Forthcoming in *Manufacturing & Service Operations Management*
manuscript MSOM-18-505

E-Companion to
Believing in Analytics: Managers' Adherence to Price Recommendations from a DSS

Felipe Caro
University of California, Los Angeles, Anderson School of Management, fcaro@anderson.ucla.edu

Anna Sáez de Tejada Cuenca
IESE Business School ASaezdetejada@iese.edu

1. Robustness Checks for Section 5

1.1. Alternative Estimation Method

Our main Heckman estimation in the paper is based on the two-step procedure detailed in Heckman (1979), in which the selection step (probit) is estimated first using maximum likelihood (MLE), and the second step is estimated using ordinary least squares (OLS) but including the inverse Mills ratio as an additional regressor. An alternative method is to estimate all coefficients (as well as the underlying correlation between the two seps' error terms) at the same time, using MLE, and compute the standard errors using the observed information matrix (OIM) (Cameron and Trivedi 2005, Greene 2012, Wooldridge 2010).

Table 1 shows the APEs of our the Heckman regression estimated using MLE. As we can see, the results do not change qualitatively from those in Table 3 of the paper: $StockoutWeek$ has a positive, significant effect on both the probability of deviating as well as on the magnitude of deviations, for all managers. In countries in which Zara owns the stores, $NumCategsDT$ has a positive, significant effect on both $ProbDeviation$ and $AbsDeviation$. In countries in which stores are franchises, $NumCategsDT$ has a negative and weakly significant effect on both $ProbDeviation$ and $AbsDeviation$ — in contrast with our main results in the paper, in which, for franchises, $NumCategsDT$ only shows a weakly significant effect on $ProbDeviation$, but insignificant on $AbsDeviation$. Moreover, the sizes of all the APEs are very similar to those in Table 3 of the paper. In addition, the correlation between the two steps (the Rho) is also strongly significant, underscoring again the need for using a selection correction model.

	Heckman APEs					
	All countries		Own stores		Franchises	
	(1)		(2)		(3)	
AbsDeviation						
StockoutWeek	0.00836^{***}	(0.000857)	0.00975^{***}	(0.000995)	0.00788^{***}	(0.00158)
NumCategsDT	0.00698^{**}	(0.00256)	0.0219^{***}	(0.00334)	-0.00896^{*}	(0.00424)
Week dummies	Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes	
ProbDeviation						
StockoutWeek	0.0115^{***}	(0.00105)	0.0119^{***}	(0.00120)	0.0127^{***}	(0.00210)
NumCategsDT	0.0128^{***}	(0.00348)	0.0304^{***}	(0.00439)	-0.0162^{**}	(0.00616)
Week dummies	Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes	
AthRho	2.845^{***}	(0.0432)	2.827^{***}	(0.0488)	2.972^{***}	(0.107)
LnSigma	-1.298^{***}	(0.0171)	-1.294^{***}	(0.0217)	-1.328^{***}	(0.0279)
N	22180		16534		5646	
N (deviated)	9775		6584		3191	

Table 1 APE estimates of the Heckman regression estimated via maximum likelihood.

Note. Top half: APEs (coefficients) of the deviation part, whose dependent variable is *AbsDeviation*; bottom half: APEs of the selection part, whose dependent variable is *ProbDeviation*. (1) includes all countries; (2), only countries in which Zara owns the stores; (3), only franchises. OIM standard errors in parentheses. Not reported: week, group, and country dummies, in both parts. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.2. Alternative Models for the Selection Part

The selection part of the Heckman model is a probit regression, as its outcome variable is binary (*Deviated*), and includes country dummies to account for heterogeneity between managers. However, the usage of fixed effects in a probit model can lead to biased estimates (Greene 2002). We repeat the probit estimation of the selection part, but use different ways of including manager heterogeneity. We show the results in Table 2. We first run a pooled probit regression, with no country effect, shown in columns (1) to (3). We second use the Chamberlain-Mundlak device (Chamberlain 1980, Wooldridge 2019), which avoids the bias from using fixed effects, while at the same time it does not require the strong assumptions needed to use random effects. This is done by including the averages by country of the independent variables ($StockoutWeek$ and $NumCategsDT$) as additional regressors. We show the estimated APEs in columns (4) to (6). Third, we report again the probit regression with country dummies, as it is in the paper, but with regular robust standard errors instead of the Heckman-specific ones proposed in Heckman (1979); we report the APEs of the probit with dummies in columns (7) to (9). Finally, columns (10) to (12) show the APEs of the probit regression with country random effects. The resulting estimates are extremely similar in sign and statistical significance, and similar in size, to those reported in the paper's main regression.

For additional robustness, we fit an alternative binary response model: the logit. The estimates are shown in Table 3. Again, we run different versions of the regression. First, without accounting for manager heterogeneity (pooled logit), shown in columns (1) to (3). Second, using the Chamberlain-Mundlak device, shown in columns (4) to (6). Third, using country dummies, shown in columns (7) to (9). Finally, using country random effects, shown in columns (10) to (12). As we can see, the APEs of the logit regression are extremely similar to those of the Heckman regression's selection part and also those of the probit regression.

	Probit APEs											
	All countries (1)	Own stores (2)	Franchises (3)	All countries (4)	Own stores (5)	Franchises (6)	All countries (7)	Own stores (8)	Franchises (9)	All countries (10)	Own stores (11)	Franchises (12)
ProbDeviation												
StockoutWeek	0.00730***	0.0103***	0.00750***	0.0109***	0.0114***	0.00863***	0.00884***	0.00911***	0.00978***	0.00909***	0.00887***	0.00960**
	(0.000933)	(0.00110)	(0.00188)	(0.00104)	(0.00120)	(0.00206)	(0.00101)	(0.00117)	(0.00199)	(0.00150)	(0.00169)	(0.00324)
NumCategsDT	0.0137***	0.0286***	-0.00452	0.0195***	0.0255***	-0.00736	0.0127***	0.0264***	-0.0150*	0.0137	0.0261***	-0.0133
	(0.00354)	(0.00448)	(0.00616)	(0.00387)	(0.00486)	(0.00688)	(0.00381)	(0.00479)	(0.00675)	(0.00729)	(0.00521)	(0.0143)
Week dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country C-M	No	No	No	Yes	Yes	Yes	No	No	No	No	No	No
Country dummies	No	No	No	No	No	No	Yes	Yes	Yes	No	No	No
Country RE	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes
LnSigma2u										-2.315***	-3.478***	-2.041***
										(0.289)	(0.376)	(0.458)
N	22180	16534	5646	22180	16534	5646	22180	16534	5646	22180	16534	5646

Table 2 APE estimates of the probit regression with different specifications.

