Summary of Yearly Comtrade Data, HS 2012

July 13, 2017

Clean data and create trade gap variables

This is for a small subset of the UN Comtrade data: specifically, for years 2012-2016, products that start with a 0 in the HS 2012 classification system, and for countries that start with the letter "A" (Afghanistan) through "M" (Myanmar).

Notes:

- *There are a lot of cases where country A has reported exports but country B has not reported anything. As much as half of the raw data becomes "missing" due to this. This seems consistent with Fisman/Wei (pg 4).*
- *The raw data contains re-exports and re-imports. These amounts are also included in a country's regular exports/imports as imports from one country to itself. For example, France has re-imports that are also included in France's imports as* `Reporter` *= France and* `Partner` *= France. This doesn't affect this analysis because there is no matched reporter/partner pair, but they might affect our trade gap measure if they're actually imports from an unforseen place. More on re-exports* ***at this link.***

```
#Keep product codes starting with "0" and country pairs starting with letters "A" through "M"
load(paste(DataPath,"Raw Data/Comtrade/Yearly/y_hs12/y_2012_hs12.Rda", sep = "/"))
y_2012_hs12 <- as.data.table(y_2012_hs12)

y_2012_hs12 <- subset(y_2012_hs12,substr(y_2012_hs12$`Commodity Code`,1,1)=='0')
y_2012_hs12 <- y_2012_hs12[grep("^[A-M]", y_2012_hs12$Reporter)]
y_2012_hs12 <- y_2012_hs12[grep("^[A-M]", y_2012_hs12$Partner)]

load(paste(DataPath,"Raw Data/Comtrade/Yearly/y_hs12/y_2013_hs12.Rda", sep = "/"))
y_2013_hs12 <- as.data.table(y_2013_hs12)

y_2013_hs12 <- subset(y_2013_hs12,substr(y_2013_hs12$`Commodity Code`,1,1)=='0')
y_2013_hs12 <- y_2013_hs12[grep("^[A-M]", y_2013_hs12$Reporter)]
y_2013_hs12 <- y_2013_hs12[grep("^[A-M]", y_2013_hs12$Partner)]

load(paste(DataPath,"Raw Data/Comtrade/Yearly/y_hs12/y_2014_hs12.Rda", sep = "/"))
y_2014_hs12 <- as.data.table(y_2014_hs12)

y_2014_hs12 <- subset(y_2014_hs12,substr(y_2014_hs12$`Commodity Code`,1,1)=='0')
y_2014_hs12 <- y_2014_hs12[grep("^[A-M]", y_2014_hs12$Reporter)]
y_2014_hs12 <- y_2014_hs12[grep("^[A-M]", y_2014_hs12$Partner)]

load(paste(DataPath,"Raw Data/Comtrade/Yearly/y_hs12/y_2015_hs12.Rda", sep = "/"))
y_2015_hs12 <- as.data.table(y_2015_hs12)

y_2015_hs12 <- subset(y_2015_hs12,substr(y_2015_hs12$`Commodity Code`,1,1)=='0')
y_2015_hs12 <- y_2015_hs12[grep("^[A-M]", y_2015_hs12$Reporter)]
y_2015_hs12 <- y_2015_hs12[grep("^[A-M]", y_2015_hs12$Partner)]

load(paste(DataPath,"Raw Data/Comtrade/Yearly/y_hs12/y_2016_hs12.Rda", sep = "/"))
y_2016_hs12 <- as.data.table(y_2016_hs12)

y_2016_hs12 <- subset(y_2016_hs12,substr(y_2016_hs12$`Commodity Code`,1,1)=='0')
y_2016_hs12 <- y_2016_hs12[grep("^[A-M]", y_2016_hs12$Reporter)]
y_2016_hs12 <- y_2016_hs12[grep("^[A-M]", y_2016_hs12$Partner)]
```

```r
