The Impact of Promotions on Store Visits: A Counterfactual Approach

by Joel Persson

May 2, 2018

Bachelor's thesis in Statistics (15 ECTS)

Supervisor: Peter Gustafsson
Supervisor at GfK: Antonio Marañon

Department of Statistics
Lund University

Abstract

This thesis empirically quantifies the impact of promotions on store visits in the Swedish grocery retailing sector with nationally representative panel data on household purchases of ground coffee. Using the potential outcomes framework, the impact is calculated as the difference between outcomes with promotion and their counterfactuals estimated with two regression models. The first model is an OLS fixed effects model used by market research firm GfK and the second is a Poisson fixed effects model. The Poisson model's identification of the promotion effect is shown to be superior by accounting for that the dependent variable is discrete, the heterogenous time effects in the cross-section, and possible brand-switching behaviour. Standard errors robust to heteroscedasticity and cross-sectional and serial correlation are estimated for inference of the promotion effect under spatio-temporal dependence. A procedure for obtaining counterfactuals with regression models under multiple concurrent and continuous treatments is presented and an estimator of the cumulative treatment effect with adjustment for spatio-temporal dependence is derived and used to estimate the promotion impact on store visits. The findings are valuable for companies in market research, retailing and consumer packaged goods. The contributions of the thesis are methods for estimating promotion impact and an improvement of GfK's methodology.

Keywords: Promotion, potential outcomes, counterfactual analysis, treatment effects, panel data econometrics, spatio-temporal

Acknowledgements

I want to thank Antonio Marañon for introducing me to GfK and the topic of this thesis and all the interesting discussions we have had about about statistics, work, and related things. I am grateful to Jonas Wallin for answering my many questions about generalised linear models and Peter Jochumzen at the Department of Economics for the advice regarding panel data econometrics. I also want to express my gratitude to my advisor Peter Gustafsson for his guidance throughout the process of writing this thesis. Lastly, I would like to thank my mother and father for their support during my studies.

Contents

1 Introduction

One of the most fundamental problems in marketing is to determine the returns of marketing activities for optimising the budget allocation across the Marketing Mix variables. The Marketing Mix consists of the 4 P's Product, Price, Place, and Promotion, a set of variables used for deciding which activities the firms should use to meet its' marketing objectives (Kotler & Keller, 2012). Product refers to aspects of the good that is sold, price is what the consumer pays for the product, place relates to the ways of providing the product to the consumer, and promotion are actions used to increase demand, differentiate, and present information for a product by means such as advertising and campaigns.

Firms carry out activities related to the 4 P's to obtain a market response. An example is the use of sales promotions, such as temporary price-discounts, to increase sales or store visits (Teunter, 2002). Despite its popularity, the effects of sales promotion are not well understood (Persson, 1995). A central reason is that its effects are unobservable, since for any given outcome from promotions, the counterfactual outcome without promotion is unknown. Historically, it has been difficult to estimate the effects of marketing activities due to a lack of methods for collecting and analysing data, but this has changed in the last decades with the development of measurement methods such as product scanners and web tracking together with the application of econometric and statistical analyses. This progress has resulted in a greater responsibility for marketing to accurately show its impact on business results, and this is heard across both industry and academia. Marketing Science Institute, a leading research-based organisation for marketing academics and global businesses, list understanding and modelling the effect, value and causal impact of marketing actions one of five research priorities from 2016 to 2018 (Marketing Science Institute, 2016).

Existing research on promotions have focused on the mechanisms that explain why and how consumers' purchasing behaviour respond to promotions. There is less research on the outcome of promotion on business metrics such as store visits. From a business standpoint, knowing the outcome is arguably of higher importance. Market research firm GfK have conducted such studies in Sweden and Germany. They applied an econometric model to household scanner data on grocery purchases from retailers to estimate the effect of promotions on store visits. GfK is interested in improving the model and better understand its limitations.

As such, the aim of this thesis is to quantify the unobservable impact of sales promotions on the number of store visits. GfK's current model is examined and an alternative approach is presented that address its limitations. The research questions of this thesis are:

1. How can the unobservable impact of promotions on store visits be quantified?

2. What is the impact of promotions on store visits?

3. How does the promotion effect differ across retailers and brands?

This study differ from previous research on promotion effects by using the potential outcomes framework, in which the number of store visits in a given time period is viewed as a potential number determined by whether there was promotion or not. A count data regression model is proposed for identifying the intrinsic brand effects of promotion on store visits. The model accounts for heterogenous time trends in the cross-section, is robust to heteroscedasticity and spatio-temporal dependence among observations, and consumer brand-switching behaviour that arise during promotions, none of which are not taken into account in GfK's original model. An algorithm for obtaining the counterfactual number of store visits had there not been promotion is shown and the incremental number of store visits are derived using the counterfactuals. The results indicate that GfK's original model provide too optimistic, biased, and inconsistent estimates of the promotion effect.

The thesis have three major contributions. First, it provides valuable estimates for managers in grocery retailing and consumer packaged goods firms of how many store visits promotions can generate. This is crucial for firms' decisions of which products to promote for optimal returns. Secondly, an easy to implement procedure for estimating effects from multiple concurrent and continuous treatments in the absence of matching covariates is presented. Third, GfK's current methodology for estimating promotion impact is improved, thereby increasing the value GfK can offer their clients by providing more accurate estimates.

The remainder of thesis is structured as follows. Section 2 provides the economic theory that explain why and how consumers respond to promotions. The section contains a review of previous research on promotion effects. Section 3 explains the statistical theory for the empirical analysis. Section 4 introduces the data and the empirical context. The regression models, the algorithm for obtaining the counterfactual outcomes, and the estimator of the promotion impact are shown in section 5. The empirical results are provided in section 6. In section 7, the results are discussed and related to the thesis' aim and research questions. Section 8 concludes the thesis with some remarks on the practical and theoretical contributions, the limitations, and possibilities for future research. Supplementary information, statistical tests, and a proof that validates the methods and findings are provided in the appendix.

2 Theoretical Background

This section serves as an introduction to promotion effects. It begins with the economic intuition for why and how promotions affect consumer behaviour. This is followed by the role of market structure on the use and effects of promotions. The third section discusses the causes and mechanisms of promotional behaviour and their link to store visits.

2.1 Consumer Choice Theory

Three principles from consumer choice, the microeconomic field devoted to the choices consumers make regarding their consumption, are of use for understanding promotion response. First, consumers have preference profiles that determine which goods they chose to purchase. Secondly, consumers' consumption is limited by their budget constraint. Third, consumers seek to maximise their utility obtained from consumption given their budget constraint (Perloff, 2014). The principles are described in the following paragraphs.

The utility a consumer obtains from a good is a function of two inputs, the pleasure it provides them and its cost. Thereby, consumers can rank goods according to their perceived utility. Consider a rational consumer with a certain preference for a good. Utility maximisation implies that the consumer will prefer to purchase the good at the lowest available price. Similarly, given two goods from which the consumer derives equal pleasure, the consumer will prefer the cheaper good. If the difference in price between two similar goods is large enough, the utility of the cheaper good may be higher even though it provides less pleasure. Consumers can thus maximise their utility by choosing the good they prefer most given an expenditure, or by choosing a sufficiently satisfying good while minimising their expenditure. This has implications for consumers response to sales promotion.

Since prices explain purchasing choices, sales promotions shift demand. How much depends on the good's *price elasticity of demand*, defined as the percentage change in quantity demanded given a percentage change in price. *Cross-price elasticity of demand*, in turn, is a measure of the percentage change in quantity demanded of good given a percentage change in price of another good (Perloff, 2014). If the price change of one good is sufficiently large for the consumer to purchase another good, there exists a *substitution effect.* Price elasticities of demand is a useful tool to understand promotion effects. With a sufficient price reduction, consumers may change their planned behaviour and purchase the good with the reduced price.

Another factor that determines the demand for a good is *budget share*, defined as the share of consumer expenditures attributed to a good. Naturally, consumers are more price-sensitive towards goods that constitute a large share of their expenditures. This does not imply that the good is necessarily expensive, since a good that is bought with sufficient frequency or in sufficient quantities will have a substantial financial impact over a longer time period.

2.2 Oligopolistic Markets

Why do retailers use promotions? What determines how large the promotion response might be? A potential answer may lie in the market structure. Asplund and Friberg (2002) studied the Swedish grocery retail market and found that the market structure had significant effect on the price level. Specifically, higher local concentration of stores, higher regional wholesaler concentration and a lower market share of large stores were correlated with higher prices. The Swedish grocery retail market is an example of an *oligopoly* with only six retailers whose sales in 2016 accounted for 87 percent of the market (HUI Research, 2017). Technically, an oligopoly is a market characterised by few firms with dominant positions. This has implications for the price level and how the market functions (Perloff, 2014).

The low number of firms enable them to directly or indirectly agree on a high price level. Thus the long-run profit tend to be positive. The low number of firms allows each firm to observe competitors' behaviours and act strategically, and it makes it easier for consumers to compare prices across firms. This creates interdependent behaviour where actions of one firm have immediate impact on the others (Pepall, Richards, & Norman, 2014). Because of the low number of firms, each firm captures a large market share and is more vulnerable to competitors' actions.

According to economic theory, an oligopoly firm with lower regular prices than its competitors for the same products will get more customers. Thus undercutting competitors' prices may start a price war benefitting none of the firms. As a consequence, regular prices for undifferentiated products tend to be similar in oligopolies. Nonetheless, the positive long-run profit and high market shares in oligopolies enable firms to reduce prices temporally to increase short-term sales. When a firm launches a sales promotion, customers will migrate to this firm and its short-term store visits and market share increase (Hirshleifer, Glazer, & Hirshleifer, 2005).

2.3 Promotion Effects on Purchasing Behaviour

A common belief among marketing managers is that sales promotions increase sales volume and market share in the short-term. This has empirical support with a large body of marketing literature showing that consumers change their behaviour in response to promotions. Some identified mechanism are increased consumption, unplanned purchasing, repeat purchasing, brand-switching, store-switching, and purchase acceleration (Neslin, Henderson, & Quelch, 1985; Gupta, 1988; Teunter, 2002; Sun, Neslin, & Srinivasan, 2003; Sun, 2005). An influential paper is Gupta (1988) that decomposes the sales increase from sales promotions into those caused by brand-switching, purchase-time acceleration, and stockpiling. He finds that 84 percent of the sales increase during sales promotion is attributable to brand-switching, less than 14 percent due to purchase-time acceleration, and that stockpiling accounts only for the remaining less than 2 percent. It is thereby important to consider which mechanism explain the promotion effect, whether the mechanism implies incremental purchases or planned purchases shifted in time, and whether the effect occur at the consumer, store, or firm level.

Brand-switching arises as a consequence of substitution effects and is by definition incremental to the brand but not for the individual or the store, since the consumer had planned to purchase the type of product. Brand-switching may induce the repeat purchasing effect, which refer to that a consumer that has tried a new brand because of the low price may continue to purchase the brand thereafter. The repeat purchasing effect is also incremental at the brand level.

The store-switching effect can also be understood as a substitution effect but at the retailer level. It occurs when a consumer choses to shop at another store that has promotion to profit from the price difference. The store-switching effect is believed to be higher for homogenous products with high demand to which consumers are price-sensitive (Teunter, 2002). Retailers infer these shopping trips to be incremental since the consumer had planned their purchase at another store. Store-switching purchases are however not incremental at the consumer level and will only be incremental for brands if combined with brand-switching, for instance because the stores carry different brands. Store-switching can generate complementary purchases, meaning that consumers will purchase other goods during their shopping trip and thereby generate additional sales apart from the product that gave rise to the store-switch (Sun, 2005).

Purchase acceleration means that consumers shift planned purchases earlier in time during promotions or buy an increased quantity, thereby stockpiling. It arises as a consequence of the relaxed budget constraint when prices are lowered, and implies no incremental purchases at any level. Blattberg, Eppen, and Lieberman

(1981) argue that stockpiling and purchase acceleration may be a desired outcome of promotions for retailers. The reason is that retailers and households have *differential inventory costs*. Holding large inventory is costly for the retailer, but rarely for the household. Retailers may thereby use promotions to off-load unsold inventory so the inventory cost is transferred to the household. From this perspective, retailers may deliberately use promotions to cause purchase acceleration and stockpiling behaviour.

There is also a substantial amount of research on statistical modelling of purchasing behaviour from the early decades of quantitative empirical marketing research. Since the outcome variable has often been in counts, previous studies on the effects of promotion and their explanatory behavioural mechanisms have typically used a model based on the Poisson or a closely related discrete distribution (Chatfield, Ehrenberg, & Goodhardt, 1966; Chatfield & Goodhardt, 1973; Paul, 1978; Frisbie, 1980; Dunn, Seader, & Wrigley, 1983). The next section explains the statistical theory for the analysis in this thesis.

3 Statistical Theory

This section explains the theory of the methods used in the empirical analysis. First, the framework for estimating treatment effects from observational data is presented. This is followed by a review of the standard model used for identification of treatment effects and its variant for count data. The section ends with methods for robust inference under heteroscedasticity and cross-sectional and serial dependence.

3.1 Causal Inference and Counterfactual Analysis

The effect of promotions on store visits can be determined by decomposing the number of store visits into two outcomes; the baseline, which is the amount without promotion, and the observed amount with promotion. The difference between the outcomes will then be the number of store visits attributed to promotions in excess of the baseline during the time period. To determine the effects of promotion on store visits is thereby a problem of *causal inference*, since the goal is to directly attribute the outcome to promotion.

The potential outcomes framework (Rubin, 1974; Holland, 1986) is the dominant framework for estimating causal effects from observational data (Imbens & Wooldridge, 2017). In this framework, the causal effect of a treatment is interpreted as the difference between a pair of *potential outcomes* for the same unit at a given point in time. Here, treatment (or intervention, manipulation) refers to any action applied to the unit, which may for instance be a consumer, household, or geographical region (Imbens & Rubin, 2015).

Let Y_{it} be the outcome for unit i , $i = 1, \dots, n$, at time t, $t = 1, \dots, T$. Each observation is associated with the binary treatment variable $X_{it} = x$, $x = 0, 1$, where 1 indicates that the unit received the treatment, and 0 that it did not. The realised outcome for a given unit i at time t is then given by $Y_{it}^{obs} = Y_{it}(X_{it}) = Y_{it}(x)$, where $Y_{it}(0)$ denotes the outcome under no promotion, i.e. the baseline, and $Y_{it}(1)$ denotes the outcome given promotion, i.e. the baseline plus the increment due to promotion. The potential outcomes for each unit are thereby

$$Y_{it}(X_{it}) = \begin{cases} Y_{it}(1) \text{ if } X_{it} = 1, \\ Y_{it}(0) \text{ if } X_{it} = 0. \end{cases} \tag{3.1}$$

Define the *treatment effect* for unit i in time period t as $\tau_{it} = Y_{it}(1) - Y_{it}(0)$. The Rubin Causal Model (Holland, 1986) states that the population level *average treatment effect for the treated* (ATT) is calculated as

$$\begin{aligned} \tau_{ATT} &= \mathbb{E}[\tau_{it} \mid X_{it} = 1] = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid X_{it} = 1] \\ &= \mathbb{E}[Y_{it}(1) \mid X_{it} = 1] - \mathbb{E}[Y_{it}(0) \mid X_{it} = 1] \end{aligned} \tag{3.2}$$

with the sample analogue

$$\begin{aligned} \hat{\tau}_{ATT} &= \frac{1}{n_1} \sum_{i=1}^{n} [\hat{\tau}_{it} \mid X_{it} = 1] \\ &= \frac{1}{n_1} \sum_{i=1}^{n} \left(y_{it}(1) - y_{it}(0) \mid X_{it} = 1 \right) \end{aligned} \tag{3.3}$$

where

$$n_1 = \sum_{i=1}^{n} X_{it}.$$

The ATT measures the effect of promotion over the time periods and units that had promotion in comparison to if they would not have promotion. The problem with the estimator is that the first term in the right hand side in (3.2) can be observed but never the second term, and τ is always unobservable since both potential outcomes cannot occur simultaneously for the same unit. What is observable for each unit at any given point in time is

$$Y_{it}^{obs} = Y_{it}(1)X_{it} + Y_{it}(0)(1 - X_{it}). \tag{3.4}$$

Since it is only possible to observe half of the outcomes of interest, causal inference is a *missing data problem*; for any observed outcome y_{it}^{obs}, the counterfactual potential outcome y_{it}^{c} required to determine the treatment effect τ is missing. This is referred to as the *fundamental problem of causal inference* (Imbens & Rubin, 2015).

To compare the observed treated and observed non-treated outcomes, the *treatment assignment mechanism* must be considered, that is, which units and time periods receive the treatment. If the assignment mechanism is random, meaning

$$\big(\mathbf{Y}_i(1), \mathbf{Y}_i(0)\big) \perp \mathbf{X}_i, \tag{3.5}$$

the setting is identical to an experiment and equation (3.2) will yield unbiased estimates of the effect. But the treatment assignment cannot be assumed to be random in observational data. It is reasonable to assume that certain weeks are selected to have promotion due to a confounding variable $\mathbf{Z}$, that is, a variable that affects both the promotion assignment $\mathbf{X}_i$ and the number of store visits $\mathbf{Y}_i$. As an example, brands are promoted certain times of the year due to seasonality in demand. Then, estimating the effect by taking the difference between the observed outcomes with promotion and those without promotion gives a biased estimate that partially reflects the effect of seasonality. This leads to a set of assumptions needed for the ATT to yield unbiased estimates of the treatment effect.

The *Stable Unit Treatment Value Assumption* (SUTVA) consists of two components: no interference, and no hidden variations in treatment. No interference means that the treatment assignment to a unit do not affect the potential outcome for other units. This is equivalent to the assumption that observations are independent and identically distributed samples from the population. No hidden variations in treatment states that for each treatment, each unit can only receive a single form of the treatment. It does not require that all units are exposed to the same treatments, but only that the treatment does not change within units over the time period (Imbens & Rubin, 2015). If either SUTVA component is not satisfied the potential outcomes are not uniquely defined (Imbens & Rubin, 2015).

The third assumption is *unconfoundedness*, also known as conditional independence (Cameron & Trivedi, 2005). It means that conditional on relevant confounders, the treatment assignment is random. Unconfoundedness for the ATT requires that

$$\mathbf{Y}_i(0) \perp \mathbf{X}_i \mid \mathbf{Z}_i. \tag{3.6}$$

If the confounder is seasonality, unconfoundedness implies that promotions are randomly assigned to units and time periods after conditioning on that seasonality affects the assignment.

The partial *overlap*, also know as matching or common support, assumption states that for all $\mathbf{z}$ in the support $\mathcal{Z}$ of $\mathbf{Z}$,

$$\mathbb{P}(\mathbf{z}) < 1 \text{ for all } \mathbf{z} \in \mathcal{Z}. \tag{3.7}$$

This means that each unit that did not receive the treatment had some probability of receiving it. It implies that for each treated unit there is a non-treated unit with similar value on the confounding variable. Thus the distributions of covariates for treated and non-treated observations overlap.

Under unconfoundedness and overlap, the treatment assignment mechanism is *strongly ignorable*, meaning that the assignment mechanism can be ignored since it is independent of the unobserved potential outcomes (Rosenbaum & Rubin, 1983). With observational data, this is unlikely to hold due to the non-random assignment mechanism. As a solution, *matching methods* have been developed that seek to equate the distribution of covariates among treated and non-treated observations so their only difference is the treatment by matching them on the values of pre-treatment confounding covariates (Stuart, 2010). Matching methods take an observation that is similar to another observation on relevant confounders but has the alternative treatment status and impute it as the counterfactual potential outcome. This results in complete time series for each treatment status, which solves the missing data problem. Identification of the ATT only requires that the counterfactual $(Y_{it}(0) \mid X_{it} = 1) = Y_{it}^{c}(0)$ is identified and imputed for each $Y_{it}(1)$ since $Y_{it}(1) \mid X_{it} = 1$ is already observed.

If strong ignorability on the other is valid, there is no omitted variable bias after conditioning on the confounder and no confounding of the treatment effect τ (Cameron & Trivedi, 2005). Then, the treatment variable X_{it} can be considered exogenous and a regression function can be used to adjust for confounders and estimate the treatment effect. Imbens and Wooldridge (2017) show this by defining the ATT conditional on $\mathbf{z}$ and note that it is identified for $\mathbf{x}$ in $\mathcal{Z}$:

$$\begin{aligned}
\mathbb{E}[\tau_{it}(z)] &\equiv \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid Z_{it} = z] \\
&= \mathbb{E}[Y_{it}(1) \mid Z_{it} = z] - \mathbb{E}[Y_{it}(0) \mid Z_{it} = z] \\
&= \mathbb{E}[Y_{it}(1) \mid X_{it} = 1, Z_{it} = z] - \mathbb{E}[Y_{it}(0) \mid X_{it} = 0, Z_{it} = z] \\
&= \mathbb{E}[Y_{it} \mid X_{it} = 1, Z_{it} = z] - \mathbb{E}[Y_{it} \mid X_{it} = 0, Z_{it} = z] \\
&= \mu_1(z) - \mu_0(z)
\end{aligned} \tag{3.8}$$

Here, $\mu_x(z) \equiv \mathbb{E}[Y_{it}(x) \mid Z_{it} = z]$, $x = 0, 1$ are the regression functions for the potential outcomes. The third equality follows by unconfoundedness, since $\mathbb{E}[Y_{it}(x) \mid X_{it} = 0, Z_{it}]$ is independent of x, and the terms in the second to last equation can be estimated with the regression functions by the overlap assumption. By the law of total expectation,

$$\begin{aligned} \tau_{ATT} &= \mathbb{E}[\tau_{it} \mid X_{it} = 1] = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid X_{it} = 1] \\ &= \mathbb{E}[\mu_1(Z_{it}) - \mu_0(Z_{it}) \mid X_{it} = 1] \end{aligned} \tag{3.9}$$

will hold without any assumptions, and the ATT can be estimated if $\mu_1(\cdot)$ and $\mu_0(\cdot)$ are identified. Under unconfoundedness, $\mu_1(\cdot)$ and $\mu_0(\cdot)$ are in fact identified, since

$$\begin{aligned} \mathbb{E}[Y_{it} \mid Z_{it} = z, X_{it} = 1] = \mu_1(z), \\ \mathbb{E}[Y_{it} \mid Z_{it} = z, X_{it} = 0] = \mu_0(z). \end{aligned}$$

Regression adjustment (RA) estimators use this result. By fitting separate regressions for each treatment status x with the confounder as an explanatory variable, the potential outcomes for each treatment status are obtained while adjusting for the influence of the confounder on the treatment effect. Specifically, obtain $\hat{\mu}_1(z)$ by fitting the regression $\mathbb{E}[Y_{it} \mid Z_{it}, X_{it} = 1]$ and obtain $\hat{\mu}_0(z)$ by fitting $\mathbb{E}[Y_{it} \mid Z_{it}, X_{it} = 0]$. Given that $\mu_1(\cdot)$ and $\mu_0(\cdot)$ are consistent estimators, the treatment effect is consistently estimated as

$$\hat{\tau}_{ATT,\ RA} = [\hat{\tau}_{it} \mid X_{it} = 1] = \frac{1}{n_1} \sum_{i=1}^{n} X_{it} \big(\hat{\mu}_1(Z_{it}) - \hat{\mu}_0(Z_{it}) \big). \tag{3.10}$$

Since X_{it} is a binary treatment variable that only take values $0, 1$, the regression coefficient $\hat{\beta}$ is the estimated treatment effect $\hat{\tau}_{it}$. The identification of the treatment effect requires that the regression is correctly specified and controls for omitted variable bias and confounding. Further, the potential outcomes framework requires panel data, also known as longitudinal data, meaning data observed for the same units over time. The following two sections presents the panel data regression model most commonly used for estimating treatment effects from observational data.

3.2 Fixed Effects Model

A simple panel data model is the static linear unit-specific and unobserved effects model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_i\gamma + \epsilon_{it}, \ \epsilon_{it} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\epsilon) \tag{3.11}$$

where y_{it} is the observed outcome y for unit i, $i = 1, \dots, n$ at time t, $t = 1, \dots, T$, $\mathbf{z}_i$ is a $(c \times 1)$ row-vector of c unobserved variables, $\mathbf{x}_{it}$ is a time-variant $(k \times 1)$ row-vector of k regressors, and the error term ϵ_{it} is assumed to follow a zero mean process with constant variance $\mathbb{V}[\epsilon_{it}|\mathbf{x}_{it}] = \sigma^2_\epsilon$ without serial correlation for all t. The parameters of interest are $\boldsymbol{\beta}$ while γ are nuisance parameters for potential confounding variables $\mathbf{z}_i$. Assuming that $\mathbf{z}_i$ are time-invariant, the model can be rewritten as

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \epsilon_{it} \tag{3.12}$$

where $\alpha_i = \mathbf{z}'_i\gamma$, so α reflects the joint impact of unit-specific unobserved effects $\mathbf{z}$ on y_{it}.

Essentially, equation (3.12) is a linear model with different intercepts α_i for each unit i, which can be written as a regression model with dummy variables,

$$y_{it} = \sum_{i=1}^{n} \alpha_i d_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \epsilon_{it} \tag{3.13}$$

in which $d_i = 1$ if an observation corresponds to unit i, and 0 otherwise. Specification (3.13) is called the least squares dummy variable (LSDV) estimator (Verbeek, 2008), and estimates n dummy variables in addition to the remaining parameters. This model is feasible to estimate in panels where the cross-sectional dimension is small, since it quickly becomes computationally demanding as $n \to \infty$.

