Replication material for "Spraying for the Beauty: Pesticide Use for Visual Appearance in Apple Production"

Software

Author(s):
Zachmann, Lucca [iD]

Publication date:
2024-05-03

Permanent link:
https://doi.org/10.3929/ethz-b-000668506

Rights / license:
Apache License 2.0

Code and Replication for 'Spraying for the Beauty: Pesticide Use for Visual Appearance in Apple Production'

Lucca Zachmann,* Chloe McCallum, and Robert Finger

Agricultural Economics and Policy Group, ETH Zurich

2024-05-02 | validate version 1.1.3

Contents

---

*Corresponding e-mail: lzachmann@ethz.ch

1 Licence

This script can be used for use, reproduction, and distribution under the Apache License 2.0 terms and conditions.

2 Loading data

The data used for this article comes from https://www.research-collection.ethz.ch/handle/20.500.11850/602779 and is publicly available.

```
path <-
"https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/602779/rc_dataset.csv"
data <- read.table(path, sep = ";", header = TRUE, na.string = "NA")
```

However, for the regression analysis we rely on data that is not publicly available in order to maintain the farms privacy (i.e. variables related to a farms' location). Thus, we provide the regression sample (regression_data.dta) separately with all variables needed to conduct the regression analysis. Note that the main regressions are conducted in Stata.

3 Loading packages

Several packages are used.

```
if (!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse, stargazer, conflicted, DirectEffects, table1, multiwayvcov,
lmtest, ggcorrplot, foreign, ggpubr, ggrepel, kableExtra, margins)

knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE, booktabs = TRUE)

# Conflicts
conflict_prefer("filter", "dplyr")
conflict_prefer("select", "dplyr")
```