Note. The dependent variable is *ProbDeviation*. (1)-(3) include no country effects; (4)-(6) include the Chamberlain-Mundlak device to account for country manager heterogeneity; (7)-(9) include country dummies; (10)-(12) include country random effects. (1), (4), (7) and (10) include all countries; (2), (5), (8) and (11) only countries in which Zara owns the stores; (3), (6), (9) and (12) only franchises. Robust standard errors in parentheses. Not reported: week and group dummies. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Logit APEs											
	All countries (1)	Own stores (2)	Franchises (3)	All countries (4)	Own stores (5)	Franchises (6)	All countries (7)	Own stores (8)	Franchises (9)	All countries (10)	Own stores (11)	Franchises (12)
ProbDeviation												
StockoutWeek	0.00729***	0.0102***	0.00779***	0.0109***	0.0113***	0.00902***	0.00892***	0.00904***	0.0102***	0.00918***	0.00884***	0.0100**
	(0.000934)	(0.00110)	(0.00195)	(0.00105)	(0.00120)	(0.00217)	(0.00102)	(0.00116)	(0.00213)	(0.00154)	(0.00170)	(0.00350)
NumCategsDT	0.0135***	0.0282***	-0.00452	0.0194***	0.0251***	-0.00722	0.0126**	0.0260***	-0.0150*	0.0136	0.0258***	-0.0133
	(0.00354)	(0.00448)	(0.00618)	(0.00388)	(0.00486)	(0.00690)	(0.00383)	(0.00479)	(0.00680)	(0.00737)	(0.00525)	(0.0145)
Week dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country C-M	No	No	No	Yes	Yes	Yes	No	No	No	No	No	No
Country dummies	No	No	No	No	No	No	Yes	Yes	Yes	No	No	No
Country RE	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes
LnSigma2u										-1.336***	-2.512***	-1.046*
										(0.297)	(0.376)	(0.467)
N	22180	16534	5646	22180	16534	5646	22180	16534	5646	22180	16534	5646

Table 3 APE estimates of the logit regression with different specifications.

Note. The dependent variable is *ProbDeviation*. (1)-(3) include no country effects; (4)-(6) include the Chamberlain-Mundlak device to account for country manager heterogeneity; (7)-(9) include country dummies; (10)-(12) include country random effects. (1), (4), (7) and (10) include all countries; (2), (5), (8) and (11) only countries in which Zara owns the stores; (3), (6), (9) and (12) only franchises. Robust standard errors in parentheses. Not reported: week and group dummies. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.3. Alternative Models for the Deviation Part

The deviation part of the Heckman regression is estimated using OLS with a correction term, the inverse Mills ratio (or *Lambda*). As mentioned in Section 5.1 of the paper, this could lead to identification problems due to collinearity because *Lambda* is a quasi-linear function around its center. According to Puhani (2000), if identification is an issue due to collinearity, much more robust estimation can be achieved using simple OLS but using only the observations with positive selection. Although in Section 5.2 of the paper we have shown that, in our Heckman regressions, the argument of the inverse Mills ratio has enough spread to mitigate the concern of collinearity causing identification issues, we additionally run two sets of linear OLS estimation with *AbsDeviation* as dependent variable, which we report in Table 4. In the first set — columns (1) to (3) — we use the full set of observations; in the second set — columns (4) to (6) — we only use the subset recommended in Puhani (2000), i.e., observations for which the managers decided to deviate from the DSS's recommended price. All the linear regressions include country, group, and week dummy variables. Although in both cases (full sample and subsample) the coefficient estimates are qualitatively extremely similar to those reported in the paper's Table 3 (and in this document's Table 1), the estimates are closer in size to those in the paper when we run the linear regression using the subsample. The fact that our results using the Heckman and the linear regressions are so similar provides further evidence against the existence of identification problems due to collinearity.

	Linear coefficients											
	All countries		Own stores		Franchises		All countries		Own stores		Franchises	
	(1)		(2)		(3)		(4)		(5)		(6)	
AbsDeviation												
StockoutWeek	0.00651^{***}	(0.000595)	0.00715^{***}	(0.000650)	0.00683^{***}	(0.00126)	0.00804^{***}	(0.000921)	0.00991^{***}	(0.00105)	0.00785^{***}	(0.00170)
NumCategsDT	0.00660^{***}	(0.00164)	0.0160^{***}	(0.00193)	-0.00630^{*}	(0.00318)	0.00514^{*}	(0.00218)	0.0172^{***}	(0.00281)	-0.00172	(0.00355)
Week dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes		Yes		Yes		Yes	
N	22180		16534		5646		9775		6584		3191	

Table 4 Coefficients of the linear regression estimated using different subsamples.

Note. The dependent variable is the deviation magnitude *AbsDeviation*. (1)-(3) include all observations; (4)-(6) include only observations in which the managers deviated from the recommended price, i.e., the dependent variable is the conditional $CondAbsDeviation = (AbsDeviation|DeviatedDSS = 1)$. (1) and (4) include all countries; (2) and (5), only countries in which Zara owns the stores; (3) and (6), only franchises. Robust standard errors in parentheses. Not reported: week, group and country dummies. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.4. Alternative Measure of Inventory and Speed of Sales

Our paper's first hypothesis posits that managers' tendency to deviate and magnitude of deviations were higher when there was more inventory and it was selling more slowly. To test that, in the paper we used a variable based on one of the measures of inventory and speed of sales that managers at Zara used — the *rotation*. We now repeat the Heckman regression, but test Hypothesis 1 using a variable based on Zara's other measure of inventory and speed of sales — the *success*. As defined in the company, the success equals the proportion of the initial inventory of a category that has already been sold by a given week w. This measure, in the broader retailing world, is often called the sellthrough rate, and is computed as $Sellthrough_{igwc} = \frac{\sum s<wSales_{igsc}}{Inventory_{ig0c}}$.

We compute this variable for each one of our observations and, in case the company made inventory adjustments between countries on a given week, we take that into account in the denominator. However, such adjustments were typically very small compared to the inventory levels in absolute terms, and in consequence the *Sellthrough* is almost always non-decreasing over time for all the categories of a given country and group. For this reason, it is extremely likely to be collinear with the week dummies and, more generally, to be reflecting only a linear time trend and not a meaningful measure. For this reason, we detrend it using the procedure described in Section 4.2 of the paper, and include in our regression the detrended $SellthroughDT_{igw} = Sellthrough_{igw} - \widehat{Sellthrough}_w$. Note that this variable behaves in the opposite direction than the one we use in the paper, $StockoutWeek$: when a product is selling out quickly, $StockoutWeek$ is smaller (there are fewer weeks left for it to be completely sold out), and *Sellthrough* is higher (a larger proportion of it has been already sold). For this reason, we expect the APE estimates of *SellthroughDT* to have the opposite sign to those of $StockoutWeek$ and, indeed, that is what Table 5 shows. Taking this into account, our results are qualitatively identical to the main ones in the paper: when items are selling out quickly, managers were less likely to deviate from the DSS's price recommendations and, even if they did deviate, they did so by smaller amounts.

