Online Supplement to Discretion in Automated Supermarket Replenishment: Censorship Bias and Self-inflicted Stockouts

Bengü Özdemir
College of Business and Economics, University of Wisconsin-Whitewater

Antti Tenhiälä
IE Business School, IE University

1. Robustness Checks

1.1. Alternative Estimation Strategies

We employ various alternative model specifications to test the robustness of our results. In our main specification, the moderation relationship between censorship bias and downward deviations is constructed via an interaction term; as an alternative approach, moderation can also be examined using selection-based models. To this end, we first run an endogenous switching regression model, also known as Tobit type 5 (Amemiya 1984), which has been used in earlier decision-making research by Freeman et al. (2021) and Liu et al. (2015). This model has two second-stage outcome equations, enabling us to estimate the coefficients under both situations: when there is no deviation and when there is a downward deviation. Accordingly, the estimation sample for the former equation excludes cases with upward deviations. The following two structural equations are jointly estimated via the Newton–Raphson maximization method by assuming that ξ_i, ζ_i and ϵ_i have a trivariate normal distribution (Lokshin and Sajaia 2011, Marra et al. 2022).

$$\begin{aligned} Stockout_i^* &= \gamma_0 + \gamma_1 ASRProposalSize_i + \gamma_2 RelativeProposalSize_i \\ &\quad + \gamma_3 RecentStockout_i + \mathbf{\Gamma} \boldsymbol{X_i} + \zeta_i \\ Stockout_i &= \mathbb{I}[Stockout_i^* > 0] \qquad\qquad if\ DownwardDeviation_i = 0 \end{aligned}$$

and

$$\begin{aligned} Stockout_i^* &= \gamma_4 + \gamma_5 ASRProposalSize_i + \gamma_6 DeviationSize_i + \gamma_7 RelativeProposalSize_i \\ &\quad + \gamma_8 RecentStockout_i + \mathbf{\Gamma} \boldsymbol{X_i} + \xi_i \\ Stockout_i &= \mathbb{I}[Stockout_i^* > 0] \qquad\qquad if\ DownwardDeviation_i = 1 \end{aligned}$$

where

$$\begin{aligned} DownwardDeviation_i^* &= \beta_0 + \beta_1 ASRProposalSize_i + \beta_2 CasePackCoverage_i + \beta_3 OnhandInventory_i \\ &\quad + \beta_4 ItemSize_i + \beta_5 Margin_i + \beta_6 Variety_i + \beta_7 ForecastDispersion_i \\ &\quad + \beta_8 SeasonalityError_i + \beta_9 Sales_i + \beta_{10} Price_i + \beta_{11} Discount_i \\ &\quad + \beta_{12} ColdDay_i + \beta_{13} HotDay_i + \beta_{14} DayBeforeHoliday_i + \beta_{15} Holiday_i \\ &\quad + \beta_{16} RelativeProposalSize_i + \beta_{17} RecentStockout_i + \boldsymbol{BX_i} + \epsilon_i \\ DownwardDeviation_i &= \mathbb{I}[DownwardDeviation_i^* > 0] \end{aligned}$$

Table OS.1 shows the results of the endogenous switching regression model's second stage. (The first stage is substantively similar to Table 2 of the main manuscript.) The positive and significant coefficient of *Recent Stockout* for no deviation cases signals that recent stockouts indeed lead to new stockouts irrespective of downward deviations. To compare this effect and the effects of the other predictors under each condition (i.e., no deviation and downward deviation), Figure OS.1 shows the predicted probability changes of a future stockout. For continuous variables, we take their 10% and 90% quantiles to calculate the probability changes and confidence intervals. For binary variables, the figure shows the change when they move from 0 to 1. We observe that the likelihood of a stockout increases by 6.5 percentage points after a recent stockout even if no deviation is made. Yet, when a downward deviation is made after a recent stockout, the likelihood of a stockout increases by 22.2 percentage points. This substantive increase in the estimated effect further supports Hypothesis 2.

Table OS.1. Endogenous Switching Regression

	Dependent variable: Stockout	
	No Deviation	Downward Deviation
Constant	-1.897^{***} (0.115)	-1.855^{*} (0.880)
ASR Proposal Size	-0.041^{***} (0.006)	-0.052^{***} (0.013)
Deviation Size		0.096^{***} (0.016)
Relative Proposal Size	-0.169^{***} (0.028)	0.092 (0.071)
Recent Stockout	0.553^{***} (0.036)	0.802^{***} (0.084)
N	18,207	1779

*p<0.05; **p<0.01; ***p<0.001
Specification includes fixed effects to control for locations, product groups, proposal weekdays, weeks, and the interactions of locations and product groups.