4 Data preparation

```
# Total land
rep_data <- data %>%
    select(id, starts_with("VarArea."), -contains("oth"))

# Reshape wide -> long
rep_data_long <- rep_data %>%
    gather(apple, value, -c(id))

rep_data_long$value <- as.numeric(rep_data_long$value)

rep_data_long <- rep_data_long %>%
```

```
    group_by(id) %>%
    summarise(tot_apple_land = sum(value, na.rm = TRUE))

data <- merge(data, rep_data_long, by = "id")

## Total land
dataset_sub <- subset(data, tot_apple_land < 40000)  # Drop three observations that show
impossibly large land

# Total land under apples
sum(dataset_sub$tot_apple_land)/100
```

```
## [1] 899.1413
```

```
# Share of land under apples from the survey compared to
# total land under apples
(sum(dataset_sub$tot_apple_land)/100)/3685.16
```

```
## [1] 0.2439898
```

```
## Female farmers
data$female <- 0
data$female[data$gender == "f"] <- 1

## Rescale farmsize
data$farmsize_scaled <- data$farmsize/100

## Share of female growers
sum(data$female)/nrow(data)
```

```
## [1] 0.06938776
```

```
## Share of organic producers
sum(data$production.A002.)/nrow(data)
```

```
## [1] 0.2
```

```
## Age
data$age <- 2022 - data$DOB
data$above_50 <- 0
data$above_50[data$age >= 50] <- 1

## Share of producers older than 50 years of age
sum(data$above_50)/nrow(data)
```

```
## [1] 0.555102
```

```r
## Data prep
data <- data %>%
    filter(tot_apple_land < 40000) %>%
    mutate(NumverOfClubVarieties = VarGrown.wellant + VarGrown.tessa +
        VarGrown.tentation + VarGrown.sweettango + VarGrown.magic_star +
        +VarGrown.sonnenglanz + VarGrown.scilate + VarGrown.rubens +
        VarGrown.redlove + VarGrown.red_boy + VarGrown.rave +
        VarGrown.parsi + VarGrown.kanzi + VarGrown.natyra + VarGrown.mairac +
        VarGrown.kissabel + VarGrown.kalei + VarGrown.greenstar +
        VarGrown.grafin_von_goldach + VarGrown.gradyrose + VarGrown.galiwa +
        VarGrown.fuji + VarGrown.evelina + VarGrown.enterprise +
        VarGrown.delcorf + VarGrown.coralise + VarGrown.cameo +
        VarGrown.cabarette + VarGrown.belchard + VarGrown.antares +
        VarGrown.ambassy + VarGrown.alegro + VarGrown.Jazz. +
        VarGrown.CrippsPink. + VarGrown.Milwa.)

data <- data %>%
    mutate(AreaOfClubVarieties = ifelse(!is.na(VarArea.wellant),
        VarArea.wellant, 0) + ifelse(!is.na(VarArea.tessa), VarArea.tessa,
        0) + ifelse(!is.na(VarArea.tentation), VarArea.tentation,
        0) + ifelse(!is.na(VarArea.sweettango), VarArea.sweettango,
        0) + ifelse(!is.na(VarArea.magic_star), VarArea.magic_star,
        0) + ifelse(!is.na(VarArea.sonnenglanz), VarArea.sonnenglanz,
        0) + ifelse(!is.na(VarArea.scilate), VarArea.scilate,
        0) + ifelse(!is.na(VarArea.rubens), VarArea.rubens, 0) +
        ifelse(!is.na(VarArea.redlove), VarArea.redlove, 0) +
        ifelse(!is.na(VarArea.red_boy), VarArea.red_boy, 0) +
        ifelse(!is.na(VarArea.rave), VarArea.rave, 0) + ifelse(!is.na(VarArea.parsi),
        VarArea.parsi, 0) + ifelse(!is.na(VarArea.kanzi), VarArea.kanzi,
        0) + ifelse(!is.na(VarArea.natyra), VarArea.natyra, 0) +
        ifelse(!is.na(VarArea.mairac), VarArea.mairac, 0) +
        ifelse(!is.na(VarArea.kissabel),
        VarArea.kissabel, 0) + ifelse(!is.na(VarArea.kalei),
        VarArea.kalei, 0) + ifelse(!is.na(VarArea.greenstar),
        VarArea.greenstar, 0) + ifelse(!is.na(VarArea.grafin_von_goldach),
        VarArea.grafin_von_goldach, 0) + ifelse(!is.na(VarArea.gradyrose),
        VarArea.gradyrose, 0) + ifelse(!is.na(VarArea.galiwa),
        VarArea.galiwa, 0) + ifelse(!is.na(VarArea.fuji), VarArea.fuji,
        0) + ifelse(!is.na(VarArea.evelina), VarArea.evelina,
        0) + ifelse(!is.na(VarArea.enterprise), VarArea.enterprise,
        0) + ifelse(!is.na(VarArea.delcorf), VarArea.delcorf,
        0) + ifelse(!is.na(VarArea.coralise), VarArea.coralise,
        0) + ifelse(!is.na(VarArea.cameo), VarArea.cameo, 0) +
        ifelse(!is.na(VarArea.cabarette), VarArea.cabarette,
            0) + ifelse(!is.na(VarArea.belchard), VarArea.belchard,
        0) + ifelse(!is.na(VarArea.antares), VarArea.antares,
        0) + ifelse(!is.na(VarArea.ambassy), VarArea.ambassy,
        0) + ifelse(!is.na(VarArea.alegro), VarArea.alegro, 0) +
        ifelse(!is.na(VarArea.Jazz.), VarArea.Jazz., 0) +
        ifelse(!is.na(VarArea.CrippsPink.),
        VarArea.CrippsPink., 0) + ifelse(!is.na(VarArea.Milwa.),
        VarArea.Milwa., 0))
```

```r
# Share of land under club varieties
sum(data$AreaOfClubVarieties)/sum(data$tot_apple_land)
```

```
## [1] 0.1981964
```

```r
data <- data %>%
    mutate(club_IS = if_else(VarGrown.Jazz. == 1 | data$VarGrown.Milwa. ==
        1 | VarGrown.CrippsPink. == 1 | VarGrown.greenstar ==
        1 | VarGrown.kanzi == 1 | VarGrown.mairac == 1 | VarGrown.sweettango ==
        1 | VarGrown.tentation == 1 | VarGrown.fuji == 1 | VarGrown.rubens ==
        1 | VarGrown.scilate == 1 | VarGrown.cameo == 1 | VarGrown.evelina ==
        1 | VarGrown.antares == 1 | VarGrown.redlove == 1 | VarGrown.rockit ==
        1, 1, 0))

data <- data %>%
    mutate(club_IS_area = ifelse(!is.na(VarArea.Jazz.), VarArea.Jazz.,
        0) + ifelse(!is.na(VarArea.Milwa.), VarArea.Milwa., 0) +
        ifelse(!is.na(VarArea.CrippsPink.), VarArea.CrippsPink.,
            0) + ifelse(!is.na(VarArea.greenstar), VarArea.greenstar,
        0) + ifelse(!is.na(VarArea.kanzi), VarArea.kanzi, 0) +
        ifelse(!is.na(VarArea.mairac), VarArea.mairac, 0) +
        ifelse(!is.na(VarArea.sweettango),
        VarArea.sweettango, 0) + ifelse(!is.na(VarArea.tentation),
        VarArea.tentation, 0) + ifelse(!is.na(VarArea.fuji),
        VarArea.fuji, 0) + ifelse(!is.na(VarArea.rubens), VarArea.rubens,
        0) + ifelse(!is.na(VarArea.scilate), VarArea.scilate,
        0) + ifelse(!is.na(VarArea.cameo), VarArea.cameo, 0) +
        ifelse(!is.na(VarArea.evelina), VarArea.evelina, 0) +
        ifelse(!is.na(VarArea.antares), VarArea.antares, 0) +
        ifelse(!is.na(VarArea.redlove), VarArea.redlove, 0) +
        ifelse(!is.na(VarArea.rockit), VarArea.rockit, 0))

data_red <- data %>%
    select(id, starts_with("VarArea.")) %>%
    replace(is.na(.), 0) %>%
    mutate(TotalAppleArea = rowSums(select(., starts_with("VarArea.")))) %>%
    select(id, TotalAppleArea)

data <- merge(data, data_red, by = "id")

data$share_club_area <- data$AreaOfClubVarieties/data$TotalAppleArea *
    100

# Own consumption of table apples
own_consumption <- data %>%
    filter(Marketing.A1. == 1) %>%
    select(MarketingApples.A6.)

# Share of farmers using table apples for own consumption
1 - nrow(subset(own_consumption, MarketingApples.A6. == 0))/nrow(own_consumption)
```

```
## [1] 0.2091837
```

```
# Average of tables apples used for own consumption
mean(own_consumption$MarketingApples.A6.)
```

```
## [1] 2.911224
```

```
# Other pesticides use
data$herbicides <- ifelse(data$WeedCon.A3. == 1 & data$WeedCon.A1. ==
    0 & data$WeedCon.A2. == 0, 1, 0)
data$fungicides <- ifelse(data$FungCon.A8. == 1 & data$FungCon.A7. ==
    0, 1, 0)
data$insecticides <- ifelse(data$InsectCon.A6. == 1 & data$InsectCon.A8. ==
    0, 1, 0)

# Herbicides
sum(data$herbicides)/nrow(data) * 100
```

```
## [1] 33.8843
```

```
# Fungicides
sum(data$fungicides)/nrow(data) * 100
```

```
## [1] 46.69421
```

```
# Insecticides
sum(data$insecticides)/nrow(data) * 100
```

```
## [1] 28.5124
```

```
# Quality of the apples
data$premium <- ifelse(is.na(data$MarketingQuality.A1.), 0, data$MarketingQuality.A1.)
data$class_one <- ifelse(is.na(data$MarketingQuality.A2.), 0,
    data$MarketingQuality.A2.)
data$class_two <- ifelse(is.na(data$MarketingQuality.A3.), 0,
    data$MarketingQuality.A3.)

data$class_one_new <- data$class_one + data$premium

# Cosmetic applications
data$cosmetic_pesticide <- 0
data$cosmetic_pesticide <- data$CosmeticApp.A2.

# Make regression sample (Only table apples)
reg_data <- data %>%
    filter(Marketing.A1. == 1) %>%
    select(id, cosmetic_pesticide, MarketingApples.A1., MarketingApples.A2.,
        MarketingApples.A3., MarketingApples.A4., MarketingApples.A5.,
        everything())

# Share of farms using cosmetic pesticide applications
round(sum(reg_data$cosmetic_pesticide)/nrow(reg_data) * 100,
    1)
```

```
## [1] 23.5
```

```r
# Share of farms using chemical thinning for size and color
reg_data$ChemicalThinning4sizecolor <- 0
reg_data$ChemicalThinning4sizecolor[(reg_data$ThinningReason.A1. ==
    1 | reg_data$ThinningReason.A2. == 1) & reg_data$Thinning.A3. ==
    1] <- 1
round(sum(reg_data$ChemicalThinning4sizecolor)/nrow(reg_data) *
    100, 1)
```

```
## [1] 59.2
```

```r
# Training
reg_data$training <- ifelse(reg_data$education1.A7. == 1, 1,
    0)

# Private extension
reg_data$private_extension <- ifelse(reg_data$PPPinfo.A4. ==
    1 | reg_data$PPPinfo.A8. == 1, 1, 0)
reg_data$private_extension[reg_data$PPPinfo.A1. == 1 | reg_data$PPPinfo.A2. ==
    1 | reg_data$PPPinfo.A3. == 1 | reg_data$PPPinfo.A5. == 1 |
    reg_data$PPPinfo.A6. == 1 | reg_data$PPPinfo.A7. == 1 | reg_data$PPPinfo.A9. ==
    1 | reg_data$PPPinfo.A10. == 1] <- 0

# Organic (and biodynamic) producers
reg_data$organic <- ifelse(reg_data$production.AOO2. == 1 | reg_data$production.AOO4. ==
    1, 1, 0)

# Optical appearance specified in contracts
reg_data$contract_optical_appearance <- 0
reg_data$contract_optical_appearance[reg_data$ContractSpec.app. ==
    1] <- 1

# Insurance
reg_data$insurance <- ifelse(reg_data$insurance.A4. == 0, 1,
    0)

# Machinery
reg_data$hand_application <- ifelse(reg_data$PPPapply.A1. ==
    1 | reg_data$PPPapply.A6. == 1, 1, 0)
reg_data$hand_application[reg_data$PPPapply.A2. == 1 | reg_data$PPPapply.A3. ==
    1 | reg_data$PPPapply.A4. == 1 | reg_data$PPPapply.A5. ==
    1 | reg_data$PPPapply.A7. == 1] <- 0

# Pest pressure (change 5 erroneous entries)
reg_data$avg_simscab[reg_data$id == 446] <- 1.477755
reg_data$avg_simscab[reg_data$id == 390] <- 2.386572
reg_data$avg_simscab[reg_data$id == 295] <- 2.375607
reg_data$avg_simscab[reg_data$id == 67] <- 2.375607
reg_data$avg_simscab[reg_data$id == 66] <- 2.386572

## Export to Stata format Classify marketing channels
reg_data$direct <- ifelse(!is.na(reg_data$MarketingApples.A4.),
```

```r
    reg_data$MarketingApples.A4./100, 0)  # Direct

# Traders, cooperatives and shops, supermarkets
reg_data$indirect <- ifelse(!is.na(reg_data$MarketingApples.A1.),
    reg_data$MarketingApples.A1./100, 0) + ifelse(!is.na(reg_data$MarketingApples.A2.),
    reg_data$MarketingApples.A2./100, 0) + ifelse(!is.na(reg_data$MarketingApples.A3.),
    reg_data$MarketingApples.A3./100, 0)

# Binary variable for the main marketing channel
reg_data$direct_bin <- 0
reg_data$direct_bin[reg_data$direct > 0.5] <- 1
reg_data$indirect_bin <- 0
reg_data$indirect_bin[reg_data$indirect > 0.5] <- 1

# Treatment variable 1
reg_data$treatment_simple <- 0  #'Direct marketing'
reg_data$treatment_simple[reg_data$direct < 0.75 & reg_data$direct >
    0.25] <- 1  #'Middle'
reg_data$treatment_simple[reg_data$direct < 0.25] <- 2  #'Indirect marketing'

# Some numbers on the marketing channel
nrow(subset(reg_data, treatment_simple == 0))  # Number of farms mainly direct marketing
```

```
## [1] 50
```

```r
nrow(subset(reg_data, treatment_simple == 1))  # Number of farms mainly mixed marketing
```

```
## [1] 32
```

```r
nrow(subset(reg_data, treatment_simple == 2))  # Number of farms mainly marketing via
intermediaries
```

```
## [1] 114
```

```r
# Treatment variable 2
reg_data$treatment_extreme <- 0  #'Direct marketing'
reg_data$treatment_extreme[reg_data$direct < 1 & reg_data$direct >
    0] <- 1  #'Middle'
reg_data$treatment_extreme[reg_data$direct == 0] <- 2  #'Indirect marketing'

# Treatment variable 3
reg_data$treatment_third <- 0  #'Direct marketing'
reg_data$treatment_third[reg_data$direct <= 0.66 & reg_data$direct >
    0.33] <- 1  #'Middle'
reg_data$treatment_third[reg_data$direct <= 0.33] <- 2  #'Indirect marketing'

# Export to stata format (.dta)
stata_data <- reg_data %>%
    select(id, ChemicalThinning4sizecolor, cosmetic_pesticide,
        treatment_simple, treatment_extreme, treatment_third,
```

```
        herbicides, insecticides, fungicides, age, female, farmsize_scaled,
        canton, organic, riskpref.prodrisk., riskpref.markrisk.,
        private_extension, marketingexp.price., marketingexp.qual.,
        marketingexp.trust., labels.Club., contract_optical_appearance,
        earnings2, training, insurance, avg_simscab, LRP, hand_application,
        class_one, class_two, class_one_new, premium) %>%
    rename(marketing_riskpref = riskpref.markrisk., production_riskpref =
    riskpref.prodrisk.,
        exp_prices = marketingexp.price., exp_quality = marketingexp.qual.,
        exp_trust = marketingexp.trust., club = labels.Club.)

write.dta(stata_data, "regression_data_replicated.dta")
```

5   Figures and Tables in the Manuscript

Figure 3 is not replicated here due to data privacy (i.e. farm locations).