	Heckman APEs					
	All countries		Own stores		Franchises	
	(1)		(2)		(3)	
AbsDeviation						
SellthroughDT	-0.238***	(0.0267)	-0.446***	(0.0417)	-0.142***	(0.0380)
NumCategsDT	0.00799**	(0.00307)	0.0321***	(0.00492)	-0.0102*	(0.00487)
Week dummies	Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes	
ProbDeviation						
SellthroughDT	-0.431***	(0.0266)	-0.454***	(0.0340)	-0.450***	(0.0436)
NumCategsDT	0.0136***	(0.00377)	0.0308***	(0.00478)	-0.0174**	(0.00665)
Week dummies	Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes	
Lambda	0.318***	(0.0195)	0.380***	(0.0247)	0.296***	(0.0255)
N	22180		16534		5646	
N (deviated)	9775		6584		3191	

Table 5 **APE estimates of the Heckman regression estimated via the two-step method, using** *SellthtroughDT* **instead of** $StockoutWeek$ **to measure inventory and speed of sales.**

Note. Top half: APEs (coefficients) of the deviation part, whose dependent variable is *AbsDeviation*; bottom half: APEs of the selection part, whose dependent variable is *ProbDeviation*. (1) includes all countries; (2), only countries in which Zara owns the stores; (3), only franchises. Robust standard errors in parentheses. Not reported: week, group, and country dummies, in both parts. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.5. Country Type as a Moderating Variable

In our paper, we formulate competing hypothesis for H2. In particular, we posit that managers might deviate *more* when there are more price categories for which they have to choose new prices (to simplify the number of pricing decisions to make the following weeks); alternatively, we posit that they might deviate *less* when there are more price categories, because they understand that the DSS is a tool meant to simplify their work in times of such information overload. Empirically, we find that managers in countries where Zara owns the stores exhibit pricing behavior compatible with the first sub-hypothesis, and managers in countries where stores are franchises set prices in a way that is compatible with the second sub-hypothesis.

Another way to test whether country type is associated with different behavior (deviating more, or deviating less, in the presence of many price categories) is to use country type as a moderating variable of the number of categories. Table 6 shows our main Heckman estimates (as shown in the paper's main results, Table 3), plus a Heckman regression containing an interaction term between country type ($Franchises$) and number of price categories ($NumCategoriesDT$). Note that, for the interaction model, the APEs cannot be computed.

The moderation analysis supports our main results. Each additional category is associated with higher probability of deviating (by 6.71 percentage points), and with deviations larger in magnitude (by 0.149). However, in franchises, each additional category is related to lower probability of deviating (by 9.1 percentage points) and smaller deviations (by 0.0231), although franchises have overall higher probability of deviating and deviation sizes. The results of this moderation analysis is consistent with our main results, estimated using own-store countries and franchises separately.

1.6. Sensitivity to the Heckman Correction

Our Heckman regression model is meant to account for for selection bias. In the first stage, managers select themselves into deviating from the DSS's price recommendation; in the second stage, if they have chosen to deviate, they decide by how much. However, is it necessary to correct for selection bias at all? A way to test it is a sensitivity analysis as described in Smith (2014). In it, the parameter ρ (the correlation coefficient between the first-stage and second-stage standard errors) is not estimated from the data, but instead it is fixed to certain (incremental) values and, for each one, the model's coefficients and rest of the parameters are calculated. Similar sensitivity analyses have been used in the Economics literature (Altonji et al. 2005, Ichino et al. 2008, Imbens 2003, Rosenbaum and Rubin 1983, Rosenbaum 1987, 2002).

In our paper's main estimates using the two-step method, $\rho = 1.0$ for own-store countries and $\rho = 0.982$ for franchises. Using MLE (Section 1.1 of this Appendix), these numbers are $\rho = 0.993$

	Heckman							
	All countries		All countries		Own stores		Franchises	
	(1)		(2)		(3)		(4)	
AbsDeviation								
StockoutWeek	0.0104***	(0.000758)	0.0103***	(0.000753)	0.0125***	(0.00101)	0.00898***	(0.00121)
NumCategoriesDT	0.00746**	(0.00285)	0.0169***	(0.00339)	0.0262***	(0.00418)	-0.00868	(0.00451)
Franchise			0.149***	(0.0278)				
Franchise×NumCategoriesDT			-0.0231***	(0.00456)				
Week dummies	Yes		Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes		Yes	
ProbDeviation (coefficients)								
StockoutWeek	0.0239***	(0.00268)	0.0237***	(0.00268)	0.0246***	(0.00317)	0.0268***	(0.00508)
NumCategoriesDT	0.0343***	(0.0103)	0.0671***	(0.0119)	0.0715***	(0.0130)	-0.0410*	(0.0184)
Franchise			0.548***	(0.112)				
Franchise×NumCategoriesDT			-0.0910***	(0.0169)				
Week dummies	Yes		Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes		Yes	
ProbDeviation (APEs)								
StockoutWeek	0.00884***	(0.000985)			0.00911***	(0.00117)	0.00978***	(0.00184)
NumCategoriesDT	0.0127***	(0.00379)			0.0264***	(0.00479)	-0.0150*	(0.00670)
Lambda	0.291***	(0.0193)	0.287***	(0.0193)	0.327***	(0.0223)	0.268***	(0.0258)
Rho	0.990		0.985		1.000		0.982	
Sigma	0.294		0.292		0.327		0.273	
N	22180		22180		16534		5646	
N (deviated)	9775		9775		6584		3191	

Table 6 Coefficient and average partial effects estimates of the Heckman regression.

Note. Top panel: coefficients of the deviation step with dependent variable *AbsDeviation*; second panel: coefficients of the selection step, with dependent variable *ProbDeviation*; third panel: APEs of the selection step (except for interaction terms); fourth panel: additional parameter estimates of the Heckman regression: Lambda, Rho and Sigma. (1) includes all countries (equivalent to (1) in the paper's Table 3); (2) includes all countries, and interactions between country type and *NumCategoriesDT*; (3), only own-store countries (equivalent to (2) in the paper's Table 3); (4), franchises only (equivalent to (3) in the paper's Table 3). Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

and $\rho = 0.995$, respectively. In our sensitivity analysis, we fix ρ to values from -0.9 to +0.9, in increments of 0.2. We then estimate the model's coefficients and parameters using MLE (because Stata does not allow constrained estimation using the two-step method).

The results of this test are reported in Table 7 (own-store countries) and 8 (franchises). As we can see, our APE estimates are, indeed, sensitive to changes to ρ, and these changes are more pronounced for the regressions using franchises only. Column (11) of both tables reports the APEs of the unconstrained estimation (the same as in this Appendix's Table 1, columns 2 and 3). As the constrained ρ approaches the unconstrained estimated one (i.e., as ρ increases), the log-likelihood function increases very quickly. In particular, the log-likelihood is much higher for the unconstrained ρ than it would be for $\rho = 0$, which would correspond to a model that does not account for selection bias. This proves that a model that corrects for managers' selection into adhering or deviating is necessary.