As a solution to estimating n dummy variables, it is possible to obtain the same model as the LSDV estimator if each variable in equation (3.12) is subtracted with their within-unit time-series average, that is

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\epsilon_{it} - \bar{\epsilon}_i) \tag{3.14}$$

where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, $\bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T} \mathbf{x}_{it}$ and $\bar{\epsilon}_i = T^{-1}\sum_{t=1}^{T} \epsilon_{it}$. This transformation is known as the *within transformation*, and yields the within, or *fixed effects* (FE), estimator. The FE estimates of $\boldsymbol{\beta}$ are obtained with ordinary least squares (OLS) as

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i). \tag{3.15}$$

The β parameter in the static FE model is interpreted as the average effect of a one unit increase in $\mathbf{x}_{it}$ on y_{it} within the same time period t. If $\mathbf{x}_{it}$ is a binary treatment variable, then $\beta = \tau$, the treatment effect.

Mundlak (1978) showed that the LSDV and FE model yield identical estimates of $\boldsymbol{\beta}$[1]. While the LSDV specification controls for the unit-specific unobserved effects α_i with dummy variables, the FE specification controls for α_i by cancelling them by subtraction, since the mean of a constant is the constant itself. Thus, the FE estimator do not estimate α_i, but only the within-unit differences from the mean.

The necessary condition for unbiased estimates with FE is that the regressors are *strictly exogenous*, meaning $\mathbb{E}[\mathbf{x}_{it} \mid \epsilon_{it}, \alpha_i] = 0$, for all t and lags s. This holds if the panel is stationary[2]. For consistent estimates of $\boldsymbol{\beta}$ with FE it is further assumed that $\mathbb{E}[(\mathbf{x}_{it} - \bar{\mathbf{x}}_{it})\epsilon_{it}] = 0$, for which it is sufficient that $\mathbb{C}ov[\mathbf{x}_{it}, \epsilon_{it}] = 0$, and $\mathbb{C}ov[\bar{\mathbf{x}}_i, \epsilon_{it}] = 0$, and those conditions follow if $\mathbf{x}_{it}$ is strictly exogenous (Cameron & Trivedi, 2013).

Time effects can be estimated by including time-specific fixed effects as dummy variables for each time period t. This will capture shocks to the outcome attributable to each time period and is computationally feasible in short panels since the number of time dummies increase with T. Another method is to include a trend variable. This is suitable when the outcome exhibits a deterministic trend over time.

The FE model is unique among panel data models in that it allows α_i to be correlated with $\mathbf{x}_{it}$ (Winkelmann, 2008). This is practical since the assumption that they are uncorrelated often does not hold empirically. In this case, it means that the promotion treatment or the confounders can be correlated with the units. To see the implication of this property, consider again the general individual-effects model in equation (3.12). Since α_i are unobserved, the OLS estimator will not be able to distinguish them from the error term. Thereby, (3.12) is equivalent to

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + (\alpha_i + \epsilon_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it} \tag{3.16}$$

where $u_{it} = (\alpha_i + \epsilon_{it})$ is a composite error term. If α_i are correlated with $\mathbf{x}_{it}$, $\hat{\boldsymbol{\beta}}$ are biased estimators of $\boldsymbol{\beta}$, and the model is subject to omitted variables bias. The within transformation in which α_i are eliminated solves this problem. Omitted time-invariant effects captured by α_i are controlled for, as are time-invariant effects contained in ϵ_{it} since they are constant and perfectly collinear with α. Thus, FE allows the researcher to control for unit-specific time-invariant bias caused by omitted variables which may confound the estimate of the treatment effect even if the variables are unobserved.

[1] A proof of this is given in section A.1 in the appendix.

[2] See section A.7 in the appendix for discussion and tests of stationarity.

3.3 Poisson Fixed Effects Model

Store visits are discrete events measured as non-negative integers. In such data effects are typically multiplicative. A common way to model it is to log-transform the dependent variable so the model becomes linear in parameters and OLS can be applied. The problems is that OLS can predict non-integers and negative values. If the data contain zeros, the log-transformation will not work since the logarithm of zero is undefined. A common ad-hoc solution is to add a constant prior to taking the log, but this makes it difficult to obtain the fitted values on the original scale. Another issue is that models with multiplicative effects have multiplicative errors,

$$y = \exp(\mathbf{x}'\boldsymbol{\beta})\eta \tag{3.17}$$

so that

$$\log y = \mathbf{x}'\boldsymbol{\beta} + \log \eta \tag{3.18}$$

where η is the multiplicative error term. Consider the mean function $\mathbb{E}[y \mid \mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta})$. It implies that $\mathbb{E}[\eta \mid \mathbf{x}] = 1$. The problem is that for the linearised model, $\mathbb{E}[\log \eta \mid \mathbf{x}]$ is only constant if η is independent of $\mathbf{x}$. If the variance of η depends on $\mathbf{x}$, the expectation of $\log \eta$ will depend on the regressors. Thus, the log-linear OLS model will suffer from endogeneity and have inconsistent estimates of $\boldsymbol{\beta}$ if the errors are heteroscedastic (Winkelmann, 2008). Count models based on discrete distributions that model the multiplicative effects directly are thereby appropriate. Count data regressions are represented with the *generalised linear models* (GLM) framework. The following section provides an introduction to GLMs.

GLMs model the dependence of a scalar variable y_i on a vector of regressors $\mathbf{x}_i$. The conditional distribution of $y_i \mid \mathbf{x}_i$ belongs to the linear exponential family of distributions with probability density function

$$f(y; \lambda, \theta) = \exp\left(\frac{y \cdot \lambda - b(\lambda)}{\theta} + c(y, \theta)\right) \tag{3.19}$$

where λ is a parameter that depend on the regressors through a linear predictor, and θ is a dispersion parameter for the variance. Functions $b(\cdot)$ and $c(\cdot)$ determine which distribution in the family to use for y. The conditional mean of y_i is given by $\mathbb{E}[y_i \mid \mathbf{x}_i] = \mu_i = b'(\lambda_i)$ and the conditional variance is $\mathbb{V}[y_i \mid \mathbf{x}_i] = \theta \cdot b''(\lambda_i)$. Thereby the distribution of y_i is determined by its mean up to a dispersion parameter θ. Also, $\mathbb{V}[y_i \mid x_i] \propto \mathbb{V}[\mu] = b''(\lambda(\mu))$, the variance function. The regression function of the dependence of $\mathbb{E}[y_i \mid \mathbf{x}_i] = \mu_i$ on $\mathbf{x}_i$ is specified as

$$g(\mu_i) = \mathbf{x}_i'\beta \tag{3.20}$$

in which $g(\cdot)$ is a link function.

The most simple distribution for modelling count data is the Poisson distribution with density function

$$f(y;\mu) = \frac{\exp(-\mu)\cdot\mu^y}{y!},\ \ y = 0, 1, 2, \ldots, \tag{3.21}$$

and the canonical link function $g(\mu) = \log(\mu)$. Thus, Poisson regression models a log-linear relationship between the mean and the predictors. The Poisson distribution assumes equidispersion; that $\mathbb{E}[Y] = \mu$ and $\mathbb{V}[Y] = \mu$. Thus, θ is fixed at 1 and the variance function is $\mathbb{V}[\mu] = \mu$.

In Poisson regression the unit-specific fixed effect α_i enter multiplicatively, so the *Poisson fixed effects* (PFE) model is given by

$$\begin{aligned}\mathbb{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i] &= \mu_{it} \\ &= \alpha_i\lambda_{it}\end{aligned} \tag{3.22}$$

where $\lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$, and has conditional probability function

$$f(y_{it} \mid \mathbf{x}_{it}, \alpha_i) = \frac{\exp(-\alpha_i\lambda_{it})(\alpha_i\lambda_{it})^{y_{it}}}{y_{it}!}. \tag{3.23}$$

Since $\mathbb{V}[y_{it} \mid \mathbf{x}_{it}, \alpha_i] = \alpha_i\exp(\mathbf{x}_{it}'\boldsymbol{\beta})$ by equidispersion and equation (3.22), the standard errors in the PFE model are by default heteroscedastic (Cameron & Trivedi, 2013). Although α_i enter multiplicatively, they may still be interpreted as a shift in intercept for unit i if the exponential conditional mean is modelled, since

$$\mu_{it} = \alpha_i\exp(\mathbf{x}_{it}'\boldsymbol{\beta}) = \exp(\log(\alpha_i)\mathbf{x}_{it}'\boldsymbol{\beta}).$$

Estimating $\boldsymbol{\beta}$ will jointly estimate α_i since they have a multiplicative effect. Like the OLS FE model, the PFE model can be estimated with dummy variables for α_i. However, since there are n fixed effects α_i to estimate from T observations in unit i, there are in total $n+k$ parameters to estimate from nT observations, so the number of dummy variables grows as $n \to \infty$. This is known as the *incidental parameters problem* and causes $\hat{\boldsymbol{\beta}}$ to be inconsistent unless n is small and T is large. The OLS FE model shown in section 3.2 does not suffer from this problem since it cancels α_i by subtraction, and inspecting the *concentrated* log likelihood function of the PFE model reveals that it neither suffers from it (Winkelmann, 2008). To do this, first assume that $\mathbf{x}_{it}$ are strictly exogenous, so that conditional on α_i the T observations for unit i are independent. Then,

$$f(y_{i1}, \ldots, y_{iT} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \alpha_i) = f(y_{i1} \mid \mathbf{x}_{i1}, \alpha_i) \ldots f(y_{iT} \mid \mathbf{x}_{iT}, \alpha_i).$$

Define $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ and $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})'$. Using (3.23), the likelihood of unit i is obtained as

$$\begin{aligned} f(\mathbf{y}_i \mid \mathbf{x}_i, \alpha_i) &= \prod_{t=1}^{T} \exp(-\alpha_i \lambda_{it})(\alpha_i \lambda_{it})^{y_{it}} / y_{it}! \\ &= \exp\Big(- \alpha_i \sum_{t=1}^{T} \lambda_{it} \Big) \prod_{t=1}^{T} \alpha_i^{y_{it}} \prod_{t=1}^{T} \lambda_{it}^{y_{it}} / \prod_{t=1}^{T} y_{it}! \end{aligned} \tag{3.24}$$

The log likelihood contribution of unit i is thereby

$$\mathcal{L}_i(\alpha_i, \boldsymbol{\beta}) = -\alpha_i \sum_{t=1}^{T} \lambda_{it} + \log(\alpha_i) \sum_{t=1}^{T} y_{it} + \sum_{t=1}^{T} y_{it} \log(\lambda_{it}) - \sum_{t=1}^{T} \log(y_{it})! \tag{3.25}$$

which has first derivate

$$\frac{\partial \mathcal{L}_i(\alpha_i, \boldsymbol{\beta})}{\partial \alpha_i} = -\sum_{t=1}^{T} \lambda_{it} + \alpha_i^{-1} \sum_{t=1}^{T} y_{it}.$$

Setting this to zero, the maximum likelihood (ML) estimator of α_i is given by

$$\hat{\alpha}_i = \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} \lambda_{it}} = \frac{\bar{y}_i}{\bar{\lambda}_i}. \tag{3.26}$$

Just as the linear OLS FE estimator uses subtraction in the within transformation to estimate its additive fixed effects α_i, the PFE model uses the ratio $\bar{y}_i / \bar{\lambda}_i$ to estimate its corresponding multiplicative fixed effects.

Substituting this estimator of α_i into (3.25) yields the concentrated log likelihood function, that is, the log likelihood function that does not depend on α_i. Taking the sum over all n units, it is given by

$$\begin{aligned} \mathcal{L}^c(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \left[- \sum_{t=1}^{T} y_{it} + \left(\log \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} \lambda_{it}} \right) \sum_{t=1}^{T} y_{it} + \sum_{t=1}^{T} y_{it} \log(\lambda_{it}) - \sum_{t=1}^{T} \log(y_{it})! \right] \\ &= C + \sum_{i=1}^{n} \left[\sum_{t=1}^{T} y_{it} \log(\lambda_{it}) - \sum_{t=1}^{T} y_{it} \log \sum_{t=1}^{T} \lambda_{it} \right] \end{aligned} \tag{3.27}$$

where C is a constant containing all terms not depending on $\boldsymbol{\beta}$. The first-order condition for $\boldsymbol{\beta}$ is then

$$
\begin{aligned}
\frac{\partial \mathcal{L}^c(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n}\left[\sum_{t=1}^{T} y_{it}\mathbf{x}_{it} - \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T}\lambda_{it}}\sum_{it}^{T}\lambda_{it}\mathbf{x}_{it}\right] \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it} - \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T}\lambda_{it}}\lambda_{it}\right) \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it} - \frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\right) = \mathbf{0}. \qquad (3.28)
\end{aligned}
$$

The ML estimator for $\boldsymbol{\beta}_{PFE}$ is the value $\boldsymbol{\beta}$ that solves (3.28). Since this estimator of $\boldsymbol{\beta}_{PFE}$ is independent of α_i, the PFE model does not suffer from the incidental parameters problem. An important feature is that consistency of $\hat{\boldsymbol{\beta}}_{PFE}$ does not require that the dependent variable is truly Poisson distributed, but only that the moment condition

$$\mathbb{E}[y_{it} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \alpha_i] = \alpha_i\lambda_{it}$$

holds (Winkelmann, 2008).

Since Poisson regression models the exponential conditional mean, $\boldsymbol{\beta}$ is not interpreted as the marginal effect of the regressors $\mathbf{x}$ on the conditional mean of y. Instead, differentiation with respect to regressor x_j yields

$$\frac{\partial \mathbb{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i]}{\partial x_{ijt}} = \alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})\beta_j = \beta_j \mathbb{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i]. \qquad (3.29)$$

Unlike with linear models, the marginal effect of x_j depends on α_i in PFE models. Consequently, consistent estimation of $\boldsymbol{\beta}$ is not sufficient for identifying the marginal effect. Still, the last equality in equation (3.36) implies that β_j measures the relative change in $\mathbb{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i]$ given a one-unit change in x_j. So if x is measured in percentages or is log-transformed, then β measures the semi-elasticity of $\mathbb{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i]$, analogous to its interpretation in log-linear OLS.

If $\mathbb{V}[Y] > \mathbb{E}[Y]$, the Poisson distribution's equidispersion property does not hold and the data is said to be *overdispersed*. If the conditional mean is specified correctly, the Poisson MLE is still consistent, but the standard errors for $\boldsymbol{\beta}$ will be wrong. This will have similar consequences as heteroscedasticity in OLS. There are two solution for this. The first is to use another model such as quasi-Poisson or negative binomial, or simply use the Poisson regression with robust standard errors. A quasi-Poisson or negative binomial model will guard against over-dispersion by estimating an extra parameter for the variance, whereas robust standard errors will guard against more departures from the Poisson model's assumption by allowing for different variances for each observation. In practice, a Poisson model with robust standard errors will most often yield very similar standard errors to a quasi Poisson or negative binomial with robust standard errors (Kleiber & Zeileis, 2008). Additionally, Greene and Hensher (2010) note that the Poisson model will be consistent even when the negative binomial is appropriate. The robust standard error approach is used in this thesis and is explained in the next section.

3.4 Robust Covariance Matrix Estimation

The parameter of greatest interest for inference is $\boldsymbol{\beta}$, since it is the coefficient for the promotion effect. The OLS estimator of $\hat{\boldsymbol{\beta}}_{FE}$ is obtained with

$$\hat{\boldsymbol{\beta}}_{FE} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \boldsymbol{\beta} + (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{e}}$$

with the corresponding residuals

$$\tilde{\mathbf{e}} = (\mathbf{I}_n - \mathbf{P})\tilde{\mathbf{y}} = (\mathbf{I}_n - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}')\tilde{\mathbf{y}}.$$

Here, $\mathbf{Y}$ and $\mathbf{e}$ are $(nT \times 1)$, $\mathbf{X}$ is $(k+1) \times nT$, $\boldsymbol{\beta}$ is $(k+1) \times 1$, $\mathbf{I}_n$ is the $(n \times n)$ identity matrix, and $\mathbf{P}$ is the $(n \times n)$ projection matrix, also known as the hat matrix. The tilde symbol denote the within transformation, $\tilde{y}_{it} = (y_{it} - \bar{y}_i)$, $\tilde{\mathbf{x}}_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$, and $\tilde{e}_{it} = (e_{it} - \bar{e}_i)$.

For inference of $\hat{\boldsymbol{\beta}}_{FE}$, the covariance matrix $\mathbf{V}[\hat{\boldsymbol{\beta}}_{FE}]$ is needed. It is derived as

$$\begin{aligned}
\mathbf{V}[\hat{\boldsymbol{\beta}}_{FE}] &= \mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] \\
&= \mathbb{E}[(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{e}}'\tilde{\mathbf{e}}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}] \\
&= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}'\mathbb{E}[\tilde{\mathbf{e}}'\tilde{\mathbf{e}}]\tilde{\mathbf{X}})(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \\
&= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}\boldsymbol{\Omega}\tilde{\mathbf{X}}')(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \qquad (3.30)
\end{aligned}$$

with $\boldsymbol{\Omega}$ being the diagonal error covariance matrix. Under the assumption of homoscedasticity, $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$, the error covariance matrix is

$$\boldsymbol{\Omega} = \mathbb{E}[\tilde{\mathbf{e}}'\tilde{\mathbf{e}} \mid \tilde{\mathbf{X}}] = \begin{bmatrix} \sigma_e^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_e^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_e^2 \end{bmatrix}$$

that is $\mathbf{\Omega} = \sigma_e^2\mathbf{I}_n = \sigma_e^2$, constant variance for all observations. Then,

$$\begin{aligned}
\mathbf{V}[\hat{\boldsymbol{\beta}}_{FE}] &= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}'\sigma_e^2\mathbf{I}_n\tilde{\mathbf{X}})(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \\
&= \sigma_e^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \\
&= \sigma_e^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \\
&= \sigma_\epsilon^2\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\right)^{-1} \qquad (3.31)
\end{aligned}$$

is the OLS FE covariance matrix, and both $\mathbf{\Omega}$ and $\mathbf{V}$ can be estimated by using the standard OLS estimator $\hat{\sigma}_{e_i}^2 = (1/(n-k))\sum_{i=1}^{n} e_i^2$ of the estimated error variance.

The procedure is different for the PFE model since it is estimated with ML. Let $\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log f_i(\mathbf{y}_i \mid \boldsymbol{\beta})$ denote the log likelihood function over all units. First obtain the gradient by taking its first derivative with respect to $\boldsymbol{\beta}$,

$$\mathcal{L}'(\boldsymbol{\beta}) = \sum_{i=1}^{n} g_i(\mathbf{y}_i \mid \boldsymbol{\beta}) \qquad (3.32)$$

where

$$g_i(\mathbf{y}_i \mid \boldsymbol{\beta}) = \left[\frac{\partial \log f_i(\mathbf{y}_i \mid \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]. \qquad (3.33)$$

Then take the second derivative to get the Hessian,

$$\mathcal{L}''(\boldsymbol{\beta}) = \sum_{i=1}^{n} h_i(\mathbf{y}_i \mid \boldsymbol{\beta}) \qquad (3.34)$$

where

$$h_i(\mathbf{y}_i \mid \boldsymbol{\beta}) = \left[\frac{\partial^2 \log f_i(\mathbf{y}_i \mid \boldsymbol{\beta}}{\partial \boldsymbol{\beta}^2}\right]. \qquad (3.35)$$

For the PFE model specifically, (3.34) is obtained by taking the first derivative of the first-order conditions of (3.28),

$$\mathcal{L}''(\boldsymbol{\beta}_{PFE}) = \sum_{i=1}^{n}\left(\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}'\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it} - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_{it}\mathbf{x}_{is}'\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\lambda_{is}\right) \qquad (3.36)$$

for $s \neq t$. The outer products of this on taking expectations and eliminating cross-products between units i and j for $i \neq j$ due to assumed independence across units yields the PFE model's corresponding estimator for (3.32) as

$$\mathcal{L}'(\boldsymbol{\beta}_{PFE}) = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_{it}\mathbf{x}_{is}'\left(y_{it} - \frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\right)\left(y_{is} - \frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{is}\right)'. \qquad (3.37)$$

Assuming correct model specification, i.e. that the true value $\boldsymbol{\beta}_0$ for $\boldsymbol{\beta}$ is obtained, Taylor approximation can be used to estimate the log likelihood around $\boldsymbol{\beta}_0$:

$$\mathcal{L}(\boldsymbol{\beta}) = \mathcal{L}(\boldsymbol{\beta}_0) + \mathcal{L}'(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathcal{L}''(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \dots$$

Setting $\mathcal{L}'(\boldsymbol{\beta}) = 0$, the maximum of the log likelihood function is found,

$$\mathcal{L}'(\boldsymbol{\beta}_0) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathcal{L}''(\boldsymbol{\beta}_0) = 0,$$

which yields the score

$$\mathcal{U}(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = [-\mathcal{L}''(\boldsymbol{\beta}_0)]^{-1}\mathcal{L}'(\boldsymbol{\beta}_0)'. \tag{3.38}$$

The asymptotic covariance matrix estimator of $\hat{\boldsymbol{\beta}}_{PFE}$ is then given by

$$\begin{aligned}
\mathbf{V}[\hat{\boldsymbol{\beta}}_{PFE}] &= \mathcal{U}(\boldsymbol{\beta})\mathcal{U}(\boldsymbol{\beta})' = [-\mathcal{L}''(\boldsymbol{\beta}_0)]^{-1}\mathbb{C}ov[\mathcal{L}'(\boldsymbol{\beta}_0)][-\mathcal{L}''(\boldsymbol{\beta}_0)]^{-1} \\
&= \Big[-\sum_{i=1}^{n} h_i(\mathbf{y}_i \mid \hat{\boldsymbol{\beta}})\Big]^{-1}\Big[\sum_{i=1}^{n} g_i(\mathbf{y}_i \mid \hat{\boldsymbol{\beta}})' g_i(\mathbf{y}_i \mid \hat{\boldsymbol{\beta}})\Big]\Big[-\sum_{i=1}^{n} h_i(\mathbf{y}_i \mid \hat{\boldsymbol{\beta}})\Big]^{-1} \\
&= \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}.
\end{aligned} \tag{3.39}$$

If the model is correctly specified, the expectation of the outer product of the score (3.38) is the Fisher information matrix $\mathcal{I}(\boldsymbol{\beta})$ (Zeileis, 2006), and (3.39) simplifies to its inverse,

$$\mathcal{I}^{-1}(\boldsymbol{\beta}) = -\mathbb{E}\big[h_i(\mathbf{y}_i \mid \boldsymbol{\beta})\big]^{-1}. \tag{3.40}$$

However, this simplification of $\mathbf{V}[\hat{\boldsymbol{\beta}}_{PFE}]$ will not hold under heteroscedastic or correlated errors within or across units[3], and although (3.39) is robust to overdispersion, it is not consistent for correlated errors. Similarly, if the errors in a model estimated with OLS are heteroscedastic or correlated, the simplification of $\mathbf{V}[\hat{\boldsymbol{\beta}}_{FE}]$ to (3.31) will neither hold. Thus, conditional on $\mathbf{X}$, inference of $\boldsymbol{\beta}$ requires a consistent estimator of $\mathbf{V}[\hat{\boldsymbol{\beta}}]$, which depends on the assumptions of the error covariance matrix $\boldsymbol{\Omega}$. To refer to both the OLS FE and PFE models, the general covariance matrix estimator of $\boldsymbol{\beta}$ is henceforth denoted

$$\mathbf{V}[\hat{\boldsymbol{\beta}}] = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}. \tag{3.41}$$

As Zeileis (2004) point out, most often the error covariance structure is unknown. This makes it impossible to parameterise and model it directly. As a solution,

[3] Tests of this for the empirical models are shown in Table 8-10 in Appendix

estimators for $\mathbf{\Omega}$ have been developed to plug into $\mathbf{B}$ in (3.41), which provide efficient standard errors given that the estimating functions are consistent for $\boldsymbol{\beta}$. These estimators are called *sandwich* estimators, referring to that $\mathbf{B}$ can be viewed as the meat between the two buns $\mathbf{A}^{-1}$ in (3.41), and the sandwich standard errors are in turn called *robust standard errors*, meaning that they are robust to various inference-invalidating error covariance structures. In the following paragraphs, the methods used in this thesis to obtain robust standard errors are described. The focus is on estimating the meat since all that is required is consistent pointwise estimation of the errors, and this is satisfied if the estimator of $\boldsymbol{\beta}$ is consistent (Greene, 2012).

Typical of the cross-sectional dimension is that the errors are independent but heteroscedastic, i.e $\mathbb{V}[\epsilon_t \mid \mathbf{x}_{it}] \neq \sigma_\epsilon^2$. This will be the case if ϵ_{it} are correlated with $\mathbf{x}_{it}$. Then, OLS and ML estimates of $\boldsymbol{\beta}$ are still unbiased and consistent but not efficient. The non-constant variance of the error term leads to an incorrect covariance matrix and $\mathbf{\Omega}$ is a biased estimator. Then, $\hat{\sigma}^2$ is also biased and standard errors are inconsistent and incorrect and the no interference assumption in SUTVA required for valid causal inference might not hold.

Under heteroscedasticity, $\mathbf{\Omega}$ has zero off-diagonal elements but non-constant variance terms on the diagonal, that is

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

Now, along the diagonal, $\omega_i \neq \sigma^2$ as in the ideal case of homoscedasticity. It can be shown that to consistently estimate $\mathbf{V}$, it is not necessary to estimate all $n \times (n+1)/2$ unknown elements in $\mathbf{\Omega}$, but only the $k \times (k+1)/2$ elements in

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_j', \tag{3.42}$$

that is, the OLS meat (Millo, 2014). Under heteroscedasticity and the assumption of no correlation between units i and j, the meat for OLS reduces to

$$\hat{\mathbf{\Omega}}_0 = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i'. \tag{3.43}$$

White (1980) suggested the HC0 estimator, which substitutes the unknown terms σ_i^2 along the diagonal in (3.43) with e_i^2, the squared residuals, and showed that the resulting OLS FE sandwich estimator

$$\mathbf{V}[\hat{\boldsymbol{\beta}}]_{HC0} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}\text{diag}[e_i^2]\tilde{\mathbf{X}}')(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tag{3.44}$$

is consistent for $\mathbf{V}[\boldsymbol{\beta}]$ under heteroscedasticity. For models estimated with ML such as the PFE model, the HC0 standard errors are obtained by taking the square root of the diagonal elements in (3.39). The HC0 estimator is asymptotically justified and does not assume any specific heteroscedasticity structure, as required when it is parameterised. Long and Ervin (2000) showed that the HC0 estimator needs at least 250 observations to be consistent under strong heteroscedasticity.