#Combine HS 2012 data
hs12 <- do.call("rbind", list(y_2012_hs12, y_2013_hs12, y_2014_hs12, y_2015_hs12, y_2016_hs12))

rm(y_2012_hs12, y_2013_hs12, y_2014_hs12, y_2015_hs12, y_2016_hs12)

#Create table where country is reporting imports
hs12im <- hs12[`Trade Flow Code`==1]

hs12im[ ,`:=`(Classification = NULL, Year = NULL,
                  `Period Desc.` = NULL, `Is Leaf Code` = NULL,
                  `Reporter ISO` = NULL, `Partner ISO` = NULL,
                  `Qty Unit Code` = NULL,  Flag = NULL
                  )]

hs12im <- rename(hs12im, "Import Value" = "Trade Value (US$)")
hs12im <- rename(hs12im, "Import Qty Unit" = "Qty Unit")
hs12im <- rename(hs12im, "Import Qty" = "Qty")
hs12im <- rename(hs12im, "Import Netweight (kg)" = "Netweight (kg)")

#Create table where country is reporting exports
hs12ex <- hs12[`Trade Flow Code`==2]

hs12ex[ ,`:=`(Classification = NULL, Year = NULL,
                  `Period Desc.` = NULL, `Is Leaf Code` = NULL,
                  `Reporter ISO` = NULL, `Partner ISO` = NULL,
                  `Qty Unit Code` = NULL,  Flag = NULL
                  )]

hs12ex <- rename(hs12ex, "Export Value" = "Trade Value (US$)")
hs12ex <- rename(hs12ex, "Export Qty Unit" = "Qty Unit")
hs12ex <- rename(hs12ex, "Export Qty" = "Qty")
hs12ex <- rename(hs12ex, "Export Netweight (kg)" = "Netweight (kg)")

#Merge import and export tables together
hs12 <- merge(hs12im, hs12ex,
              by.x=c("Period", "Aggregate Level",
                     "Reporter Code", "Reporter", "Partner Code", "Partner",
                     "Commodity Code", "Commodity"),
              by.y=c("Period", "Aggregate Level",
                     "Partner Code", "Partner", "Reporter Code", "Reporter",
                     "Commodity Code", "Commodity"),all=TRUE)

hs12 <- rename(hs12, "Importer" = "Reporter")
hs12 <- rename(hs12, "Exporter" = "Partner")

#Create variable of the trade value gap between what countries report
hs12$Raw_gap = hs12$`Export Value` - hs12$`Import Value`

#Create variable of the log trade value gap
hs12$Log_gap = log(hs12$`Export Value`) - log(hs12$`Import Value`)

#Create variable of the trade value gap as a ratio of total reported trade
hs12$Gap_ratio = hs12$`Raw_gap`/(hs12$`Import Value` + hs12$`Export Value`)

#Repeat created variables but for the gap in quantity reported
hs12$`Export Netweight (kg)` <- as.numeric(hs12$`Export Netweight (kg)`)
```

```r
hs12$`Import Netweight (kg)` <- as.numeric(hs12$`Import Netweight (kg)`)

hs12$Qty_raw_gap = hs12$`Export Netweight (kg)` - hs12$`Import Netweight (kg)`

hs12$Qty_log_gap = log(hs12$`Export Netweight (kg)`) - log(hs12$`Import Netweight (kg)`)

hs12$Qty_gap_ratio = hs12$`Qty_raw_gap`/
                     (hs12$`Export Netweight (kg)` + hs12$`Import Netweight (kg)`)

save(hs12,file = paste(DataPath,"Analysis Data","hs12.Rda", sep = "/"))
rm(hs12, hs12ex, hs12im)
```