5.1   Table 3: Descriptive statistics of the final sample (N=196)

```
# Descriptive statistics
desc_stats <- reg_data %>%
    dplyr::select(cosmetic_pesticide, ChemicalThinning4sizecolor,
        treatment_simple, TotalAppleArea, organic, LRP, labels.Club.,
        contract_optical_appearance, class_one_new, class_two)

desc_stats$ChemicalThinning4sizecolor <- factor(desc_stats$ChemicalThinning4sizecolor,
    levels = c(0, 1), labels = c("No", "Yes"))
desc_stats$cosmetic_pesticide <- factor(desc_stats$cosmetic_pesticide,
    levels = c(0, 1), labels = c("No", "Yes"))
desc_stats$LRP <- factor(desc_stats$LRP, levels = c(0, 1), labels = c("No",
    "Yes"))
desc_stats$treatment_simple <- factor(desc_stats$treatment_simple,
    levels = c(0, 1, 2), labels = c("Mainly direct marketing\n(>75% direct sales)",
        "Mixed marketing\n(25-75% direct sales)", "Mainly marketing via
        intermediarees\n(<25% direct sales)"))
desc_stats$organic <- factor(desc_stats$organic, levels = c(0,
    1), labels = c("No", "Yes"))
desc_stats$labels.Club. <- factor(desc_stats$labels.Club., levels = c(0,
    1), labels = c("No", "Yes"))
desc_stats$contract_optical_appearance <- factor(desc_stats$contract_optical_appearance,
    levels = c(0, 1), labels = c("No", "Yes"))

labels <- list(variables = list(cosmetic_pesticide = "Cosmetic chemical pesticide
application",
    ChemicalThinning4sizecolor = "Cosmetic chemical thinning for size and color",
    TotalAppleArea = "Total land under apples (are)", organic = "Organic production",
    LRP = "Low-residue producer", labels.Club. = "Club apple producer",
    contract_optical_appearance = "Visual appearance specified in contract",
    class_one_new = "Premium & class 1 quality (%)", class_two = "Class 2 quality (%)"),
```

```
    groups = list("", "Marketing channel", ""))

strata <- c(list(Total = desc_stats), split(desc_stats, desc_stats$treatment_simple))

my.render.cont <- function(x) {
    with(stats.apply.rounding(stats.default(x), digits = 2),
        c("", `Mean (SD)` = sprintf("%s (%s)", MEAN, SD)))
}
my.render.cat <- function(x) {
    c("", sapply(stats.default(x), function(y) with(y, sprintf("%d (%0.0f %%)",
        FREQ, PCT))))
}

rndr <- function(x, name, ...) {
    if (length(x) == 0) {
        y <- desc_stats[[name]]
        s <- rep("", length(render.default(x = y, name = name,
            ...)))
        if (is.numeric(y)) {
            p <- t.test(y ~ desc_stats$indirect_bin, var.equal = TRUE,
                alternative = "two.sided")$p.value
        } else {
            p <- chisq.test(table(y, droplevels(desc_stats$indirect_bin)))$p.value
        }
        s[2] <- sub("<", "&lt;", format.pval(p, digits = 3, eps = 0.001))
        s
    } else {
        render.default(x = x, name = name, ...)
    }
}

rndr.strat <- function(label, n, ...) {
    ifelse(n == 0, label, render.strat.default(label, n, ...))
}

kbl(table1(strata, labels, groupspan = c(1, 3, 1), render.continuous = my.render.cont,
    render.categorical = my.render.cat), longtable = T, booktabs = T,
    align = "lcccc", linesep = c("", "", "", "\\addlinespace",
        "", "", "\\addlinespace", "", "\\addlinespace", "", "",
        "\\addlinespace", "", "", "\\addlinespace", "", "", "\\addlinespace",
        "", "", "\\addlinespace", "", "\\addlinespace"), format = "latex",
    caption = "Descriptive statistics of the final sample (N=196)",
    col.names = linebreak(c("", "Total", "Direct marketing",
        "Mixed marketing", "Marketing via intermediaries"))) %>%
    add_header_above(c("", "", `Marketing channel (% direct sales)` = 3),
        bold = TRUE) %>%
    kable_styling(latex_options = "scale_down", font_size = 7)
```

Table 1: Descriptive statistics of the final sample (N=196)

| | | Marketing channel (% direct sales) | | |
	Total	Direct marketing	Mixed marketing	Marketing via intermediaries
	(N=196)	(N=50)	(N=32)	(N=114)

Cosmetic chemical pesticide application				
No	150 (77 %)	43 (86 %)	27 (84 %)	80 (70 %)
Yes	46 (23 %)	7 (14 %)	5 (16 %)	34 (30 %)
Cosmetic chemical thinning for size and color				
No	80 (41 %)	34 (68 %)	16 (50 %)	30 (26 %)
Yes	116 (59 %)	16 (32 %)	16 (50 %)	84 (74 %)
Total land under apples (are)				
Mean (SD)	410 (640)	130 (300)	160 (320)	610 (740)
Organic production				
No	160 (82 %)	37 (74 %)	27 (84 %)	96 (84 %)
Yes	36 (18 %)	13 (26 %)	5 (16 %)	18 (16 %)
Low-residue producer				
No	38 (19 %)	13 (26 %)	12 (38 %)	13 (11 %)
Yes	158 (81 %)	37 (74 %)	20 (62 %)	101 (89 %)
Club apple producer				
No	173 (88 %)	50 (100 %)	31 (97 %)	92 (81 %)
Yes	23 (12 %)	0 (0 %)	1 (3 %)	22 (19 %)
Visual appearance specified in contract				
No	162 (83 %)	50 (100 %)	29 (91 %)	83 (73 %)
Yes	34 (17 %)	0 (0 %)	3 (9 %)	31 (27 %)
Premium & class 1 quality (%)				
Mean (SD)	69 (28)	64 (37)	60 (31)	74 (21)
Class 2 quality (%)				
Mean (SD)	26 (25)	32 (36)	33 (28)	20 (15)

6 Figures and Tables in the Appendix

6.1 Table 4: Regression output estimating the effect of marketing channel choice on cosmetic pesticide use (equation 11)

Table 4 is produced in the ReplicationCodeMultinomialTreatmentEffectsModel.do file and exported to Table4_1.csv (for cosmetic pesticide use) and Table4_2.csv (for chemical cosmetic thinning) respectively.

6.2 Table A1: Pesticide loads of plant protection products allowed in Swiss orchards (N=518)

Code for this graph is available upon request. Data is publicly available from https://www.psm.admin.ch/de/produkte.

6.3 Figure B1: Frequency distributions of direct marketing shares in our sample (N=196)

```
FigureB1 <- ggplot(reg_data, aes(direct, fill = ..x..)) + geom_vline(xintercept = 0.25,
    alpha = 0.5, color = "red") + geom_vline(xintercept = 0.75,
    alpha = 0.5, color = "red") + geom_histogram(bins = 15, color = "darkgrey") +
    theme_minimal() + theme(panel.grid.minor = element_blank(),
    panel.grid.major.x = element_blank()) + scale_fill_gradient2(name = "Direct sales
    (%)\n ",
    high = "darkorange1", mid = "white", low = "darkblue", midpoint = 0.5,
    breaks = c(0, 0.25, 0.5, 0.75, 1), labels = c("0 (indirect selling)",
        "25", "50", "75", "100 (direct selling)"), limits = c(0,
```

```
        1)) + scale_x_continuous(name = "", breaks = c(0, 0.11,
    0.25, 0.5, 0.75, 0.9, 1), labels = c("0%", "\nIndirect marketing",
    "25%", "50%\nMixed marketing", "75%", "\nDirect marketing",
    "100%")) + scale_y_continuous(name = "Count", limits = c(0,
    100), breaks = c(0, 20, 40, 60, 80, 100))

FigureB1
```
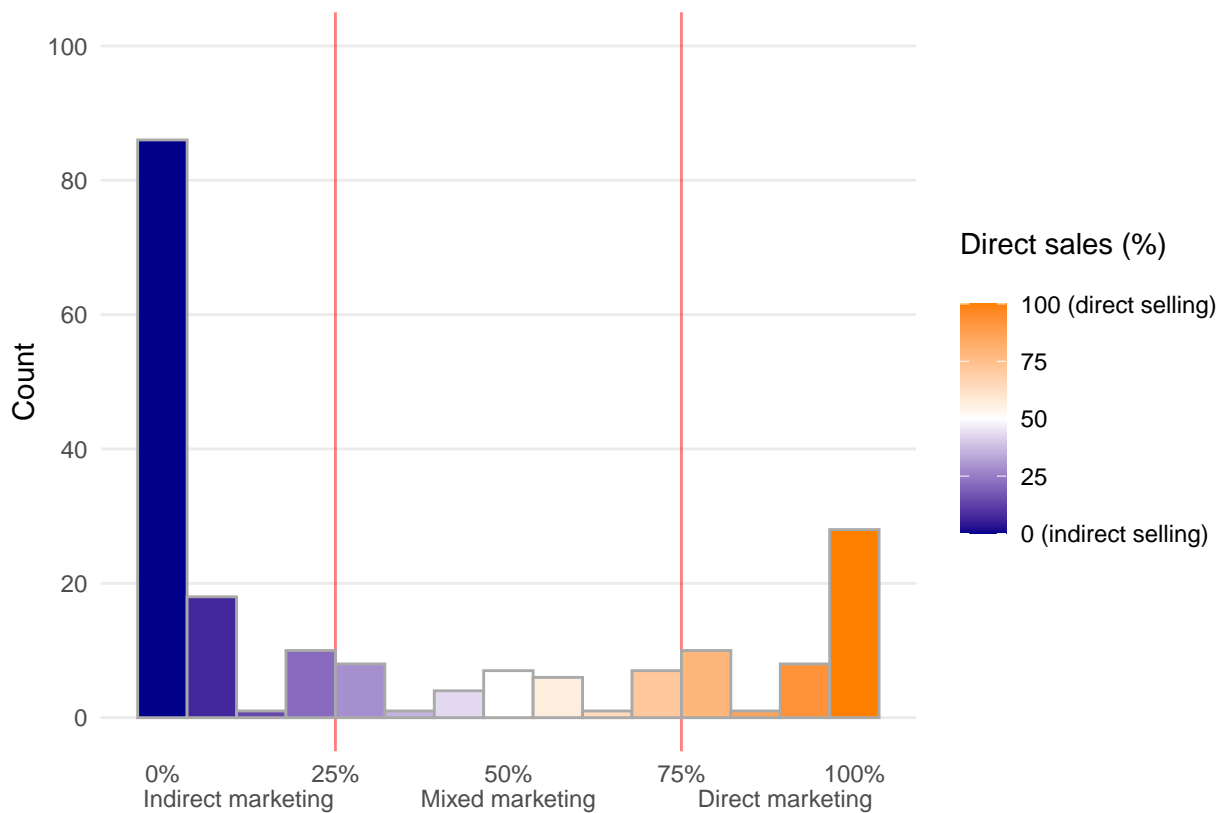


```
ggsave(FigureB1, file = "Appendix_FigureB1.png", width = 20,
    height = 10, dpi = 600, units = "cm", bg = "white")
```