The HC0 estimator does however not adjust standard errors for temporal or cross-section correlation, which is often present in panel data. For this purpose there exists *clustering* sandwich estimators that adjust for the dependence by clustering the computation of the standard errors on the dimensions. Baltagi (2005) argue that the assumption of cross-sectional independence between units i and j at a given time point t and lag s can be unrealistic and lead to misleading results even under random sampling if the cross-section is a spatial variable, such as region. Hoechle (2007) mention that this *spatial dependence* may arise as a consequence of shared social norms, psychological traits, and herd behaviour, for instance with regards to purchasing. Since such latent factors are typically not possible to measure and include as covariates, the dependence enter the error terms. According to Baltagi (2005), spatial dependence is a common problem in aggregated panels with large T. Failure to account for spatial dependence will bias the standard errors and violate the no interference assumption in the SUTVA for causal inference. Mathematically, there exists spatial dependence if $\mathbb{E}[e_{it} \mid e_{js}] \neq 0$ for all $i \neq j$.

If the observations on the other hand are correlated over time within the cross-section, i.e $\mathbb{C}ov[\epsilon_t, \epsilon_s] \neq 0$ for $t \neq s$ over each i, there is *serial correlation*. Serial correlation is usually assumed in panels since the same individuals are measured over time and individual behaviour tends to be stable. If the model does not capture this temporal dependence, it also ends up in the errors. As the case with spatial dependence, serially correlated errors becomes increasingly problematic as $T \to \infty$.

When the error terms are serially correlated, they are not independent, and thus $\boldsymbol{\Omega}$ is not diagonal. Parameterising the correlation requires knowledge of its form or imposing further assumptions, and it is typically difficult to estimate $\boldsymbol{\Omega}$ directly as when correcting for heteroscedasticy (Zeileis, 2004). The solution is to estimate the meat $\mathbf{B}$ empirically by computing weighted sums of the serial correlations of $e_i\mathbf{x}_i$ (Kleiber & Zeileis, 2008). This is what the HAC estimators do, explained next.

Driscoll and Kraay (1998) proposed a heteroscedasticity, autocorrelation and spatial correlation consistent (HACSC) covariance matrix estimator by extending the Newey-West non-parametric HAC estimator (Newey & West, 1987). The Driscoll and Kraay estimator is suitable for FE panel data models in balanced and unbalanced panels with error structure assumed to be heteroscedastic, with potential contemporaneous and lagged dependence between and within cross-sections, and with weak serial correlation up to an arbitrary lag. The following section first presents the Newey-West estimator and then the extension by Driscoll and Kraay.

The Newey-West covariance matrix estimator for the FE model, which assumes cross-sectional independence, $\mathbb{E}[e_{it} \mid e_{js}] = 0$ for $j \neq i$, is given by

$$\mathbf{V}[\hat{\boldsymbol{\beta}}]_{HAC} = \hat{\mathbf{A}}^{-1}\hat{\boldsymbol{S}}_T\hat{\mathbf{A}}^{-1} \tag{3.45}$$

where $\hat{\mathbf{A}}^{-1}$ is the bread and $\hat{\boldsymbol{S}}_T$ is the empirically computed Newey-West non-parametric estimator of the meat calculated as

$$\hat{\boldsymbol{S}}_T = \hat{\boldsymbol{\Omega}}_0^{(i)} + \sum_{l=1}^{m} w(l,m)[\hat{\boldsymbol{\Omega}}_l^{(i)} + \hat{\boldsymbol{\Omega}}_l'^{(i)}] \quad \text{where} \quad \hat{\boldsymbol{\Omega}}_l^{(i)} = \frac{1}{T}\sum_{t=l+1}^{T} \hat{v}_{it}\hat{v}_{it-l}' \tag{3.46}$$

with $\hat{v}_{it} = \tilde{\mathbf{x}}_{it}e_{it}$. This expression shows that the Newey-West estimator takes the White estimator $\hat{\boldsymbol{\Omega}}_0$ given in (3.43) and adds a sum of covariances between the different residuals and a Bartlett weight function

$$w(l,m) = 1 - \frac{l}{m+1} \tag{3.47}$$

with linearly decaying weights, where the argument m is the length of lag $l = |i-j|$ of the serial correlation among the residuals. This weight function guarantees that $\hat{\boldsymbol{S}}_T$ is positive semi-definite in every sample and that the sample serial correlation function is smoothed. That way the higher order lags are given less weight in the correction for serial correlation, since it is assumed that the temporal dependence decreases with increasing lags. If the computation of the standard errors are clustered on the cross-section, the HAC standard errors are robust vs. temporal heteroscedasticity and cross-sectional correlation. Clustering on the time dimension provides robustness against spatial heteroscedasticity and serial correlation.

Driscoll and Kraay adjusted the Newey-West meat estimator $\hat{\boldsymbol{S}}_T$ by taking the HAC's standard errors of the cross-section averages instead of the cross-section averages of the HAC's standard errors (Vogelsang, 2012). It thereby averages the product $\tilde{\mathbf{x}}_t\mathbf{e}_t$ over the dependent clusters and may viewed as a time-clustered version of Newey-West standard errors. The Driscoll and Kraay meat estimator is given by

$$\hat{\hat{\boldsymbol{S}}}_T = \hat{\hat{\boldsymbol{\Omega}}}_0 + \sum_{l=1}^{m} w(it, js)[\hat{\hat{\boldsymbol{\Omega}}}_l + \hat{\hat{\boldsymbol{\Omega}}}_l'], \text{where } \hat{\hat{\boldsymbol{\Omega}}}_l = \frac{1}{T} \sum_{t=l+1}^{T} \hat{\hat{v}}_t \hat{\hat{v}}_{t-l}'. \tag{3.48}$$

Here, $\hat{\hat{v}}_t = \sum_{i=1}^{n} \hat{v}_{it}$ is obtained by stacking the $(k \times 1)$ vectors $[\hat{v}_{1t}, \ldots, \hat{v}_{nt}]$ into a $(nk \times 1)$ vector $\hat{\mathbf{v}}_t$ with transpose $\hat{\mathbf{v}}_t' = [\hat{v}_{1t}', \ldots, \hat{v}_{1n}']$ and then defining $\hat{\hat{v}}_t = \sum_{i=1}^{n} \hat{v}_{it}$. That is, $\hat{\hat{v}}_t$ is the cross-section average of $\hat{v}_{it}$ times n.

Note that in equation (3.48) of the Driscoll and Kraay meat estimator, the weight function is

$$w(it, js) = 1 - \frac{d(it, js)}{m+1}, \tag{3.49}$$

so the summation is over i, j, s and t where i and j refer to cross-sectional units, t and s are time points, and $d(it, js) = |t - s|$ if $[t - s] \leq m$ and 0 otherwise (Cameron & Miller, 2015). This extension make the meat estimator spatial correlation consistent. Thus, the full HACSC covariance matrix estimator is given by

$$\mathbf{V}[\hat{\boldsymbol{\beta}}]_{HACSC} = \hat{\mathbf{A}}^{-1} \hat{\hat{\boldsymbol{S}}}_T \hat{\mathbf{A}}^{-1} \tag{3.50}$$

Taking the square root of the diagonal elements in $\hat{\hat{\boldsymbol{S}}}_T$ yields the Driscoll and Kraay standard errors for $\hat{\boldsymbol{\beta}}$.

As a consequence of using the HAC estimator of the cross-section averages, HACSC standard errors are consistent regardless of the cross-sectional dimension of the panel. Vogelsang (2012) show that for FE models in panels with large T and fixed n, the estimator only requires weak exogeneity, $\mathbb{E}[\mathbf{x}_{it} \mid \epsilon_{it}] = 0$ for $s = t$, $t - 1$, in the cross-section and time dimension. It does not assume any particular distribution of the errors and is consistent for weak serial correlation and unknown spatial correlation at the same point in time and at different lags without a measure of the spatial distance itself, given that the time series is stationary. The spatial correlation is however not adjusted according to physical distance. Since the HACSC standard errors are computed on the empirical x_{it}, e_{it}, the PFE model's standard errors are also adjusted to the empirical dispersion. Thereby the PFE model's equidispersion property is not a limitation. Regarding OLS, Hoechle (2007) finds that the HACSC standard errors are better calibrated than the usual non-robust and Newey-West HAC standard errors under cross-sectional dependence in a FE model.

4 Data

The dataset consists of grocery purchases in Sweden from week 36 in 2015 to week 35 in 2017 obtained from GfK[4] Consumer Scan; a nationally representative panel of 3000 households that continosly record their grocery purchases using an app or in-home scanner in exchange for rewards (GfK, 2018). The following section describes the processing of the data and the variables of interest. The sampling methodology for Consumer Scan can be found in section A.2 in the appendix.

Prior to analysis, GfK have anonymised the household level data per the General Data Protection Regulation and aggregated it at the weekly, regional and retailer level and multiplied it by a factor to be numerically representative for the regional populations. Thus, each observation is a weekly estimate for the respective region's population level purchases at a retailer. A benefit of the aggregation is that the panel is balanced with $T = 105$ weeks for each region-retailer combination, but the number of households included in a given observation vary and is unknown. According to Cooper and Nakanishi (1988), promotions are usually managed on a weekly basis and thus the weekly aggregation may not be a significant limitation.

The two variables that define the cross-section are retailer and geographic region. The seven regions are contiguous and cover the whole of Sweden. The regions are listed in Table 6 in Appendix. The retailers are Coop Extra, ICA Maxi, ICA Supermarket and Willys, which belong to the Coop, ICA and Axfood group respectively. In 2016, Coop's market share accounted for 19 percent of the Swedish grocery market. Coop Extra are Coop's profile stores in small cities offering simple and fast shopping at low prices (Coop, 2018). ICA is Sweden's largest retailer group with a market share of 50.8 percent in 2016 (HUI Research, 2017). ICA Supermarket are ICA's mid-size stores meant for quick daily grocery purchases (ICA, 2018b). ICA Maxi are their largest supermarkets covering the full product range at lower prices (ICA, 2018a). Willys is part of the Axfood group and is positioned as their low-price chain (Axfood, 2018). In 2016, Axfood's sales were 16.4 percent of the market while Willys alone had 10.6 percent (HUI Research, 2017). Together ICA, Coop and Axfood have 90 percent of the market share combined (Gullstrand, Olofsdotter, & Karantininis, 2011). This strongly indicates an oligopoly market structure with price competition and strategic interaction as consequences. It is thereby expected that sales promotion increase store visits.

The data is also aggregated at the national level for the region and retailer

[4]GfK (short for "Society for Consumer Research" in German) is Germany's largest market research institute and the fourth largest market research organisation globally, specialising in providing data-driven insights and decision support within marketing.

variables[5]. The national level for the region variable is the sum of the regional data for each retailer, whereas the retailer national level is the sum of the data for all grocery retailers in Sweden, not only those included in this dataset.

The variables of interest are:

> *Shopping trips*: the number of store visits with a purchase in week t.
>
> *Coffee shopping trips*: the number of store visits with a purchase of ground coffee in week t.
>
> *Promotion trips share*: the share of shopping trips in week t in which a brand was purchased on promotion.

These variables are chosen because they have been used in GfK's previous analysis of promotion effects. In the dataset, promotion trips share is included for the 11 most frequently promoted product categories. For each product category, it is available for the four most popular brands plus the aggregate of the remaining brands, from here on called Other brands. Data on the number of shopping trips and the number of promotional shopping trips for each brand are also available.

The ground coffee category is analysed in this thesis. It consists of the four most popular domestic brands Gevalia, Löfbergs, Zoegas, Classic, and the aggregate of the remaining brands. These four brands' market share is around 85 percent in Sweden (Durevall, 2017), indicating an oligopoly within the product category. The oligopoly structure at the retailer level implies that the retailers will compete for customers with sales promotion, potentially causing store-switching, and the oligopoly structure within the product category facilitates price competition between the brands' products within stores, potentially causing brand-switching.

Coffee is well suited for analysing promotion effects. Coffee consumption in Sweden is high and widespread in the population, which means that there are less weeks with no purchases in the dataset than for some other categories. Because of the widespread use, the data may be representative of the population even though coffee consumption is not a stratification variable in the sampling of households. Additionally, the Swedish coffee market is highly competitive with frequent promotions (Durevall, 2017). This means that the dataset contain a large number of treated outcomes so the effect of promotions can be identified with high precision relative less promoted categories. Finally, coffee is believed to be a so called *loss leader* (Persson, 1995), that is, a product that attracts consumers to the store, causing store-switching. However, the coffee price elasticity of demand is low. A commonly used figure is 0.25, meaning that a 10 percent decrease in price corresponds to a 2.5

[5]Summary statistics on the region-retailer aggregate level is provided in A.9 in Appendix.

percent increase in quantity demanded. The response in shopping trips from coffee promotion may thereby not be as strong as for more price elastic groceries.

The shopping trips variables are in thousands and are thus non-negative continuous variables although they are discrete on their original scale. The promotion trips share variables are in percentages and thereby continuous on the interval 0 to 100. The promotion trips share may be viewed as continuous treatment variable that reflect the share of households that received the treatment of being able to purchase on promotion. Promotion trips share is calculated for each brand by dividing the number of promotional coffee shopping by the total number of coffee shopping trips. In 37 percent of the weeks, Gevalia was not purchased. For Löfbergs, the corresponding share is 53 percent, for Zoegas, 38 percent, for Classic it is 61 percent, and for the remaining ground coffee brands it is 38 percent of the weeks. As a consequence, the promotion trips share is undefined for these weeks since it is not possible to divide by zero. The promotion trips share for these observations were recoded with a zero to not omit them from the analysis. The reasoning of the imputation is that the share of purchases made on promotion is zero if there are no purchases. On the regional level, 45 percent of the weeks across the brands have no promotional purchases, and thereby a promotion trips share of zero. The promotion trips share and shopping trip variables have no missing values.

Regionally aggregated descriptive statistics for the retailers are provided in Table 1. The table show that the national average in weekly shopping trips differ between retailers and the mean promotion trips share of the brands is similar. The median promotion trips share is zero or a small in a couple of entries. This indicates that the retailer have few weeks with promotion, which increases the uncertainty of the estimates of the promotion effect. The retailers' mean and variance in shopping trips and coffee shopping trips are not equal, so the equidispersion property of the Poisson distribution is violated and robust standard errors are needed.

Figure 1 show the number of shopping trips to retailer for each region over the time period. Apart from Coop's decreasing shopping trips in Stockholm, most region's have no apparent trend. Figure 2 show the corresponding information for coffee shopping trips. Likewise, most coffee shopping trips time series do not have any trend, apart from Coop in Stockholm. The peaks in the time series appear to follow the same pattern across regions for some retailers, most notably for Willys in Figure 2. This indicates cross-sectional dependence. Figure 3 displays how the distribution of shopping trips and coffee shopping trips vary across regions. The figure shows that zeros are abundant in some regions for coffee shopping trips.

Table 1: Regionally aggregated descriptive statistics

Variable	Mean	St. Dev.	Min	Median	Max
Coop					
Shopping trips	1,258.81	123.88	958.63	1,269.92	1,500.76
Coffee shopping trips	60.95	26.01	20.96	56.48	167.63
Promotion trips share					
Gevalia	27.23	31.02	0.00	14.54	100.00
Löfbergs	33.04	36.12	0.00	22.43	100.00
Classic	35.33	33.07	0.00	32.05	100.00
Löfbergs	46.84	46.48	0.00	38.96	100.00
Other brands	30.44	26.36	0.00	26.78	100.00
ICA Maxi					
Shopping trips	1,478.41	100.01	1,218.55	1,492.48	1,686.29
Coffee shopping trips	151.73	61.77	38.10	149.28	336.90
Promotion trips share					
Gevalia	32.75	29.80	0.00	24.46	96.40
Löfbergs	32.83	38.89	0.00	9.58	100.00
Zoegas	25.29	31.28	0.00	14.64	97.94
Classic	30.68	39.90	0.00	0.00	100.00
Other brands	43.28	26.13	0.00	45.28	96.55
ICA Supermarket					
Shopping trips	1,613.46	111.56	1,289.84	1,634.62	1,878.93
Coffee shopping trips	102.43	63.02	29.55	82.51	343.80
Promotion trips share					
Gevalia	36.71	29.22	0.00	31.42	100.00
Löfbergs	24.27	34.00	0.00	0.00	100.00
Zoegas	43.23	31.40	0.00	44.14	100.00
Classic	29.83	39.20	0.00	0.00	100.00
Other brands	39.07	28.89	0.00	38.69	100.00
Willys					
Shopping trips	1,517.51	107.01	1,249.91	1,521.12	1,992.17
Coffee shopping trips	122.18	121.85	25.11	69.72	553.88
Promotion trips share					
Gevalia	25.30	31.60	0.00	11.58	100.00
Löfbergs	17.79	31.29	0.00	0.00	100.00
Zoegas	28.05	34.50	0.00	12.50	100.00
Classic	22.61	36.66	0.00	0.00	100.00
Other brands	24.54	30.41	0.00	12.48	100.00

Panel dimensions: balanced with n=7, T=105. Number of observation per retailer: nT = 735. The table show regionally aggregated descriptive statistics for the retailers. Shopping trips and coffee shopping trips are in 1000's. Promotion trips shares are in percentages.

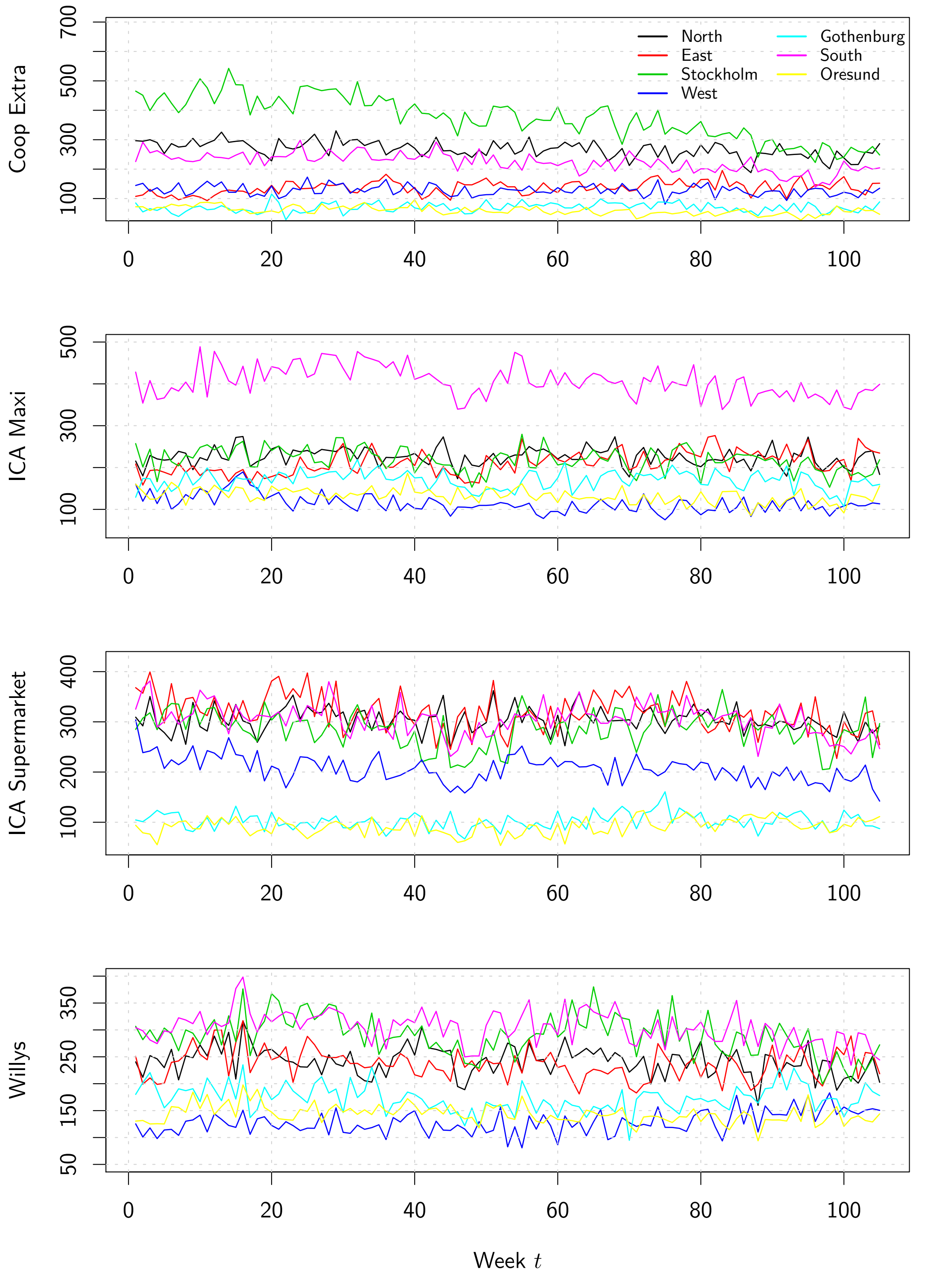


Figure 1: Shopping trips (in 1000) per retailer and region. Week t is relative to week 36 in 2015 and ranges to week 35 in 2017. The regions are contiguous and cover Sweden. Each observation is an estimate for the regional population.

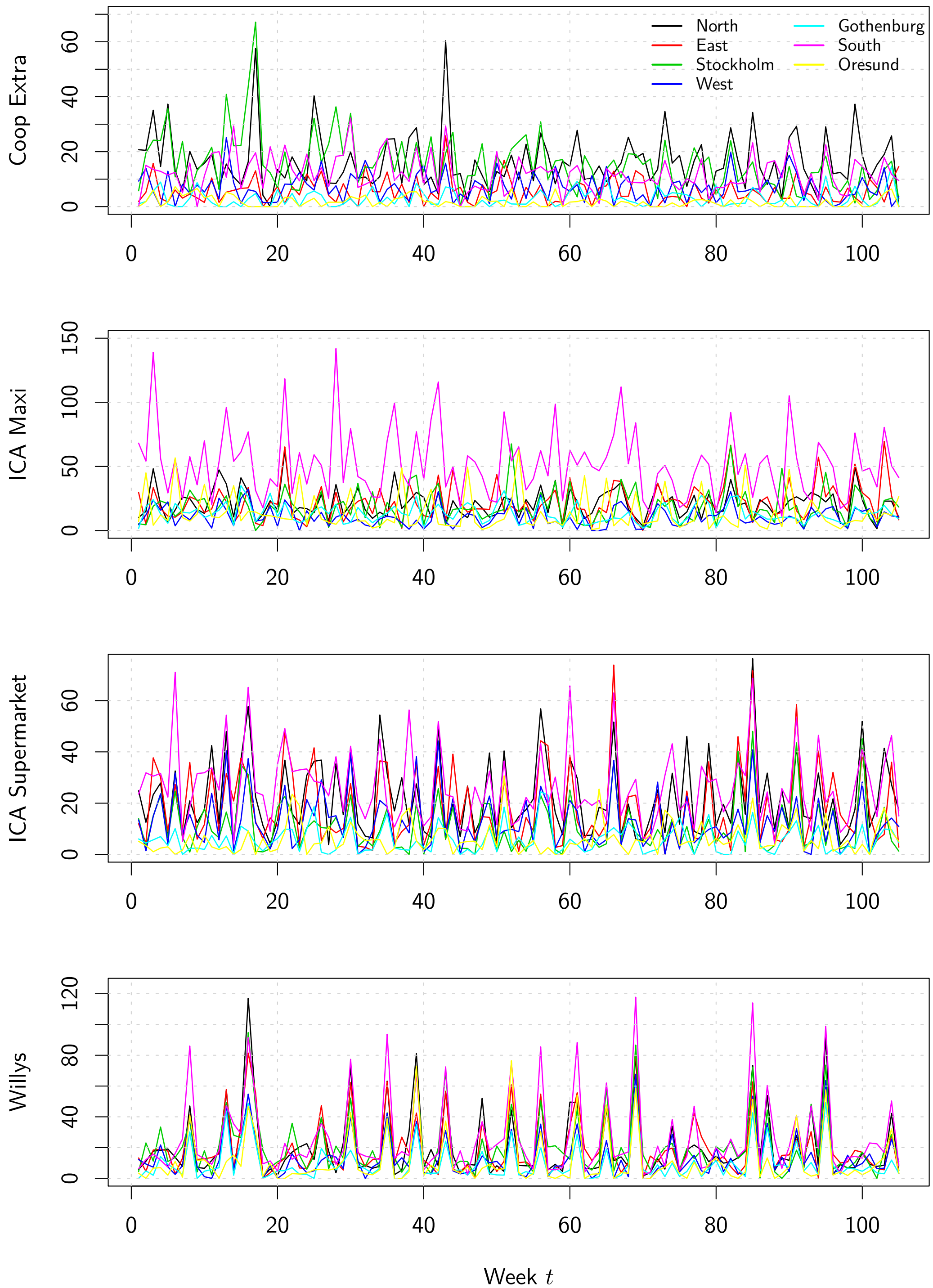


Figure 2: Coffee shopping trips (in 1000) per retailer and region. Week t is relative to week 36 in 2015 and ranges to week 35 in 2017. The regions are contiguous and cover Sweden. Each observation is an estimate for the regional population.

