0.0966	(0.0792)	0.0460	(0.1384)
Review length	0.0021	(0.0010) *	0.0001	(0.0008)	0.0020	(0.0013)
Time to purchase (in years)	-0.0174	(0.0705)	0.0461	(0.0649)	0.0216	(0.1103)
In-store purchase	-0.3577	(0.0691) ***	-0.8930	(0.0766) ***	-1.2030	(0.1122) ***
List Price			0.0073	(0.0004) ***		
Discount	-0.4419	(0.0814) ***	-0.9305	(0.0879) ***	-0.9017	(0.1376) ***
Holiday	-0.0515	(0.0473)	-0.0415	(0.0490)	-0.0565	(0.0735)
Repeated browses	0.0079	(0.0164)	-0.0641	(0.0161) ***	-0.0927	(0.0237) ***
Products browsed	0.0020	(0.0006) ***	0.0026	(0.0007) ***	0.0031	(0.0011) **
Keyword searches	-0.0004	(0.0075)	-0.0088	(0.0087)	-0.0103	(0.0131)
Recency	0.0008	(0.0002) ***	-0.0003	(0.0004)	-0.0003	(0.0006)
Another return	2.3201	(0.0407) ***	0.1722	(0.0395) ***	0.4539	(0.0591) ***
Promo	0.0060	(0.0453)	-0.0504	(0.0521)	-0.0860	(0.0792)
Offline shopper	-0.3344	(0.0419) ***				
Frequency (per year)	-0.0027	(0.0004) ***				
Age > 36	-0.1146	(0.0394) **				
Male	-0.2795	(0.0835) ***				
Distance	-0.0003	(0.0001) *				
Accessories (baseline)						
Clothing			0.3541	(0.0492) ***		
Home item or furniture			-0.7885	(0.0883) ***		
Misc			0.2263	(0.1913)		
Brand A (baseline)						
Brand B			-0.1729	(0.1082)		
Brand C			-0.2294	(0.0896) *		
Number of consumers	5003		1889		1739	
Number of products	2998		6318		2634	
Number of transactions	24450		19764		15976	

Significance codes are 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 4 Effect of return of another product in the same order

	(1)		(2)		(3)	
	Product Fixed Effect		Customer Fixed Effect		Customer & Product Fixed Effects	
Variables	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Number of reviews	-0.1472	(0.0664) *	-0.2061	(0.0565) ***	-0.2237	(0.0895) *
Number of reviews^2	0.0187	(0.0106) .	0.0216	(0.0093) *	0.0206	(0.0141)
Valence of reviews	-0.1408	(0.0260) ***	-0.1400	(0.0242) ***	-0.2085	(0.0374) ***
Dispersion of ratings	0.0097	(0.0380)	0.0352	(0.0348)	0.1029	(0.0533) .
Helpful reviews	-0.2514	(0.0598) ***	-0.1073	(0.0537) *	-0.2206	(0.0836) **
By top reviewers	0.1058	(0.0737)	-0.0616	(0.0656)	0.2028	(0.1019) *
Review length	0.0020	(0.0007) **	-0.0002	(0.0006)	0.0013	(0.0010)
Time to purchase (in years)	-0.3418	(0.0489) ***	0.0177	(0.0527)	-0.0200	(0.0790)
In-store purchase	-0.7970	(0.0435) ***	-0.9512	(0.0500) ***	-1.2320	(0.0670) ***
List Price			0.0056	(0.0003) ***		
Discount	-0.4860	(0.0602) ***	-0.9653	(0.0708) ***	-1.0867	(0.1002) ***
Holiday	0.0096	(0.0349)	-0.0136	(0.0387)	0.0389	(0.0518)
Total returns	0.0270	(0.0008) ***	-0.0015	(0.0012)	-0.0014	(0.0015)
Repeated browses	-0.0374	(0.0112) ***	-0.0598	(0.0120) ***	-0.0750	(0.0154) ***
Products browsed	0.0043	(0.0004) ***	0.0035	(0.0006) ***	0.0037	(0.0008) ***
Keyword searches	-0.0111	(0.0060) .	-0.0091	(0.0073)	-0.0156	(0.0100)
Recency	0.0008	(0.0002) ***	-0.0008	(0.0003) *	-0.0011	(0.0004) *
Promo	0.1508	(0.0361) ***	-0.0042	(0.0436)	-0.0211	(0.0598)
Offline shopper	-0.4239	(0.0326) ***				
Frequency (per year)	-0.0063	(0.0004) ***				
Age > 36	-0.1525	(0.0291) ***				
Male	-0.3263	(0.0634) ***				
Distance	-0.0004	(0.0001) **				
Accessories (baseline)						
Clothing			0.4164	(0.0417) ***		
Home item or furniture			-0.8367	(0.0727) ***		
Misc			0.3718	(0.1517) *		
Brand A (baseline)						
Brand B			-0.4328	(0.0860) ***		
Brand C			-0.3492	(0.0716) ***		
Number of consumers	8614		2693		2441	
Number of products	3971		7936		3436	
Number of transactions	47392		36749		27587	

Significance codes are 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.

Table 5 Effect of number of products returned.

The number of purchases is correlated with the number of returns at 0.86. So, we include only one of them. The effects of the key review variables are similar to the results presented in Table 5 after controlling for either total number of purchases or total number of returns. However, unlike the total number of returns that has no significant effect, the total number of purchases has a significant positive effect (p-value 0.037) on the probability of return with both consumer and product fixed effects. The table that is otherwise very similar to the Table 13, is omitted to conserve space.

C.2 Additional Tests in Regression Discontinuity Analysis

Dependent variable: ret	Model 4	(1)	(2)	(3)	(4)
Discontinuity	0.186 (0.077)	0.188 (0.077)	0.193 (0.083)	0.085 (0.076)	0.086 (0.095)
Quadratic av_rating		X			
Full set of controls			X		
Cut point shifted +0.02				X	
Cut point shifted -0.02					X
Number of consumers	1340	1340	1340	1386	1185
Number of products	1485	1485	1485	1588	1279
Number of purchases	9236	9236	9236	9811	7616

Table 6 The effect of discontinuity treatment when average rating crosses the x.y5 thresholds.

The first two robustness checks show that the effect of the discontinuity is neither due to non-linearity across the cut-point or due to co-occurrence of the discontinuity with other factors that could alter probability of return. The last two falsification checks show that the effects are not due to comparison of products with higher and lower average ratings around *any* threshold, but only around the cut point that causes rounding up of the average ratings.