	(1) Rho=-0.9	(2) Rho=-0.7	(3) Rho=-0.5	(4) Rho=-0.3	(5) Rho=-0.1	(6) Rho=0.1	(7) Rho=0.3	(8) Rho=0.5	(9) Rho=0.7	(10) Rho=0.9	(11) Unconstrained Rho
AbsDeviation											
StockoutWeek	0.0103^{***}	0.00931^{***}	0.00931^{***}	0.00951^{***}	0.00977^{***}	0.0101^{***}	0.0103^{***}	0.0104^{***}	0.0103^{***}	0.0100^{***}	0.00975^{***}
	(0.00141)	(0.00114)	(0.00107)	(0.00105)	(0.00104)	(0.00105)	(0.00105)	(0.00104)	(0.00101)	(0.000975)	(0.000995)
NumCategoriesDT	0.0101^{*}	0.0130^{***}	0.0145^{***}	0.0157^{***}	0.0167^{***}	0.0177^{***}	0.0187^{***}	0.0196^{***}	0.0206^{***}	0.0215^{***}	0.0219^{***}
	(0.00417)	(0.00324)	(0.00297)	(0.00285)	(0.00280)	(0.00280)	(0.00282)	(0.00285)	(0.00288)	(0.00299)	(0.00334)
ProbDeviation											
StockoutWeek	0.0127^{***}	0.0108^{***}	0.00990^{***}	0.00940^{***}	0.00914^{***}	0.00916^{***}	0.00961^{***}	0.0103^{***}	0.0112^{***}	0.0119^{***}	0.0119^{***}
	(0.00129)	(0.00122)	(0.00119)	(0.00117)	(0.00117)	(0.00117)	(0.00118)	(0.00120)	(0.00122)	(0.00123)	(0.00120)
NumCategoriesDT	0.0267^{***}	0.0267^{***}	0.0265^{***}	0.0264^{***}	0.0264^{***}	0.0265^{***}	0.0269^{***}	0.0275^{***}	0.0282^{***}	0.0293^{***}	0.0304^{***}
	(0.00457)	(0.00469)	(0.00475)	(0.00478)	(0.00479)	(0.00479)	(0.00479)	(0.00479)	(0.00476)	(0.00464)	(0.00439)
AthRho	-1.472	-0.867	-0.549	-0.310	-0.100	0.100	0.310	0.549	0.867	1.472	2.827^{***}
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(0.0488)
LnSigma	-1.128^{***}	-1.469^{***}	-1.612^{***}	-1.684^{***}	-1.716^{***}	-1.719^{***}	-1.699^{***}	-1.660^{***}	-1.598^{***}	-1.482^{***}	-1.294^{***}
	(0.0287)	(0.0269)	(0.0263)	(0.0261)	(0.0261)	(0.0261)	(0.0259)	(0.0252)	(0.0240)	(0.0216)	(0.0217)
LogLikelihood	-9733.039	-9021.099	-8827.982	-8747.486	-8707.885	-8678.841	-8626.371	-8501.436	-8228.598	-7584.869	-6926.655
N	16534	16534	16534	16534	16534	16534	16534	16534	16534	16534	16534

Table 7 Coefficients estimates of the Heckman regression using data from own-store countries only, as a sensitivity analysis to selection bias.

Note. Top half: APEs of the deviation part; bottom half: APEs of the selection part. (1) to (10) are estimated by fixing ρ to values from -0.9 to +0.9 in increments of 0.2; (11) is equivalent to Table 1's column (2). Estimated via MLE. OIM standard errors in parentheses. Not reported: week, group, and country dummies, in both parts. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

	(1) Rho=-0.9	(2) Rho=-0.7	(3) Rho=-0.5	(4) Rho=-0.3	(5) Rho=-0.1	(6) Rho=0.1	(7) Rho=0.3	(8) Rho=0.5	(9) Rho=0.7	(10) Rho=0.9	(11) Unconstrained Rho
AbsDeviation											
StockoutWeek	0.0100^{***}	0.00803^{***}	0.00767^{***}	0.00766^{***}	0.00777^{***}	0.00793^{***}	0.00805^{***}	0.00809^{***}	0.00803^{***}	0.00785^{***}	0.00788^{***}
	(0.00239)	(0.00190)	(0.00176)	(0.00171)	(0.00169)	(0.00169)	(0.00169)	(0.00167)	(0.00163)	(0.00158)	(0.00158)
NumCategoriesDT	0.00744	0.00291	0.00106	-0.000197	-0.00123	-0.00221	-0.00324	-0.00442	-0.00589	-0.00806^{*}	-0.00896^{*}
	(0.00504)	(0.00404)	(0.00372)	(0.00359)	(0.00353)	(0.00353)	(0.00356)	(0.00362)	(0.00372)	(0.00390)	(0.00424)
ProbDeviation											
StockoutWeek	0.0183^{***}	0.0136^{***}	0.0116^{***}	0.0104^{***}	0.00985^{***}	0.00987^{***}	0.0105^{***}	0.0113^{***}	0.0123^{***}	0.0131^{***}	0.0127^{***}
	(0.00277)	(0.00232)	(0.00211)	(0.00202)	(0.00199)	(0.00201)	(0.00205)	(0.00210)	(0.00215)	(0.00216)	(0.00210)
NumCategoriesDT	-0.0119	-0.0141^{*}	-0.0147^{*}	-0.0149^{*}	-0.0150^{*}	-0.0149^{*}	-0.0148^{*}	-0.0145^{*}	-0.0140^{*}	-0.0135^{*}	-0.0162^{**}
	(0.00668)	(0.00668)	(0.00671)	(0.00674)	(0.00675)	(0.00675)	(0.00675)	(0.00674)	(0.00669)	(0.00653)	(0.00616)
AthRho	-1.472	-0.867	-0.549	-0.310	-0.100	0.100	0.310	0.549	0.867	1.472	2.972^{***}
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(0.107)
LnSigma	-1.159^{***}	-1.454^{***}	-1.578^{***}	-1.641^{***}	-1.669^{***}	-1.671^{***}	-1.655^{***}	-1.622^{***}	-1.568^{***}	-1.471^{***}	-1.328^{***}
	(0.0374)	(0.0343)	(0.0332)	(0.0328)	(0.0327)	(0.0327)	(0.0325)	(0.0317)	(0.0302)	(0.0277)	(0.0279)
LogLikelihood	-3340.731	-2969.86	-2863.962	-2818.832	-2796.977	-2782.628	-2760.937	-2713.848	-2616.103	-2399.768	-2176.03
N	5646	5646	5646	5646	5646	5646	5646	5646	5646	5646	5646

Table 8 Coefficient estimates of the Heckman regression using data from franchise countries only, as a sensitivity analysis to selection bias.