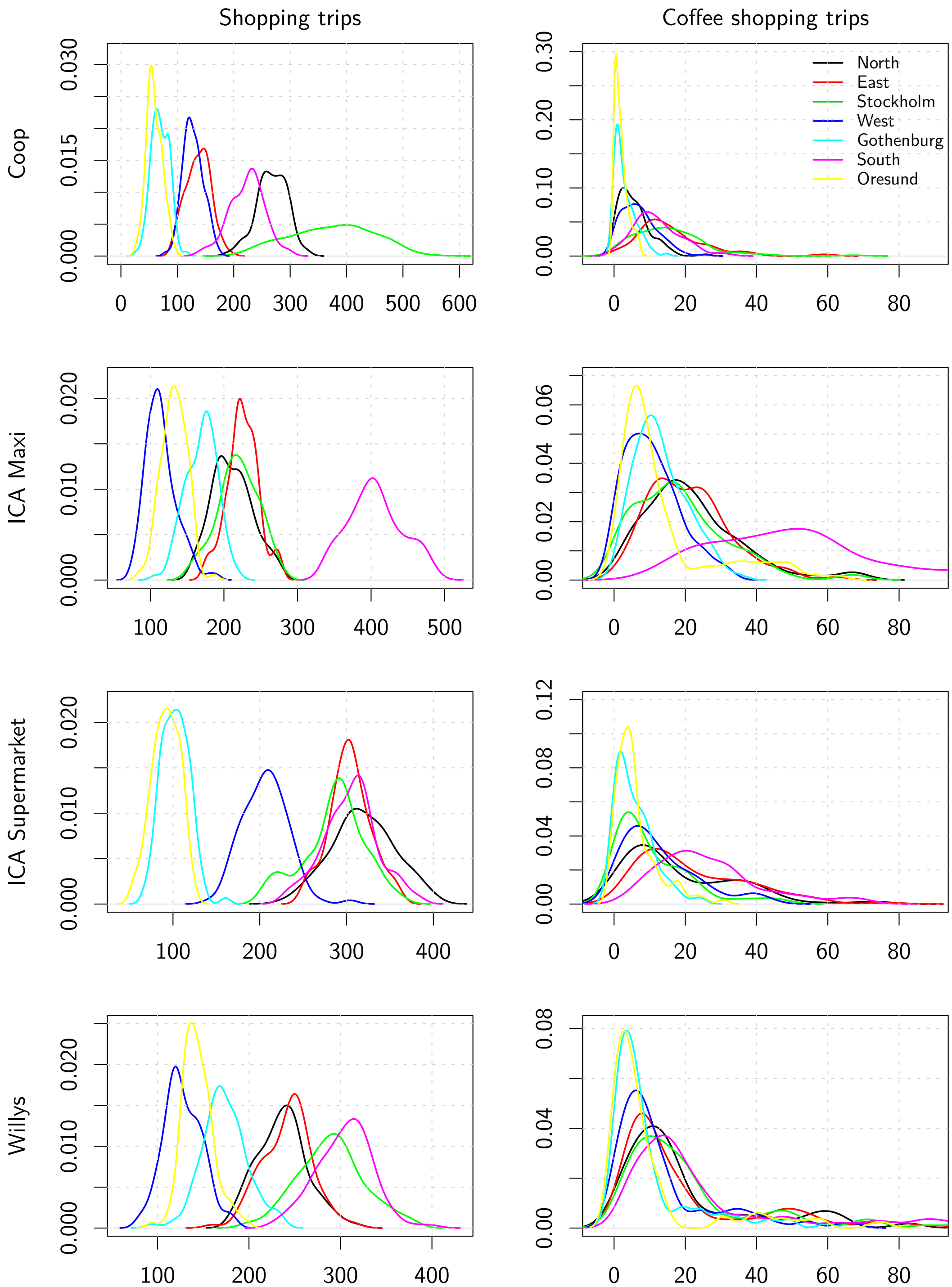


Figure 3: Kernel density estimates of shopping trips and coffee shopping trips per retailer and region. The x-axis is the dependent variable (in 1000) and the y-axis is the density.

5 Method

This section explains the method used to quantify the promotion impact. First, the regression models are presented. This is followed with the procedure for estimating the counterfactual potential outcomes of no promotion for weeks that had promotion. Finally, the estimator used to calculate the number of shopping trips due to promotion is shown.

5.1 Model Specification

Two fixed effects (FE) regression models are estimated for each retailer. The first is a log-linear FE model estimated with OLS currently used by GfK. Since n is small, the region fixed effects are estimated with dummy variables. The model is given by

$$\log(y_{it}) = \sum_{i=1}^{7} \alpha_i d_i + \sum_{j=1}^{4} \beta_j x_{ijt} + \delta \log(z_t) + \epsilon_{it} \tag{5.1}$$

in which $i = 1, \ldots, 7$, $t = 1, \ldots, 105$, y_{it} are the number of shopping trips in region i at week t, the dummy d_i takes the value 1 if an observation is for region i, and 0 otherwise, α_i captures time-invariant unobserved effects for region i, and ϵ_{it} is the error term following some distribution. The overall intercept is suppressed so all region fixed effects are estimated. Regressors x_j, $j = 1, \ldots, 4$, are the time-variant promotions trips shares that reflect the share of purchases made on promotion for the ground coffee brands Gevalia, Löfbergs, Zoegas, Classic. GfK do not include the remaining coffee brands captured by the Other brands variable in this model. The variable z_t is a normalised time index calculated as $\mathbf{y}_t/\bar{y}_t$ for each t on the national industry level of $\mathbf{y}_t$. It thereby captures the overall nation-wide industry trend, seasonality and shocks in shopping trips among all retailers that the households purchased at over the time period, including other retailers than the four included in this dataset. Estimating the time index only consumes one degree of freedom instead of $T + 1$ degrees of freedom as time dummies plus a time trend would. The time index is log-transformed to bring it on the same scale as the dependent variable.

In the OLS FE model, the dependent variable is log-transformed to estimate the multiplicative effects with OLS. A problem with the log-transformation is that the coffee shopping trips variable contain zeros, and the log of zero is undefined. GfK have previously removed these observations to fit the model, but then information is lost since zero is the mode value of the dependent variable in some regions. Additionally, the time effects may not be correctly estimated since observations are left out from a possible trend. To fit the model to all observations, coffee shopping trips observations with a value of zero are imputed with half of the smallest positive value before taking the logarithm. Although this imputation is preferable to omitting the observations, it suffers from that it not apparent how $\mathbb{E}[y \mid \mathbf{x}, \alpha, \delta]$ should be recovered on the non-log scale given the transformation. For this the exponent is taken of the fitted values. To adjust for the bias induced by the imputation, the dependent variable with the imputed values are differenced to the fitted values from the OLS FE regression when estimating the treatment effect.

The second model is a count data PFE variant estimated as a GLM with log-link and dummy variables for the region fixed effects, specified as

$$\log(\mu_{it}) = \log(\mathbb{E}[y_{it}]) = \sum_{i=1}^{7} \alpha_i d_i + \sum_{j=1}^{5} \beta_j x_{ijt} + \sum_{i=1}^{7} \delta_i \alpha_i d_i \log(z_t). \tag{5.2}$$

Since the dependent variable y_{it} is in thousand with decimals and thereby continuous, y_{it} are multiplied with 1000 to obtain y_{it} as integers prior to fitting the model. The fitted values are divided with 1000 after taking the exponent to obtain them on their original scale.

In the PFE model, the promotion effects of Other brands are estimated, resulting in $k = 5$. This makes the estimated promotion effects of the main brands Gevalia, Löfbergs, Zoegas, and Classic robust against brand-switching behaviour. To see how, consider a premium coffee brand that belongs to Other brands. The premium brand is more expensive than the main brands at their respective regular prices. If the promotion price of the premium brand is less or close to that of the main brands, customers might brand-switch to the premium brand. The purchases of the premium brand will appear in the dependent variable regardless of whether the promotion effect of Other brands is estimated, but if the Other brands promotion effect is not estimated, the purchase and subsequent increase in the dependent variable will be incorrectly attributed to promotion on the main brands if any of them were promoted that week. Thus, the OLS FE model's estimated promotion effects of the main brands are likely biased upwards by not including Other brands.

Another feature of the PFE model is that its time index z_t is retailer-specific, meaning that for each retailer's model, it is calculated on the retailer's national level data instead of on the industry national level. The justification is that while seasonality can be common among all retailers in the industry, their trend in shopping trips and the shocks they are subject to cannot be be assumed to be equal. The trend may vary among retailers due to unobserved factors such as changes in customer bases, competition, or advertising during the time period. Shocks that affect one retailer may not affect the others, for instance product launches and firm entry and exit into a market rarely affect all firms equally. The retailer-specific time index is interacted with the region fixed effects α_i to estimate a different coefficient δ_i for the time index for each region. In other words, the interaction allows the time effects in each region to relate differently to the retailer's national level time effects, since the trend and shocks may also vary across regions. If a retailer opens a supermarket it only affects the number of shopping trips of the retailers in that region. As an example, Figure 1 shows that Coop had a decreasing time trend in Stockholm but the other regions did not, and that the frequency and intensity of peaks in shopping trips can vary across regions. The alternative method to model regionally heterogenous time effects would be to interact region dummies with week dummies, but that is not possible to estimate since it would consume nT degrees of freedom, i.e. as many as the number of observations, apart from those required for the estimation of the other parameters.

It should be noted that the modelling approaches of the log-linear OLS FE and the PFE model differ. In a log-linear model estimated via OLS, the left hand side of the model equation is $\mathbb{E}[\log(\mathbf{y}) \mid \mathbf{x}]$. In the PFE model, it is $\log(\mathbb{E}[\mathbf{y} \mid \mathbf{x}])$. Now, $\mathbb{E}[\log(\mathbf{y}) \mid \mathbf{x}] \neq \log(\mathbb{E}[\mathbf{y} \mid \mathbf{x}])$, unless $\mathbf{y}$ is fully determined by $\mathbf{x}$, which is not the case.

The models are fitted separately for each retailer using both shopping trips and coffee shopping trips as dependent variable to obtain their respective promotion effects. The estimated models are identical for each retailer apart from the time index which changes with the retailer and dependent variable. Driscoll and Kraay HACSC robust standard errors are estimated for all parameters for valid inference under the empirical cross-sectional and temporal correlation. This adjusts the PFE model's restricted variance according to the empirical dispersion. The lag length up to which the residuals may be autocorrelated is set to the heuristic $m = \text{floor}[4(T/100)^{2/9}]$ taken from Newey-West's plug-in procedure (Newey & West, 1994) for finding the optimum lag length. Constraining the lag length imposes that the residuals follow a MA(q) process, but this is not necessarily a problem since MA(q) processes can approximate AR(p) processes. The robust standard errors are clustered on the region variable to adjust for the regional dependence between observations captured by the errors and the White HC0 weighting scheme is applied to adjust for time-wise heteroscedasticity. HC0 requires at least 250 observations and HACSC requires large T for consistency. Since $nT = 735$ and $T = 105$, the robust standard errors are assumed to be consistent. GfK have previously not considered any correction for dependence among the errors in the OLS FE model, and thus inference would be incorrect with these data.

5.2 Estimation of Counterfactuals

This thesis' impact evaluation problem differ from the typical in the literature. With these data, standard matching methods cannot be used since there is no pre-treatment period with pre-treatment covariates that can be used to match observations to obtain the counterfactual outcomes. The literature mostly considers a single binary treatment that start at the same time for all treated units. Here, there are multiple treatments starting and ending at different time points that are active for a varying number of time periods across units. Since the treatment variable $\mathbf{x}$ is continuous, the treatment effect τ_{it} is not the same for all treated units and time periods as it is when the treatment variable is binary. Further, there likely exist non-observed confounders not captured by the time index variable that invalidates taking the difference between outcomes for different units and weeks.

As a solution, each observation is used as its own match for its counterfactual outcome. This solves the problem of unobserved time-varying confounders since they are per definition equal for the same observation in each week and region. Remaining potential time-invariant confounders are eliminated by the fixed effects in the models. Consequently, the unconfoundedness assumption is more likely to hold and the overlap assumption holds per construction, since the distribution of covariates is identical across treated and control units if the same observations are used for both. Thereby, the treatment assignment mechanism is ignorable and the regression models can be used to obtain the counterfactuals. Separate counterfactuals are estimated in each instance when one of brands was not promoted while the other brands were, and another counterfactual is estimated for the alternative situation if none of the brands would have been promoted. The counterfactual estimation procedure is as follows:

> *Step 1.* Fit the regression to obtain the predicted outcomes $\hat{y}_{it}^{obs}$, the estimated coefficients $\hat{\beta}_j$ for brand j, and the coefficients for the remaining covariates, given the observed promotion data x_{ijt}.
>
> *Step 2.* If $\hat{\beta}_j \neq 0$, meaning that brand j's promotion had an effect, replace all promotion trips share data x_{ijt} with zeros while keeping the remaining brands' promotion trips share data as observed. Estimate the predicted outcomes given these data using the coefficients obtained in Step 1. This drops β_j from the regression and estimates $\hat{y}_{it}^{c,j}(0)$, the weekly regional counterfactual number of shopping trips if the share of purchases of brand j made on promotion had been zero, which would be the case it was not promoted over the time period. Redo this step

for each $\hat{\beta}_j \neq 0$, $j = 1, \ldots, k$. The results are counterfactual outcomes $\hat{y}_{it}^{c,j}(0)$ for each brand j.

Step 3. Replace all promotion trips share data x_{ijt} with zeros for all brands whose $\hat{\beta}_j \neq 0$. Then estimate the predicted outcomes given these data using the estimated coefficients from Step 1. This drops all $\beta_j \neq 0$ from the regression and yields $\hat{y}_{it}^{c}(0)$, the estimated counterfactual number of shopping trips if no brand had been promoted.

The procedure is expressed algorithmically in Algorithm 1 with an arbitrary regression model. In contrast to standard methods, it can estimate counterfactuals of several continuous treatments with multiple treatment periods that vary in the cross-section and does not require pre-treatment matching covariates. The procedure is run for all retailers with both the OLS FE and PFE models with shopping trips and coffee shopping trips as dependent variables to obtain each model's counterfactuals for each retailer.

Algorithm 1: Counterfactual estimation procedure

Input : Regression function and panel data on observed outcomes, observed treatments, and confounding covariates.

Output: Estimated outcomes given the observed treatments and counterfactual outcomes of no treatment for each treatment and counterfactual outcomes of no treatment on any treatment.

Obtain $(\hat{y}_{it}^{obs}, \hat{\alpha}_i, \hat{\boldsymbol{\beta}}, \hat{\delta})$ by fitting $y_{it}^{obs} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \delta \log(z_t) + \epsilon_{it}$
foreach $x_{ijt} \in \mathbf{x}_{it}, j = 1, \ldots, k$ **do**
    Set $x_{ijt}^c = x_{ijt}$
    if $\hat{\beta}_j \in \hat{\boldsymbol{\beta}} \neq 0$ **then**
        Set $x_{ijt}^c \leftarrow 0, \forall i, t, \; i = 1, \ldots, n, \; t = 1, \ldots, T$
    end
    $\hat{y}_{it}^{c,j}(0) = \hat{\alpha}_i + \mathbf{x}_{it}'^{c}\hat{\boldsymbol{\beta}} + \hat{\delta}\log(z_t)$
end
foreach $x_{ijt} \in \mathbf{x}_{it}, j = 1, \ldots, k$ **do**
    Set $x_{ijt}^c = x_{ijt}$
    if $\hat{\beta}_j \in \hat{\boldsymbol{\beta}} \neq 0$ **then**
        $x_{ijt}^c \leftarrow 0, \forall i, t, \; i = 1, \ldots, n, \; t = 1, \ldots, T$
    end
end
$\hat{y}_{it}^{c}(0) = \hat{\alpha}_i + \mathbf{x}_{it}'^{c}\hat{\boldsymbol{\beta}} + \hat{\delta}\log(z_t)$

5.3 Estimation of Promotion Impact

Having obtained the counterfactuals, it is possible to estimate the unobservable promotion impact. Retailers and producers of consumer packaged goods are interested in the cumulative impact over the promotion period, not the average regional impact given by the ATT. Thereby, the ATT estimator is adapted and the sample *cumulative treatment effect for the treated* (CTT) for each retailer is calculated as the sum across weeks and regions of the difference between the estimated outcomes given the observed data and the estimated counterfactual outcomes under no promotion. This is a measure of the total impact of promotion over all regions and weeks that had promotion in comparison to if there would have been no promotion.

The estimated joint impact of promotion on all brands is calculated as

$$\hat{\tau}_{CTT} = \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{y}_{it}^{obs} - \hat{y}_{it}^{c}(0)\right) \times \left(\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}^{obs} \Big/ \sum_{i=1}^{n}\sum_{t=1}^{T}\hat{y}_{it}^{obs}\right) \tag{5.3}$$

while the estimated impact of promotion of brand j is calculated analogously with

$$\hat{\tau}_{CTT}^{j} = \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{y}_{it}^{obs} - \hat{y}_{it}^{c,j}(0)\right) \times \left(\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}^{obs} \Big/ \sum_{i=1}^{n}\sum_{t=1}^{T}\hat{y}_{it}^{obs}\right) \tag{5.4}$$

for $n = 7$, $T = 105$, $j = 1, \ldots, 4$ for OLS FE, and $j = 1, \ldots, 5$ for the PFE model.

Note that $\hat{y}_{it}^{obs}$, $\hat{y}_{it}^{c}(0)$ and $\hat{y}_{it}^{c,j}(0)$ are simply the left hand sides of the regression functions. The replacement of x_{ijt} with zeros drops $\hat{\beta}_j$ from the right hand side of the regression function of $\hat{y}_{it}^{c,j}(0)$ and drops all $\hat{\beta}_j$, $j = 1, \ldots, k$ from $\hat{y}_{it}^{c}(0)$, since a parameter cannot be estimated from only zeros. As the specification of the regressions used to obtain the counterfactuals is otherwise identical to that of $\hat{y}_{it}^{obs}$, the only difference between the terms in the parenthesis in (5.4) is that the right hand side of the regression function of $\hat{y}_{it}^{obs}$ includes $\hat{\beta}_j$ and x_{ijt} for brand j. For equation (5.3), the right hand side of the regression function of $\hat{y}_{it}^{obs}$ includes $\hat{\beta}_j$ and x_{ijt} for all k brands. As such, the estimator takes the difference between the estimated outcomes with promotion and the estimated outcomes without promotion.

In comparison to the RA estimator of the ATT given by equation (3.10) on page 11, the estimator takes the sums instead of the average to obtain the cumulative effect. The spatio-temporal dependence in the errors causes $\sum^{n}\sum^{T}\hat{y}_{it}^{obs}$ to not equal $\sum^{n}\sum^{T} y_{it}^{obs}$. The adjustment factor $\sum^{n}\sum^{T} y_{it}^{obs} / \sum^{n}\sum^{T}\hat{y}_{it}^{obs}$ corrects for this so the impact is estimated for actual observed outcomes y_{it}^{obs} rather than $\hat{y}_{it}^{obs}$. The difference $|\sum^{n}\sum^{T}(y_{it}^{obs} - \hat{y}_{it}^{obs})|$ is larger for the OLS FE model than the PFE model, indicating worse fit and consequently greater need for the adjustment.

6 Empirical Results

Table 2 and Table 3 show the regression results. In both models, the β parameters measure the within-region average percentage change in shopping trips in a given week associated with a one percent increase in the number of households that purchased the brand on promotion. Although the $\hat{\beta}$ values are small, the impact over the time period is substantial since the dependent variable is in thousands, the promotion trips share ranges from 0 to 100, and the effects adds up over the 7 regions each with 105 weeks of data.

A central finding is that few β parameters are significant in the models with general shopping trips as the dependent variable, see Table 2. Notably, the OLS FE model estimates no statistically significant effect of promotion for Coop, while in the PFE model, promotion on Classic and Other brands has significant effects. For ICA Maxi, Gevalia has significant effects in the OLS FE model while Classic is the brand with significant effects in the PFE model. The OLS FE model also estimates that Classic has a significant effect for ICA Supermarket and that Zoegas is the single brand with significant effect for Willys. The PFE model estimates no significant effects for any coffee brand for ICA Supermarket and Willys.

The interpretation of the δ parameters for the time index is not important. What is interesting is that the retailer-region specific time indices captured by the interaction terms in the PFE model are significant and that the δ_i coefficients are not equal, since it indicates that the time effects differ between regions. The industry time index used in the OLS FE model does not account for this and may thereby not sufficiently capture the time effects on the dependent variable for each region, leading to spurious results with too high estimates of the β parameters.

Table 3 show the regression results with coffee shopping trips as the dependent variable. In comparison to the models with shopping trips as dependent variable, a higher number of the brands have statistically significant and stronger effects, indicated by the higher $\hat{\beta}$ values. For Coop and ICA Maxi, all brands have significant effects in the OLS FE and PFE model. For Willys, the OLS FE model estimates significant effects for all brands, but only Zoegas have significant effects in the PFE model. A common pattern across retailers is that the OLS FE model estimates larger effects. A reason may be its brand-switching bias and that its industry time index does not sufficiently account for time effects, which are then incorrectly attributed to promotions.

Two measures of model fit are provided. The root mean square error (RMSE) measures the sample standard deviation of the difference between the observed values on the outcome variable and the predicted values, and is calculated as

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_t^{obs} - \hat{\mathbf{y}}_t)^2} \tag{6.1}$$

where $\mathbf{y}_t$ are each retailer's stacked region time series. The RMSE is computed on the estimation data and thereby reflects the size of the residuals, where a smaller value indicate better fit. Since the residuals are squared prior to averaging, RMSE penalises large deviations. To illustrate why this is useful, consider a retailer that launches a campaign with coordinated sales promotions across its stores for a given time period to increase purchases. The retailer advertises this campaign so consumers are aware of the reduced prices. The promotion campaign causes a peak in store visits and a higher than usual share of promotional purchases. Other times of the year, the retailer offers non-planned and non-advertised sales promotion to get rid of excess inventory, as argued by Blattberg et al. (1981). This will reasonably not have as large effect on store visits. Since the retailer is primarily interested in the promotion impact during the campaign whose promotions caused the strongest effect, the predictive error for the observations that span the planned promotions with large effect should be given higher weight when evaluating model fit. The PFE model has smallest RMSE with both dependent variables for all retailers, see Table 2 and 3.

The bias statistic is also provided, computed as the average amount (in 1000's) by which the observed number of shopping trips is greater than the predicted,

$$\frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_t^{obs} - \hat{\mathbf{y}}_t). \tag{6.2}$$

Table 2 and 3 show that the PFE model yield less biased estimates for all retailers and both dependent variables. The pseudo R^2 is obtained by fitting the models using the **R** packages **plm** for the OLS FE model **pglm** for the PFE model. The pseudo R^2 statistic can only be used for within model assessments of fit.

Additional model diagnostics are performed but not shown to save space. There are no signs of multicollinearity among covariates and QQ-plots of the residuals indicate superior fit of the PFE model. The Vuong test for non-nested models is used to test for misspecification and which model has superior fit. The test's results provided in Table 7 in section A.3 indicate that the PFE model has superior fit.

Table 2: Estimated shopping trips models

	Coop		ICA Maxi		ICA Supermarket		Willys	
	(OLS FE)	(PFE)	(OLS FE)	(PFE)	(OLS FE)	(PFE)	(OLS FE)	(PFE)
α_i Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\hat{\delta}$ Industry index	1.1053***		1.2521***		0.8402***		0.9821***	
	(0.2255)		(0.1277)		(0.2361)		(0.1279)	
$\hat{\beta}_1$ Gevalia	0.0002	0.0002*	0.0002**	0.0001	0.0001	−0.00002	0.0002	0.00001
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\hat{\beta}_2$ Löfbergs	0.0002	−0.00003	0.00002	0.0001	0.00001	0.00001	0.0001	−0.00001
	(0.0002)	(0.0001)	(0.0002)	(0.00007)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\hat{\beta}_3$ Zoegas	0.00004	0.00003	−0.0001	−0.00001	0.0001	0.0001	0.0002*	0.0001
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\hat{\beta}_4$ Classic	−0.0002	0.0002**	0.0001	0.0002***	0.0002**	0.0001	0.0003	0.00004
	(0.00017)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
$\hat{\beta}_5$ Other brands		−0.0002**		0.0001		0.0001		0.0001
		(0.0001)		(0.0001)		(0.0001)		(0.0001)
$\hat{\delta}_1$ Retailer × North index		0.6830***		0.7922***		0.5595***		0.9319***
		(0.0765)		(0.0923)		(0.0672)		(0.1310)
$\hat{\delta}_2$ Retailer × East index		−0.1039		0.6664***		1.1698***		0.8988***
		(0.1649)		(0.1875)		(0.1247)		(0.1589)
$\hat{\delta}_3$ Retailer × Stockholm index		1.8368***		1.4480***		1.2263***		1.1841***
		(0.0719)		(0.0961)		(0.1550)		(0.1665)
$\hat{\delta}_4$ Retailer × West index		0.7168***		1.1779***		1.0985***		0.6933***
		(0.1045)		(0.2125)		(0.1488)		(0.2572)
$\hat{\delta}_5$ Retailer × Gothenburg index		0.3717*		0.9466***		0.7517***		1.0554***
		(0.2005)		(0.1755)		(0.1922)		(0.1510)
$\hat{\delta}_6$ Retailer × South index		1.1106***		1.0295***		1.1474***		0.9744***
		(0.1312)		(0.0807)		(0.1051)		(0.0787)
$\hat{\delta}_7$ Retailer × Öresund index		1.3154***		0.8575***		0.6430**		0.9889***
		(0.2568)		(0.2129)		(0.3123)		(0.1158)
Observations	735	735	735	735	735	735	735	735
RMSE	32.012	19.763	22.301	19.702	25.801	20.680	24.150	21.294
Bias	-2.276	<0.001	-1.216	<0.001	-1.452	<0.001	-1.384	<0.001
Adj. pseudo R^2	0.052	0.642	0.156	0.383	0.069	0.411	0.123	0.352

Significance levels: ***, $p < 0.01$, **, $p < 0.05$, * $p < 0.1$

Notes: This table show the results from the estimated log-linear fixed effects (OLS FE) and the Poisson fixed effects (PFE) models with shopping trips as dependent variable. Heteroscedasticity, autocorrelation and spatial correlation consistent (HACSC) robust standard errors in parentheses. Region fixed effects are included to capture region-specific unobserved effects. The OLS FE model's industry index captures regionally aggregated (industry level) seasonality, time trend and shocks in shopping trips for all grocery retailing firms that the sample of households purchased from. The index used in the PFE model is retailer-specific and interacted the region fixed effects to capture each retailer's region-specific seasonality, time trend and shocks.