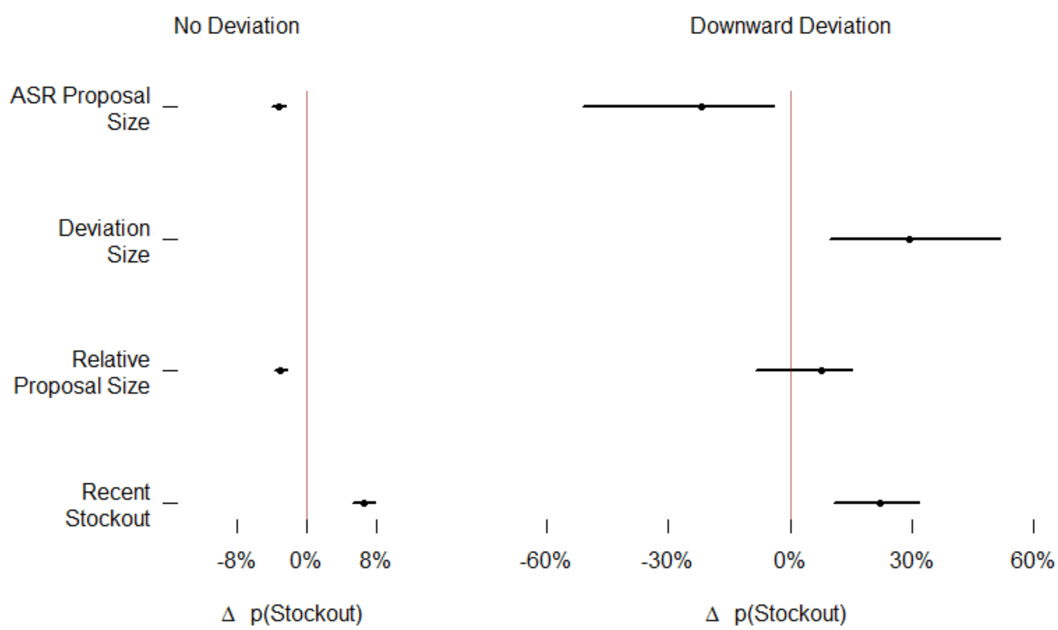

Figure 1: Endogenous Switching Regression

One half of the endogenous switching model is effectively equivalent to a probit model with sample selection. This specification is closely related to the Heckman selection framework used by Caro and Saez de Tejada Cuenca (2023), with the main distinction that our second-stage outcome is binary rather than continuous. Therefore, the consistency of the findings under the endogenous switching model also supports the robustness of our results to a sample-selection-type specification.

Next, following Elmaghraby et al. (2015), we consider the deviation decision making as a stepwise process. To do so, we model the process in three stages: first the decision to make a *Downward Deviation*, then the magnitude of the deviation: *Deviation Size*, and finally, the outcome: *Stockout.* To estimate this process, we utilize a seemingly unrelated regression (SUR) model (Zellner 1962). Apart from allowing us to test our hypotheses within the stepwise process of Elmaghraby et al. (2015), this model also helps us alleviate the concerns related to having fixed effects in a maximum likelihood framework with binary outcomes. Specifically, our main model, recursive bivariate probit model, uses the full information maximum likelihood estimator and it has been argued that fixed effects (such as product groups, weeks, weekdays, locations, and interactions of product groups and locations in our model) create a risk of making the maximum likelihood estimator inconsistent in nonlinear models (Greene 2004). The first stage of the model is the linear form of the first equation of our main model to predict *Downward Deviation.* The second and third stages are:

$$\begin{aligned}
DeviationSize_i &= \alpha_0 + \alpha_1 DownwardDeviation_i + \alpha_2 ASRProposalSize_i \\
&\quad + \alpha_3 RelativeProposalSize + \alpha_4 RecentStockout_i + \boldsymbol{BX_i} + \epsilon_i \\
Stockout_i &= \gamma_0 + \gamma_1 DeviationSize_i + \gamma_2 ASRProposalSize_i \\
&\quad + \gamma_3 RelativeProposalSize_i + \gamma_4 RecentStockout_i \\
&\quad + \gamma_5 DeviationSize_i * RelativeProposalSize_i \\
&\quad + \gamma_6 DeviationSize_i * RecentStockout_i + \boldsymbol{\Gamma X_i} + \epsilon_i
\end{aligned}$$

The first-stage results of the SUR model, reported in Table OS.2, show that *Recent Stockout* is positive and significant ($p<0.001$), supporting Hypothesis 1. The second- and third-stage results are provided in Table OS.3 Columns 1 and 2, respectively. Although *Recent Stockout* predicts *Downward Deviation*, we observe that it is not a significant predictor of *Deviation Size*; yet, the interaction term between *Recent Stockout* and *Deviation Size* predicts *Stockout* ($p<0.001$), supporting Hypothesis 2. Following Elmaghraby et al. (2015), we compare the goodness of fit and predictive power of this more detailed model to those of our main model. The Bayesian information criterion (BIC) is 17,756 and 90,787 for our main model and the SUR model, respectively. The higher BIC score of the SUR model indicates poorer fit of the stepwise modeling of the decision process. We examine the predictive power of our model and the SUR model by randomly dividing our data into training and testing sets, estimating the models on the training data, and using the coefficients to predict the outcomes in the testing data. Via this prediction, we obtain the area under the receiver operating characteristic curve, which is 0.79 and 0.78 for the recursive bivariate probit model and the SUR model, respectively, indicating that the stepwise modeling is also not better in predictive power.

Table OS.2. SUR: Linear Effect of Censorship Bias on Downward Deviations

	Coefficients
Constant	-0.049^{**} (0.018)
ASR Proposal Size	0.006^{***} (0.0003)
Case Pack Coverage	-0.0003 (0.0003)
On-hand Inventory	0.003^{***} (0.002)
Item Size	-0.0005^{***} (0.0001)
Margin	-0.002 (0.007)
Variety	-0.001^{***} (0.0003)
Forecast Dispersion	-0.0002 (0.001)
Seasonality Error	0.024 (0.017)
Sales	-0.003^{***} (0.001)
Price	0.008^{***} (0.002)
Discount	-0.056^{**} (0.018)
Cold Day	0.004 (0.007)
Hot Day	0.021^{***} (0.005)
Day Before Holiday	-0.049^{**} (0.015)
Holiday	0.020 (0.016)
Relative Proposal Size	0.053^{***} (0.001)
Recent Stockout	0.023^{***} (0.005)
Adjusted R^2	0.21
N	20,787

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Specification includes fixed effects to control for locations, product groups, proposal weekdays, weeks, and the interactions of locations and product groups.