Value Trade Gap

The difference between what the exporting country reports and what the importing country reports in US dollars.

Coverage

```r
load(paste(DataPath,"Analysis Data/hs12.Rda", sep = "/"))
hs12 <- as.data.table(hs12)

options(digits=2)

#Remove observations where one or more countries do not report imports/exports.
#699,887 rows deleted.
hs12 <- hs12[!is.na(Log_gap)]

#For each year, how many product*country pairs / all possible product*country pairs?

product <- hs12[, uniqueN(`Commodity Code`)]
product_year <- hs12[, uniqueN(`Commodity Code`), by=Period]
product_year <- rename(product_year, Products = V1)

pair <- unique(setDT(hs12), by = c("Importer", "Exporter"))
pair <- pair[, .N]
pair_year <- unique(setDT(hs12), by = c("Importer", "Exporter", "Period"))
pair_year <- pair_year[, .N, by=Period]
pair_year <- rename(pair_year, Pairs = N)

year_coverage <- merge(product_year, pair_year)
year_coverage$Total_products <- product
year_coverage$Total_pairs <- pair

year_coverage$Coverage <- (year_coverage$Products*year_coverage$Pairs)/
                          (year_coverage$Total_products*year_coverage$Total_pair)

year_coverage
```

```
##    Period Products Pairs Total_products Total_pairs Coverage
## 1:   2012      614  1715            617        3423     0.50
## 2:   2013      614  2224            617        3423     0.65
## 3:   2014      614  2539            617        3423     0.74
## 4:   2015      612  2633            617        3423     0.76
## 5:   2016      615  1842            617        3423     0.54
```

```
rm(pair_year, product_year, year_coverage)

#For each product, how many year*country pairs / all possible year*country pairs?

year <- hs12[, uniqueN(`Period`)]
year_product <- hs12[, uniqueN(`Period`), by=`Commodity Code`]
year_product <- rename(year_product, Years = V1)

pair_product <- unique(setDT(hs12), by = c("Importer", "Exporter", "Commodity Code"))
pair_product <- pair_product[, .N, by= .(`Commodity Code`)]
pair_product <- rename(pair_product, Pairs = N)

product_coverage <- merge(year_product, pair_product)
product_coverage$Total_years <- year
product_coverage$Total_pairs <- pair

product_coverage$Coverage <- (product_coverage$Years*product_coverage$Pairs)/
  (product_coverage$Total_years*product_coverage$Total_pairs)

product_coverage[order(Coverage)][1:10]
```

```
##     Commodity Code Years Pairs Total_years Total_pairs Coverage
##  1:         010231     2     2           5        3423  0.00023
##  2:         030356     2     2           5        3423  0.00023
##  3:         010633     2     3           5        3423  0.00035
##  4:         020830     4     3           5        3423  0.00070
##  5:         030446     3     4           5        3423  0.00070
##  6:         030455     3     4           5        3423  0.00070
##  7:         030564     4     5           5        3423  0.00117
##  8:         010239     5     5           5        3423  0.00146
##  9:         020840     5     5           5        3423  0.00146
## 10:         021091     5     5           5        3423  0.00146
```

```
product_coverage[order(-Coverage)][1:10]
```

```
##     Commodity Code Years Pairs Total_years Total_pairs Coverage
##  1:             09     5  2006           5        3423     0.59
##  2:             08     5  1929           5        3423     0.56
##  3:             07     5  1796           5        3423     0.52
##  4:             03     5  1741           5        3423     0.51
##  5:             04     5  1601           5        3423     0.47
##  6:             06     5  1242           5        3423     0.36
##  7:           0901     5  1239           5        3423     0.36
##  8:             02     5  1106           5        3423     0.32
##  9:             01     5  1105           5        3423     0.32
## 10:             05     5  1104           5        3423     0.32
```

```
rm(year_product, pair_product, product_coverage)

#For each country pair, how many year*product / all possible year*product?

product_pair <- hs12[, uniqueN(`Commodity Code`), by = c("Importer", "Exporter")]
product_pair <- rename(product_pair, Products = V1)

year_pair <- hs12[, uniqueN(`Period`), by = c("Importer", "Exporter")]
year_pair <- rename(year_pair, Years = V1)

pair_coverage <- merge(product_pair, year_pair, by = c("Importer", "Exporter"))
```

```
pair_coverage$T_products <- product
pair_coverage$T_years <- year

pair_coverage$Coverage <- (pair_coverage$Products*pair_coverage$Years)/
  (pair_coverage$T_products*pair_coverage$T_years)

pair_coverage$Exporter <- strtrim(pair_coverage$Exporter, 15)
pair_coverage[order(Coverage)][1:10]
```

```
##       Importer        Exporter Products Years T_products T_years Coverage
##  1:    Albania       Australia        1     1        617       5  0.00032
##  2:    Albania Bolivia (Plurin        1     1        617       5  0.00032
##  3:    Albania China, Hong Kon        1     1        617       5  0.00032
##  4:    Albania  Dominican Rep.        1     1        617       5  0.00032
##  5:    Albania         Morocco        1     1        617       5  0.00032
##  6:    Algeria China, Hong Kon        1     1        617       5  0.00032
##  7:    Algeria  Dominican Rep.        1     1        617       5  0.00032
##  8:  Argentina         Austria        1     1        617       5  0.00032
##  9: Azerbaijan         Morocco        1     1        617       5  0.00032
## 10:    Bahrain  CÃ´te d'Ivoire        1     1        617       5  0.00032
```

```
pair_coverage[order(-Coverage)][1:10]
```