Table 3: Estimated coffee shopping trips models

	Coop		ICA Maxi		ICA Supermarket		Willys	
	(OLS FE)	(PFE)	(OLS FE)	(PFE)	(OLS FE)	(PFE)	(OLS FE)	(PFE)
α_i Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\hat{\delta}$ Industry index	0.494***		0.417***		0.626***		1.477***	
	(0.098)		(0.096)		(0.107)		(0.184)	
$\hat{\beta}_1$ Gevalia	0.0060***	0.0022***	0.0066***	0.0019***	0.0066***	0.0011***	0.0032***	0.0003
	(0.0008)	(0.0004)	(0.0008)	(0.0005)	(0.0009)	(0.0004)	(0.0011)	(0.0003)
$\hat{\beta}_2$ Löfbergs	0.0058***	0.0021***	0.0042***	0.0012***	0.0058***	0.0007**	0.0038**	0.0004
	(0.0008)	(0.0003)	(0.0007)	(0.0003)	(0.0013)	(0.0003)	(0.0015)	(0.0003)
$\hat{\beta}_3$ Zoegas	0.0063***	0.0026***	0.0068***	0.0020***	0.0045***	0.0010**	0.0064***	0.0007**
	(0.0008)	(0.0005)	(0.0007)	(0.0005)	(0.0009)	(0.0004)	(0.0011)	(0.0003)
$\hat{\beta}_4$ Classic	0.0053***	0.0022***	0.0054***	0.0015***	0.0079***	0.0007*	0.0032***	0.0002
	(0.0008)	(0.0005)	(0.0007)	(0.0003)	(0.0009)	(0.0004)	(0.0011)	(0.0004)
$\hat{\beta}_5$ Other brands		0.0026***		0.0010***		0.0005		0.0005
		(0.0006)		(0.0003)		(0.0004)		(0.0003)
$\hat{\delta}_1$ Retailer × North index		0.9102***		0.8801***		0.9331***		0.9681***
		(0.0896)		(0.1034)		(0.0448)		(0.0442)
$\hat{\delta}_2$ Retailer × East index		0.7217***		0.8574***		1.1022***		0.8743***
		(0.1599)		(0.1382)		(0.0855)		(0.0431)
$\hat{\delta}_3$ Retailer × Stockholm index		0.8492***		1.0809***		1.2671***		0.8080***
		(0.1219)		(0.1359)		(0.0727)		(0.0419)
$\hat{\delta}_4$ Retailer × West index		0.5233***		0.9588***		1.0711***		0.9516***
		(0.0996)		(0.1185)		(0.0902)		(0.0389)
$\hat{\delta}_5$ Retailer × Gothenburg index		0.8025***		0.7383***		1.0294***		1.0708***
		(0.2024)		(0.0890)		(0.1255)		(0.0494)
$\hat{\delta}_6$ Retailer × South index		0.6070***		0.9121***		0.7054***		0.9019***
		(0.1068)		(0.08007)		(0.04617)		(0.02925)
$\hat{\delta}_7$ Retailer × Öresund index		0.9619***		0.5632***		0.4334***		1.2653***
		(0.23767)		(0.1678)		(0.1234)		(0.0750)
Observations	735	735	735	735	735	735	735	735
RMSE	6.363	4.692	16.842	10.229	12.417	6.339	12.417	8.977
Bias	-1.262	<0.001	-2.336	<0.001	-1.857	<0.001	-1.552	<0.001
Adj. pseudo R^2	0.227	0.483	0.263	0.493	0.298	0.662	0.466	0.826

Significance levels: ***, $p < 0.01$, **, $p < 0.05$, * $p < 0.1$

Notes: This table show the results from the estimated log-linear fixed effects (OLS FE) and the Poisson fixed effects (PFE) models with the coffee shopping trips as dependent variable. Heteroscedasticity, autocorrelation and spatial correlation consistent (HACSC). Region fixed effects are included to capture region-specific unobserved effects. The OLS FE model's industry index captures regionally aggregated (industry level) seasonality, time trend and shocks in shopping trips for all grocery retailing firms that the sample of households purchased from. The index used in the PFE model is retailer specific and interacted the region fixed effects to capture each retailer's region-specific seasonality, time trend and shocks.

Table 4 shows $\hat{\tau}_{CTT}$, the estimated unobservable number of shopping trips due to promotion on all ground coffee brands, estimated as in equation (5.3). The column Percent of trips shows the estimate of the share of observed shopping trips that were due to promotion, calculated as $\hat{\tau}_{CTT}/y^{obs}$, where y^{obs} is the sum of y_{it}^{obs} over i and t. For general shopping trips, Coop has the smallest effect among the retailers according to both models. The PFE model estimates about half the effect of the OLS FE model, 0.16 versus 0.32 percent of all shopping trips. According to the OLS FE model, Willys have received the largest effects among all retailers with 1.22 percent of the observed shopping trips being due to coffee promotion, but The PFE model estimates Willys' share to be only 0.28 percent. Percent of trips are similar for ICA Maxi and ICA Supermarkets in the OLS FE model, but with the PFE model ICA Maxi has received almost twice the relative effect.

The Percent of trips are higher for coffee shopping trips since their promotion effects were larger. For coffee shopping trips, the PFE model estimates smaller effects of promotion than the OLS FE model. Another difference between the models is that the OLS FE model estimates about equal Percent of trips for the retailers while the PFE estimates differ substantially across the retailers, with Coop having the largest relative effect and Willys the smallest. A reason may be that the PFE model estimates the promotion effects of all remaining brands in coffee category by including Other brands, and this changes the estimates of the main brand's effects more when the estimates are larger.

Table 5 shows $\hat{\tau}_{CTT}^{j}$, the estimated unobservable number of shopping trips due to promotion per ground coffee brand, estimated as in equation (5.4). The table also shows for which brands promotion had a significant impact, where the significance asterisks are taken from the $\hat{\beta}$ coefficients in Table 2 and 3. A couple of findings are worth mentioning. The increase in shopping trips associated with promotion on Other brands is large, particularly for coffee shopping trips. This shows that the promotion effect of the remaining brands in the category should be estimated. However, their effect is not significant for all retailers. For some of the brands, the estimated promotion effect on general shopping trips is negative. In particular, the OLS FE model estimates that Coop's sales promotion on Classic have led to half a million less shopping trips, while the PFE model estimates about half a million increase. It should be noted that the OLS FE models estimate is insignificant while the PFE model's is significant. The cause for these differences in the estimates may that the OLS FE model's estimates of the promotion effects are biased due to brand-switching, inconsistent due to its heteroscedastic errors, and potentially spurious by not considering retailer-region specific time effects, and the fact that $\mathbb{E}[\log(\mathbf{Y}) \mid \mathbf{X}] \neq \log(\mathbb{E}[\mathbf{Y} \mid \mathbf{X}])$, as previously discussed.

Table 4: Estimated number of shopping trips due to promotion

Outcome	Model	Retailer	Incremental, $\hat{\tau}_{CTT}$	Baseline, $\hat{y}^c$	Observed, y^{obs}	Percent of trips
Shopping trips	OLS FE	Coop Extra	428 046	131 746 671	132 174 717	0.32
		ICA Maxi	1 116 678	154 116 625	155 233 303	0.72
		ICA Supermarket	1 269 612	168 143 898	169 413 510	0.75
		Willys	1 943 890	157 394 977	159 338 866	1.22
	PFE	Coop Extra	211 069	131 963 648	132 174 717	0.16
		ICA Maxi	1 440 408	153 792 895	155 233 303	0.93
		ICA Supermarket	945 406	168 468 104	169 413 510	0.56
		Willys	447 026	158 891 840	159 338 866	0.28
Coffee shopping trips	OLS FE	Coop Extra	2 762 602	3 637 177	6 399 779	43.17
		ICA Maxi	7 136 908	8 795 204	15 932 113	44.80
		ICA Supermarket	5 645 379	5 126 266	10 754 960	52.41
		Willys	5 602 262	7 226 795	12 829 060	43.67
	PFE	Coop Extra	1 508 738	4 891 041	6 399 779	23.57
		ICA Maxi	2 875 773	13 056 340	15 932 113	18.05
		ICA Supermarket	1 176 487	9 578 475	10 754 960	10.94
		Willys	885 006	11 944 051	12 829 060	6.90

Notes: This table shows the estimates of the impact of promotion on ground coffee on shopping trips. Incremental is the estimated joint impact of all ground coffee brand's promotion effects on shopping trips for the weeks that had promotion, calculated as equation (5.3). Baseline are the estimated counterfactual number of shopping trips if none of the brands had been promoted. Observed are the observed number of shopping trips, equal to the sum of Baseline and Incremental. Percent of trips is the share of the observed shopping trips due to promotion, calculated as Incremental divided by Observed. The OLS FE model only estimates the promotion effect of the brands Gevalia, Löfbergs, Zoegas, and Classic. The PFE model estimates the promotion effect of all ground coffee brands that were purchased by the sample of households.

Table 5: Estimated number of shopping trips due to promotion per brand

Outcome	Model	Retailer	Gevalia	Löfbergs	Zoegas	Classic	Other
			Brand				
Shopping trips	OLS FE	Coop Extra	334 794	541 685	87 453	-542 898	
		ICA Maxi	884 604**	66 903	-203 513	362 773	
		ICA Supermarket	3 561	32 713	300 255	927 388**	
		Willys	459 937	121 290	781 938*	587 387	
	PFE	Coop Extra	297 738*	-71 940	62 144	483 132**	-563 271**
		ICA Maxi	396 372	257 638	-22 253	576 357***	236 462
		ICA Supermarket	-103 126	28 219	449 316	196 051	376 901
		Willys	55 821	-18 160	177 632	92 260	140 236
Coffee shopping trips	OLS FE	Coop Extra	645 898***	785 252***	727 845***	713 392***	
		ICA Maxi	2 600 490***	1 259 502***	1 989 515***	1 807 006***	
		ICA Supermarket	1 989 693***	1 200 417***	1 256 424***	2 068 558***	
		Willys	1 423 231***	1 436 021**	2 988 150***	1 229 099***	
	PFE	Coop Extra	276 142***	331 553***	317 347***	352 882***	396 671*
		ICA Maxi	916 736***	366 144***	673 440***	576 581***	536 923***
		ICA Supermarket	397 434***	182 635**	280 777**	255 109*	117 506
		Willys	134 681	158 162	379 225**	74 382	175 022

Significance levels: ***, $p < 0.01$, **, $p < 0.05$, * $p < 0.1$

Notes: This table shows $\hat{\tau}^{j}_{CTT}$, the estimates of each brand's impact of promotion on shopping trips. The estimates are calculated as equation (5.4), that is, by taking the sum across regions and weeks of the difference between the predicted number of shopping trips given the observed promotion and an estimated counterfactual of no promotion on the brand for the weeks that had promotion. The Other column show the impact of promotion on the remaining ground coffee brands that were purchased by households in the sample.

Figure 4 and 5 show the time series of the observed shopping trips and coffee shopping trips against their counterfactuals under no promotion on any brand. The regional time series are aggregated to show the total impact on the national level over time. The time series of the counterfactuals are multiplied with the adjustment factor so the counterfactuals can be compared against the actual observed outcomes rather than the predicted outcomes given the observed promotion.

The difference between the time series of the observed and counterfactual outcomes corresponds to the impact of promotion on all brands, and the sum of the differences over all weeks corresponds to $\hat{\tau}_{CTT}$ derived in equation (5.3) and the estimates shown in column Incremental in Table 4. The weeks where the observed and the counterfactual time series intersect are the weeks with no promotion since for those weeks the observed outcome equals the estimated counterfactual multiplied with the adjustment factor. To save space the estimated counterfactual time series per brand are not shown. The time series of the joint impact of all brands per region are shown in Figure 8a to 11b in section A.10 in the appendix.

In Figure 4, the counterfactual time series follow the seasonal variation and trend over time of the observed time series, meaning that the index variable captures time effects. The differences between the estimated counterfactuals from the models are more apparent in Figure 5. The OLS counterfactual is visibly lower than the PFE counterfactual for each retailer.

Figure 6 and 7 show $\hat{\tau}_{CTT,t}$, the regionally aggregated estimated joint brand effects of promotion for each retailer and model over time in relation to the baseline, calculated as

$$\hat{\tau}_{CTT,t} = \sum_{i=1}^{n} \left(\hat{y}_{it}^{obs} - \hat{y}_{it}^{c}(0) \right) \times \left(\sum_{i=1}^{n} y_{it}^{obs} \Big/ \sum_{i=1}^{n} \hat{y}_{it}^{obs} \right). \tag{6.3}$$

Figure 6 shows for which weeks the OLS FE model's estimates negative joint brand effects for Coop and ICA Maxi. The PFE model also estimates negative joint effects for Coop and ICA Supermarket for some weeks. Further, Figure 6 shows that the models' weekly estimates are about equal for ICA Maxi, confirming the numerical estimates in Table 4. The differences between the models' estimates of the impact are more apparent in Figure 7, especially for Willys and ICA Supermarket. Since the models' time series of the effect are not less than zero in any given week, there were no negative joint brand effects of coffee promotion on coffee shopping trips.

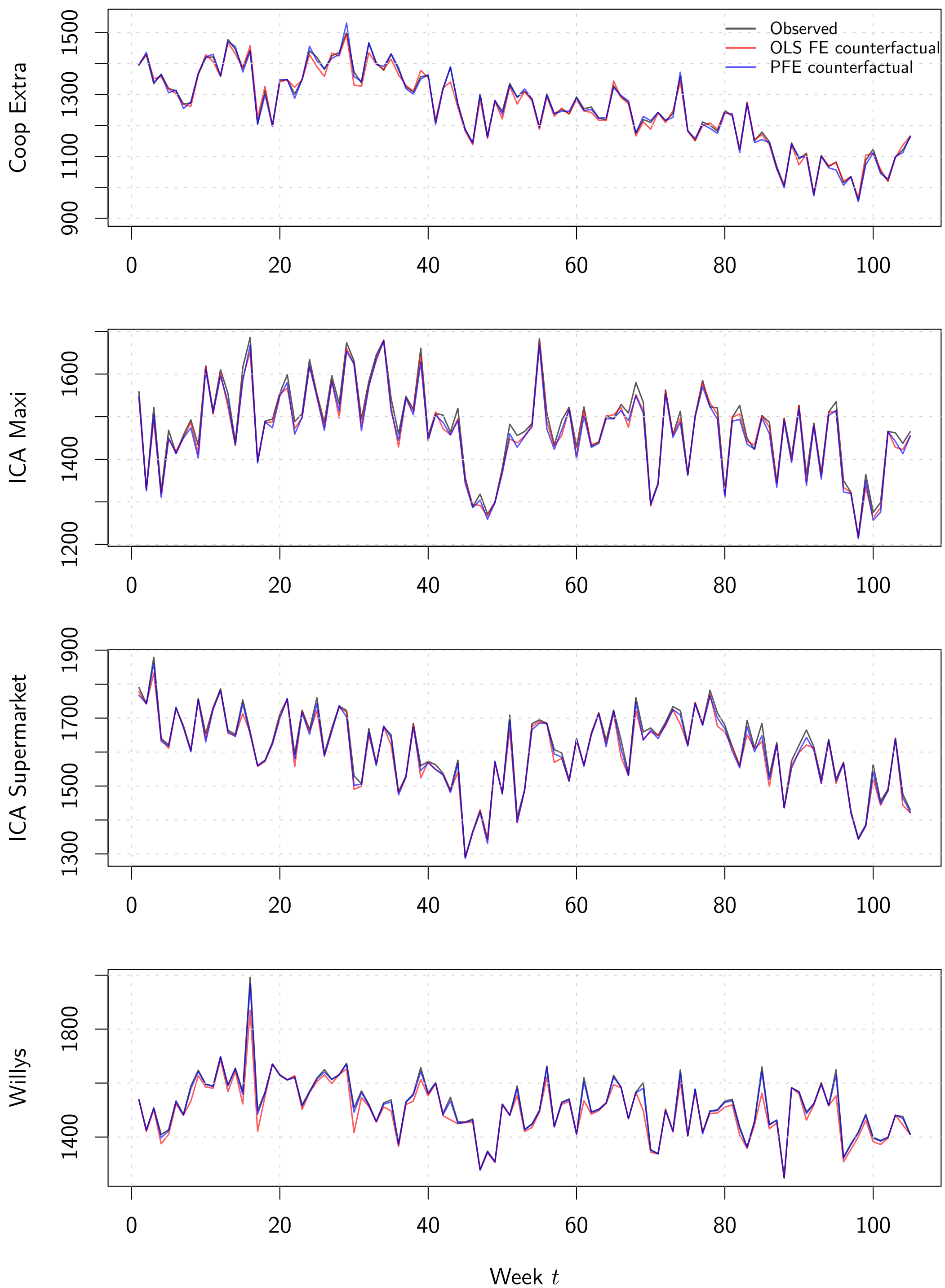


Figure 4: Shopping trips: Regionally aggregated observed and estimated counterfactual number of shopping trips (in 1000) under no promotion on any ground coffee brand per retailer. The difference between the observed and the counterfactual time series is the corresponding model's estimated joint impact of promotions on the brands.

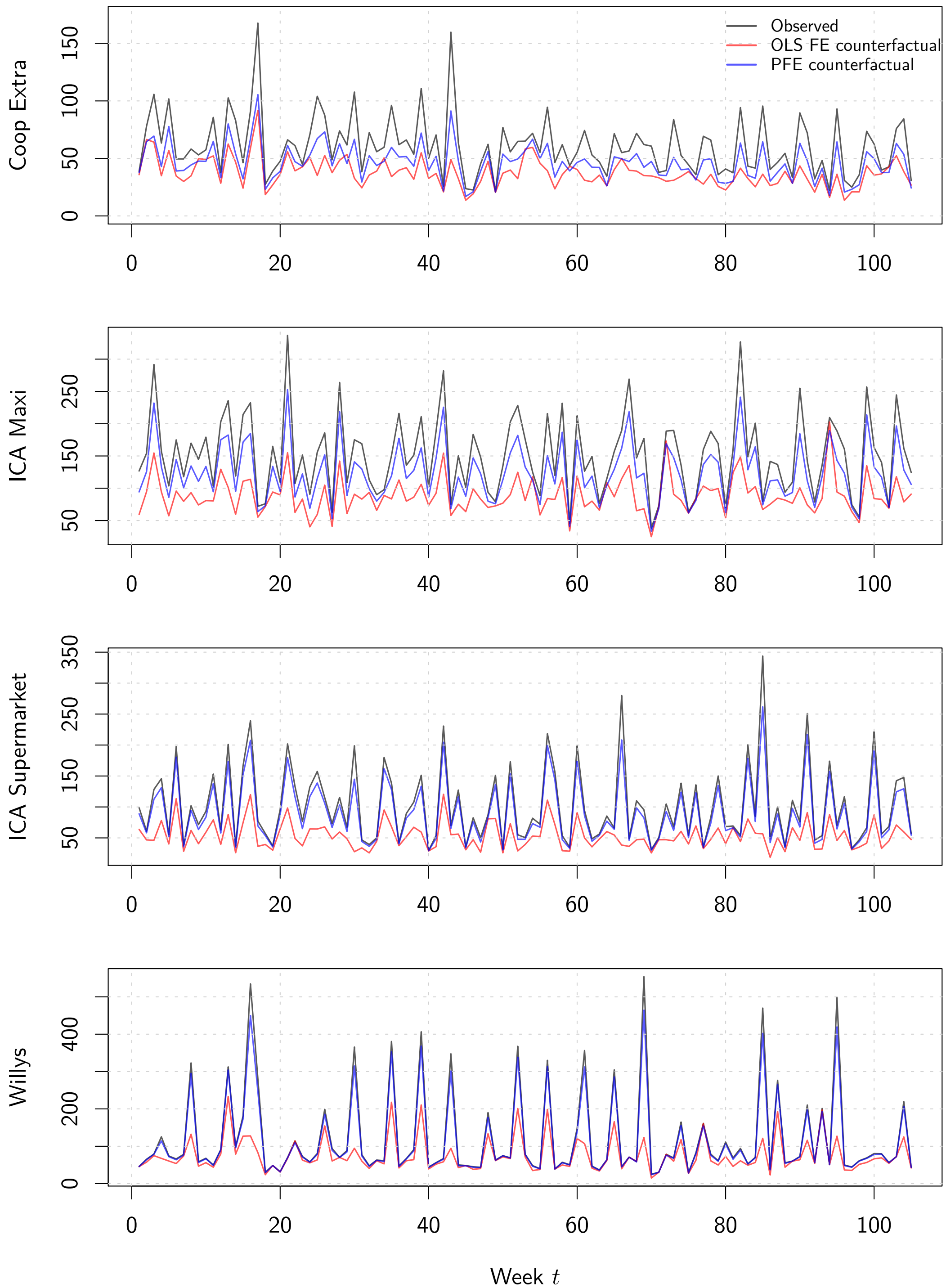


Figure 5: Coffee shopping trips: Regionally aggregated observed and estimated counterfactual number of coffee shopping trips (in 1000) under no promotion on any ground coffee brand per retailer. The difference between the observed and the counterfactual time series is the corresponding model's estimated joint impact of promotions on the brands.

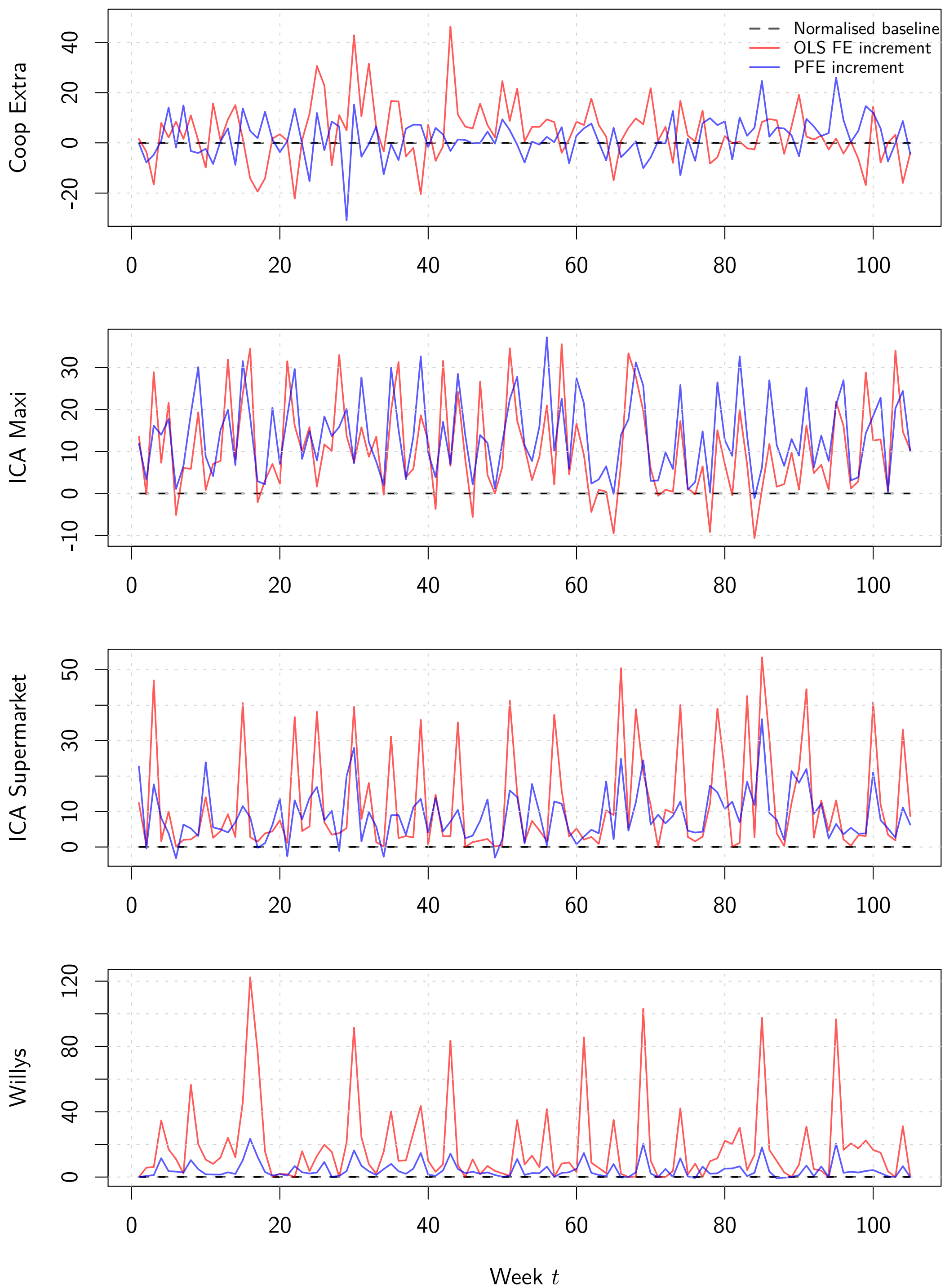


Figure 6: Shopping trips: Regionally aggregated estimated joint impact $\hat{\tau}_{CTT,t}$ of ground coffee promotion on the number of shopping trips (in 1000), calculated as in equation (6.3). The time series are the estimated incremental number of shopping trips due to promotion in relation to the baseline of no promotion in any week.