The recursive bivariate probit model assumes bivariate normality, and under distributional misspecification, the estimates may be inconsistent (Wojtyś et al. 2016) or biased (Park et al. 2022). To alleviate these concerns, we run a copula approach that relaxes the joint normality assumption (Hasebe 2013). As there are several copulas available, we run our model with 22 different copulas. The likelihood-ratio-based test for model selection developed by Vuong (1989) did not favor any particular copula; thus here, we discuss the results produced by the Clayton copula. However, we note that all 22 copulas produce similar coefficient estimates, demonstrating the robustness of the results. The coefficient estimates of the selection equation with the Clayton copula support Hypothesis 1, as *Recent Stockout* is positive and significant ($p<0.001$). Table OS.3 Column 3 shows that relaxing the joint normality assumption of the error terms still leads to significant coefficient estimates for the interaction terms between *Downward Deviation* and *Relative Proposal Size* (anchoring bias), and *Downward Deviation* and *Recent Stockout* (censorship bias).

We then consider the dependent variable of the first stage: *Downward Deviation*. In the main analysis, this variable takes the value of 1 if the purchase order is less than the ASR proposal, meaning that an upward deviation or no deviation is coded as 0. Lumping those two cases together may affect the coefficient estimation. To alleviate such concerns, we run a multinomial logit model where the dependent variable takes the value of 1, 2, and 3 for downward deviation, no deviation, and upward deviation cases, respectively. The results indicate that, when there is a recent stockout, a downward deviation is more likely relative to no deviation ($p<0.001$), but an upward deviation is not ($p>0.1$). These results are in line with censorship bias, supporting Hypothesis 1.

Table OS.3. Alternative Estimation Strategies

	1. SUR 2nd Stage	2. SUR 3rd Stage	3. Copula
Constant	−0.826***	0.039**	−1.875***
	(0.106)	(0.013)	(0.131)
ASR Proposal Size	0.044***	−0.002***	−0.050***
	(0.002)	(0.0003)	(0.006)
Deviation Size		0.009***	0.092***
		(0.001)	(0.008)
Downward Deviation	7.702***		0.428***
	(0.056)		(0.089)
Relative Proposal Size	−0.055***	−0.009***	−0.162***
	(0.012)	(0.002)	(0.027)
Recent Stockout	0.046	0.091***	0.547***
	(0.038)	(0.005)	(0.085)
Downward Deviation * Relative Proposal Size			0.226***
			(0.025)
Downward Deviation * Recent Stockout			0.247***
			(0.058)
Deviation Size * Relative Proposal Size		0.002***	
		(0.0003)	
Deviation Size * Recent Stockout		0.022***	
		(0.002)	
N	20,787	20,787	20,787

*p<0.05; **p<0.01; ***p<0.001
Specification includes fixed effects to control for locations, product groups, proposal weekdays, weeks, and the interactions of locations and product groups. Standard errors are in parentheses. Columns (1) and (4) use robust standard errors multiway clustered at the location and product-group levels. Columns (2) and (3) report SUR standard errors (not clustered).

1.2. Alternative Operationalizations

As discussed in Section 4.2.2, we examine censorship bias under the condition where it can exist: *Recent Stockout*. Our main measure is based on the previous seven days. To test robustness to alternative definitions of "recent," we redefine *Recent Stockout* using windows of 1–6 and 8–14 days. When we use one or two days, *Recent Stockout* is not significant in the first stage, but the signs and significance of the estimated coefficients in the second stage remain the same. The first-stage outcome is unsurprising since experiencing a recent stockout a day or two days ago would make the stockout incident very fresh in the minds of decision-makers. Thus, deviating downward from the ASR proposal would unlikely happen. Another possible approach to operationalize censorship bias is to focus on stockout occurrences in the most recent order coverage period. When we run the analysis based on whether the inventory has gone to 0 in the most recently observed order coverage period that matches the same weekdays as the order coverage period of the ASR proposal at hand, we find that *Recent Stockout* is significant ($p<0.01$) and positive in the first equation, and the interaction between *Downward Deviation* and *Recent Stockout* is significant ($p<0.001$) and positive in the second equation, providing further support for both hypotheses.

We also test the robustness of the results to the operationalization of anchoring bias, our benchmark bias. In the main analysis, we used the mean of last seven days' purchase orders to calculate the tendency to anchor. To be sure that our results do not change depending on the time window, we also created this variable by using the mean purchase orders of the previous 14 days. We used *Recent Stockout* using a 14-day window to ensure temporal alignment across variables in this analysis. The use of these alternative operationalizations did not change the results related to anchoring and censorship biases. Although we have argued that past purchase orders are more salient than past demand to anchor on, we are aware that most of past research has used the latter to operationalize the anchoring bias. Hence, we use the mean of past seven days' sales as an alternative operationalization in otherwise the same recursive bivariate probit model, and observe that this operationalization does not change the results.

The next robustness check relates to how we operationalize our main dependent variable. In the main analysis, a binary variable, *Stockout*, is used to assess the performance implications of deviations susceptible to censorship bias. As an alternative dependent variable, we compute the number of days without any stock relative to the order coverage period of an SKU. This dependent variable captures the percent of the order coverage period without the SKU at hand. The results of a linear regression model with endogenous treatment in which the second stage is linear show support for both hypotheses: *Recent Stockout* is significant ($p<0.05$) and positive in the first stage, and the interaction between *Recent Stockout* and *Downward Deviation* is significant ($p<0.001$) and positive in the second stage. Results suggest that a downward deviation made after a recent stockout increases the percent of the order coverage period without the SKU at hand by 19 percentage points.

Next, we examine whether the impact of downward deviations triggered by censorship bias is different in terms of having a self-inflicted stockout earlier or later. To do so, we create two binary variables: *Stockout Early* and *Stockout Late*. *Stockout Late* takes the value of 1 if stockout occurred in the last day of the order coverage period, whereas *Stockout Early* takes the value of 1 if stockout occurred before the last day of the order coverage period. We run two recursive bivariate probit models with these dependent variables in the second stage. We find that the interaction between *Recent Stockout* and *Downward Deviation* is not statistically significant in predicting *Stockout Late*. In contrast, the interaction is positive and significant ($p<0.001$) in predicting *Stockout Early*, which is in line with our argument that censorhsip bias leads to self-inflicted stockouts.