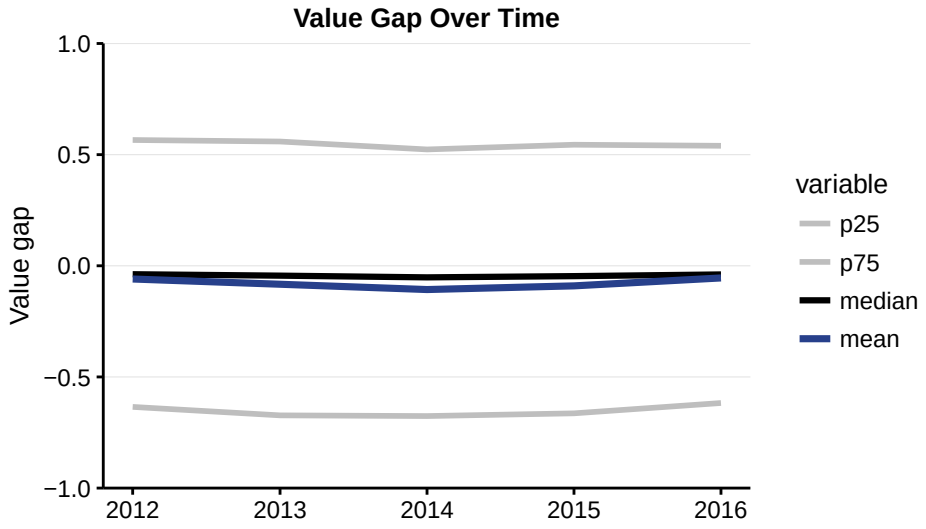
```
##       Importer Exporter Products Years T_products T_years Coverage
##  1:    Belgium   France      581     5        617       5     0.94
##  2:      Italy   France      551     5        617       5     0.89
##  3:    Germany   France      546     5        617       5     0.88
##  4:     France  Belgium      545     5        617       5     0.88
##  5: Luxembourg   France      545     5        617       5     0.88
##  6: Luxembourg  Belgium      537     5        617       5     0.87
##  7:     France    Italy      534     5        617       5     0.87
##  8:     France  Germany      528     5        617       5     0.86
##  9:    Belgium  Germany      518     5        617       5     0.84
## 10:      Italy  Germany      516     5        617       5     0.84
```

```
rm(product_pair, year_pair, pair_coverage, pair, product, year)
```

Trade gap over time

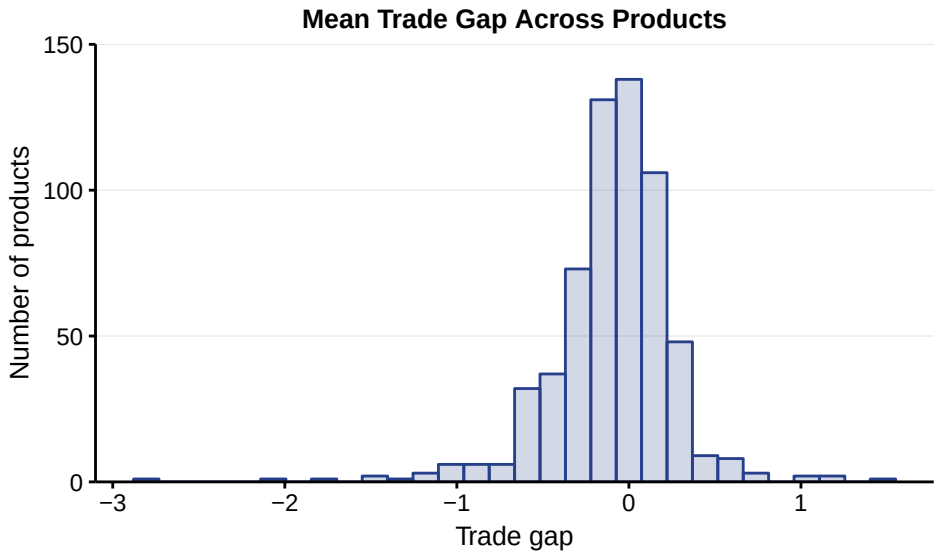
```
#How has the trade gap changed over time?

hs12$Period <- as.Date(hs12$Period, "%Y")
hs12$Period <- floor_date(hs12$Period,"year")

periods <- hs12[, .(mean = as.double(mean(Log_gap)),
                    median = as.double(median(Log_gap)),
                    p25 = as.double(quantile(Log_gap,.25)),
                    p75 = as.double(quantile(Log_gap,.75))
),
by=Period]

periods <- melt(periods, id = 'Period')
periods$variable <- factor(periods$variable, levels = c("p25","p75","median","mean"))

ggplot(data=periods ) +
  geom_line(data=periods, aes(x = Period, y = value, colour = variable, size=variable)) +
  scale_colour_manual(values=c("grey","grey","black","royalblue4")) +
  background_grid(major = 'y', minor = "none") +
  scale_size_manual(values = c(1,1,1.1,1.25)) +
```

```
  scale_y_continuous(expand = c(0, 0), limits = c(-1,1), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Value gap") +
  labs(title="Value Gap Over Time")
```