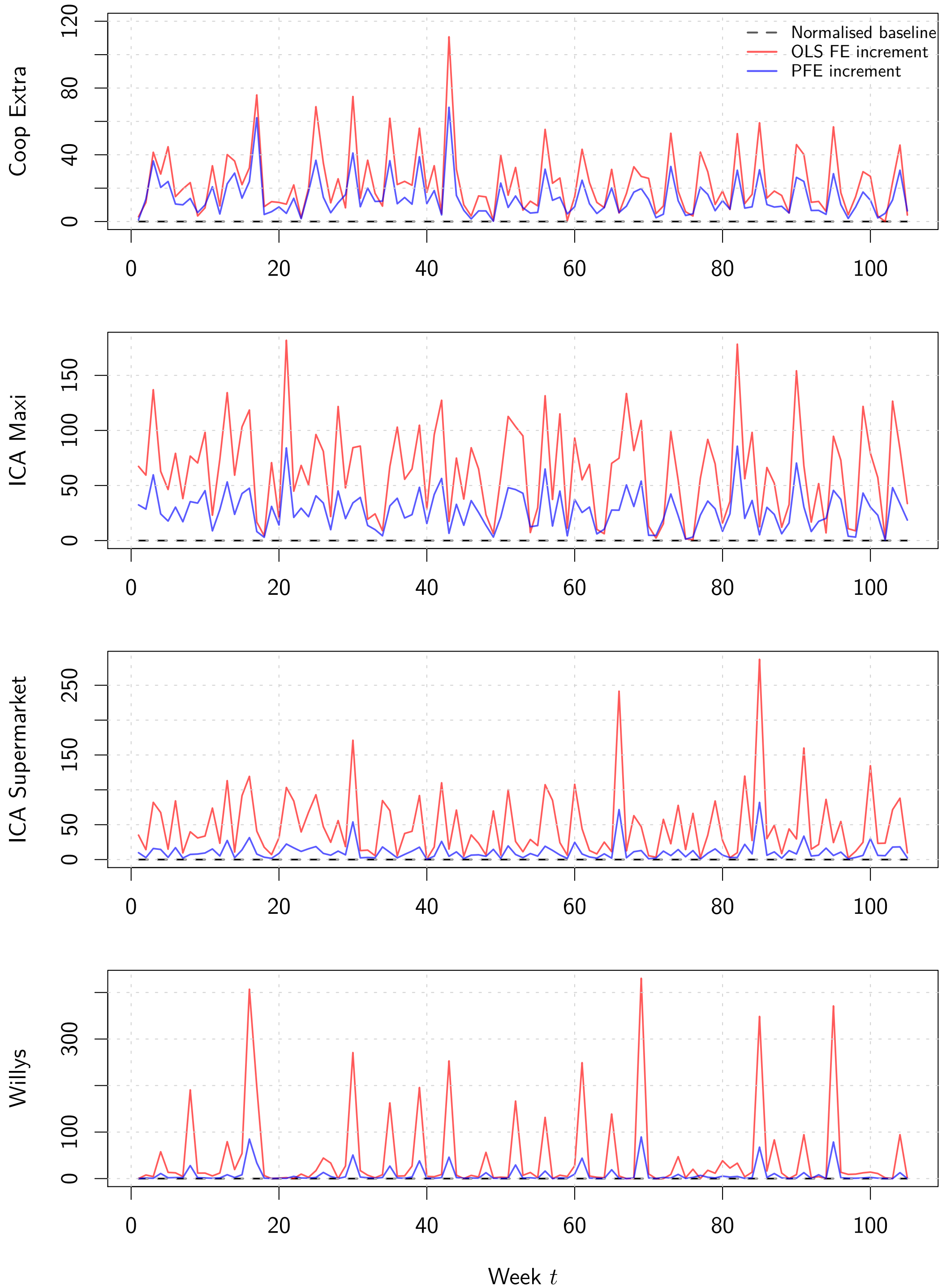


Figure 7: Coffee shopping trips: Regionally aggregated estimated joint impact $\hat{\tau}_{CTT,t}$ of ground coffee promotion on the number of coffee shopping trips (in 1000), calculated as in equation (6.3). The time series are the estimated incremental number of coffee shopping trips due to promotion in relation to the baseline of no promotion in any week.

7 Discussion

The first research question of this thesis is how to quantify the unobservable impact of promotions on store visits. Using the potential outcomes framework, the impact is quantified as the difference between the number of store visits for observations with promotion and their counterfactual outcomes without promotion. A procedure that estimates the counterfactual outcomes with a regression model is shown and expressed algorithmically for ease of implementation. In contrast to existing matching methods, the procedure does not require pre-treatment covariates and works with multiple continuous and concurrent treatments with varying starts, lengths and frequencies in the cross-section. In comparison to the standard regression adjustment estimator, the counterfactuals are estimated using coefficients obtained from a single regression on all observed data, not with one regression per treatment status. Thus the time effects on each observation are accounted for. An estimator of the cumulative treatment effect is derived from the regression adjustment estimator of the average treatment effect. Due to spatio-temporal dependence among the residuals, the sum of the fitted values are not equal to the sum of the observed outcomes. The proposed estimator correct for this dependence with an adjustment factor.

The second research question is to determine the impact of promotions on store visits empirically. Using the proposed method, the results show that the estimates depend on the regression model used, see Table 4 and 5. Both GfK's log-linear OLS fixed effects (OLS FE) model and the Poisson fixed effects (PFE) model estimates small and mostly insignificant effects of ground coffee promotions on general grocery shopping trips, see Table 4. It seems that promotions on ground coffee have minor ability to generate incremental general grocery shopping trips. The impact of coffee promotion on coffee shopping trips is larger and significant, as shown in Table 5. There is thus support for that coffee promotions increase the number of coffee purchases, which indicates that sales promotions are most effective at increasing purchases of the product that is promoted.

Research questions three is to determine how the effect of promotion on store visits differ between retailers and brands. This is achieved by fitting separate regressions per retailer and decomposing the promotion effect per brand. The PFE model generally estimates smaller effects, see Table 4 and 5. For instance, the OLS FE model estimates that 1.22 percent of Willys' shopping trips during the time period were due to promotion according, whereas the PFE model's estimate is 0.28 percent. In numbers, this corresponds to about 1.9 million versus 450 thousand shopping trips. For coffee shopping trips, Coop has had the largest relative effect according to the PFE model. This is intuitive from an economic perspective, since

Coop have the highest regular prices. Some of the retailers had a significant increase in shopping trips from promoting other coffee brands, see Table 5. This shows that their effect should be considered.

GfK's OLS FE model have several weaknesses that explain why it yields different estimates of the promotion impact. The model has log-transformed dependent variable to model the multiplicative effects. This is problematic since coffee shopping trips contains zeros, which cannot be log-transformed. A constant is added to these observations to fit the model, but this introduces bias. Further, the OLS FE model's brand level promotion effects suffer from bias by not accounting for that during promotions consumer brand-switch, which Gupta (1988) show is the dominant behavioural change of consumers during promotions. Another weakness is that it uses the industry level time trend and seasonality to model time effects although these are not common among retailers and regions. The regression results confirm that the predicted outcomes from the OLS FE model has higher bias than those of the PFE model, see the bias statistic in Table 2 and 3. Also, estimates of β in log-linear OLS models are inconsistent if the standard errors are heteroscedastic, which Breusch-Pagan tests in Table 8 in the appendix indicate is the case. The proposed PFE model addresses these weaknesses by being able to handle outcomes with a value of zero, by estimating the promotion effects of all ground coffee brands that were purchased, and by accounting for region-retailer specific seasonality and time trends. HACSC robust standard errors are estimated to adjust for heteroscedasticity, spatio-temporal dependence and the equidispersion property of the PFE model for correct inference of the promotion effect.

The validity of causal claims depend on whether the necessary assumptions for causal inference in the potential outcomes framework are satisfied. It is questionable whether the no-interference assumption of the stable unit treatment value assumption (SUTVA) is satisfied. Likely, the assignment of promotions to regions are not independent of each other. The no-hidden variations in treatment may be satisfied since the type of treatment is held constant within regions and different counterfactuals are estimated for each brand. Promotions are likely planned in a pre-specified pattern over the calendar year according to seasonality in demand. By using a time index variable in the regressions that capture these effects it is more probable that unconfoundedness is fulfilled, and the PFE model control for these effects on the retailer-region level instead of on just the industry level. The overlap assumption is satisfied since the difference is taken between each observation's own counterfactual for the same unit and time period, thereby cancelling possible confounders of the estimated impact. Since a nationally represented sample is used the external validity towards the Swedish population can be considered satisfactory.

8 Concluding Remarks

The practical contribution of this thesis is an improvement of GfK's evaluation of promotion effects. This has been accomplished by proposing a new model that address several limitations of their original model. The findings suggest that the original model overestimated the impact of promotion on store visits. Being able to provide accurate information is crucial for GfK and their clients. For retailers, accurate information on marketing performance is critical for planning and forecasting business performance. For the promoted brand, being able to show the brand's ability to draw consumers to the store offers an advantage in negotiations with retailers, for instance with regards to shelf placement of the brand's products.

The thesis contributes to the existing literature on promotion effects. It shows how to estimate the unobservable promotion impact using the potential outcomes framework. This approach has only recently been adopted in marketing (see Rossi (2017) and Varian (2016)). Secondly, the thesis suggests which models may be appropriate for modelling such data. Third, an algorithm for estimating counterfactual outcomes under multiple continuous and concurrent treatments that vary in the cross-section without using matching covariates is presented and an estimator of the cumulative treatment effect with adjustment for spatio-temporal dependence is derived.

The study has a few limitations. One concern is whether the effect truly is incremental. It is possible that consumers found out about the promotion at the store and would purchase anyway. The results may suffer from measurement bias since the data are self-reported, not transaction level data. The analysis has not considered the magnitude of the price reduction. It is likely that the effect of promotion on store visits is stronger the deeper the price reduction. However, according to economic theory on revealed preferences it is reasonable to assume that the price discount is indirectly reflected in the purchasing frequency. It is further possible that there is simultaneity bias, i.e. that the number of shopping trips affect the promotion assignment and vice versa. Throughout the thesis it is argued that the time index variable captures some of this bidirectional dependence, but the promotion effect may be interpreted as associative rather than causal to not overstate the findings.

For future research, it would be interesting to use data on the individual household level. Another possibility would be to study the impact on revenue or profits. A bayesian approach could be beneficial by being able to use results from previous studies and managerial experience to set priors.

References

Asplund, M., & Friberg, R. (2002). Food Prices and Market Structure in Sweden. *Scandinavian Journal of Economics*, *104*(4), 547–566. Retrieved from http://doi.wiley.com/10.1111/1467-9442.00301 doi: 10.1111/1467-9442.00301

Axfood. (2018). *Axfoodfamiljen: Willys.* Retrieved 2018-03-25, from https://www.axfood.se/om-axfood/bolagsfamilj/willys/

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3rd ed.). Hoboken, New Jersey: John Wiley & Sons Ltd.

Blattberg, R. C., Eppen, G. D., & Lieberman, J. (1981). A Theoretical and Empirical Evaluation of Price Deals for Consumer Nondurables. *Journal of Marketing*, *45*(1), 116–129. Retrieved from http://www.jstor.org/stable/1251725 doi: 10.2307/1251725

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287–1294. doi: 10.2307/1911963

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies*, *47*(1), 239. Retrieved from https://academic.oup.com/restud/article-lookup/doi/10.2307/2297111 doi: 10.2307/2297111

Cameron, A. C., & Miller, D. M. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, *50*(2), 317–372. doi: 10.3368/jhr.50.2.317

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications* (Cambridge ed.). Cambridge: Cambridge University Press.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data* (2nd ed.). Cambridge: Cambridge University Press.

Chatfield, C., Ehrenberg, A. S. C., & Goodhardt, G. J. (1966). Progress on a Simplified Model of Stationary Purchasing Behaviour. *Journal of the Royal Statistical Society. Series A (General)*, *129*(3), 317–367. Retrieved from http://www.jstor.org/stable/2343502

Chatfield, C., & Goodhardt, G. J. (1973). A Consumer Purchasing Model with Erland Inter-Purchase Times. *Journal of the American Statistical Association*, *68*(344, December), 828–835. Retrieved from http://www.jstor.org/stable/2284508

Coop. (2018). *Coop - Våra butiker.* Retrieved 2018-03-26, from https://www.coop.se/butiker-erbjudanden/vara-butiker/

Cooper, L. G., & Nakanishi, M. (1988). *Market Share Analysis: Evaluating Competitive Marketing Effectiveness.* Norwell, Massachusetts: Kluwer Academic Publishers.

Croissant, Y., & Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, *27*(2), 1–43. doi: 10.1186/1478-7954-4-13

Driscoll, J., & Kraay, A. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *The Review of Economics and Statistics*, *80*(4), 549–560.

Dunn, R., Seader, R., & Wrigley, N. (1983). An Investigation of the Assumption of the NBD Model as Applied to Purchasing at Individual Stores. *Journal of the Royal Statistical Society*, *32*(3), 249–259. Retrieved from `http://www.jstor.org/stable/2347947`

Durevall, D. (2017). Cost Pass-Through in the Swedish Coffee Market. *Working Papers in Economics (S-WoPEc), University of Gothenburg, Department of Economics*(697). Retrieved from `http://hdl.handle.net/2077/51976`

Frasco, M. (2017). *Package 'Metrics': Evaluation Metrics for Machine Learning.* Retrieved from `https://cran.r-project.org/web/packages/Metrics/Metrics.pdf`

Frisbie, G. A. (1980). Ehrenberg's Negative Binomial Model Applied to Grocery Store Trips. *Journal of Marketing Research*, *17*(3), 385. Retrieved from `http://www.jstor.org/stable/3150539?origin=crossref` doi: 10.2307/3150539

GfK. (2018). *GfK Consumer Scan.* Retrieved 2018-03-23, from `http://ecpo-se.gfk.com/`

Greene, W. H. (2012). *Econometric Analysis* (7th ed., Vol. 97). London: Pearson Education. doi: 10.1198/jasa.2002.s458

Greene, W. H., & Hensher, D. A. (2010). *Modelling Ordered Choices - A Primer* (1st ed.). Cambridge: Cambridge University Press.

Gullstrand, J., Olofsdotter, K., & Karantininis, K. (2011). *Konkurrens och Makt i den Svenska Livsmedelskedjan* (Tech. Rep.). Agrifood Economics Centre.

Gupta, S. (1988). Impact of Sales Promotions on When, What, and How Much to Buy. *Journal of Marketing Research*, *25*(4), 342. Retrieved from `http://www.jstor.org/stable/3172945?origin=crossref` doi: 10.2307/3172945

Hadri, K. (2000). Testing for Stationarity in Heterogenous Panel Data. *The Econometrics Journal*, *3*(2), 148–161. doi: 10.1111/1368-423X.00043

Hirshleifer, J., Glazer, A., & Hirshleifer, D. (2005). *Price Theory and Applications* (7th ed.). Cambridge: Cambridge University Press.

Hlavac, M. (2015). Stargazer: Well-Formated Regression and Summary Statistics.

R package version 5.2., 1–11. Retrieved from http://cran.r-project.org/package=stargazer

Hlavac, M. (2018). *Package 'stargazer': Well-Formatted Regression and Summary Statistics Tables.* Retrieved from https://cran.r-project.org/web/packages/stargazer

Hoechle, D. (2007). Robust Standard Errors for Panel Regressions with Cross-sectional Dependence. *Stata Journal*, *7*(3), 281–312.

Holland, P. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960. doi: 10.2307:2289064

HUI Research. (2017). *Dagligvarukartan 2017* (Tech. Rep.). Handelns Utredningsinstitut. Retrieved from http://www.hui.se/statistik-rapporter/index-och-barometrar/

ICA. (2018a). *ICA Maxi.* Retrieved 2015-03-25, from https://www.ica.se/butiker/maxi/

ICA. (2018b). *ICA Supermarket.* Retrieved 2018-03-25, from https://www.ica.se/butiker/supermarket/

Im, K. S., Pesaran, M. H., & Shin, Y. (2005). Testing for Seasonal Unit Roots in Heterogeneous Panels. *Journal of Econometrics*, *115*(1), 53–74. doi: 10.1016/j.econlet.2004.06.018

Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences* (1st ed.). Cambridge: Cambridge University Press.

Imbens, G. W., & Wooldridge, J. M. (2017). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, *47*(1), 5–86. doi: 10.1257/jel.47.1.5

Kleiber, C., & Zeileis, A. (2008). *Applied Econometrics with R* (1st ed.). New York: Springer. doi: 10.1007/978-0-387-77318-6

Kotler, P., & Keller, K. L. (2012). *Marketing Management* (14th ed., Vol. 22) (No. 4). London: Pearson Education. doi: 10.1080/08911760903022556

Marketing Science Institute. (2016). Research Priorities 2016-2018. *Marketing Science Institute (MSI)*, 21. Retrieved from http://www.msi.org/uploads/articles/MSI{_}RP16-18.pdf

Merkle, E. C., & You, D. (2018a). *Getting Started with nonnest2* (No. 1989). Retrieved from https://cran.r-project.org/web/packages/nonnest2

Merkle, E. C., & You, D. (2018b). *Package 'nonnest2': Tests of Non-Nested Models.* Retrieved from https://cran.r-project.org/web/packages/nonnest2 doi: 10.1037/met0000038>a

Millo, G. (2014). Robust Standard Error Estimators for Panel Models: a Unifying Approach. *Munich Personal RePEc Archive*(54954). doi: 10.18637/jss.v000

.i00
Millo, G. (2017). *Package 'pglm': Panel Generalized Linear Models.* Retrieved from https://cran.r-project.org/package=pglmAuthor
Millo, G., Tappe, K., & Croissant, M. Y. (2017). *Package 'plm': Linear Models for Panel Data* (No. 1). Retrieved from https://cran.r-project.org/package=plm
Mundlak, Y. (1978). Pooling of Time Series and Cross-section Data. *Econometrica*, *46*(1), 69–86. doi: 10.2307/1913646
Neslin, S., Henderson, C., & Quelch, J. (1985). Consumer Promotions and the Acceleration of Product Purchases. *Marketing Science*, *4*(2), 147–165. Retrieved from https://doi.org/10.1287/mksc.4.2.147
Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708. Retrieved from http://www.jstor.org/stable/1913610
Newey, W. K., & West, K. D. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *The Review of Economic Studies*, *61*(4), 631–653. Retrieved from https://academic.oup.com/restud/article-lookup/doi/10.2307/2297912 doi: 10.2307/2297912
Paul, A. E. (1978). A Generalized Compound Poisson Model for Consumer Purchase Panel Data. *Journal of the American Statistical Association*, *73*(364), 706–713. Retrieved from http://www.jstor.org/stable/2286262
Pepall, L., Richards, D., & Norman, G. (2014). *Industrial Organization: Contemporary Theory and Empirical Applications* (No. January). New Jersey: Wiley.
Perloff, J. M. (2014). *Microeconomics* (7th ed.). London: Pearson Education.
Persson, P.-G. (1995). *Modeling the Impact of Sales Promotion on Store Profits* (Doctoral dissertation, Stockholm School of Economics). Retrieved from https://ex.hhs.se/dissertations/221940-FULLTEXT01.pdf
Pesaran, M. H. (2004). General Diagnostic Tests for Cross-section Dependence in Panels. *Faculty of Economics*(1240), 1–39. Retrieved from http://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe0435.pdf doi: 10.17863/CAM.5113
Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41–55. doi: 10.1093/biomet/70.1.41
Rossi, P. E. (2017). Causality and Marketing Analytics. In *Msi analytics conference, february 8-9 2017.*
Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and

Nonrandomized Studies. *Journal Of Educational Psychology*, *66*(5), 688–701. doi: 10.1037/h0037350

Sharpsteen, C., & Bracken, C. (2018). *LaTeX Graphics for R.* Retrieved from `https://github.com/yihui/tikzDevice`

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science - a Review Journal of the Institute of Mathematical Statistics*, *25*(1), 1–21. doi: 10.1214/09-STS313

Sun, B. (2005). Promotion Effect on Endogenous Consumption. *Marketing Science*, *24*(3), 430–443. doi: 10.1287/mksc.1040.0110

Sun, B., Neslin, S. A., & Srinivasan, K. (2003). Measuring the Impact of Promotions on Brand Switching When Consumers Are Forward Looking. *Journal of Marketing Research*, *40*(4), 389–405. Retrieved from `http://www.jstor.org/stable/30038874`

Teunter, L. H. (2002). *Analysis of Sales Promotion Effects on Household Purchase Behavior* (Doctoral dissertation, Erasmus University Rotterdam). Retrieved from `http://hdl.handle.net/1765/868`

Varian, H. R. (2016). Causal Inference in Economics and Marketing. *Proceedings of the National Academy of Sciences*, *113*(27), 7310–7315. doi: 10.1073/pnas.1510479113

Verbeek, M. (2008). *A Guide to Modern Econometrics* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Ltd.

Vogelsang, T. J. (2012). Heteroskedasticity, Autocorrelation, and Spatial Correlation Robust Inference in Linear Panel Models with Fixed-Effects. *Journal of Econometrics*, *166*(2), 303–319. doi: 10.1016/j.jeconom.2011.10.001

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, *57*(2), 307–333. Retrieved from `http://www.jstor.org/stable/1912557`

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817. Retrieved from `http://www.jstor.org/stable/1912934?origin=crossref` doi: 10.2307/1912934

Winkelmann, R. (2008). *Econometric Analysis of Count Data* (5th ed.). Berlin: Springer. doi: 10.1007/978-3-540-78389-3

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, Massachusetts: MIT Press. doi: 10.1515/humr.2003.021

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, *11*(10), 1–17. doi: 10.18637/

jss.v011.i10

Zeileis, A. (2006). Object-oriented Computation of Sandwich Estimators. *Journal of Statistical Software*, *16*(9), 17. doi: 10.18637/jss.v081.b02

A Appendices

A.1 Proof of LSDV and Fixed Effects Estimators

Throughout this thesis the LSDV and fixed effects specifications are used interchangeably. A necessary condition for this to be valid is that the specifications give the same estimates of the main parameter of interest $\boldsymbol{\beta}$. This section provides a proof following Greene (2012) that the specifications coincide. Only the OLS case is proven here but the result also holds for the Poisson fixed effects model (see Cameron and Trivedi (2013)).

Matrix algebra is used for more compact notation. The LSDV estimator is first derived from the pooled estimator, then the FE estimator is derived, and lastly it is proven that they coincide.

The pooled panel data model is given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon, \tag{A.1}$$

where $\mathbf{Y}$ and ϵ are $(nT \times 1)$, $\mathbf{X}$ is $(k+1) \times nT$ and $\boldsymbol{\beta}$ is $(k+1) \times 1$. For the LSDV specification, break up $\mathbf{X}$ in two components:

1. $\mathbf{W}$, which is $\mathbf{X}$ less its column of ones for each intercept, $[\mathbf{X}_{01}, \dots, \mathbf{X}_{0N}]'$ where $\mathbf{X}_{0i}$ is a $(T \times 1)$ column vector of ones, and

2. $$\mathbf{X}_0 = \begin{bmatrix} \mathbf{X}_{01} & 0 & \dots & 0 \\ 0 & \mathbf{X}_{02} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X}_{0N}, \end{bmatrix}$$

   where $\mathbf{W}$ is $(nT \times k)$ and $\mathbf{X}_0$ is $(nT \times n)$. Then the LSDV estimator is

   $$\mathbf{Y} = \mathbf{X}_0\alpha + \mathbf{W}\gamma + \epsilon, \tag{A.2}$$

   in which α is the $(n \times 1)$ column vector of the n intercepts, and γ is $(k \times 1)$ and has the k slope coefficients.