6.4   Figure C1: Identification of correlated variables

Note that variables that relate to a farms location (e.g. distance to intermediaries and distance to direct buyers) are not included here for data privacy reasons.

```
### Outcome specification
cor_data1 <- stata_data %>%
    select(avg_simscab, hand_application, farmsize_scaled, organic,
        training, insurance, private_extension, LRP, club, production_riskpref,
        contract_optical_appearance, age, female) %>%
    drop_na()
```

```
corr1 <- round(cor(cor_data1, method = "pearson"), 2)
p.mat1 <- cor_pmat(cor_data1, conf.level = 0.95)

corr1 <- data.frame(corr1)
p.mat1 <- data.frame(p.mat1)

corr_long1 <- corr1 %>%
    mutate(row = rownames(corr1)) %>%
    gather(key, variable, -row)

p.mat_long1 <- p.mat1 %>%
    mutate(row = rownames(p.mat1)) %>%
    gather(key, variable, -row)

long_cmd1 <- merge(corr_long1, p.mat_long1, by = c("key", "row"))

long_cmd1 <- long_cmd1 %>%
    filter(variable.x != 1 & variable.y < 0.05)

# Renaming
colnames(corr1)[colnames(corr1) == "ChemicalThinning4sizecolor"] <- "Chemical cosmetic
thinning"
colnames(corr1)[colnames(corr1) == "cosmetic_pesticide"] <- "Chemical cosmetic\npesticide
use"
colnames(corr1)[colnames(corr1) == "avg_simscab"] <- "Pest pressure"
colnames(corr1)[colnames(corr1) == "hand_application"] <- "Hand aplication"
colnames(corr1)[colnames(corr1) == "farmsize_scaled"] <- "Farm size"
colnames(corr1)[colnames(corr1) == "organic"] <- "Organic"
colnames(corr1)[colnames(corr1) == "training"] <- "Training"
colnames(corr1)[colnames(corr1) == "insurance"] <- "Insurance"
colnames(corr1)[colnames(corr1) == "private_extension"] <- "Private extension"
colnames(corr1)[colnames(corr1) == "LRP"] <- "Low-residue production"
colnames(corr1)[colnames(corr1) == "club"] <- "Club apple production"
colnames(corr1)[colnames(corr1) == "production_riskpref"] <- "Production\nrisk
preferences"
colnames(corr1)[colnames(corr1) == "contract_optical_appearance"] <- "Contract"
colnames(corr1)[colnames(corr1) == "age"] <- "Age"
colnames(corr1)[colnames(corr1) == "female"] <- "Female"

rownames(corr1)[rownames(corr1) == "ChemicalThinning4sizecolor"] <- "Chemical cosmetic
thinning"
rownames(corr1)[rownames(corr1) == "cosmetic_pesticide"] <- "Chemical cosmetic\npesticide
use"
rownames(corr1)[rownames(corr1) == "avg_simscab"] <- "Pest pressure"
rownames(corr1)[rownames(corr1) == "hand_application"] <- "Hand aplication"
rownames(corr1)[rownames(corr1) == "farmsize_scaled"] <- "Farm size"
rownames(corr1)[rownames(corr1) == "organic"] <- "Organic"
rownames(corr1)[rownames(corr1) == "training"] <- "Training"
rownames(corr1)[rownames(corr1) == "insurance"] <- "Insurance"
rownames(corr1)[rownames(corr1) == "private_extension"] <- "Private extension"
rownames(corr1)[rownames(corr1) == "LRP"] <- "Low-residue production"
rownames(corr1)[rownames(corr1) == "club"] <- "Club apple production"
rownames(corr1)[rownames(corr1) == "production_riskpref"] <- "Production\nrisk
preferences"
```

```r
rownames(corr1)[rownames(corr1) == "contract_optical_appearance"] <- "Contract"
rownames(corr1)[rownames(corr1) == "age"] <- "Age"
rownames(corr1)[rownames(corr1) == "female"] <- "Female"

corr_plot1 <- ggcorrplot(corr1, hc.order = FALSE, legend.title = "Pearson\ncorrelation",
    p.mat = p.mat1, insig = "blank", lab = TRUE, lab_size = 2,
    outline.color = "grey") + ggtitle("A) Outcome specification") +
    theme(panel.grid.major.x = element_blank(), panel.grid.major.y = element_blank(),
        axis.text.x = element_text(size = 8, angle = 90, vjust = 0.5,
            hjust = 1), axis.text.y = element_text(size = 8),
        legend.position = "Bottom")

### Treatment selection
cor_data2 <- stata_data %>%
    select(earnings2, farmsize_scaled, organic, training, marketing_riskpref,
        exp_prices, exp_quality, exp_trust, age, female) %>%
    drop_na()

corr2 <- round(cor(cor_data2, method = "pearson"), 2)
p.mat2 <- cor_pmat(cor_data2, conf.level = 0.95)

corr2 <- data.frame(corr2)
p.mat2 <- data.frame(p.mat2)

corr_long2 <- corr2 %>%
    mutate(row = rownames(corr2)) %>%
    gather(key, variable, -row)

p.mat_long2 <- p.mat2 %>%
    mutate(row = rownames(p.mat2)) %>%
    gather(key, variable, -row)

long_cmd2 <- merge(corr_long2, p.mat_long2, by = c("key", "row"))

long_cmd2 <- long_cmd2 %>%
    filter(variable.x != 1 & variable.y < 0.05)

# Renaming
colnames(corr2)[colnames(corr2) == "farmsize_scaled"] <- "Farmsize"
colnames(corr2)[colnames(corr2) == "earnings2"] <- "Specialization"
colnames(corr2)[colnames(corr2) == "training"] <- "Training"
colnames(corr2)[colnames(corr2) == "insurance"] <- "Insurance"
colnames(corr2)[colnames(corr2) == "marketing_riskpref"] <- "Marketing\nrisk preferences"
colnames(corr2)[colnames(corr2) == "exp_prices"] <- "Prices"
colnames(corr2)[colnames(corr2) == "exp_quality"] <- "Quality"
colnames(corr2)[colnames(corr2) == "exp_trust"] <- "Trust"
colnames(corr2)[colnames(corr2) == "age"] <- "Age"
colnames(corr2)[colnames(corr2) == "female"] <- "Female"

rownames(corr2)[rownames(corr2) == "earnings2"] <- "Specialization"
rownames(corr2)[rownames(corr2) == "farmsize_scaled"] <- "Farmsize"
rownames(corr2)[rownames(corr2) == "training"] <- "Training"
rownames(corr2)[rownames(corr2) == "insurance"] <- "Insurance"
```

```r
rownames(corr2)[rownames(corr2) == "marketing_riskpref"] <- "Marketing\nrisk preferences"
rownames(corr2)[rownames(corr2) == "exp_prices"] <- "Prices"
rownames(corr2)[rownames(corr2) == "exp_quality"] <- "Quality"
rownames(corr2)[rownames(corr2) == "exp_trust"] <- "Trust"
rownames(corr2)[rownames(corr2) == "age"] <- "Age"
rownames(corr2)[rownames(corr2) == "female"] <- "Female"

corr_plot2 <- ggcorrplot(corr2, hc.order = FALSE, legend.title = "Pearson\ncorrelation",
    p.mat = p.mat2, insig = "blank", lab = TRUE, lab_size = 2,
    outline.color = "grey") + ggtitle("B) Treatment specification") +
    theme(panel.grid.major.x = element_blank(), panel.grid.major.y = element_blank(),
        axis.text.x = element_text(size = 8, angle = 90, vjust = 0.5,
            hjust = 1), axis.text.y = element_text(size = 8),
        legend.position = "Bottom")

FigureC1 <- ggarrange(corr_plot1, corr_plot2, ncol = 1, common.legend = TRUE,
    legend = "bottom")

FigureC1
```
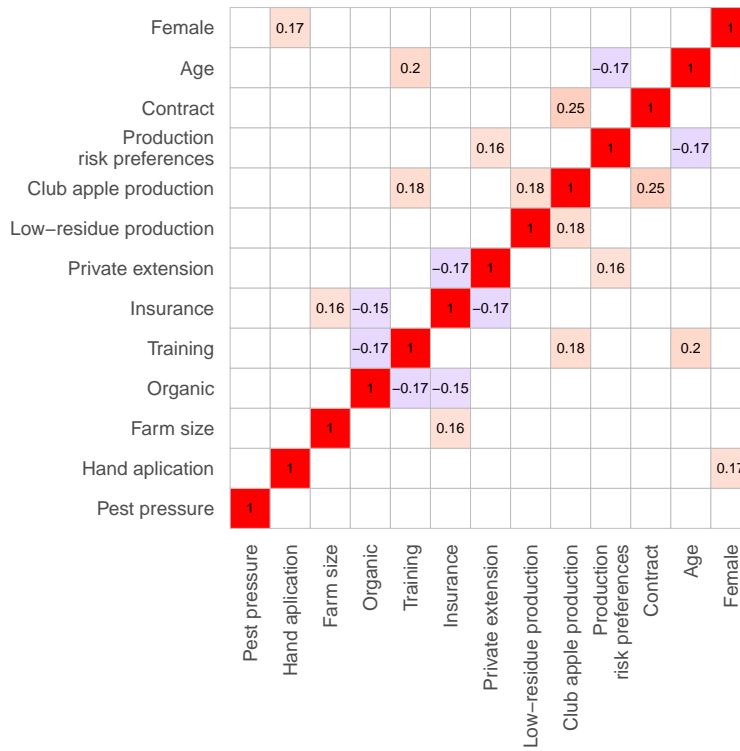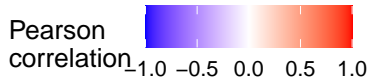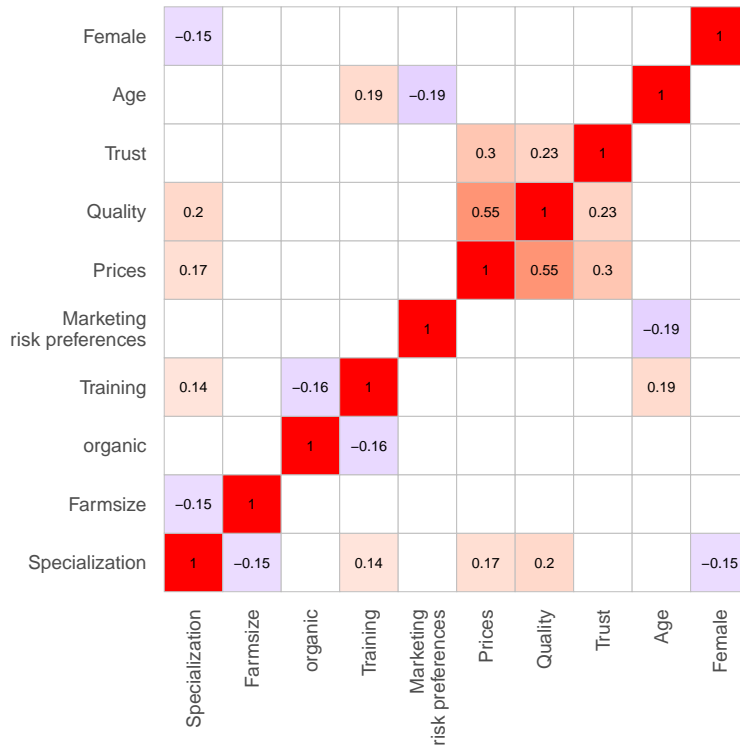
A) Outcome specification