Note. Top half: APEs of the deviation part; bottom half: APEs of the selection part. (1) to (10) are estimated by fixing ρ to values from -0.9 to +0.9 in increments of 0.2; (11) is equivalent to Table 1's column (3). Estimated via MLE. OIM standard errors in parentheses. Not reported: week, group, and country dummies, in both parts. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

2. Robustness Checks for Section 6

2.1. Effect on Marking Down When the DSS Recommended Staying Put

As a complement to the paper's Section 6, we also want to measure the effect of the interventions on managers' likelihood to mark products down when the DSS recommended keeping their price unchanged, as this was a very common way in which deviations occurred. We compute the probability of each manager marking a product down conditional on the DSS recommending keeping its price unchanged as

$$CMarkdown_{itg} = \frac{\sum_{w\in\mathcal{W}(t)}\sum_{c\in\mathcal{C}(g,w)}\mathbb{I}\left(Price_{itgwc} < Price_{itg,w-1,c}\right)\times\mathbb{I}\left(PriceDSS_{itgwc} = Price_{itg,w-1,c}\right)}{\sum_{w\in\mathcal{W}(t)}\sum_{c\in\mathcal{C}(g,w)}\mathbb{I}\left(PriceDSS_{itgwc} = Price_{itg,w-1,c}\right)},$$

and repeat the paper's DiD analyses with this new dependent variable. Note that, for some countries and groups, the DSS always recommended marking down, so $CMarkdown_{itg}$ is not defined, and there are fewer observations of this variable than there are of $Adherence_{itg}$.

As shown in Figure 1, after Intervention 1 franchise managers marked products down when the DSS did not recommend it less often, but the drop is much larger for Intervention 2.

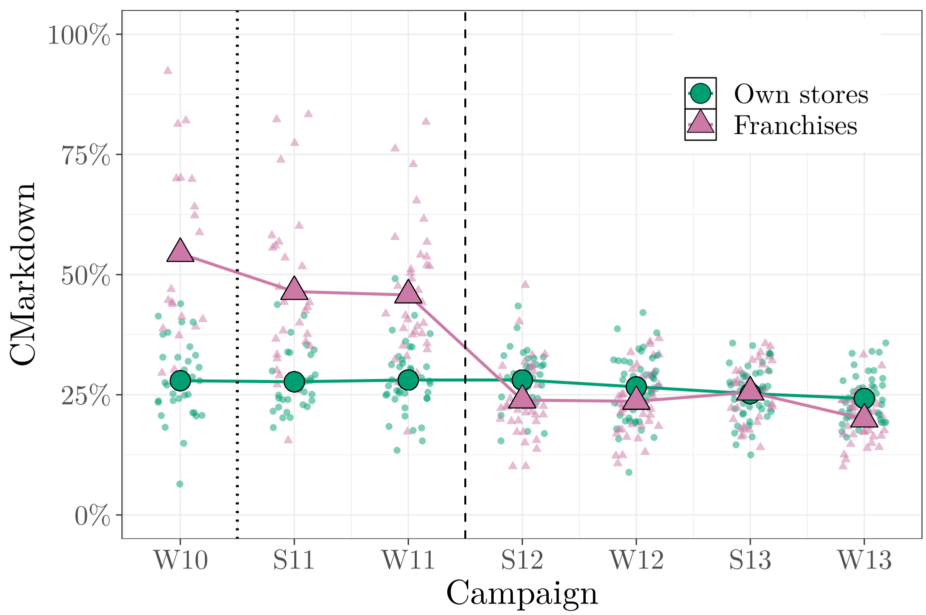

Figure 1 Scatter plot of managers' probability of marking a product down when the DSS recommended keeping its price unchanged by campaign, for own-store countries (green, circle) and franchises (pink, triangle).

Note. Each small point corresponds to the value of the variable $CMarkdown$ of a country manager in a given campaign (horizontally jittered within a campaign for clarity). The large points united by lines correspond to the average $\overline{CMarkdown_t}$ of own-store country managers (green) and franchises (pink) in each campaign. The dotted line marks when Intervention 1 took place (S2011), and the dashed line marks Intervention 2 (S2012).

In Table 9 we can see that Intervention 1's coefficient is only statistically significant ($p < 0.05$) in model (1), but not in the more nuanced regressions (2) and (4). Intervention 2, on the contrary, has a large, statistically significant effect in all specifications, ranging from 20.9 to 21.5 percentage points. This effect is 27.6 percentage points for managers who had the lowest adherence before this intervention took place (first quartile); 16.1 and 11.6 for the two middle quartiles; and 3.9 percentage points for the top quartile.

	DiD coefficients CMarkdown							
	(1)		(2)		(3)		(4)	
Int1×Franchise	-0.0858*	(0.0332)	-0.0422	(0.0334)			-0.0429	(0.0314)
Int2×Franchise	-0.214***	(0.0266)	-0.209***	(0.0276)	-0.215***	(0.0354)		
Int1×Franchise×AdhPreInt1Q1					-0.119**	(0.0400)		
Int1×Franchise×AdhPreInt1Q2					0.00392	(0.0341)		
Int1×Franchise×AdhPreInt1Q3					0.0849**	(0.0255)		
Int1×Franchise×AdhPreInt1Q4					0.161***	(0.0260)		
Int2×Franchise×AdhPreInt2Q1							-0.276***	(0.0433)
Int2×Franchise×AdhPreInt2Q2							-0.161***	(0.0369)
Int2×Franchise×AdhPreInt2Q3							-0.116***	(0.0293)
Int2×Franchise×AdhPreInt2Q4							-0.0390***	(0.00802)
Int1	0.0143	(0.0114)						
Int2	0.00851	(0.00983)						
Franchise	0.234***	(0.0413)						
ExperienceWithDSS	-0.00867***	(0.00224)	0.190**	(0.0600)	0.0374	(0.120)	0.122	(0.0959)
LogNumStores	-0.0264***	(0.00337)	-0.0110	(0.0255)	0.0531	(0.0514)	0.0178	(0.0409)
Constant	0.362***	(0.0148)						
Group dummies	No		Yes		Yes		Yes	
Campaign dummies	No		Yes		Yes		Yes	
Country dummies	No		Yes		Yes		Yes	
N	10275		10275		7766		9835	

Table 9 Difference-in-differences coefficient estimates, where the dependent variable is $CMarkdown$.

Note. This table is equivalent Table 4 of the paper, but with the new dependent variable $CMarkdown$. Robust standard errors in parentheses. Not reported: group, campaign, and country dummy variables, in (2), (3) and (4).