If n is large it can be computationally demanding and difficult to interpret the n intercepts. The FE estimator express observations as deviations from the variables' time average for unit i. Note that $\mathbf{X}_0'\mathbf{X}_0$ is the $(n \times n)$ diagonal matrix, i.e

$$\mathbf{X}_0'\mathbf{X}_0 = \begin{bmatrix} T & 0 & \dots & 0 \\ 0 & T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T, \end{bmatrix}$$

whose inverse is the $(n \times n)$ matrix

$$(\mathbf{X}_0'\mathbf{X}_0)^{-1} = \begin{bmatrix} T^{-1} & 0 & \dots & 0 \\ 0 & T^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T^{-1}, \end{bmatrix}$$

and that

$$\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0' = \begin{bmatrix} T^{-1}1_T & 0 & \dots & 0 \\ 0 & T^{-1}1_T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T^{-1}1_T, \end{bmatrix}$$

in which 1_T is a $(T \times T)$ matrix with all elements being 1, and $\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'$ is a symmetric $(nT \times nT)$ matrix. For all Y_i

$$T^{-1}1_T\mathbf{Y}_i = \begin{bmatrix} T^{-1}\sum_{t=1}^{T} Y_{it} \\ T^{-1}\sum_{t=1}^{T} Y_{it} \\ \vdots \\ T^{-1}\sum_{t=1}^{T} Y_{it} \end{bmatrix} \tag{A.3}$$

is a $(T \times 1)$ matrix denoted $\bar{Y}_i$, and

$$\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{Y} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_n \end{bmatrix} \tag{A.4}$$

is the $(nT \times 1)$ matrix $\bar{Y}$ which contain the mean values on Y for each unit i. Also,

$$\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{W} = \begin{bmatrix} \bar{W}_1 \\ \bar{W}_2 \\ \vdots \\ \bar{W}_n \end{bmatrix} \tag{A.5}$$

is the $(nT \times k)$ matrix called $\bar{W}$, where $\bar{W}_i$ is a $(T \times k)$ matrix where the columns are the means of the corresponding regressor for the ith unit. We can then define the $(nT \times nT)$ matrix

$$\mathbf{D} = \mathbf{I}_{nT} - \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0' \tag{A.6}$$

and $\mathbf{DW}$ becomes the $(nT \times k)$ matrix, which contains the deviation of each regressor from the regressor's mean for the observation's unit. $\mathbf{DY}$ is in turn the $(nT \times 1)$ matrix of each observations deviation in Y from the observation's unit's mean Y. Pre-multiplying the LSDV model (A.2) with $\mathbf{D}$,

$$\mathbf{DY} = \mathbf{DX}_0\alpha + \mathbf{DW}\gamma + \mathbf{D}\epsilon = 0 + \mathbf{DW}\gamma + \mathbf{D}\epsilon. \tag{A.7}$$

$\mathbf{D}\epsilon$ contain the deviation of each observations error from the observation's unit's mean error. Estimating (A.8) with OLS yields

$$\begin{aligned} \hat{\gamma}_{FE} &= (\mathbf{W}'\mathbf{D}'\mathbf{DW})^{-1}\mathbf{W}'\mathbf{D}'\mathbf{DY} \\ &= (\mathbf{W}'\mathbf{DW})^{-1}\mathbf{W}'\mathbf{DY}. \end{aligned} \tag{A.8}$$

Now, to show that FE and LSDV coincide, first recall that OLS has the property

$$\mathbf{X}'\mathbf{e} = 0,$$

where $\mathbf{e}$ is the column vector of the residuals, since the regressors are orthogonal to the residuals. If $\mathbf{X}$ is $(\mathbf{X}_0\mathbf{W})$, then $\mathbf{X}'\mathbf{e}$ is a $(n \times k)$ column vector with elements equal to zero. Thus, the slope estimates of the FE estimator will be equal to those from those estimated with OLS if $\mathbf{X}$ is orthogonal to the residuals in the FE estimator. To prove that LSDV and FE coincides it thereby suffices to show that for the FE estimator, $\mathbf{X}$ is orthogonal to the residuals.

Theorem 1 *The fixed effects estimator and least square dummy variable estimator yield identical estimates of* $\boldsymbol{\beta}$.

Proof. Recall that the FE estimator is obtained by fitting (A.8) with OLS. Therefore $\mathbf{DX}$ is orthogonal to the residuals from the OLS regression. Define the residuals from (A.8) as

$$\mathbf{d} = \mathbf{DY} - \mathbf{DW}\hat{\gamma}_{FE}, \tag{A.9}$$

which correspond to the residuals from the LSDV estimator (A.2). This holds since (A.4) and (A.5) together with (A.6) lets us rewrite (A.9) as

$$\begin{aligned}\mathbf{d} &= \mathbf{Y} - \bar{\mathbf{Y}} - (\mathbf{W} - \bar{\mathbf{W}}\hat{\gamma}, \\ &= \mathbf{Y} - (\bar{\mathbf{Y}} - \bar{\mathbf{W}}\hat{\gamma}) - \mathbf{W}\hat{\gamma}.\end{aligned}$$

Also, $\bar{\mathbf{Y}} - \bar{\mathbf{W}}\hat{\gamma}$ can be rewritten as

$$\begin{bmatrix}\bar{\mathbf{Y}}_1 - \bar{\mathbf{W}}_1\hat{\gamma} \\ \bar{\mathbf{Y}}_2 - \bar{\mathbf{W}}_2\hat{\gamma} \\ \vdots \\ \bar{\mathbf{Y}}_n - \bar{\mathbf{W}}_n\hat{\gamma}\end{bmatrix} = \mathbf{X}_0 \begin{bmatrix} T^{-1}\sum_{t=1}^{T}(\mathbf{Y}_{1t} - \mathbf{W}_{1t}\hat{\gamma}) \\ T^{-1}\sum_{t=1}^{T}(\mathbf{Y}_{2t} - \mathbf{W}_{2t}\hat{\gamma}) \\ \vdots \\ T^{-1}\sum_{t=1}^{T}(\mathbf{Y}_{nt} - \mathbf{W}_{nt}\hat{\gamma})\end{bmatrix} = \mathbf{X}_0\hat{\alpha}_{FE}$$

where Y_{it} is the ith unit's value in the tth time period, $\mathbf{W}_{it}$ is a row vector of the k values on the regressors in the tth time period for unit i, and

$$\hat{\alpha}_{FE} = \begin{bmatrix} T^{-1}\sum_{t=1}^{T}(\mathbf{Y}_{1t} - \mathbf{W}_{1t}\hat{\gamma}) \\ T^{-1}\sum_{t=1}^{T}(\mathbf{Y}_{2t} - \mathbf{W}_{2t}\hat{\gamma}) \\ \vdots \\ T^{-1}\sum_{t=1}^{T}(\mathbf{Y}_{nt} - \mathbf{W}_{nt}\hat{\gamma})\end{bmatrix}.$$

Thus, equation (A.9) can be expressed as

$$\mathbf{d} = \mathbf{Y} - \mathbf{X}_0\hat{\alpha}_{FE}\mathbf{W}\hat{\gamma}_{FE}.$$

This shows that the vector $\mathbf{d}$ is the result from estimating α and γ in(A.2) with $\hat{\alpha}_{FE}$ and $\hat{\gamma}_{FE}$. Now, the FE estimator for the slope coefficients will equal the OLS estimator of γ if $\mathbf{X} = (\mathbf{X}_0\mathbf{W})$, which occur when $\mathbf{X}'\mathbf{d} = 0$.

To show that $\mathbf{X}'\mathbf{d} = 0$, first rewrite $\mathbf{d}$ as

$$\begin{aligned}\mathbf{d} = \mathbf{DY} - \mathbf{DW}\hat{\gamma}_{FE} &= \mathbf{DY} - \mathbf{DW}((\mathbf{W}'\mathbf{DW})^{-1}\mathbf{W}'\mathbf{DY}) \\ &= (\mathbf{I}_{nT} - \mathbf{DW}(\mathbf{W}'\mathbf{DW})^{-1}\mathbf{W}')\mathbf{DY}.\end{aligned}$$

Therefore,

$$\mathbf{X}'\mathbf{d} = (\mathbf{X}_0\mathbf{W})'\mathbf{d} = (\mathbf{X}_0\mathbf{W})'(\mathbf{I}_{nT} - \mathbf{DW}(\mathbf{W}'\mathbf{DW})^{-1}\mathbf{W}')\mathbf{DY}, \tag{A.10}$$

in which the first n rows of $\mathbf{X'd}$ are

$$\mathbf{X}_0'(\mathbf{I}_{nT} - \mathbf{DW}(\mathbf{W'DW})^{-1}\mathbf{W'})\mathbf{DY} = (\mathbf{X}_0'\mathbf{I}_{nT} - \mathbf{X}_0'\mathbf{DW}(\mathbf{W'DW})^{-1}\mathbf{W'})\mathbf{DY}.$$

Since $\mathbf{D} = \mathbf{D}'$ by symmetry,

$$(\mathbf{X}_0'\mathbf{I}_{nT} - \mathbf{X}_0'\mathbf{DW}(\mathbf{W'DW})^{-1}\mathbf{W'})\mathbf{DY} = (\mathbf{X}_0'\mathbf{I}_{nT} - \mathbf{X}_0'\mathbf{D'W}(\mathbf{W'DW})^{-1}\mathbf{W'})\mathbf{D'Y}.$$

Because $\mathbf{X}_0'\mathbf{D}' = (\mathbf{DX})' = 0'$, it follows that

$$\begin{aligned}\mathbf{X'd} = (\mathbf{X}_0'\mathbf{I}_{nT} - \mathbf{X}_0'\mathbf{D'W}(\mathbf{W'DW})^{-1}\mathbf{W'})\mathbf{D'Y} &= (\mathbf{X}_0'\mathbf{I}_{nT} - 0')\mathbf{D'Y} \\ &\equiv (\mathbf{X}_0'\mathbf{I}_{nT})\mathbf{D'Y} \\ &= \mathbf{X}_0'\mathbf{D'Y} = 0\end{aligned}$$

which is an $(n \times 1)$ column of zeros.

Thus, the following k rows in $\mathbf{X'd}$, equation (A.10), are

$$\begin{aligned}&\mathbf{W}'(\mathbf{I}_{nT} - \mathbf{DW}(\mathbf{W'DW})^{-1}\mathbf{W'})\mathbf{DY} \\ &= (\mathbf{W'I}_{nT} - \mathbf{W'DW}(\mathbf{W'DW})^{-1}\mathbf{W'})\mathbf{DY} \\ &= (\mathbf{W}' - \mathbf{I}_k\mathbf{W}')\mathbf{DY} = (\mathbf{W}' - \mathbf{W}')\mathbf{DY} = 0,\end{aligned}$$

and thereby $\mathbf{X}$ is orthogonal to $\mathbf{d}$ and it follows that the FE and LSDV estimators yield identical estimates of $\boldsymbol{\beta}$. ■

A.2 Sampling Methodology

GfK Sweden's Consumer Scan is based on a sample of 3000 households. The sample of households and weighting structures are selected to achieve representativity and market coverage with respect to consumption and purchasing behaviour for selected product categories.

The variables used for stratification are *region* and *household size*. The region variable ensure geographical spread and coverage. GfK have found household size to be the most important predictor of consumption and purchasing behaviour on the household level. The groups of the stratification variables are shown in Table 6.

Larger Stockholm, Larger Gothenburg and Öresund (Helsingborg and Malmö) are metropolitan areas while the remaining four are traditionally defined market regions. With 7 regions and five household classes, a total of 35 strata are created from which households are randomly sampled.

Table 6: Stratification variables for sampling of panelists

Stratification Variable	
Region	Household size
North	1 person household
East	2 person household
Larger Stockholm	3 person household
West	4 person household
Larger Gothenburg	5+ person household
South	
Öresund	

Notes: The regions are mutually exclusive in spatial coverage and cover the whole of Sweden.

GfK have observed that the variance of the estimators of purchasing behaviour increase for larger households. The reason may be that individual differences in purchasing are accumulated for larger households. Since the standard errors are directly proportionate to the stratum variances, a consequence is that larger household contribute more to the overall standard error of the mean for a given estimator across all household groups. If the panel size of 3000 households would be proportionally allocated to the 35 strata according to Sweden's actual distributions of household size, single person households would make up a large share of the sample. The low share of large households would increase their and the overall estimator variances further. Thus, GfK use *optimum allocation*, which is an allocation rule that minimises the standard error of the mean for a given estimator. The optimum allocation regarding the sample size n for stratum h, $h = 1, \ldots, L$, is given by

$$n_h = \frac{(N_h/N) \times S_h}{\sum_{h=1}^{L}(N_h/N) \times S_h} \times n \tag{A.11}$$

where

$$\begin{aligned} L &= \text{count of strata} \\ N &= \text{sum of all stratum sample sizes} \\ n &= \text{total sample size } h \\ N_h &= \text{number of households in stratum } h \\ s_h &= \text{standard deviation in stratum } h \end{aligned}$$

where $n = n_1 + \ldots + n_H$. Note that n_h is different from the real proportion of group h, given by $P_h = N_h/N$, so equation (A.11) equals

$$n_h = \frac{P_h \times S_h}{\sum_{h=1}^{5} P_h \times S_h} \times n. \tag{A.12}$$

The effect is that the number of single person households in the sample are reduced while additional households are selected from the household groups with large variance. Thereby the variances for the estimators are decreased.

The distributions of auxiliary variables not used for stratification for the obtained sample are checked to ensure representativity towards the population. Weighting variables are applied to adjust the panel composition with reference to the actual population distributions, obtained from SCB. The weighting variables used for the sample of households are *region*, *household size*, *age of reference person*, *having children 0-6 years* and *having children 0-12 years*.

A.3 Tests for Model Selection and Misspecification

Since the OLS FE model and the PFE model are not nested, standard methods to compare models such as the classical likelihood ratio (LR) test cannot be used for comparison. Vuong (1989) developed a two-step testing procedure for comparison of non-nested models. The first step is a variance test of *indistinguishability* of the models fits to the focal population. The second step is a LR test of model comparison. The LR test's test statistic is

$$\text{LR}_{NN} = \frac{1}{\sqrt{n}\omega}(\mathcal{L}_f(\hat{\alpha}) - \mathcal{L}_g(\hat{\beta})) \tag{A.13}$$

where

$$\omega^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{L}_f(y_i \mid x_i, \hat{\alpha}) - \mathcal{L}_g(y_i \mid x_i, \hat{\beta})\right)^2 - \Big(\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_f(y_i \mid x_i, \hat{\alpha}) - \mathcal{L}_g(y_i \mid x_i, \hat{\beta})\Big)^2. \tag{A.14}$$

The test statistic is computed by dividing the average difference of the log likelihood functions at their ML estimates with the estimated standard error of the average difference. The latter is simply the standard deviation of the unit differences in log likelihood divided by the square root of n. Under H_0, the test statistic is asymptotically standard normal distributed (Winkelmann, 2008).

The variance test aims to select the model that is closest to the true conditional distribution by using the null hypothesis that the models are indistinguishably close,

$$H_0 = \mathbb{E}_0\big[\mathcal{L}_f(y_i \mid x_i, \hat{\alpha}) - \mathcal{L}_g(y_i \mid x_i, \hat{\beta})\big] = 0. \tag{A.15}$$

The LR test then shows if one of the model fits the data better. The test cannot determine if the closest model is the true model.

Since the computation of the test statistic requires that the models are estimated with ML, the OLS FE model is refitted as a GLM with an identity link function. This approach is valid since OLS and ML estimates are equivalent for Gaussian models. A finite sample correction is made to the test statistic based on models' AIC. Two tests are made: 1) a variance test that the models are indistinguishable, and 2) a non-nested likelihood ratio test with H_0 that the model fits are equal for the focal population, and H_{1A} that the OLS FE model fits the data better than the PFE model and H_{1B} that the PFE model fits the data better than the OLS FE model. The results in Table 7 show that variance test's hypothesis that the models are indistinguishable are not rejected for any of the retailers' models but the LR test results indicate that the PFE model has a superior fit to the data. The p-values of the LR test are for H_{A1} since in all cases where H_{A1} was rejected, H_{1B} was not, and their test statistics z are the same.

A.4 Tests for Heteroscedasticity

The Breusch-Pagan test (Breusch & Pagan, 1979) is a Lagrange Multiplier (LM) test for evaluating if the errors in a linear regression model are heteroscedastic. The test considers the errors heteroscedastic if the variance of the errors are dependent on independent variables' values. The procedure of the test for panel data is as

Table 7: Vuong test for model selection and misspecification

Variable	Retailer	Variance test		Non-nested LR test	
		p-value	ω^2	p-value	z
Shopping trips	Coop Extra	0.5	2,844,378	<0.0001***	26.212
	ICA Maxi	0.5	2,032,063	<0.0001***	25.975
	ICA Supermarket	0.5	1,949,271	<0.0001***	26.562
	Willys	0.5	2,566,125	<0.0001***	25.280
Coffee shopping trips	Coop Extra	0.5	2,665,746	<0.0001***	27.392
	ICA Maxi	0.5	13,009,981	<0.0001***	23.610
	ICA Supermarket	0.5	4,294,820	<0.0001***	26.150
	Willys	0.5	11,064,615	<0.0001***	25.092

Significance levels: ***, $p < 0.01$, **, $p < 0.05$, * $p < 0.1$

Notes: This table show the results from the Vuong test of misspecification for OLS fixed effects (OLS FE) models and the Poisson fixed effects (PFE) models. Two tests are made: 1) a variance test that the models are indistinguishable, and 2) a non-nested likelihood ratio test with H_0 that the model fits are equal for the focal population, and H_A that the OLS FE model fits the data better than the PFE model. The variable column refer to the dependent variable used in the model since separate models are fitted for shopping trips and coffee shopping trips.

follows: Fit a pooled linear regression model and obtain the residuals

$$e_{it} = y_{it} - \hat{\alpha}_i - \hat{\beta}_1 \mathbf{x}_{1it} + \ldots + \hat{\beta}_k \mathbf{x}_{kit},\ j = 1, \ldots, k \tag{A.16}$$

which will have a zero mean. If the residuals' variance does not depend on $\mathbf{x}_{it}$, their variance can be estimated as the average squared values of the residuals. If their variance does depend on $\mathbf{x}_{it}$, their variance can be modelled as a linear function of $\mathbf{x}_{it}$ by using the auxiliary regression

$$e_{it}^2 = \gamma_0 + \gamma_1 \mathbf{x}_{1it} + \ldots + \gamma_k \mathbf{x}_{kit} + v_{it}. \tag{A.17}$$

The LM test statistic is given by the sample size n times the coefficient of determination for the auxiliary regression,

$$LM = nR_{e^2}^2. \tag{A.18}$$

The test statistic is asymptotically χ_k^2-distributed under the null hypothesis of homoscedasticity,

$$\begin{aligned} &H_0 : \gamma_1 = \ldots = \gamma_k = 0 \\ &H_A : \gamma_j \neq 0,\ \text{for any } j, j = 1, \ldots, k. \end{aligned}$$

The results in Table 8 show that heteroscedasticity robust standard errors are needed.

A.5 Tests for Serial Correlation

Wooldridge (2002) provide general tests for serial correlation in FE models. If the pooled model's errors are uncorrelated, the FE model will have negatively serially correlated errors with $\mathbb{C}orr[e_{it}, e_{is}] = -1/(T-1)$ for each t, s. The correlation decrease for $T \to \infty$ (Croissant & Millo, 2008) and is therefore appropriate for panels with large T. Durbin-Watson and Breusch-Godfrey LM tests can thereby be made on the time demeaned data to test for serial correlation in FE models.

The Durbin-Wu statistic d_p for panel data for $H_0 : \rho = 0$ is given by

$$d_p = \frac{\sum_{i=1}^{n} \sum_{t=2}^{T} (\tilde{e}_{it} - \tilde{e}_{i,t-1})^2}{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{e}_{it}^2} \tag{A.19}$$

where $\tilde{e}_{it}$ denote the within transformed idiosyncratic error component. Under the assumption of $e_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$, the test statistic follows a linear combination of chi-

Table 8: Test of heteroscedastic errors

Variable	Model	Retailer	p-value	df	LM
Shopping trips	OLS FE	Coop Extra	$< 0.0001^{***}$	11	74.168
		ICA Maxi	$< 0.0001^{***}$	11	69.417
		ICA Supermarket	$< 0.0001^{***}$	11	86.628
		Willys	$< 0.0001^{***}$	11	67.058
	PFE	Coop Extra	$< 0.0001^{***}$	18	110.77
		ICA Maxi	$< 0.0001^{***}$	18	55.585
		ICA Supermarket	$< 0.0001^{***}$	18	57.695
		Willys	$< 0.0001^{***}$	18	54.735
Coffee shopping trips	OLS FE	Coop Extra	$< 0.0001^{***}$	11	63.944
		ICA Maxi	$< 0.0001^{***}$	11	50.533
		ICA Supermarket	$< 0.0001^{***}$	11	41.917
		Willys	$< 0.0001^{***}$	11	64.032
	PFE	Coop Extra	$< 0.0001^{***}$	18	164.28
		ICA Maxi	$< 0.0001^{***}$	18	108.04
		ICA Supermarket	$< 0.0001^{***}$	18	130.48
		Willys	$< 0.0001^{***}$	18	127.55

Significance levels: ***, $p < 0.01$, **, $p < 0.05$, * $p < 0.1$

Notes: This table show the results from the Breusch-Pagan test of heteroscedasticity in the error terms in the OLS fixed effects (OLS FE) and Poisson fixed effects (PFE) models. The variable column refer to the dependent variable used in the model since separate models are fitted for shopping trips and coffee shopping trips.

squared variables. The Durbin-Wu test assumes that e_{it} follows an AR(1) process and requires the regressors to be non-stochastic.

The Breusch-Godfrey LM test is more general as it does not make the AR(1) or non-stochastic regressors assumptions. It uses the obtained within transformed residuals from a fitted model to test whether the auxiliary model

$$\tilde{e}_{it} = \gamma_0 + \gamma_1 \tilde{\mathbf{x}}_{1it} + \rho_1 \tilde{e}_{i,t-1} + \gamma_2 \tilde{\mathbf{x}}_{2it} + \rho_2 \tilde{e}_{i,t-2} + \ldots + \gamma_k \tilde{\mathbf{x}}_{kit} + \rho_p \tilde{e}_{p,t-2} + v_{it} \quad (A.20)$$

holds. The test has the standard LM test statistic as shown in equation (A.18) and the same null hypothesis as the Durbin-Wu test. The number of observations available for the auxiliary regression (A.20) is $n = T - p$ so n depends on the number of lags tested p.

Table 9 presents the results from the Breusch-Godfrey (BG) LM test and the Durbin-Wu (DW) test. The BG test has a max order of 4, corresponding to a test of serial correlation up to a month since it is assumed that some households purchase behaviour follow a monthly pattern. The DW test is two-sided with normal approximated p-values.

The results in Table 9 indicate there is no serial correlation in the errors in the models for coffee shopping trips. The test results are inconclusive for shopping trips models. The Breusch-Godfrey test rejects the null of no serial correlation in all models, but the panel Durbin-Watson only rejects the null for the OLS FE model for Coop's and the PFE mode for Willys. The Durbin-Watson test assumes AR(1) errors, and an inspection show that the models' errors follow different ARIMA structures. Thus, the Breusch-Godfrey test results may be more accurate.

An inspection of the residual autocorrelation functions reveal that the OLS FE model's residuals for Coop's shopping trips are serially correlated in regions East, Stockholm, South and Öresund. The OLS FE residuals are correlated for Willys shopping trips in Stockholm while the PFE residuals are correlated in West. ICA Maxi have correlated residuals in East in both models. No regions have significantly serially correlated residuals in both shopping trips models for ICA Supermarket. There is no significant serial correlation in the residuals for any coffee shopping trips model.

Overall, there are indications that autocorrelation consistent robust standard errors are need for the shopping trips models, but necessarily for the coffee shopping trips models.

Table 9: P-values from tests of serial correlation in the errors

Variable	Model	Retailer	BG	DW
Shopping trips	OLS FE	Coop Extra	<0.0001***	<0.0001***
		ICA Maxi	<0.0001***	0.1826
		ICA Supermarket	<0.0001***	0.5560
		Willys	<0.0001***	0.2696
	PFE	Coop Extra	0.0017***	0.3464
		ICA Maxi	<0.0001***	0.2200
		ICA Supermarket	<0.0001***	0.3004
		Willys	<0.0001***	0.0535*
Coffee shopping trips	OLS FE	Coop Extra	0.8703	0.5297
		ICA Maxi	0.3784	0.7040
		ICA Supermarket	0.7938	0.4578
		Willys	0.2448	0.0715*
	PFE	Coop Extra	0.0985*	0.6146
		ICA Maxi	0.0481**	0.0415**
		ICA Supermarket	0.0291**	0.0299**
		Willys	0.0499**	0.2802

Significance levels: ***, $p < 0.01$, **, $p < 0.05$, * $p < 0.1$

Notes: This table show p-values from Breusch-Godfrey (BG) and Durbin-Watson (DW) tests for serial correlation in the error terms of the OLS fixed effects (OLS FE) and Poisson fixed effects (PFE) models. The variable column refer to the dependent variable used in the model since separate models are fitted for shopping trips and coffee shopping trips.

A.6 Tests for Cross-Sectional Dependence

In spatial statistics literature, cross-sectional dependence is typically measured with a spatial matrix that defines the correlation structure. This method is not useful in economic applications where a measure of space is not given and the correlation is unknown. Econometricians have developed tests of cross-sectional dependence that do not require an a priori specification of the correlation structure (Pesaran, 2004).

Breusch and Pagan (1980) proposed a LM test with the null hypothesis of zero cross equation error term correlations, based on the average pairwise correlation of OLS residuals. Specifically, define the LM statistic

$$CD_{lm} = T \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{\rho}_{ij}^2, \tag{A.21}$$

where $\hat{\rho}_{ij}$ is the sample estimate of the pairwise correlation of the residuals (Pesaran, 2004), i.e

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} \tilde{e}_{it}\tilde{e}_{jt}}{\sqrt{\left(\sum_{t=1}^{T} \tilde{e}_{it}^2\right)\left(\sum_{t=1}^{T} \tilde{e}_{it}\tilde{e}_{jt}^2\right)}}. \tag{A.22}$$

Breusch-Pagan showed that under the null hypothesis

$$H_0 : Cov(\epsilon_{it}, \epsilon_{jt} = 0, \text{ for all } i \neq j,$$

where i, j refer to cross-sectional units, the CD_{LM} statistic is asymptotically chi-square distributed with $n(n-1)/2$ degrees of freedom. The Breusch-Pagan test is valid for panels with small n and large T and is thus suitable for the dataset.

The results in Table 10 show that the null hypothesis of no spatial correlation in the errors are rejected for all models except for Coop's coffee shopping trips OLS FE model. The OLS FE models errors are positively correlated across regions while the correlation is negative in the PFE model, except for one case. The results show that spatial correlation consistent robust standard errors are needed for inference.