Our next robustness check concerns how discounts enter the analysis. Following Khosrowabadi et al. (2022) in our main analysis, *Discount* takes the value of 1 if an SKU will be discounted during the order coverage period. However, one could argue that, if an SKU was discounted before the order, it is likely that the decision-maker would tend to order less of this product for the next period. To incorporate this idea, we add *Prior Discount* in the first equation. This variable takes the value of 1 in the first ordering decision after an SKU has been discounted. This variable is not significant in the first stage to predict downward deviations. *Recent Stockout* remains significant ($p<0.001$) and positive in the first equation, and its interaction with *Downward Deviation* remains significant ($p<0.001$) and positive in the second equation. Furthermore, if we remove observations with *Discount*=1 or *Prior Discount*=1, *Recent Stockout* remains positive and significant in the first stage, and its interaction with *Downward Deviation* remains positive and significant in the second stage.

1.3. Other Modifications to the Empirical Strategy

Apart from anchoring bias, the newsvendor literature has examined demand chasing, a behavior in which decision-makers anchor on the previous order quantity and adjust orders toward the last realized demand (Lau and Bearden 2012). The reason why such heuristic is considered biased is based on an experimental design in which each period's demand is considered independent from the previous period's demand. Such a premise is tricky for field research in retail operations, where forecasting models incorporating information from previous periods, namely time series models, show consistently superior performance (Gür Ali and Gürlek 2020). It is therefore unclear whether the demand-chasing heuristic should be considered a bias in this setting. Nevertheless, as a robustness check, we add a variable to our model to capture demand chasing to see if it influences the effect of censorship bias. We follow Kirshner and Moritz (2021) and operationalize demand chasing as *Sales Trend* (regression slope of daily sales over time). When we add this variable to our model, it has a negative coefficient in the first equation (Table OS.4 Column 1), meaning that the likelihood of a downward deviation increases when the sales trend is decreasing (negative), which is consistent with demand-chasing behavior. However, this behavior does not have negative operational consequences in terms of stockouts, as the interaction between *Downward Deviation* and *Sales Trend* is not significant in the second equation (Table OS.5 Column 1). We still observe that *Recent Stockout* is significant and positive in the first equation (Table OS.4 Column 1) and its interaction with *Downward Deviation* is significant and positive in the second equation (Table OS.5 Column 1), showing further support for our two hypotheses.

Table OS.4. Other Modifications to the Empirical Strategy: Predicting Downward Deviations

	1. Demand chasing	2. Holiday week removed	3. Second sample included	4. No winsorization	5. Conventional standard errors
Constant	−3.019***	−2.854***	−3.097***	−2.582***	−3.014***
	(0.238)	(0.194)	(0.270)	(0.254)	(0.198)
ASR Proposal Size	0.041***	0.041***	0.044***	0.015*	0.041***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.002)
Case Pack Coverage	−0.007	-0.005	-0.002	−0.012*	−0.007*
	(0.005)	(0.006)	(0.005)	(0.005)	(0.003)
On-hand Inventory	0.022	0.015	0.020	-0.001	0.023
	(0.015)	(0.014)	(0.016)	(0.006)	(0.021)
Item Size	−0.007**	−0.008***	−0.006***	−0.005*	−0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Margin	-0.026	0.006	0.031	0.078	-0.029
	(0.128)	(0.128)	(0.107)	(0.092)	(0.067)
Variety	−0.015***	−0.012**	−0.013***	−0.016**	−0.015***
	(0.004)	(0.004)	(0.003)	(0.005)	(0.003)
Forecast Dispersion	−0.012	-0.009	−0.032*	−0.016	−0.011
	(0.016)	(0.015)	(0.016)	(0.008)	(0.011)
Seasonality Error	0.273	0.014	0.578**	0.133	0.271
	(0.146)	(0.189)	(0.183)	(0.158)	(0.149)
Sales	−0.003	-0.009	−0.015*	-0.001	−0.008
	(0.007)	(0.009)	(0.006)	(0.001)	(0.006)
Price	0.092***	0.098***	0.074**	0.042*	0.093***
	(0.020)	(0.020)	(0.026)	(0.019)	(0.017)
Discount	−0.228	-0.241	−0.348*	−0.300	-0.231
	(0.131)	(0.157)	(0.154)	(0.158)	(0.175)
Cold Day	-0.079	0.052	-0.111	-0.074	−0.078
	(0.107)	(0.087)	(0.102)	(0.108)	(0.058)
Hot Day	0.236**	0.217*	0.136	0.230**	0.236***
	(0.082)	(0.091)	(0.084)	(0.082)	(0.043)
Day Before Holiday	−0.708		-0.262	−0.726	−0.703***
	(0.369)		(0.329)	(0.383)	(0.201)
Holiday	0.492		0.290	0.594	0.490*
	(0.413)		(0.279)	(0.411)	(0.200)
Relative Proposal Size	0.335***	0.332***	0.345***	0.376***	0.335***
	(0.024)	(0.022)	(0.015)	(0.023)	(0.011)
Recent Stockout	0.170**	0.163*	0.131*	0.166**	0.172***
	(0.063)	(0.066)	(0.056)	(0.057)	(0.040)
Sales Trend	−0.101*				
	(0.050)				
N	20,787	18,675	38,225	20,787	20,787

*p<0.05; **p<0.01; ***p<0.001. Specification includes fixed effects to control for locations, product groups, proposal weekdays, weeks, and the interactions of locations and product groups. Robust standard errors are multiway clustered on location and product group, except for Column 5.

Next, we exclude a week around the only holiday in our data collection period. Although the main analyses already control for *Holiday* and *Day Before Holiday*, this check removes observations around the holiday altogether. The results hold (Table OS.4, Column 2; Table OS.5, Column 2).