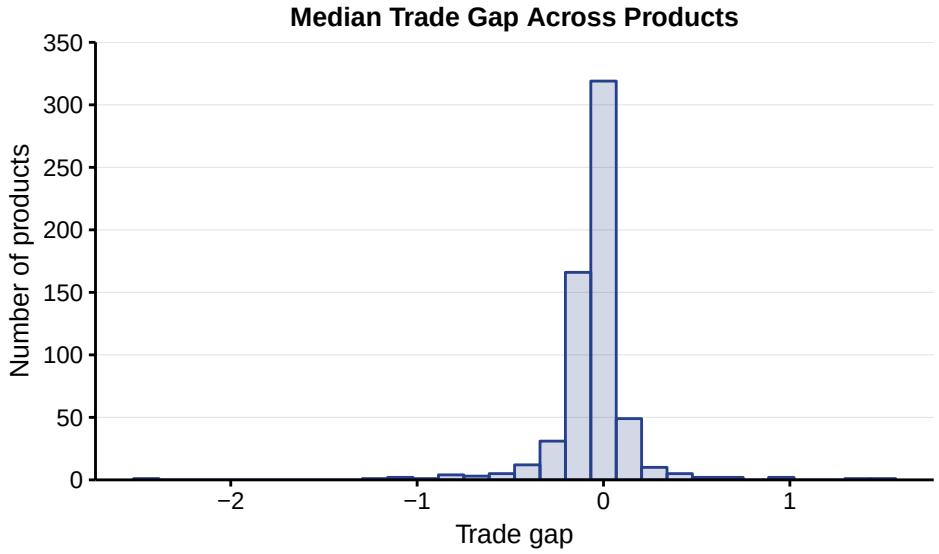


Trade gap across products

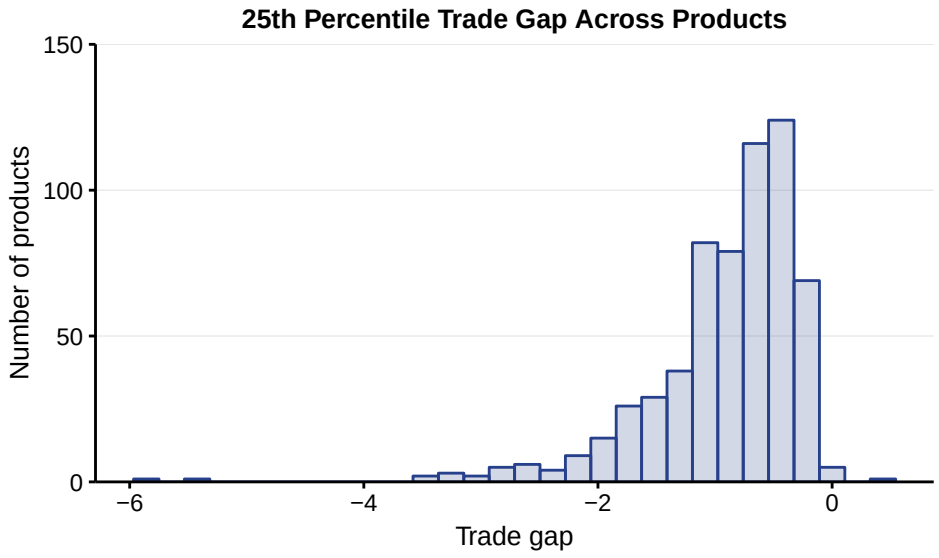
```
#Across products?
products <- hs12[, .(mean = as.double(mean(Log_gap)),
                     median = as.double(median(Log_gap)),
                     p25 = as.double(quantile(Log_gap,.25)),
                     p75 = as.double(quantile(Log_gap,.75))
),
by= `Commodity Code`]

ggplot(data=products, aes(mean)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,150), minor_breaks = NULL) +