A.7 Tests for Panel Unit Root

A necessary condition for valid time series regressions is stationarity. A process Y_t is *strongly stationary* if the cumulative distribution function $F_{\mathbf{Y}}(t_1 + s, \ldots, t_n + s) = F_{\mathbf{Y}}(t_1, \ldots, t_n)$ for any time indices $(t_1, \ldots, t_n)$ and all time lags s. Thus, it is strongly stationary if the joint distribution of the vector $(y_{t_1+s}, \ldots, y_{t_{n+s}})$ equals

Table 10: Tests of cross-sectional dependence in the errors

Variable	Model	Retailer	p-value	CD	df	ρ
Shopping trips	OLS FE	Coop Extra	$< 0.0001^{***}$	158.29	21	0.0826
		ICA Maxi	$< 0.0001^{***}$	54.002	21	0.0249
		ICA Supermarket	$< 0.0001^{***}$	76.03	21	0.1389
		Willys	$< 0.0001^{***}$	58.582	21	0.0789
	PFE	Coop Extra	$< 0.0001^{***}$	117.82	21	-0.1446
		ICA Maxi	$< 0.0001^{***}$	97.25	21	-0.1549
		ICA Supermarket	$< 0.0001^{***}$	79.87	21	-0.1556
		Willys	$< 0.0001^{***}$	108.87	21	-0.1581
Coffee shopping trips	OLS FE	Coop Extra	0.5416	19.68	21	0.0375
		ICA Maxi	$< 0.0001^{***}$	124.37	21	0.1414
		ICA Supermarket	$< 0.0001^{***}$	108.63	21	0.1564
		Willys	$< 0.0001^{***}$	114.95	21	0.1958
	PFE	Coop Extra	0.0187^{**}	36.61	21	-0.0971
		ICA Maxi	$< 0.0001^{***}$	121.50	21	-0.0625
		ICA Supermarket	$< 0.0001^{***}$	66.18	21	-0.0139
		Willys	$< 0.0001^{***}$	147.39	21	0.2330

Significance levels: ***, $p < 0.01$, **, $p < 0.05$, * $p < 0.1$

Notes: This table show the results from the Breusch-Pagan test of cross-sectional (spatial) dependence in the error terms of the OLS fixed effects (OLS FE) and Poisson fixed effects (PFE) models. The variable column refer to the dependent variable used in the model since separate models are fitted for shopping trips and coffee shopping trips.

the joint distribution of $(y_{t_1}, \ldots, y_{t_n})$ for all $(t_1, \ldots, t_n)$ and s. This means that the distribution of the sequence of random variables $(y_{t_1}, \ldots, y_{t_n})$ is independent of time. A process y_t is *weakly stationary* if its mean and variance remain unchanged over time, that is $\mathbb{E}[y_t] = \mu_y(t)$ and $\mathbb{V}[y_t] = \sigma^2$, and its covariance $\mathbb{C}ov[y_t, y_{t-s}]$ depends only on lag s, not t. A process is *trend stationary* if there exists a β such that $y_t - \beta t$ is stationary. A trend stationary process has a deterministic trend, which upon removal makes the process stationary.

If the process' characteristic function has a root of 1 or larger, the process has a *unit root*. Such non-stationary processes are denoted $I(1)$, while processes without a unit root are $I(0)$. While both unit root processes and trend stationary processes are non-stationary, trend stationary processes need not have a unit root. In presence of a shock, a trend stationary process will converge to its mean not affected by the shock, while the mean of a unit root process will be permanently changed as they do not converge over time.

This section explain common panel unit roots test that apart from the time dimension also consider the panel's cross-section, since the way in which n and T converge to infinity is important for the asymptotic behaviour of the estimators and tests for non-stationarity (Baltagi, 2005).

Either one can test if each unit's time series are non-stationary or if all units are non-stationary. The Im, Pesaran, and Shin (2005) (IPS) test's null hypothesis is that all units have a unit root but allow heterogenous coefficients. The IPS test alternative hypothesis is that some of the units have unit roots:

$$H_0 : \rho_i = 1 \text{ for all } i$$
$$H_1 : \begin{cases} \rho_i < 1 \text{ for } i = 1, \ldots, N_1 \\ \rho_i = 1 \text{ for } i = N_1 + 1, \ldots, N \end{cases}$$

The test computes ADF test statistics for each unit and combine them by averaging the unit root test, i.e. $\bar{t} = (1/n) \sum_{i=1}^{n} t_{\rho_i}$ for each unit i, where t_{ρ_i} is the unit-specific t-statistic for the null hypothesis. The IPS test requires that $n/T \to 0$ for $n \to \infty$.

Hadri (2000) proposes a residual-based LM test that is a generalisation of the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for time series. The Hadri test has the null hypothesis that the time series is stationarity around a deterministic trend. It uses residuals obtained from a regression of y_{it} on a constant, or a constant plus a trend,

$$y_{it} = r_{it} + \epsilon_{it} \tag{A.23}$$
$$r_{it} = r_{i,t-1} + u_{it}. \tag{A.24}$$

where r_{it} is a random walk and $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ and $u_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_u^2)$ are mutually independent and identically distributed across i and over t. The null hypothesis is that there is no unit root in the time series for any unit i, and the alternative hypothesis is that the panel has a unit root for some i,

$$H_0 : \sigma_u^2 = 0. \tag{A.25}$$

If the variance $u_{it} = 0$, then $r_{it} = r_{i,t-1}$ is a constant and y_{it} is stationary. The Hadri test allows for heteroscedasticity adjustments. The test is well suited for large n and T.

The IPS and Hadri tests are computed for both dependent variables. The tests are for trend-stationarity since the number of shopping trips follow seasonal patterns and a deterministic trend, as seen for Coop in Stockholm in Figure 1. The time series are thereby assumed to be mean reverting. The IPS test uses the limit of 8 lags, corresponding to 8 weeks, and the number of lags for the Hadri test are computed automatically using AIC. The heteroscedasticity consistent version of the Hadri test is used since the Breusch-Pagan test in section A.4 showed signs of heteroscedasticity. The tests' results are provided in Table 11.

The LM test results indicate that one region's shopping trips time series has a unit root and that the coffee shopping trips time series are stationary in each region for all retailers. The IPS test indicate that at least one unit is not trend stationary.

A.8 Software

The statistical computing language **R** is used throughout. The **plm** package (Croissant & Millo, 2008; Millo, Tappe, & Croissant, 2017) is used for estimating the log-linear OLS FE model and **pglm** package (Millo, 2017) is used for the PFE model. The same two packages are used for the panel model tests of heteroscedasticity and cross-sectional and serial correlation in the errors. The robust covariance matrices are estimated with **plm**, **pglm**, and the **sandwich** package (Zeileis, 2004, 2006). The Vuong tests are carried out using the **nonnest2** package (Merkle & You, 2018a, 2018b) and the RMSE and bias metrics are computed using the **Metrics** package (Frasco, 2017). LaTeXformatted tables are generated in **R** with the **stargazer** package (Hlavac, 2015, 2018) and the **tikzDevice** package (Sharpsteen & Bracken, 2018) is used for high resolution figures.

Table 11: P-values from panel unit root tests

Variable	Retailer	IPS	LM
Shopping trips	Coop Extra	<0.0001***	<0.0001***
	ICA Maxi	<0.0001***	<0.0001***
	ICA Supermarket	<0.0001***	<0.0001***
	Willys	<0.0001***	0.0001***
Coffee Shopping trips	Coop Extra	<0.0001***	0.4630
	ICA Maxi	<0.0001***	0.9398
	ICA Supermarket	<0.0001***	0.9835
	Willys	<0.0001***	0.1162

Significance levels: ***, $p < 0.01$, **, $p < 0.05$, * $p < 0.1$

Notes: This table show the panel unit root test results for trend stationarity with region-specific intercepts and trend as exogenous variables for the dependent variables. The Im-Pesaran-Shin (IPS) test have the null hypothesis that each region's time series is trend stationary with the alternative hypothesis that some regions have unit roots, possibly with heterogenous coefficients. The Lagrange Multiplier (LM) test have the null hypothesis that each region's time series is trend stationary around a deterministic trend with no unit root and the alternative hypothesis that the panel has a unit root. The LM test is heteroscedasticity consistent.

A.9 National Industry Level Summary Statistics

Table 12 show descriptive statistics for the grocery retailers that the households in the panel have purchased from during the time period, including retailers not included in this dataset. The industry time index variable used in the OLS FE model is calculated on these data.

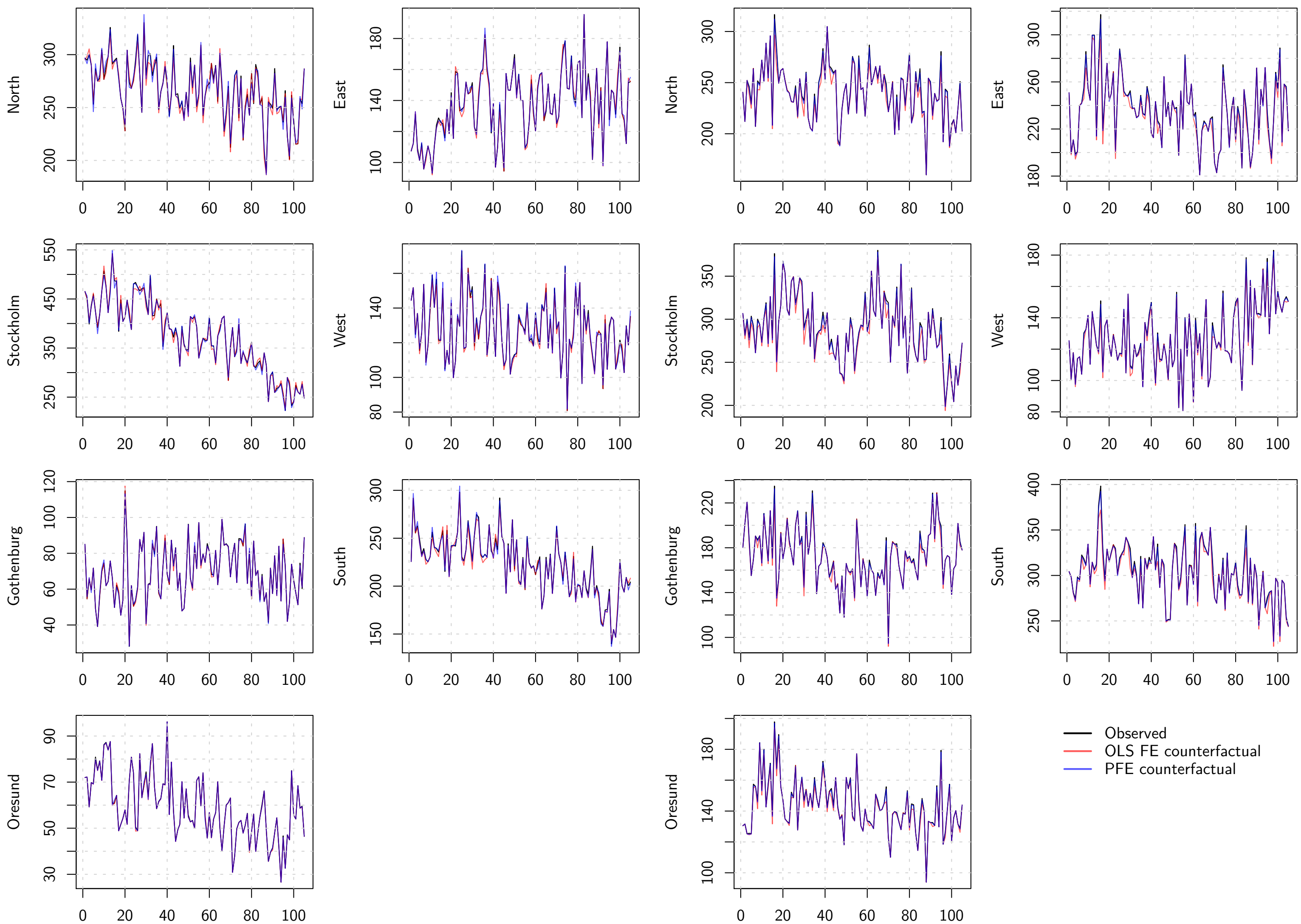
Table 12: Descriptive statistics for retailer and regional aggregates

Variable	Mean	St. Dev.	Min	Max
Shopping trips	12,879.070	522.014	11,587.800	14,427.700
Coffee shopping trips	805.211	275.432	368.740	1,652.723
Promotion trips share				
Gevalia	47.899	23.454	9.263	91.289
Löfbergs	50.597	25.229	0.000	93.888
Zoegas	52.225	20.281	8.783	87.868
Classic	57.544	27.674	0.000	97.724
Other brands	42.751	15.210	8.915	73.980

Notes: The table show the national aggregate across regions and retailers, corresponding to the national industry level data. Shopping trips and coffee shopping trips are in thousands. Promotion trips shares are in percentages.

A.10 Regional Cumulative Effects Plots

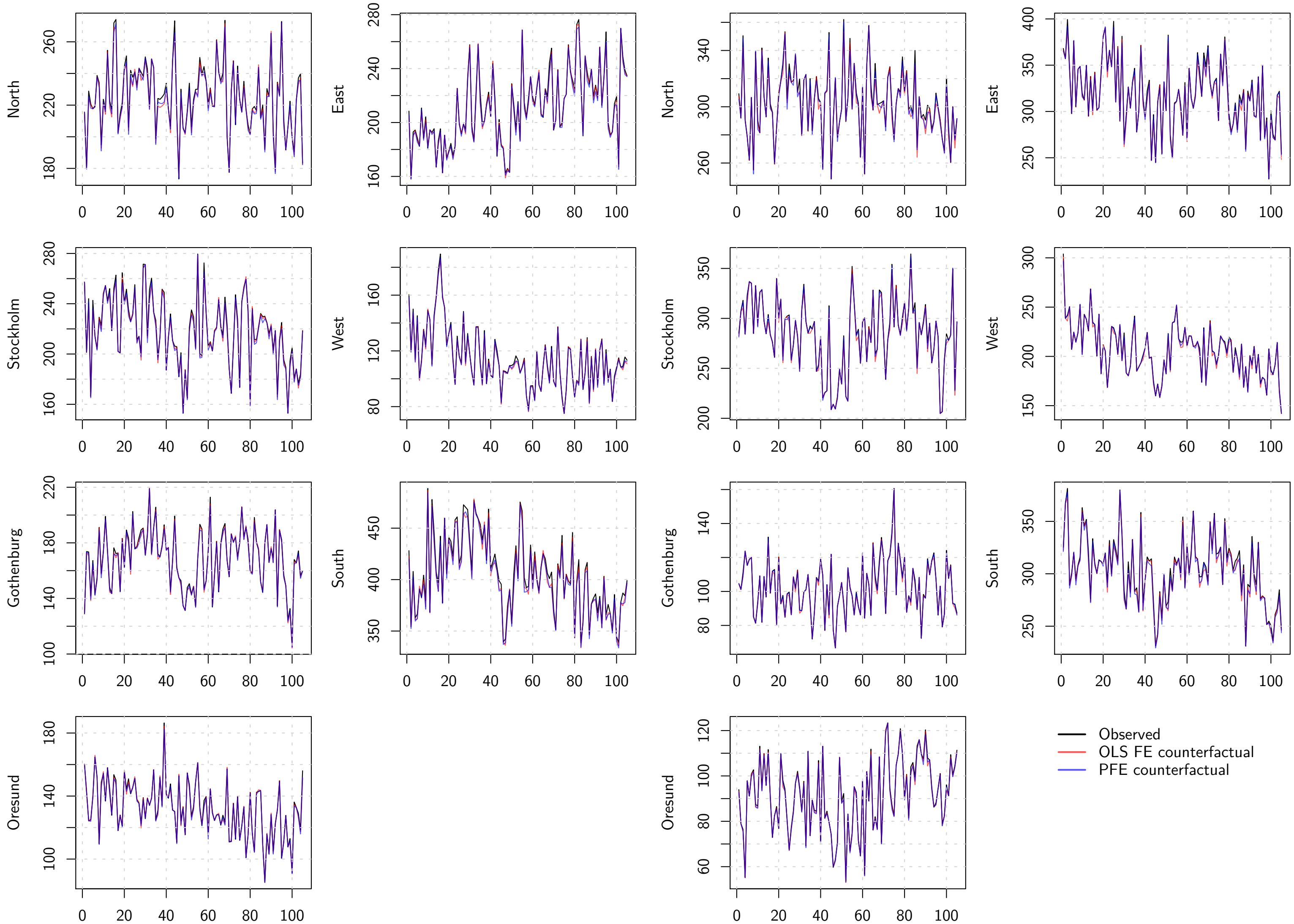
Time series of the observed and counterfactual outcomes per region are shown in Figure 8-11. The estimated counterfactual time series follow the seasonality and trend in the observed outcomes well with some variation across regions. For instance, for ICA Maxi in Figure 9a, the difference between the time series seems to be larger in North and South. The differences are even more visible for the coffee shopping trips time series in Figure 10 to 11.



(a) Coop Extra

(b) Willys

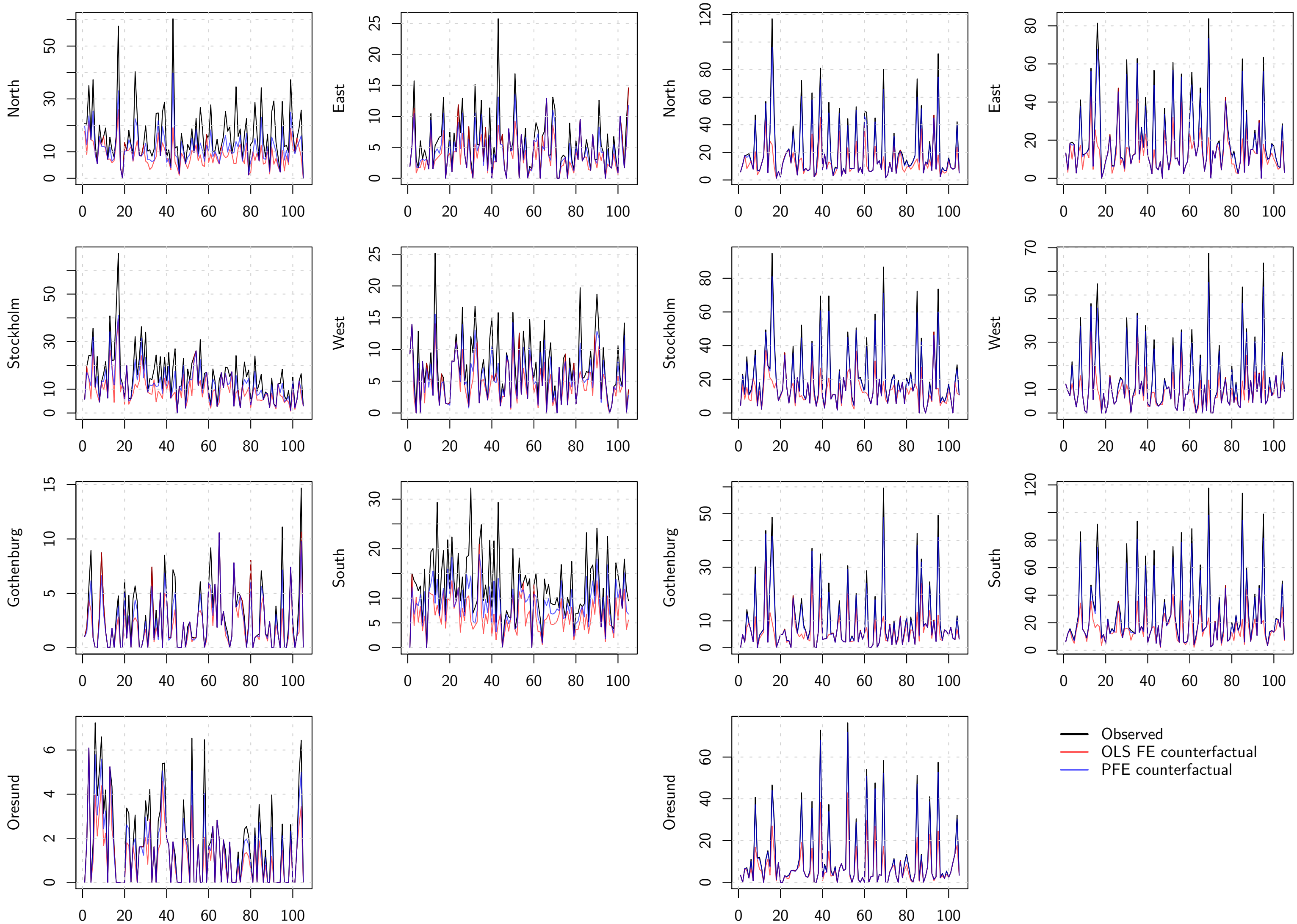
Figure 8: Shopping trips: regional observed and estimated counterfactual number of shopping trips (in 1000) under no promotion on any ground coffee brand per retailer. The difference between the observed and the counterfactual time series is the corresponding model's estimated joint impact of promotions on the brands per region. The x-axis is week t relative to week 36 in 2015 ranging to week 35 in 2017.



(a) ICA Maxi

(b) ICA Supermarket

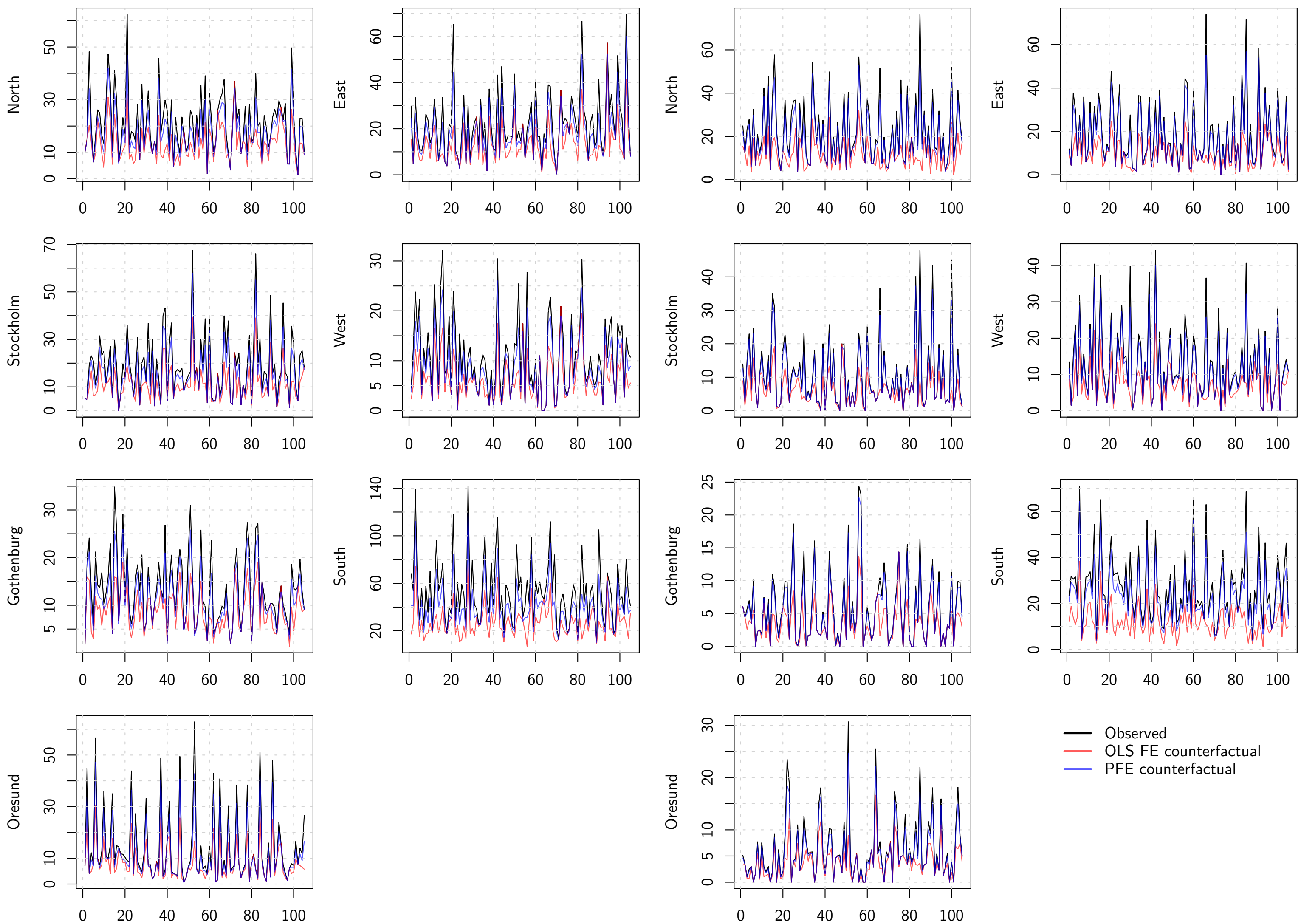
Figure 9: Shopping trips: regional observed and estimated counterfactual number of shopping trips (in 1000) under no promotion on any ground coffee brand per retailer. The difference between the observed and the counterfactual time series is the corresponding model's estimated joint impact of promotions on the brands per region. The x-axis is week t relative to week 36 in 2015 ranging to week 35 in 2017.



(a) Coop Extra

(b) Willys

Figure 10: Coffee shopping trips: regional observed and estimated counterfactual number of coffee shopping trips (in 1000) under no promotion on any ground coffee brand per retailer. The difference between the observed and the counterfactual time series is the corresponding model's estimated joint impact of promotions on the brands per region. The x-axis is week t relative to week 36 in 2015 ranging to week 35 in 2017.



(a) ICA Maxi

(b) ICA Supermarket

Figure 11: Coffee shopping trips: regional observed and estimated counterfactual number of coffee shopping trips (in 1000) under no promotion on any ground coffee brand per retailer. The difference between the observed and the counterfactual time series is the corresponding model's estimated joint impact of promotions on the brands per region. The x-axis is week t relative to week 36 in 2015 ranging to week 35 in 2017.