Our next robustness checks relate to how we utilize our data. As mentioned in Section 4.1, we have two datasets collected a year apart: one for hypothesis testing and one for testing policies to reduce self-inflicted stockouts. The former involves 12 weeks of ordering decisions from three locations, whereas the latter involves 12 weeks of ordering decisions from two additional locations. We run our main analysis by combining these two datasets to test that the results hold regardless of store selection. *Recent Stockout* is positive and significant in the estimation of the first equation (Table OS.4 Column 3), indicating that Hypothesis 1 holds in the combined data. Regarding the performance implications, we observe in Table OS.5 Column 3 that the interaction between *Downward Deviation* and *Recent Stockout* is positive and significant, supporting Hypothesis 2.

Table OS.5. Other Modifications to the Empirical Strategy: Predicting Stockouts

	1. Demand chasing	2. Holiday week removed	3. Second sample included	4. No winsorization	5. Conventional standard errors
Constant	-1.863^{***}	-1.909^{***}	-1.958^{***}	-1.880^{***}	-1.873^{***}
	(0.132)	(0.130)	(0.101)	(0.133)	(0.109)
ASR Proposal Size	-0.052^{***}	-0.051^{***}	-0.051^{***}	-0.051^{***}	-0.052^{***}
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Deviation Size	0.088^{***}	0.087^{***}	0.075^{***}	0.062^{***}	0.088^{***}
	(0.007)	(0.007)	(0.008)	(0.006)	(0.011)
Downward Deviation	0.859^{***}	0.739^{***}	0.974^{***}	1.038^{***}	0.817^{**}
	(0.145)	(0.160)	(0.261)	(0.146)	(0.271)
Relative Proposal Size	-0.181^{***}	-0.171^{***}	-0.179^{***}	-0.173^{***}	-0.171^{***}
	(0.028)	(0.029)	(0.022)	(0.029)	(0.020)
Recent Stockout	0.548^{***}	0.565^{***}	0.485^{***}	0.542^{***}	0.542^{***}
	(0.088)	(0.087)	(0.054)	(0.085)	(0.035)
Sales Trend	0.289^{**}				
	(0.111)				
Downward Deviation * Relative Proposal Size	0.200^{***}	0.207^{***}	0.201^{***}	0.192^{***}	0.194^{***}
	(0.020)	(0.024)	(0.045)	(0.023)	(0.037)
Downward Deviation * Recent Stockout	0.223^{**}	0.208^{***}	0.267^{***}	0.204^{***}	0.218^{*}
	(0.070)	(0.058)	(0.075)	(0.055)	(0.087)
Downward Deviation * Sales Trend	-0.042				
	(0.218)				
N	20,787	18,675	38,225	20,787	20,787

*p<0.05; **p<0.01; ***p<0.001. Specification includes fixed effects to control for locations, product groups, proposal weekdays, weeks, and the interactions of locations and product groups. Robust standard errors are multiway clustered on location and product group, except for Column 5.

In the main analyses, we winsorize the variables to limit the influence of extreme values. The results remain robust without winsorization (Table OS.4 Column 4 and Table OS.5 Column 4). The same is true if we replace the multiway-clustered robust standard errors used in the main analyses with conventional standard errors (Table OS.4 Column 5 and Table OS.5 Column 5).

2. Potential Confounders and Alternative Explanations

We have argued that censorship bias is the reason for the phenomenon we observe, that is, ASR system users order less than algorithmic suggestions after a recent stockout, which leads to an increased probability of future stockouts. Yet, there could be alternative explanations, other than censorship bias, causing this behavioral pattern. One such explanation is anchoring bias: After a stockout, the ASR system considers lost sales and produces a higher proposal. Users are then likely to deviate downward from the proposal since they mentally anchor on past purchase orders. To distinguish this effect from that of censorship bias, we control for anchoring bias in the main analysis as well as in all of our robustness checks. Yet, anchoring bias is only one alternative explanation. In this section, we investigate what else could explain the phenomenon we observe.

2.1. Users' Reaction to ASR System's Lost-Sales Estimation

Leading to Hypothesis 2, we argue that ASR systems are not susceptible to censorship bias thanks to the progress made in the methods of forecasting and ordering under censored demand (Shi et al. 2016). The principle underlying these methods is to forecast lost sales by extrapolating the demand rate that preceded the stockout. When the ASR system does that, the resulting proposals may seem excessive to the users, who are susceptible to censorship bias. Yet, what if the ASR system overreacts to stockouts when it performs the extrapolation to uncensor the demand? In this case, downward deviations after a recent stockout could be explained by users not agreeing with how the system estimates lost sales, instead of suffering from censorship bias. In other words, could the ASR system users be correcting for a possible over-extrapolation by the ASR system? If this were the case and decision-makers successfully corrected the ASR system's proposals, we should not see the second-stage results on the increased likelihood of future stockouts, yet this could still be an alternative explanation for the observed behavior of ordering less than the algorithmic suggestions after stockouts. As such, it could be considered less of a bias than rational, albeit unsuccessful, behavior. To test this alternative explanation, we must overcome the limitation that the proprietary algorithms of the ASR system are unknown to us. To do so, we create sales forecasts using triple exponential smoothing with uncensoring of demand. Specifically, we first estimate lost sales for stockout days based on SKU-location-specific weekly trend factors and weekday seasonality indices. We then run the triple exponential smoothing on the uncensored demand estimates for each product in each location separately. On average, our forecasts exceed ASR proposals by 0.55 units. The median difference is zero, and dispersion is limited (the interquartile range: -1.98 to 1.06), though the tails are asymmetric, with larger positive deviations (29.08 at the 99th percentile vs. -20.00 at the 1st percentile). This distribution suggests that our forecasts largely align with ASR proposals in typical cases, but occasionally predict substantially higher values. To assess the forecast accuracy of this procedure, we use the symmetric mean absolute percentage error (sMAPE):

$$= (1/n) \sum_{i=1}^{n} \frac{|Actual_i - Forecast_i|}{(Actual_i + Forecast_i)/2}$$