B) Treatment specification

Pearson correlation

```
ggsave(FigureC1, file = "Appendix_FigureC1.png", width = 40,
    height = 20, dpi = 600, units = "cm", bg = "white")
```

6.5 Table C1: Regression results when dropping correlated variables in the Outcome and Treatment specifications

Table C1 is produced in the ReplicationCodeMultinomialTreatmentEffectsModel.do file and exported to TableC1.csv.

6.6 Table C2: Linear probability estimation model on cosmetic pesticide use and chemical cosmetic thinning (equation C1)

```
reg_data$canton <- factor(reg_data$canton)
reg_data$treatment_factor <- factor(reg_data$treatment_simple)

linear_models <- list(m1 <- lm(cosmetic_pesticide ~ treatment_factor,
    data = reg_data), m2 <- lm(cosmetic_pesticide ~ treatment_factor +
    avg_simscab + canton + hand_application + farmsize_scaled +
    organic + training + insurance + private_extension + labels.Club. +
    LRP + riskpref.prodrisk. + contract_optical_appearance +
    age + female, data = reg_data), m3 <- lm(ChemicalThinning4sizecolor ~
    treatment_factor, data = reg_data), m4 <- lm(ChemicalThinning4sizecolor ~
    treatment_factor + avg_simscab + canton + hand_application +
        farmsize_scaled + organic + training + insurance + private_extension +
        labels.Club. + LRP + riskpref.prodrisk. + contract_optical_appearance +
        age + female, data = reg_data))

clwb <- list(se1 <- cluster.boot(linear_models[[1]], reg_data$canton,
    boot_type = "wild"), se2 <- cluster.boot(linear_models[[2]],
    reg_data$canton, boot_type = "wild"), se3 <- cluster.boot(linear_models[[3]],
    reg_data$canton, boot_type = "wild"), se4 <- cluster.boot(linear_models[[4]],
    reg_data$canton, boot_type = "wild"))

ct <- list(ct1 <- coeftest(linear_models[[1]], clwb[[1]])[, 2],
    ct2 <- coeftest(linear_models[[2]], clwb[[2]])[, 2], ct3 <-
coeftest(linear_models[[3]],
        clwb[[3]])[, 2], ct4 <- coeftest(linear_models[[4]],
        clwb[[4]])[, 2])

stargazer(c(linear_models), se = c(ct), dep.var.labels = c("Cosmetic pesticide use
(Yes/No)",
    "Cosmetic chemical thinning (Yes/No)"), covariate.labels = c("Mixed marketing (25-75%
    direct sales)",
    "Indirect marketing (<25% direct sales)", "Pest pressure (%)",
    "Hand application (Yes/No)", "Farmsize (in ha)", "Organic (Yes/No)",
    "Training (Yes/No)", "Insurance (Yes/No)", "Private extension (Yes/No)",
    "Club label (Yes/No)", "Low-residue (Yes/No)", "Production risk preferences",
    "Contract (Yes:1/No:0)", "Age (years)", "Female (Yes/No)"),
    omit.stat = c("LL", "ser", "f", "rsq"), no.space = FALSE,
    omit = c("[c][a][n][t][o][n]", "[C][o][n][s][t]"), add.lines = list(c("Constant",
```

```
        "Yes", "Yes", "Yes", "Yes"), c("Cantonal fixed effects",
        "No", "Yes", "No", "Yes")), notes.align = "l", digits = 3,
    single.row = TRUE, notes.append = TRUE, column.sep.width = "-25pt",
    notes = "Clustered and wild bootstrapped standard errors in parentheses (Woolridge,
    2013).",
    type = "html", font.size = "tiny", out = "Appendix_TableC2.doc")
```