2.2. Difference-in-Differences Regression of Each Intervention Separately

Our analysis of the effect of the interventions is a difference-in-differences where all available data (countries and campaigns) is included, and both interventions are studied simultaneously. This could be a problem because the DSS got implemented gradually, so new countries appear in the post-intervention periods whose managers were not using the DSS before the interventions took place. Potentially, our estimates could be only a reflection of the arrival of those new country managers, who may have been more eager to adhere to the DSS's recommendations. To rule out this possibility, we repeat our estimation with using only data from the countries whose managers were already using the DSS before each intervention took place. If we ran this robustness check studying the two interventions at the same time we would be dismissing all the countries that were added after Intervention 1 but before Intervention 2. Therefore, we study the two interventions separately. For each one of them, we take only the countries whose managers were using the DSS before the intervention, and only the campaign immediately before and immediately afer the intervention. The results are in Tables 10 (Intervention 1) and 11 (Intervention 2), and show coefficient estimates almost identical to those in our main analysis in the paper (for $Adherence$) and in this e-Companion's Section 2.1 (for $CMarkdown$).

2.3. Difference-in-Differences Regression for a No-Intervention Period

The two interventions performed by Zara took place right before a spring-summer campaign. Therefore, our DiD analyses of the effect of the interventions could potentially be capturing only

	DiD coefficients											
	Adherence (1)		Adherence (2)		Adherence (3)		CMarkdown (4)		CMarkdown (5)		CMarkdown (6)	
Int1×Franchise	0.00696	(0.0265)	0.00947	(0.0260)			-0.0660	(0.0363)	-0.0332	(0.0354)		
Int1×Franchise×AdhPreInt1Q1					0.0284	(0.0349)					-0.0841	(0.0432)
Int1×Franchise×AdhPreInt1Q2					-0.0207	(0.0540)					0.00837	(0.0478)
Int1×Franchise×AdhPreInt1Q3					0.0146	(0.00767)					0.206***	(0.0185)
Int1×Franchise×AdhPreInt1Q4					-0.0317	(0.0294)					0.0420***	(0.0111)
Int1	-0.0186*	(0.00802)					0.0141	(0.0139)				
Franchise	-0.128***	(0.0317)					0.222***	(0.0434)				
ExperienceWithDSS	0.0290***	(0.00689)	0.0384	(0.209)	0.0664	(0.212)	-0.00955	(0.00887)	-1.052	(0.621)	-0.848	(0.621)
LogNumStores	0.00164	(0.00800)	0.107	(0.0893)	0.0952	(0.0901)	-0.0341***	(0.00875)	0.536*	(0.263)	0.450	(0.263)
Constant	0.559***	(0.0258)					0.384***	(0.0259)				
Group dummies	No		Yes		Yes		No		Yes		Yes	
Campaign dummies	No		Yes		Yes		No		Yes		Yes	
Country dummies	No		Yes		Yes		No		Yes		Yes	
N	2216		2216		2216		2172		2172		2172	

Table 10 DiD estimates for Intervention 1.

Note. This table is equivalent to regressions (1), (2), and (3) in Table 4 of the paper, but here only include data from the campaigns immediately before and after Intervention 1 (W2010 and S2011), and only countries in which the DSS had already been implemented before Intervention 1 took place. In (1), (2), and (3) the dependent variable is *Adherence*; in (4), (5), and (6) it is *CMarkdown*. Robust standard errors in parentheses. Not reported: group, campaign, and country dummy variables, in (2), (3), (5), and (6).

	DiD coefficients											
	Adherence (1)		Adherence (2)		Adherence (3)		CMarkdown (4)		CMarkdown (5)		CMarkdown (6)	
Int2×Franchise	0.0910***	(0.0265)	0.0919**	(0.0270)			-0.225***	(0.0299)	-0.226***	(0.0304)		
Int2×Franchise×AdhPreInt2Q1					0.154***	(0.0391)					-0.278***	(0.0457)
Int2×Franchise×AdhPreInt2Q2					0.0234	(0.0277)					-0.189***	(0.0325)
Int2×Franchise×AdhPreInt2Q3					0.116*	(0.0549)					-0.188*	(0.0836)
Int2×Franchise×AdhPreInt2Q4					-0.126***	(0.0125)					0.0505***	(0.0140)
Int2	0.00466	(0.0104)					0.00854	(0.0126)				
Franchise	-0.113***	(0.0237)					0.140***	(0.0248)				
ExperienceWithDSS	0.0231***	(0.00553)	0.0572	(0.225)	0.169	(0.192)	-0.00958	(0.00618)	0.287	(0.245)	0.187	(0.223)
LogNumStores	0.000263	(0.00508)	0.0761	(0.190)	-0.0183	(0.162)	-0.0291***	(0.00494)	-0.177	(0.206)	-0.0935	(0.188)
Constant	0.515***	(0.0278)					0.392***	(0.0228)				
Group dummies	No		Yes		Yes		No		Yes		Yes	
Campaign dummies	No		Yes		Yes		No		Yes		Yes	
Country dummies	No		Yes		Yes		No		Yes		Yes	
N	2999		2999		2999		2972		2972		2972	

Table 11 DiD estimates for Intervention 2.

Note. This table is equivalent to regressions (1), (2), and (4) in Table 4 of the paper, but here only include data from the campaigns immediately before and after Intervention 2 (W2011 and S2012), and only countries in which the DSS had already been implemented before Intervention 2 took place. In (1), (2), and (3) the dependent variable is *Adherence*; in (4), (5), and (6) it is *CMarkdown*. Robust standard errors in parentheses. Not reported: group, campaign, and country dummy variables, in (2), (3), (5), and (6).

a seasonality trend in adherence behavior. In other words, it could happen that managers consistently adhere more in spring-summer than in fall-winter, for reasons that are unrelated to the interventions, such as reacting to different purchasing behavior from consumers in different times of the year, or others. To rule out this possibility, we run the DiD analysis including a no-intervention, i.e., we create a binary variable $Placebo_t$ that takes value 1 starting in S2013. If the results in the paper can be attributed to seasonality differences in adherence, then we will find that the no-intervention variable is also statistically significant. Table 12 confirms that this is not the case, and no significant changes in *Adherence* or *CMarkdown* occurred after the no-intervention.

	DiD coefficients							
	Adherence (1)		Adherence (2)		CMarkdown (3)		CMarkdown (4)	
Int1 ×Franchise	0.0162	(0.0269)	0.0112	(0.0216)	-0.0856*	(0.0332)	-0.0425	(0.0333)
Int2 ×Franchise	0.101***	(0.0189)	0.103***	(0.0193)	-0.222***	(0.0287)	-0.217***	(0.0293)
Placebo ×Franchise	-0.0147	(0.0115)	-0.0138	(0.0112)	0.0149	(0.00972)	0.0176	(0.00934)
Int1	-0.0311**	(0.0106)			0.0172	(0.0122)		
Int2	-0.0225*	(0.00950)			0.0132	(0.0109)		
Placebo	-0.0278**	(0.00840)			0.00223	(0.00825)		
Franchise	-0.133***	(0.0292)			0.234***	(0.0412)		
ExperienceWithDSS	0.0208***	(0.00275)	0.266***	(0.0491)	-0.0106**	(0.00363)	0.189**	(0.0609)
LogNumStores	0.00121	(0.00450)	0.0186	(0.0210)	-0.0263***	(0.00336)	-0.0107	(0.0259)
Constant	0.572***	(0.0161)			0.364***	(0.0151)		
Group dummies	No		Yes		No		Yes	
Campaign dummies	No		Yes		No		Yes	
Country dummies	No		Yes		No		Yes	
N	10431		10431		10275		10275	

Table 12 DiD estimates for a no-intervention period (S2013).