This is a standard measure of forecast accuracy in the retail industry (Brau et al. 2023). As a relative measure, it is commensurate across SKUs with different sales volumes, but it corrects for the basic MAPE's inherently asymmetric treatment of over- and under-forecasting and its over-penalization of errors when sales quantities are low (Makridakis and Hibon 2000), which they are for some of the SKUs in our data. The average sMAPE for our forecasts is 8.75%, indicating high accuracy (cf. Brau et al. 2023). We use the produced forecasts to estimate the ASR system's possible over-estimation of lost sales and create a new variable: the difference between ASR proposals and forecasts. If the ASR system is overestimating demand after a stockout and decision-makers are correcting for it, then this difference variable should have a positive and significant coefficient in the first equation. In Table OS.6 Column 1, we see that this difference is indeed positive and significant ($p<0.001$) in predicting downward deviations, but even after controlling for this effect, *Recent Stockout* stays significant and positive in the first equation, showing support for Hypothesis 1. Similarly, Table OS.7 Column 1 shows support for Hypothesis 2.

2.2. Safety Stock Ignorance

Since decision-makers generally aim to reduce the ex-post inventory error in inventory replenishment tasks (Schweitzer and Cachon 2000), ASR system users could be inclined to place their orders always to match exactly the demand that they expect. After a recent stockout, such approach to ordering would ignore that also the safety stock must be replenished. If ASR system users were indeed biased towards ordering expected demand while ignoring the safety stock requirements, they would be more likely to deviate downward after a stockout. Such tendency would be an alternative to censorship bias to explain the curious behavior of ordering less after a stockout. We note, however, that this effect should be captured by anchoring bias variables in our models. Both anchoring bias operationalizations (i.e., the anchor point being the mean purchase orders and mean sales) capture the tendency to "pull-to-center" while ignoring safety stock requirements after a stockout. It is therefore unlikely that safety stock ignorance, instead of censorship bias, explains the observed phenomenon. Yet, to further ensure that this possible alternative explanation does not confound censorship bias, we add *Expected Lead Time* as an additional independent variable in the first equation of our model to predict *Downward Deviation*. The behavioral model presented by Tong and Feiler (2017) predicts an ordering behavior that leads to having less safety stock for products with short replenishment lead times. This prediction means that when the expected lead time is shorter, ASR system users are more inclined to pursue lower safety stock levels, increasing the likelihood of deviating downward. Hence, by adding *Expected Lead Time*, we are able to check whether users' safety stock considerations cancel the impact of censorship bias. The results in Table OS.6 Column 2 show that *Expected Lead Time* is not significant in predicting *Downward Deviation* and adding this additional variable does not change the support for our hypotheses.

2.3. Previous Upward Deviations

The discretion of ASR system users is not limited to downward deviations; they may also deviate upward from the proposed orders. Thus, it could be that users have a higher tendency to deviate downward after they have previously made an upward deviation. In the study of van Donselaar et al. (2010), this behavior is called order advancement: ordering more than the algorithmic suggestions on less busy days, and then ordering less on busier days. In our study, it could be that users deviate upward immediately after a stockout but then revise subsequent orders by deviating downward to avoid accumulating too much inventory. This could be an alternative explanation to the observed phenomenon of ordering less than the proposed quantity after a recent stockout. The impact of a previous upward deviation should be carried in *ASR Proposal Size*, since the inventory level caused by the previous deviation decision is considered by the ASR system when it produces the next proposal. Yet, it may be that the ASR system user does not realize that the previous deviation has been already taken into account in the proposal, and thus, we test this alternative explanation by adding a binary variable, *Previous Up*, in the first equation of our model. This variable takes the value of 1 if the previous decision was an upward deviation. The estimated coefficient of *Previous Up* is negative and insignificant ($p>0.1$) in Table OS.6 Column 3. Thus, the downward deviations in our sample cannot be explained as ASR system users offsetting their earlier upward deviations. Both hypotheses remain supported (Table OS.6, Column 3; Table OS.7, Column 3).

2.4. Workload Leveling

One reason for the deviations from algorithmic suggestions in the study of van Donselaar et al. (2010) is that managers consider labor capacity requirements for the handling of arriving shipments when they decide on order quantities, whereas a typical ASR system does not consider them when producing proposals. Hence, ASR system users may be more inclined to deviate downward if lots of orders are about to arrive on the same day as the current proposal's projected arrival date. To test this, we include in our model delivery weekday fixed effects, which capture the average busyness of each weekday. All of the estimated coefficients are significant in Table OS.6 Column