  labs(title="Mean Trade Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```



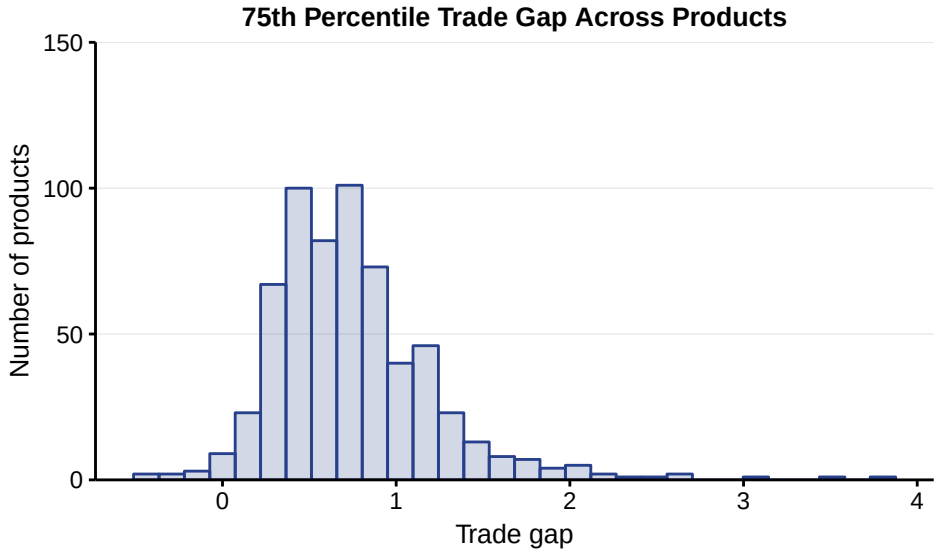

```
ggplot(data=products, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,350),  breaks=seq(0, 350, 50)) +
  labs(title="Median Trade Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```




```
ggplot(data=products, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0),  limits = c(0,150), minor_breaks = NULL) +
  labs(title="25th Percentile Trade Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```



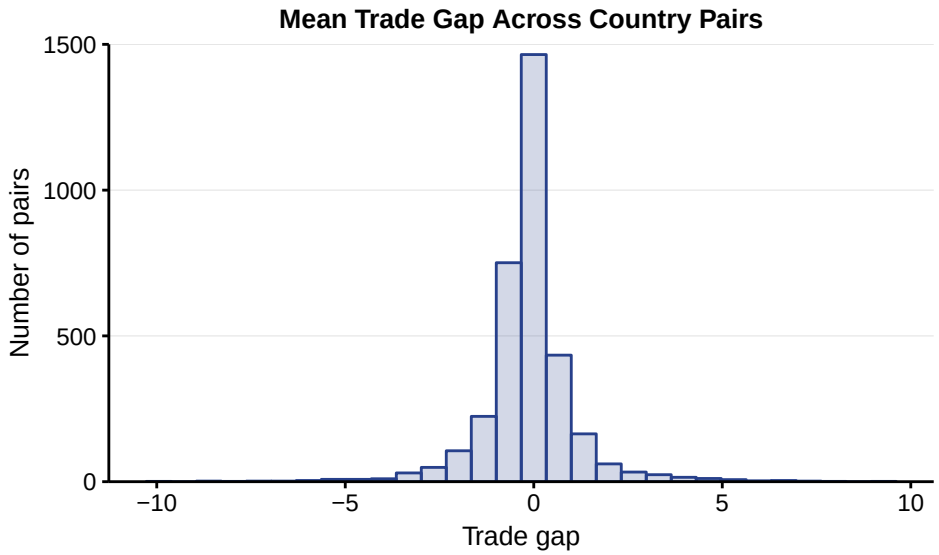