6.7 Figure C2: Average marginal effects from Logit model regression on chemical cosmetic thinning methods

```
logit_regression <- glm(formula = ChemicalThinning4sizecolor ~
    treatment_factor + avg_simscab + canton + hand_application +
        farmsize_scaled + organic + training + insurance + private_extension +
        labels.Club. + LRP + riskpref.prodrisk. + contract_optical_appearance +
        age + female, binomial(link = "logit"), data = reg_data)

ME <- margins(logit_regression)

ME_red <- summary(ME)

ME <- ME %>%
    select(contains(c("dydx")))

ME_red$CI95_new <- ME_red$SE * 3.92
ME_red$CI99_new <- ME_red$SE * 5.15
ME_red$CI95_new_upper <- ME_red$AME + (ME_red$CI95_new/2)
ME_red$CI95_new_lower <- ME_red$AME - (ME_red$CI95_new/2)
ME_red$CI99_new_upper <- ME_red$AME + (ME_red$CI99_new/2)
ME_red$CI99_new_lower <- ME_red$AME - (ME_red$CI99_new/2)

ME_red <- ME_red[!grepl("canton", ME_red$factor), ]

ME_red$sig_estimate <- ifelse(sign(ME_red$CI95_new_upper) ==
    sign(ME_red$CI95_new_lower), as.character(round(ME_red$AME,
    3)), "")

ME_red$factor[ME_red$factor == "labels.Club."] <- "Club"
ME_red$factor[ME_red$factor == "treatment_factor2"] <- "Mainly marketing via
intermediaries\n(<25% direct sales)"
ME_red$factor[ME_red$factor == "treatment_factor1"] <- "Mixed marketing\n(25-75% direct
sales)"
ME_red$factor[ME_red$factor == "LRP"] <- "Low-residue production"
ME_red$factor[ME_red$factor == "insurance"] <- "Insurance"
ME_red$factor[ME_red$factor == "contract_optical_appearance"] <- "Contract"
ME_red$factor[ME_red$factor == "avg_simscab"] <- "Pest pressure"
ME_red$factor[ME_red$factor == "age"] <- "Age"
ME_red$factor[ME_red$factor == "farmsize_scaled"] <- "Farm size"
ME_red$factor[ME_red$factor == "private_extension"] <- "Private extension"
ME_red$factor[ME_red$factor == "riskpref.prodrisk."] <- "Production\nrisk preferences"
ME_red$factor[ME_red$factor == "training"] <- "Training"
ME_red$factor[ME_red$factor == "female"] <- "Female"
```

```
ME_red$factor[ME_red$factor == "organic"] <- "Organic"

ME_red <- subset(ME_red, factor != "hand_application")

# Now plot them
FigureC2 <- ggplot(ME_red, aes(x = factor, y = AME)) + geom_hline(yintercept = 0,
    color = "red", size = 1, alpha = 0.5) + geom_linerange(aes(x = reorder(factor,
    AME), ymin = CI95_new_lower, ymax = CI95_new_upper), alpha = 0.4,
    size = 1, position = position_dodge(width = 1/2)) + geom_point(aes(x = factor,
    y = AME)) + geom_text(aes(x = factor, y = AME, label = sig_estimate),
    size = 3.5, vjust = -0.5) + xlab("") + ylab("Coefficient estimates (and 95%
    confidence interval)") +
    geom_point() + scale_y_continuous(limits = c(-0.8, 0.8)) +
    coord_flip() + theme_minimal() + theme(axis.title.x = element_text(margin = margin(t
    = 20,
    r = 0, b = 10, l = 0)), panel.grid.major.y = element_blank(),
    strip.placement = "outside", strip.background = element_rect(fill = NA,
        colour = "black"), strip.text.y = element_text(size = 10,
        face = "bold", colour = "black"), axis.text.y = element_text(size = 11),
    panel.spacing = unit(0, "cm"))

FigureC2
```


Coefficient estimates (and 95% confidence interval)

20
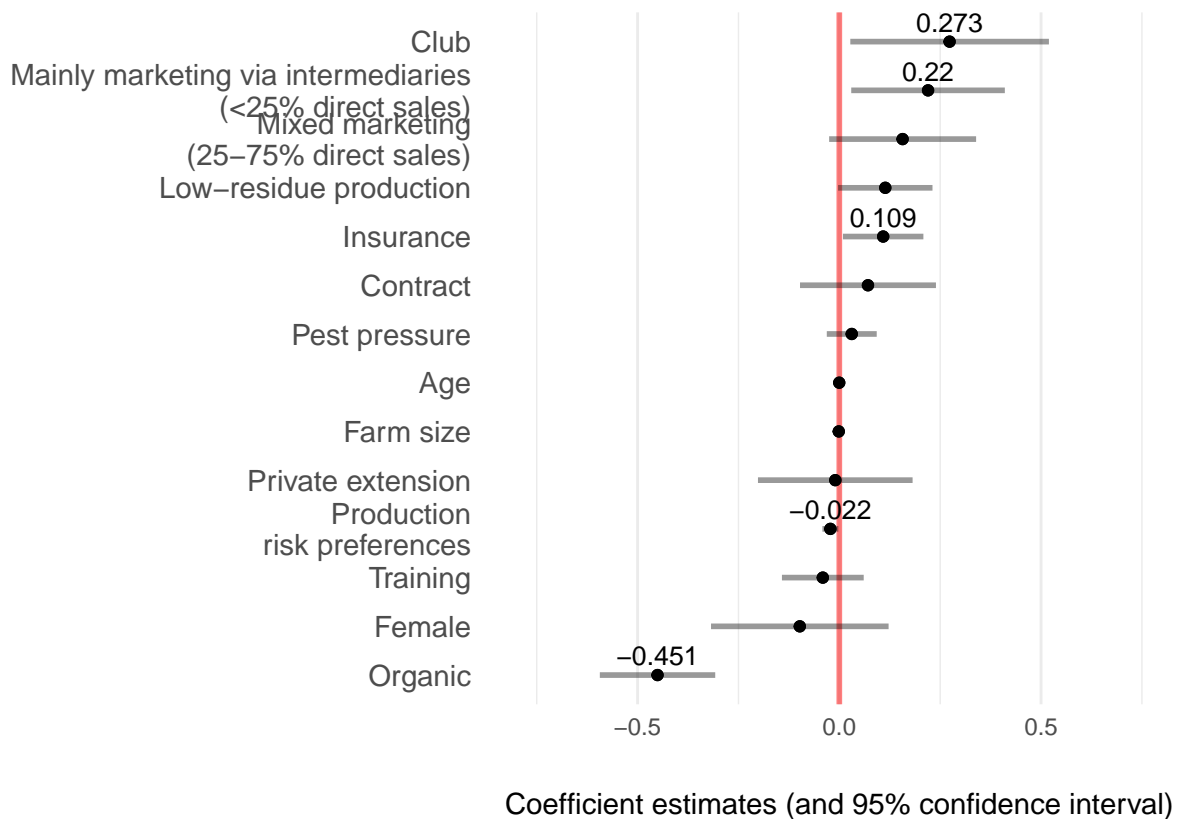
```
ggsave(FigureC2, file = "Appendix_FigureC2.png", width = 30,
    height = 15, dpi = 600, units = "cm", bg = "white")
```

6.8 Figure C3: Regressions with other chemical pesticide use as outcome variables (equation 11)

Note that the coefficient estimates and standard errors are calculated in the ReplicationCodeMultinomial-TreatmentEffectsModel.do file and exported to FigureC3.csv.

```
model_output <- data.frame(var = c("Mixed marketing\n(25-75% of direct sales)",
    "Mainly marketing via intermediaries\n(0-25% of direct sales)",
    "Mixed marketing\n(25-75% of direct sales)", "Mainly marketing via
    intermediaries\n(0-25% of direct sales)",
    "Mixed marketing\n(25-75% of direct sales)", "Mainly marketing via
    intermediaries\n(0-25% of direct sales)"),
    cat = c("Use of chemical herbicides only", "Use of chemical herbicides only",
        "Use of chemical fungicides", "Use of chemical fungicides",
        "Use of chemical insecticides", "Use of chemical insecticides"),
    estimate = c(0.149, 0.184, -0.394, 0.022, 0.247, -0.06),
    se = c(0.2435, 0.1317, 0.0259, 0.0174, 0.0206, 0.0202))

model_output$CI95_new <- model_output$se * 3.92
model_output$CI99_new <- model_output$se * 5.15
model_output$CI95_new_upper <- model_output$estimate + (model_output$CI95_new/2)
model_output$CI95_new_lower <- model_output$estimate - (model_output$CI95_new/2)
model_output$CI99_new_upper <- model_output$estimate + (model_output$CI99_new/2)
model_output$CI99_new_lower <- model_output$estimate - (model_output$CI99_new/2)

# Now plot them
FigureC3 <- ggplot(model_output, aes(x = var, y = estimate, color = cat)) +
    geom_hline(yintercept = 0, color = "red", size = 1, alpha = 0.5) +
    geom_linerange(aes(x = var, ymin = CI95_new_lower, ymax = CI95_new_upper),
        alpha = 0.8, size = 1, position = position_dodge(width = 1/2)) +
    geom_linerange(aes(x = var, ymin = CI99_new_lower, ymax = CI99_new_upper),
        alpha = 0.4, size = 1, position = position_dodge(width = 1/2)) +
    geom_point(aes(x = var, y = estimate), size = 1.2, position = position_dodge(width =
    0.5)) +
    geom_text(aes(x = var, y = estimate, label = round(estimate,
        2)), show.legend = F, position = position_dodge(width = 0.5),
        size = 3, vjust = -0.5) + xlab("") + ylab("Coefficient estimates relative to
        mainly direct marketers (75-100% direct sales)\n(99/95% confidence intervals)") +
    coord_flip() + theme_minimal() + theme(axis.title.x = element_text(margin = margin(t
    = 20,
    r = 0, b = 10, l = 0)), panel.grid.major.y = element_blank(),
    strip.placement = "outside", strip.background = element_rect(fill = NA,
        colour = "black"), strip.text.y = element_text(size = 10,
        face = "bold", colour = "black"), axis.text.y = element_text(size = 11),
    panel.spacing = unit(0, "cm"), legend.key.size = unit(1,
        "cm"), legend.position = "right") + scale_color_manual(name = "Pesticide group:",
    breaks = c("Use of chemical herbicides only", "Use of chemical fungicides",
        "Use of chemical insecticides"), labels = c("Use of chemical herbicides only",
        "Use of chemical fungicides", "Use of chemical insecticides"),
```

```
    values = c("grey40", "grey60", "grey80")) + guides(color = guide_legend(override.aes
    = list(size = 3)))
```