Table OS.6. Alternative Explanations: Predicting Downward Deviations

	1. Reaction to ASR uncensoring	2.Safety stock ignorance	3.Previous upward deviation	4.Workload leveling	5. User Attention (I)	6. User Attention (II)	7. User Attention (III)
Constant	−2.995*** (0.264)	−3.291*** (0.141)	−2.994*** (0.230)	−2.775*** (0.242)	−3.014*** (0.238)	−3.015*** (0.238)	−2.973*** (0.234)
ASR Proposal Size	0.039*** (0.006)	0.041*** (0.011)	0.041*** (0.006)	0.041*** (0.006)	0.041*** (0.006)	0.041*** (0.006)	0.044*** (0.006)
Case Pack Coverage	-0.007 (0.005)	-0.006 (0.007)	-0.007 (0.005)	-0.007 (0.005)	-0.007 (0.005)	-0.007 (0.005)	-0.008 (0.005)
On-hand Inventory	0.011 (0.012)	0.017 (0.025)	0.021 (0.015)	0.023 (0.016)	0.023 (0.015)	0.023 (0.015)	0.022 (0.015)
Item Size	−0.007*** (0.002)	−0.006** (0.002)	−0.007** (0.002)	−0.007** (0.002)	−0.007** (0.002)	−0.007** (0.002)	−0.007** (0.002)
Margin	-0.061 (0.119)	-0.017 (0.157)	-0.029 (0.126)	-0.029 (0.127)	-0.027 (0.128)	-0.026 (0.128)	-0.036 (0.125)
Variety	−0.013** (0.005)	−0.014*** (0.003)	−0.014*** (0.003)	−0.015*** (0.004)	−0.015*** (0.004)	−0.015*** (0.004)	−0.015*** (0.004)
Forecast Dispersion	-0.006 (0.016)	-0.011 (0.032)	-0.011 (0.015)	-0.011 (0.016)	-0.012 (0.016)	-0.012 (0.016)	-0.012 (0.016)
Seasonality Error	0.236 (0.136)	0.265 (0.194)	0.265 (0.150)	0.271 (0.148)	0.271 (0.148)	0.270 (0.147)	0.250 (0.153)
Sales	-0.007 (0.009)	-0.008 (0.014)	-0.007 (0.009)	-0.008 (0.009)	-0.008 (0.009)	-0.008 (0.009)	-0.001 (0.011)
Price	0.085*** (0.024)	0.089*** (0.024)	0.092*** (0.021)	0.093*** (0.021)	0.093*** (0.020)	0.092*** (0.021)	0.090*** (0.021)
Discount	−0.314* (0.138)	-0.243 (0.201)	−0.228 (0.134)	−0.231 (0.132)	−0.232 (0.130)	−0.233 (0.131)	−0.213 (0.125)
Cold Day	-0.070 (0.103)	-0.073 (0.154)	-0.078 (0.107)	-0.078 (0.106)	-0.077 (0.106)	-0.077 (0.106)	-0.080 (0.106)
Hot Day	0.239** (0.090)	0.240*** (0.035)	0.235** (0.082)	0.236** (0.082)	0.236** (0.082)	0.235** (0.082)	0.235** (0.081)
Day Before Holiday	−0.800* (0.357)	-0.700 (0.470)	−0.695 (0.363)	−0.703 (0.369)	−0.700 (0.368)	−0.701 (0.368)	−0.700 (0.370)
Holiday	0.434 (0.657)	0.489 (0.528)	0.478 (0.402)	0.490 (0.412)	0.489 (0.412)	0.491 (0.411)	0.478 (0.414)
Relative Proposal Size	0.332*** (0.028)	0.332*** (0.028)	0.333*** (0.025)	0.335*** (0.024)	0.334*** (0.025)	0.334*** (0.024)	0.326*** (0.025)
Recent Stockout	0.165** (0.055)	0.220*** (0.061)	0.171** (0.063)	0.172** (0.063)	0.173** (0.063)	0.174** (0.064)	0.159* (0.069)
Difference b/w ASR Proposal and Sales Forecast	0.002*** (0.001)						
Expected Lead Time		0.123 (0.072)					
Previous Upward Deviation			-0.200 (0.154)				
Delivery Day Fixed Effects (Monday)				−0.477*** (0.082)			
Delivery Day Fixed Effects (Wednesday)				−0.239* (0.096)			
Delivery Day Fixed Effects (Thursday)				−0.126** (0.045)			
Delivery Day Fixed Effects (Friday)				−0.167* (0.070)			
Spoiled Quantity					0.005 (0.016)		
Spoiled						0.048 (0.071)	
Sales Quantity							-0.021 (0.012)
N	19,224	20,787	20,787	20,787	20,787	20,787	20,787

*p<0.05; **p<0.01; ***p<0.001. Specifications include fixed effects to control for locations, product groups, proposal weekdays, weeks, and the interactions of locations and product groups. In Column 4, instead of proposal weekdays fixed effects, we have delivery day fixed effects. Robust standard errors are multiway clustered on location and product group.

4: β_{Monday}=-0.48 (p<0.001), $\beta_{Wednesday}$=-0.24 (p<0.05), $\beta_{Thursday}$=-0.13 (p<0.01), and β_{Friday}=-0.17 (p<0.05), indicating that the likelihood of a downward deviation is indeed affected by the expected delivery day. This signals that ASR system users in our setting may be incorporating handling capacity considerations in their replenishment decisions. Yet, even after controlling for this effect, our hypotheses are still supported (Table OS.6 Column 4 and Table OS.7 Column 4).

2.5. Users' Attention and ASR Proposals' Salience

Our results might also be affected by a systematic factor that would draw ASR system users' attention to specific ASR proposals. Users might scrutinize more closely the proposals of any SKUs where they have just had to scrap spoiled items. To check this idea, we run additional models where such events are controlled for with two alternative operationalizations: *Spoiled Quantity*, the spoiled inventory quantity of the SKU on the decision day, and *Spoiled*, a binary variable that takes the value of 1 if the SKU had spoiled inventory on decision day. Turns out that neither variable is significant (Table OS.6 Columns 5 and 6), but our hypotheses are still supported, as shown in Table OS.6 Columns 5 and 6, and Table OS.7 Columns 5 and 6.