Note. The regressions in this table are equivalent to regressions (1) and (2) in Table 4 of the paper but including the no-intervention indicator, denoted by *Placebo*. In (1) and (2) the dependent variable is *Adherence*; in (3) and (4) it is *CMarkdown*. Robust standard errors in parentheses. Not reported: group, campaign, and country dummy variables, in (2) and (4).

2.4. Synthetic Control

We have estimated the effect of the two interventions on franchise country managers (as compared to own-store countries). However, the estimates we have obtained with this method cannot be claimed as causal, due to two related reasons: first, the assignment into treatment or control groups was not random; second, the two groups are substantially different from each other, as explained in Section 3.1 on the paper, which means that own store countries may not be well-suited to serve as control group for our DiD analysis.

A possible alternative is to create a synthetic control group for franchises, and use the DiD method to estimate the effect of the interventions on franchises compared to their synthetic control. Synthetic control is a method, proposed in Abadie and Gardeazabal (2003) and popularized in Abadie et al. (2010) that is well suited for situations in which treated units do not have a comparable control group. In this method, a "synthetic control" for every treated unit is generated based on the non-treated units. Each treatment unit is expressed as a weighted average of non-treated units, where the matching is done based on pre-treatment outcomes plus some covariates. The method estimates the best set of weights for each one of the matching variables and for each one of the non-treated units, such as the pre-treatment match with the focal treated unit is the best possible (smallest quadratic error). These weights used on the non-treated units are used to compute a post-treatment outcome of each synthetic control unit, which is the counterfactual to the treatment, i.e., what the outcome of the treated units would have been absent any treatment (Cunningham 2021, Huntington-Klein 2021).

The synthetic control method requires that every unit (focal treated unit, i.e., franchise, and non-treated units, i.e., own-store countries) has exactly one observation for every time period. To do that, we aggregate our data at the country-campaign level (it), as opposed to the previous intervention analyses, which were performed at the country-campaign-group level (itg). We remove the countries that are missing observations (i.e., where the DSS was rolled out later than the W2010 campaign). Apart from pre-Intervention outcome ($Adherence_{it}$), we match our treated and non-treated units based on two covariates from our data ($ExperienceWithDSS$ and $LogNumStores$), as well as three country-level macroeconomic variables (GDP per capita, Human Development Index, and Population Size) to account for the possibility that different types of populations with different consumption patterns influence who manages each country and how. We were able to obtain macroeconomic data for 52 countries in total, 21 franchises and 31 own-store countries. Our final dataset contains 42 units (21 franchises and their 21 synthetic controls), over 7 campaigns each.

We generated two sets of synthetic controls: one for Intervention 1, using pre–Intervention 1 adherence and covariates (one campaign of data), and one for Intervention 2, with three (pre–Intervention 2) campaigns of data. We then run a DiD analysis for each one separately.

Figure 2 shows the average adherence of franchises (pink, solid line) and of their synthetic controls (green, dashed line). Table 13 shows the coefficient estimates of the DiD analysis, comparing franchises' adherence to that of their synthetic controls. Column (1) of the table is equivalent to column (1) of our main DiD analyses, i.e., contains an Intervention 1 dummy $Int1$ instead of campaign dummies, and does not contain country manager idiosyncratic effects. Column (2) does contain campaign dummies (as opposed to Intervention 1 dummy) and country dummies. As we can see in the table, when we do not control for country/campaign, Intervention 1 seems to have a 6.31–percentage point effect on franchises' adherence compared to that of their synthetic controls. However, this estimate decreases to 1.63, and is not statistically significant, when we include country and campaign dummy variables. This is consistent with our main DiD results: Intervention 1 did not have a robust effect on franchises' adherence.

Figure 3 shows the average adherence of franchises and their synthetic controls generated using pre–Intervention 2 data; the corresponding DiD coefficient estimates are in Table 14. Without country/campaign dummies, Intervention 2's estimated effect on franchises was 8.02 percentage points. When we control for country and campaign, this effect decreases to 3.91 percentage points, but is still statistically significant. We conclude that Intervention 2 did, indeed, have a positive effect on franchises' adherence.

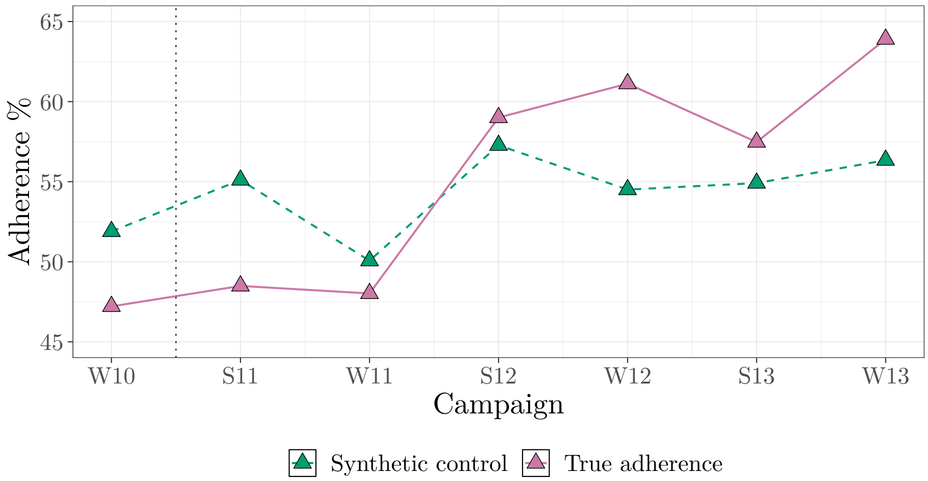

Figure 2 Average adherence of franchises (pink, solid line) and of their synthetic controls (green, dashed line).

Note. Each point corresponds to the average $\overline{Adherence_t}$ across all countries, computed from country-aggregate data $Adherence_{it}$. Each franchise's synthetic control was generated based on pre–Intervention 1 data on adherence and covariates. The dotted line marks when Intervention 1 took place (S2011).