FigureC3
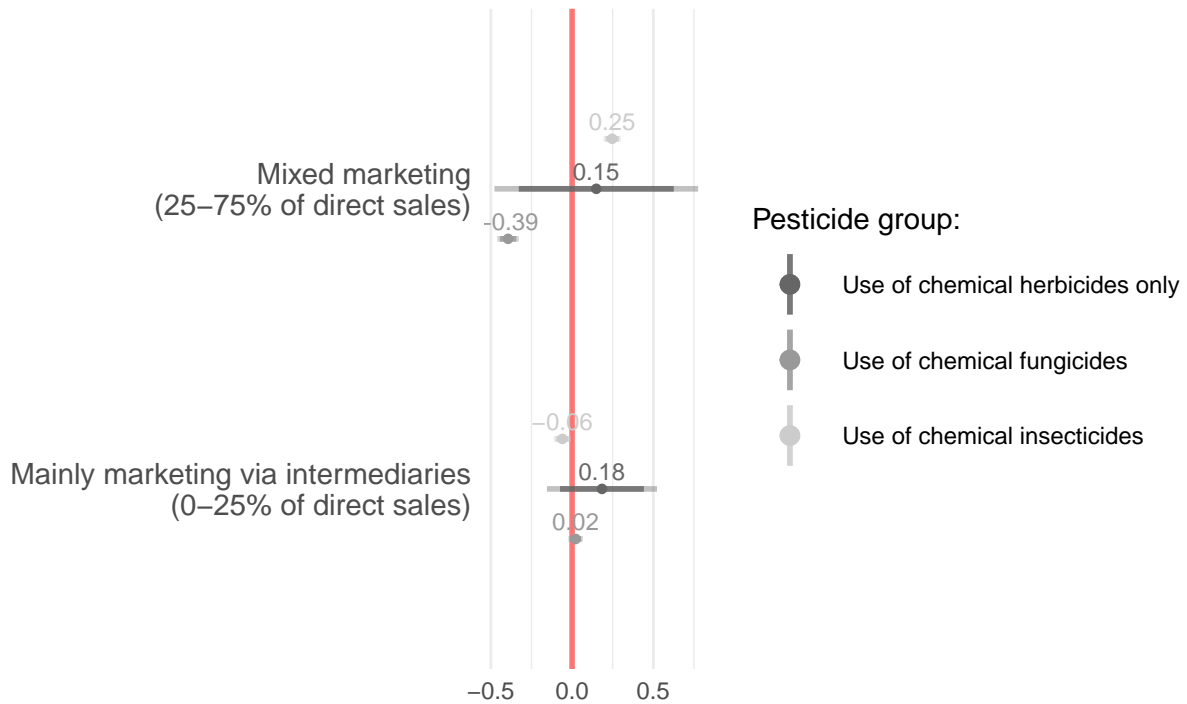
Coefficient estimates relative to mainly direct marketers (75–100% direct sales)
(99/95% confidence intervals)

```
ggsave(FigureC3, file = "Appendix_FigureC3.png", width = 25,
    height = 10, dpi = 600, units = "cm", bg = "white")
```

6.9 Table C3: Linear probability models using club variables as treatment variables (equation C1)

```
club_models <- list(m1 <- lm(formula = ChemicalThinning4sizecolor ~
    labels.Club., data = reg_data), m2 <- lm(formula = ChemicalThinning4sizecolor ~
    labels.Club. + avg_simscab + canton + hand_application +
        farmsize_scaled + organic + training + insurance + private_extension +
        LRP + riskpref.prodrisk. + contract_optical_appearance +
        age + female, data = reg_data), m3 <- lm(formula = ChemicalThinning4sizecolor ~
    club_IS, data = reg_data), m4 <- lm(formula = ChemicalThinning4sizecolor ~
    club_IS + avg_simscab + canton + hand_application + farmsize_scaled +
        organic + training + insurance + private_extension +
        LRP + riskpref.prodrisk. + contract_optical_appearance +
        age + female, data = reg_data), m5 <- lm(formula = ChemicalThinning4sizecolor ~
```

```
        club_IS_area, data = reg_data), m6 <- lm(formula = ChemicalThinning4sizecolor ~
        club_IS_area + avg_simscab + canton + hand_application +
            farmsize_scaled + organic + training + insurance + private_extension +
            LRP + riskpref.prodrisk. + contract_optical_appearance +
            age + female, data = reg_data))

clwb <- list(se1 <- cluster.boot(club_models[[1]], reg_data$canton,
    boot_type = "wild"), se2 <- cluster.boot(club_models[[2]],
    reg_data$canton, boot_type = "wild"), se3 <- cluster.boot(club_models[[3]],
    reg_data$canton, boot_type = "wild"), se4 <- cluster.boot(club_models[[4]],
    reg_data$canton, boot_type = "wild"), se5 <- cluster.boot(club_models[[5]],
    reg_data$canton, boot_type = "wild"), se6 <- cluster.boot(club_models[[6]],
    reg_data$canton, boot_type = "wild"))

ct <- list(ct1 <- coeftest(club_models[[1]], clwb[[1]])[, 2],
    ct2 <- coeftest(club_models[[2]], clwb[[2]])[, 2], ct3 <- coeftest(club_models[[3]],
        clwb[[3]])[, 2], ct4 <- coeftest(club_models[[4]], clwb[[4]])[,
        2], ct5 <- coeftest(club_models[[5]], clwb[[5]])[, 2],
    ct6 <- coeftest(club_models[[6]], clwb[[6]])[, 2])

stargazer(c(club_models), se = c(ct), dep.var.labels = c("Cosmetic chemical thinning
(Yes/No)"),
    order = c("labels.Club.", "club_IS", "club_IS_area"), covariate.labels = c("Club
    label (Yes:1/No:0)",
        "Club variety grown (Yes:1/No:0)", "Club variety share (%)",
        "Pest pressure (%)", "Hand application (Yes:1/No:0)",
        "Farmsize (in ha)", "Organic (Yes:1/No:0)", "Training (Yes:1/No:0)",
        "Insurance (Yes:1/No:0)", "Private extension (Yes:1/No:0)",
        "Low-residue (Yes:1/No:0)", "Production risk preferences",
        "Contract (Yes:1/No:0)", "Age (years)", "Female (Yes:1/No:0)"),
    omit.stat = c("LL", "ser", "f", "rsq"), no.space = FALSE,
    omit = c("[c][a][n][t][o][n]", "[C][o][n][s][t]"), add.lines = list(c("Constant",
        "Yes", "Yes", "Yes", "Yes", "Yes", "Yes"), c("Cantonal fixed effects",
        "No", "Yes", "No", "Yes", "No", "Yes")), notes.align = "l",
    digits = 3, single.row = TRUE, notes.append = TRUE, column.sep.width = "-25pt",
    notes = "Clustered and wild bootstrapped standard errors in parentheses (Woolridge,
    2013).",
    type = "html", font.size = "tiny", out = "Appendix_TableC3.doc")
```

6.10 Table C4: Regression output estimating the effect of marketing channel choice on cosmetic pesticide use (equation 11) excluding organic producers

Table C4 is produced in the ReplicationCodeMultinomialTreatmentEffectsModel.do file and exported to TableC4.csv.

6.11 Table D1: Using different numbers of simulation draws per observation based on Halton sequences

Table D1 is produced in the ReplicationCodeMultinomialTreatmentEffectsModel.do file and exported to Table4_1.csv (for cosmetic pesticide use) and Table4_2.csv (for chemical cosmetic thinning) respectively.

6.12 Table E1: Full regression output of Table 4

Table E1 which is the full output from Table 4 is produced in the ReplicationCodeMultinomialTreatmentEffectsModel.do file and exported to Table4_1.csv (for cosmetic pesticide use) and Table4_2.csv (for chemical cosmetic thinning) respectively.