Table OS.7. Alternative Explanations: Predicting Stockouts

	1. Reaction to ASR uncensoring	2.Safety stock ignorance	3.Previous upward deviation	4.Workload leveling	5. User Attention (I)	6. User Attention (II)	7. User Attention (III)
Constant	−1.881***	−1.868***	−1.872***	−1.802***	−1.873***	−1.872***	−1.873***
	(0.127)	(0.147)	(0.132)	(0.180)	(0.133)	(0.132)	(0.132)
ASR Proposal Size	−0.052***	−0.053***	−0.052***	−0.052***	−0.052***	−0.052***	−0.052***
	(0.006)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Deviation Size	0.085***	0.085***	0.088***	0.088***	0.088***	0.088***	0.088***
	(0.007)	(0.012)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
Downward Deviation	0.798***	1.096***	0.837***	0.817***	0.825***	0.839***	0.811***
	(0.213)	(0.281)	(0.158)	(0.148)	(0.151)	(0.155)	(0.145)
Relative Proposal Size	−0.181***	−0.174***	−0.171***	−0.171***	−0.171***	−0.171***	−0.170***
	(0.027)	(0.039)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Recent Stockout	0.561***	0.540***	0.542***	0.542***	0.542***	0.542***	0.542***
	(0.096)	(0.087)	(0.086)	(0.086)	(0.086)	(0.086)	(0.085)
Difference b/w ASR Proposal and Sales Forecast	0.001 (0.001)						
Downward Deviation × Relative Proposal Size	0.208*** (0.029)	0.168*** (0.023)	0.192*** (0.021)	0.194*** (0.020)	0.193*** (0.020)	0.192*** (0.021)	0.194*** (0.018)
Downward Deviation × Recent Stockout	0.268*** (0.066)	0.186*** (0.053)	0.215*** (0.062)	0.218*** (0.062)	0.217*** (0.063)	0.215*** (0.062)	0.218*** (0.063)
N	19,224	20,787	20,787	20,787	20,787	20,787	20,787

*p<0.05; **p<0.01; ***p<0.001. Specifications include fixed effects to control for locations, product groups, proposal weekdays, weeks, and the interactions of locations and product groups. In Column 4, instead of proposal weekdays fixed effects, we have delivery day fixed effects. Robust standard errors are multiway clustered on location and product group.

Users' attention may also be drawn to high-selling SKUs. Although the main model controls for relative *Sales*, we add absolute *Sales Quantity* to account for this possibility. This measure is marginally significant (p<0.1) (Table OS.6 Column 7), but its inclusion does not change the effects of *Recent Stockout* or the interaction effect of *Recent Stockout* with *Downward Deviation* in either stage (Table OS.6 Column 7 and Table OS.7 Column 7).

Finally, one more factor that could affect ASR system users' attention is their screen layout; for example, a user could sequence ASR proposals based on profit margins. While this would be captured with our control variable *Margin*, there are countless other ways to organize the screen layout. Since the layout, and thus the sequence of ASR proposals, is entirely up to the user and not recorded in the data, we are not able to control directly for its effect. However, since users' screen layout preferences may be habitual and not change so often over time, it is likely that the attention-influencing effect of the screen layout is captured by the user fixed effects (the fixed effects of product groups and locations and their interactions) that we have incorporated in all our models.

References

Amemiya T (1984) Tobit models: A survey. *Journal of Econometrics* 24(1-2):3-61.

Brau R, Aloysius J, Siemsen E (2023) Demand planning for the digital supply chain: How to integrate human judgment and predictive analytics. *Journal of Operations Management* 69(6):965-982.

Caro F, Saez de Tejada Cuenca A (2023) Believing in analytics: Managers' adherence to price recommendations from a DSS. *Manufacturing & Service Operations Management*, 25(2):524-542.

Elmaghraby W, Jank W, Zhang S, Karaesmen IZ (2015) Sales force behavior, pricing information, and pricing decisions. *Manufacturing & Service Operations Management* 17(4):495-510.

Freeman M, Robinson S, Scholtes S (2021) Gatekeeping, fast and slow: An empirical study of referral errors in the emergency department. *Management Science* 67(7):4209-4232.

Greene W (2004) The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal* 7(1):98-119.

Gür Ali Ö, Gürlek R (2020) Automatic Interpretable Retail forecasting with promotional scenarios. *International Journal of Forecasting* 36(4):1389-1406.

Hasebe T (2013) Copula-based maximum-likelihood estimation of sample-selection models. *The Stata Journal* 13(3):547-573.

Khosrowabadi N, Hoberg K, Imdahl C (2022) Evaluating Human Behaviour in Response to AI Recommendations for Judgemental Forecasting. *European Journal of Operational Research* 303(3): 1151-1167.

Kirshner SN, Moritz BB (2021) Measuring demand chasing behavior. *Decision Sciences* 52(6):1264-1281.

Lau N, Bearden JN (2013) Newsvendor demand chasing revisited. *Management Science*, 59(5), 1245-1249.

Liu A, Mazumdar T, Li B (2015) Counterfactual decomposition of movie star effects with star selection. *Management Science* 61(7):1704-1721.

Lokshin M, Sajaia Z (2011) Impact of interventions on discrete outcomes: Maximum likelihood estimation of the binary choice models with binary endogenous regressors. *The Stata Journal* 11(3):368-385.

Makridakis S, Hibon M (2000) The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* 16(4):451-476.

Marra G, Radice R, Zimmer D (2022) A Unifying Switching Regime Regression Framework with Applications in Health Economics.

Park W-Y, Nickerson J, Bigelow L (2022) Stuck in the Middle (of Time): Strategic Repositioning and Survival in Response to an Innovation Shock in a Growing Market.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science* 46(3):404-420.

Shi C, Chen W, Duenyas I (2016) Nonparametric data-driven algorithms for multiproduct inventory systems with censored demand. *Operations Research* 64(2):362-370.

Tong J, Feiler D (2017) A behavioral model of forecasting: Naive statistics on mental samples. *Management Science* 63(11):3609-3627.

van Donselaar KH, Gaur V, van Woensel T, Broekmeulen RA, Fransoo JC (2010) Ordering behavior in retail stores and implications for automated replenishment. *Management Science* 56(5):766-784.

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*:307-333.

Wojtyś M, Marra G, Radice R (2016) Copula regression spline sample selection models: the r package semiparsamplesel. *Journal of Statistical Software* 71:1-66.

Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57(298):348-368.