	DiD coefficients			
	Adherence			
	(1)		(2)	
Int1×Treated	0.0631*	(0.0249)	0.0163	(0.0128)
Int1	-0.0351***	(0.00673)		
Treated	-0.0468*	(0.0188)		
ExperienceWithDSS	0.0180***	(0.00236)	0.553***	(0.0466)
LogNumStores	0.00189	(0.0276)	-0.0674	(0.0720)
Constant	0.500***	(0.0288)		
Group dummies	No		No	
Campaign dummies	No		Yes	
Country dummies	No		Yes	
N	294		294	

Table 13 Difference-in-differences coefficient estimates of the effect of Intervention 1, where the dependent variable is $Adherence_{it}$, the treated units are franchises, and the control units are their synthetic controls.

Note. Estimation done using country-aggregate data (i.e., one observation for each country and campaign). Each franchise's synthetic control was generated based on pre–Intervention 1 data on adherence and covariates. Robust standard errors in parentheses. Not reported: campaign, and country dummy variables, in (2).

The synthetic control method's ability to generate a good control group depends on there existing similar non-treated units, in terms of the pre-treatment outcomes and covariates, to the treated ones. If this is the case, then the pre-treatment outcomes (in our case, pre-intervention adherence for each intervention) will look extremely similar between treated units and their synthetic controls — the distance between the two sets of pre-treatment outcomes is, precisely, what the method is minimizing. However, as we can see in Figures 2 and 3, this is not the case here: in each figure, before the corresponding intervention, franchises' average adherence is much lower than

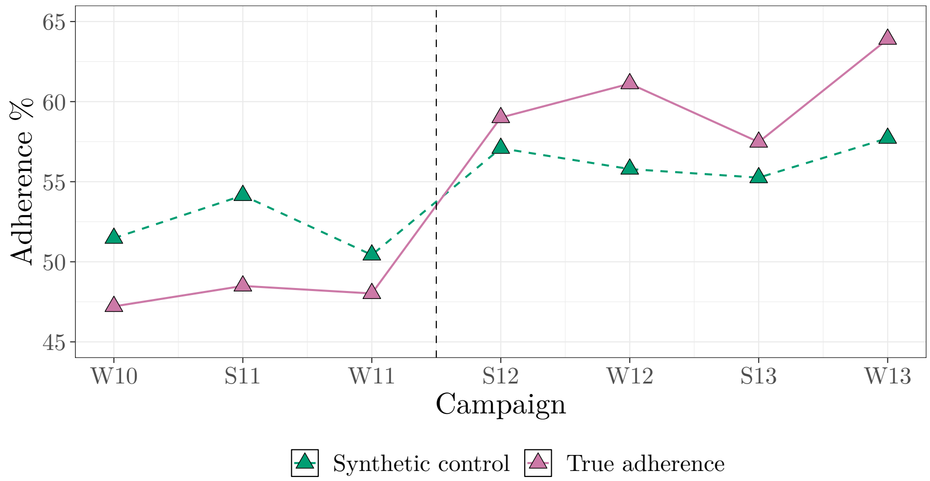

Figure 3 Average adherence of franchises (pink, solid line) and of their synthetic controls (green, dashed line).

Note. Each point corresponds to the average $\overline{Adherence_t}$ across all countries, computed from country-aggregate data $Adherence_{it}$. Each franchise's synthetic control was generated based on pre–Intervention 2 data on adherence and covariates. The dashed line marks when Intervention 2 took place (S2012).

	DiD coefficients			
	Adherence			
	(1)		(2)	
Int2×Treated	0.0802**	(0.0215)	0.0391*	(0.0167)
Int2	0.0225*	(0.0101)		
Treated	-0.0411**	(0.0108)		
ExperienceWithDSS	0.00627	(0.00336)	0.548***	(0.0517)
LogNumStores	0.00133	(0.0222)	-0.0618	(0.0783)
Constant	0.506***	(0.0217)		
Group dummies	No		No	
Campaign dummies	No		Yes	
Country dummies	No		Yes	
N	294		294	

Table 14 Difference-in-differences coefficient estimates of the effect of Intervention 2, where the dependent variable is $Adherence_{it}$, the treated units are franchises, and the control units are their synthetic controls.

Note. Estimation done using country-aggregate data (i.e., one observation for each country and campaign). Each franchise's synthetic control was generated based on pre–Intervention 2 data on adherence and covariates. Robust standard errors in parentheses. Not reported: campaign, and country dummy variables, in (2).

that of their synthetic controls. The reason for this very poor match is the fact that franchises' adherence, in general, was much lower than that of own-store countries (that is, precisely, what the interventions were meant to correct). This is most clearly seen by looking at the paper's Figure 3: before Intervention 2, many pink triangles (franchises) are strictly below most (or all) green circles (own stores); in W2010, even the average adherence across franchises is strictly below every own store country's adherence! Therefore, the quality of our synthetic controls is not very high and, although our main qualitative results still hold under this method (Intervention 1 did not have an effect on franchises' adherence, but Intervention 2 had a positive one), our estimated effect sizes differ.

References

Abadie A, Diamond A, Hainmueller J (2010) Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association* 105(490):493–505.

Abadie A, Gardeazabal J (2003) The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review* 93(1):113–132.

Altonji JG, Elder TE, Taber CR (2005) Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy* 113(1):151–184.

Cameron AC, Trivedi PK (2005) *Microeconometrics: Methods and Applications* (Cambridge University Press), 1st edition.

Chamberlain G (1980) Analysis of Covariance with Qualitative Data. *The Review of Economic Studies* 47(1):225–238.

Cunningham S (2021) *Causal Inference: The Mixtape* (Yale University Press).

Greene W (2002) The Bias of the Fixed Effects Estimator in Nonlinear Models. *NYU working paper* .

Greene WH (2012) *Econometric Analysis* (Prentice Hall), 7th edition.

Heckman JJ (1979) Sample Selection Bias as a Specification Error. *Econometrica* 47(1):153–161.

Huntington-Klein N (2021) *The Effect : An Introduction to Research Design and Causality* (Chapman and Hall/CRC).

Ichino A, Mealli F, Nannicini T (2008) From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and Their Sensitivity? *Journal of Applied Econometrics* 23(3):305–327.

Imbens GW (2003) Sensitivity to Exogeneity Assumptions in Program Evaluation. *American Economic Review* 93(2):126–132.

Puhani P (2000) The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys* 14(1):53–68.

Rosenbaum PR (1987) Sensitivity Analysis for Certain Permutation Inferences in Matched Observational Studies. *Biometrika* 74(1):13–26.

Rosenbaum PR (2002) Sensitivity to Hidden Bias. Rosenbaum PR, ed., *Observational Studies*, 105–170, Springer Series in Statistics (New York, NY: Springer).

Rosenbaum PR, Rubin DB (1983) Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. *Journal of the Royal Statistical Society: Series B (Methodological)* 45(2):212–218.

Smith J (2014) Bivariate Normal Selection Model. URL `https://econjeff.blogspot.com/2014/01/bivariate-normal-selection-model.html`.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data* (MIT Press), 2nd edition.

Wooldridge JM (2019) Correlated Random Effects Models with Unbalanced Panels. *Journal of Econometrics* 211(1):137–150.