6.13 Table F1: Direct and indirect marketing shares per canton

```r
# Direct and indirect marketing shares per region
data$direct_reg <- ifelse(!is.na(data$MarketingApples.A4.), data$MarketingApples.A4./100,
    0)  # Direct

# Traders, cooperatives, shops/Supermarkets
data$indirect_reg <- ifelse(!is.na(data$MarketingApples.A1.),
    data$MarketingApples.A1./100, 0) + ifelse(!is.na(data$MarketingApples.A2.),
    data$MarketingApples.A2./100, 0) + ifelse(!is.na(data$MarketingApples.A3.),
    data$MarketingApples.A3./100, 0)

direct_per_region <- data %>%
    filter(Marketing.A1. == 1) %>%
    group_by(canton) %>%
    summarise(n = n(), mean_indirect = mean(indirect_reg), mean_direct =
    mean(direct_reg))

direct_per_region <- direct_per_region %>%
    filter(canton != "OW")

direct_per_region$language <- ""
direct_per_region$language[direct_per_region$canton == "AG"] <- "German"
direct_per_region$language[direct_per_region$canton == "BE"] <- "German"
direct_per_region$language[direct_per_region$canton == "BL"] <- "German"
direct_per_region$language[direct_per_region$canton == "GR"] <- "German"
direct_per_region$language[direct_per_region$canton == "LU"] <- "German"
direct_per_region$language[direct_per_region$canton == "SG"] <- "German"
direct_per_region$language[direct_per_region$canton == "SO"] <- "German"
direct_per_region$language[direct_per_region$canton == "SZ"] <- "German"
direct_per_region$language[direct_per_region$canton == "TG"] <- "German"
direct_per_region$language[direct_per_region$canton == "ZG"] <- "German"
direct_per_region$language[direct_per_region$canton == "ZH"] <- "German"
direct_per_region$language[direct_per_region$canton == "GE"] <- "French"
direct_per_region$language[direct_per_region$canton == "VD"] <- "French"
direct_per_region$language[direct_per_region$canton == "VS"] <- "Mixed"
direct_per_region$language[direct_per_region$canton == "JU"] <- "Mixed"
direct_per_region$language[direct_per_region$canton == "FR"] <- "Mixed"

direct_per_region$total_land[direct_per_region$canton == "FR"] <- 16.61
direct_per_region$total_land[direct_per_region$canton == "AG"] <- 124.8
direct_per_region$total_land[direct_per_region$canton == "BE"] <- 124.35
direct_per_region$total_land[direct_per_region$canton == "BL"] <- 37.43
direct_per_region$total_land[direct_per_region$canton == "GE"] <- 51.61
direct_per_region$total_land[direct_per_region$canton == "GR"] <- 29.05
direct_per_region$total_land[direct_per_region$canton == "JU"] <- 3.34
```

Canton	Number of observations per canton	Total land under apples (ha)	Average share of direct marketing	Average share of marketing via intermediaries
AG	12	124.80	0.34	0.46
BE	18	124.35	0.31	0.53
BL	5	37.43	0.45	0.42
FR	3	16.61	0.01	0.70
GE	6	51.61	0.40	0.52
GR	2	29.05	0.05	0.95
JU	2	3.34	0.00	0.95
LU	8	125.90	0.83	0.16
SG	15	213.46	0.50	0.40
SO	6	45.88	0.24	0.67
SZ	4	15.45	0.36	0.55
TG	58	1133.19	0.84	0.05
VD	14	581.55	0.65	0.29
VS	12	952.17	0.87	0.03
ZG	5	31.21	0.48	0.40
ZH	25	150.45	0.25	0.63

```r
direct_per_region$total_land[direct_per_region$canton == "LU"] <- 125.9
direct_per_region$total_land[direct_per_region$canton == "NE"] <- 1.74
direct_per_region$total_land[direct_per_region$canton == "SG"] <- 213.46
direct_per_region$total_land[direct_per_region$canton == "SH"] <- 4.35
direct_per_region$total_land[direct_per_region$canton == "SO"] <- 45.88
direct_per_region$total_land[direct_per_region$canton == "SZ"] <- 15.45
direct_per_region$total_land[direct_per_region$canton == "TG"] <- 1133.19
direct_per_region$total_land[direct_per_region$canton == "VD"] <- 581.55
direct_per_region$total_land[direct_per_region$canton == "VS"] <- 952.17
direct_per_region$total_land[direct_per_region$canton == "ZG"] <- 31.21
direct_per_region$total_land[direct_per_region$canton == "ZH"] <- 150.45


direct_per_region$diff <- direct_per_region$mean_indirect + direct_per_region$mean_direct


export <- direct_per_region %>%
    select(canton, n, total_land, mean_indirect, mean_direct) %>%
    mutate(mean_indirect = round(mean_indirect, 2), mean_direct = round(mean_direct,
        2), total_land = round(total_land, 2))


colnames(export) <- c("Canton", "Number of observations per canton",
    "Total land under apples (ha)", "Average share of direct marketing",
    "Average share of marketing via intermediaries")


knitr::kable(export, format = "latex", booktabs = TRUE, align = "lcccc") %>%
    add_header_above(bold = TRUE) %>%
    kable_styling(latex_options = "scale_down", font_size = 7)
```

```r
write.table(export, file = "Appendix_TableF1.doc", sep = ",",
    quote = FALSE, row.names = F)
```

6.14   Figure F1: Average share of marketing via intermediaries and direct marketing per canton

```r
FigureF1 <- ggplot(direct_per_region, aes(x = mean_direct, y = mean_indirect)) +
    geom_point(aes(size = total_land)) + geom_text_repel(aes(label = canton),
    max.overlaps = 100) + scale_y_continuous(name = "Average share of marketing via
    intermediaries",
```
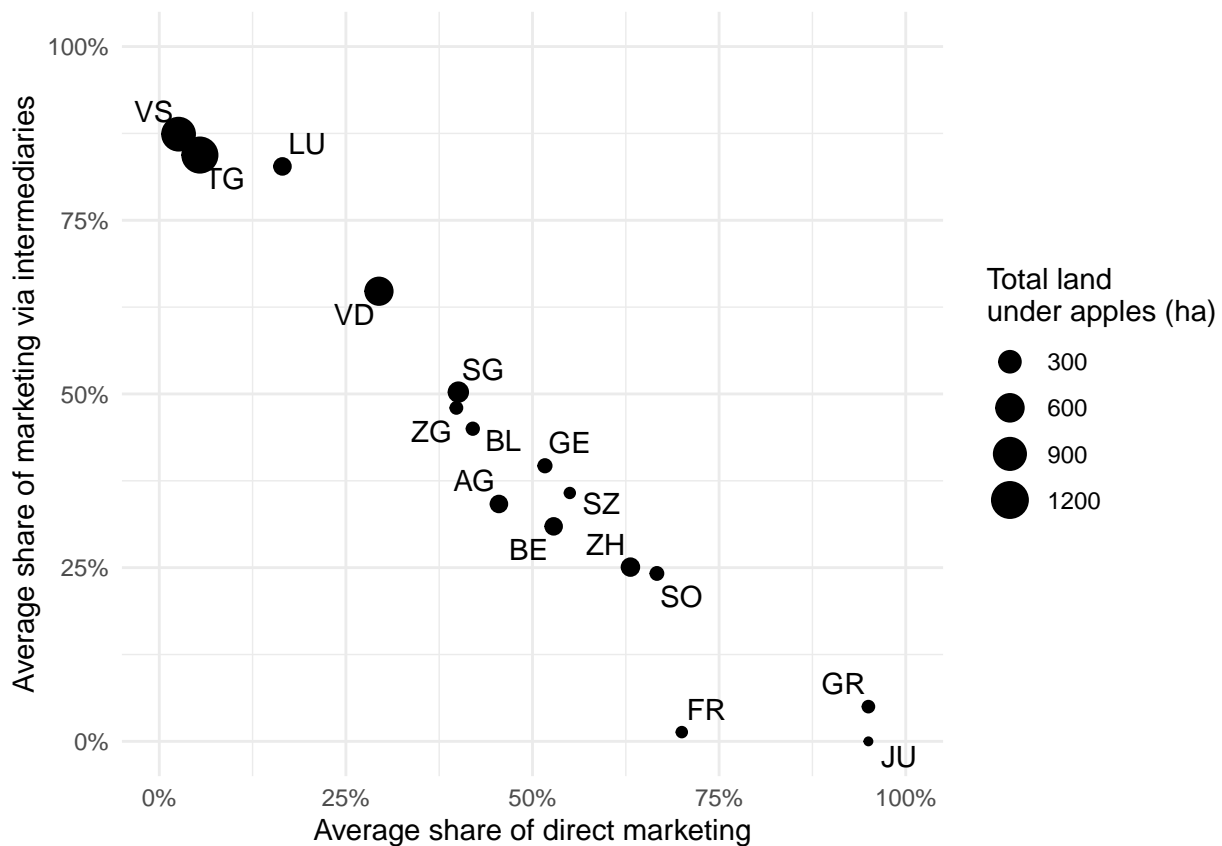
```
    limits = c(0, 1), labels = scales::percent) + scale_x_continuous(name = "Average
    share of direct marketing",
    limits = c(0, 1), labels = scales::percent) + scale_size_continuous(name = "Total
    land \nunder apples (ha)",
    limits = c(3, 1200)) + theme_minimal()

FigureF1
```



```
ggsave(FigureF1, file = "Appendix_FigureF1.png", width = 30,
    height = 15, dpi = 600, units = "cm", bg = "white")
```

6.15   Table G1: Survey representativeness

Numbers in this table are calculated in the data preparation section.

6.16   Table H1: Using thirds as alternative treatment variables

This table is calculated in the ReplicationCodeMultinomialTreatmentEffectsModel.do file.

6.17   Table H2: Using extreme cases as alternative treatment variables

This table is calculated in the ReplicationCodeMultinomialTreatmentEffectsModel.do file.