```
ggplot(data=products, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0),  limits = c(0, 150), minor_breaks = NULL) +
  labs(title="75th Percentile Trade Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```



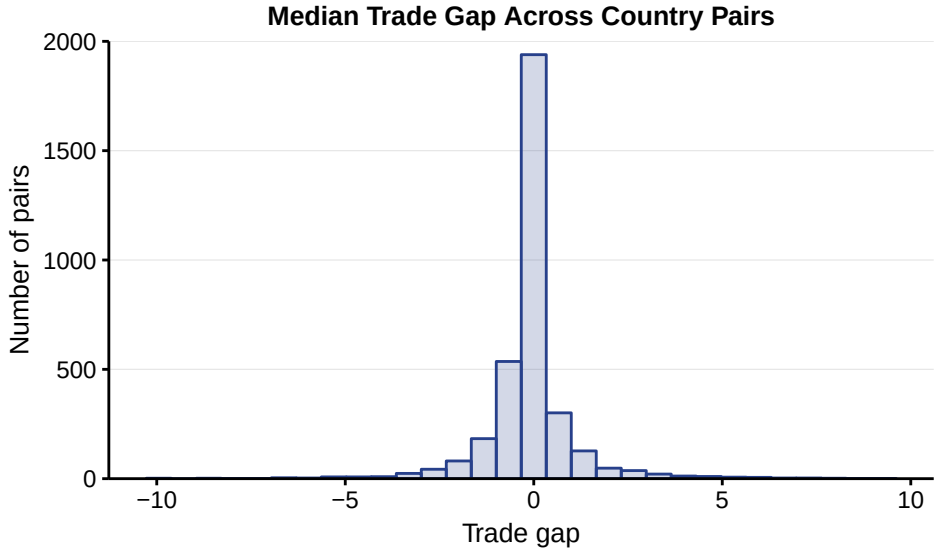
Trade gap across country pairs

```
#Across countries?
countries <- hs12[, .(mean = as.double(mean(Log_gap)),
                      median = as.double(median(Log_gap)),
                      p25 = as.double(quantile(Log_gap,.25)),
                      p75 = as.double(quantile(Log_gap,.75))
),
by= c("Importer", "Exporter")]

ggplot(data=countries, aes(mean)) +
```

```
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 1500),  minor_breaks = NULL) +
  labs(title="Mean Trade Gap Across Country Pairs") +
  labs(x="Trade gap", y="Number of pairs")
```

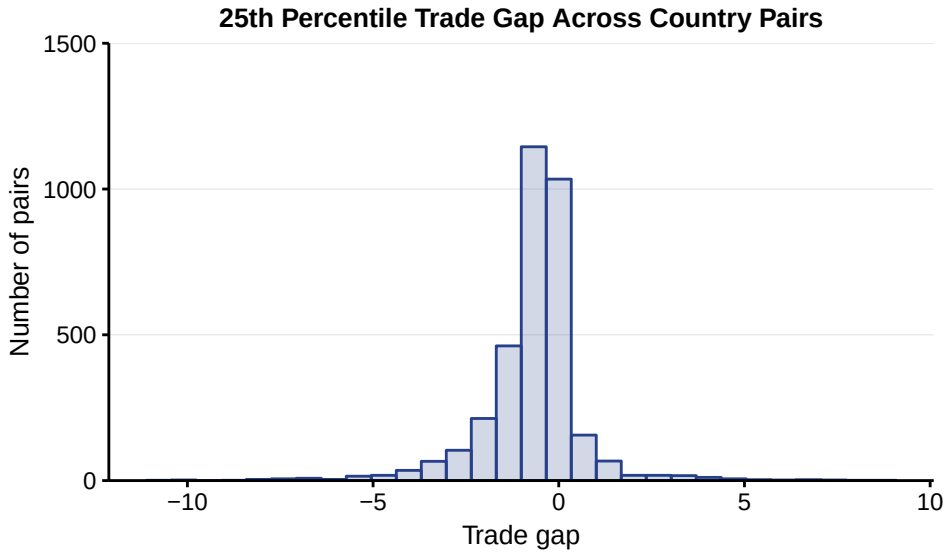



```
ggplot(data=countries, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 2000),  minor_breaks = NULL) +
  labs(title="Median Trade Gap Across Country Pairs") +
  labs(x="Trade gap", y="Number of pairs")
```

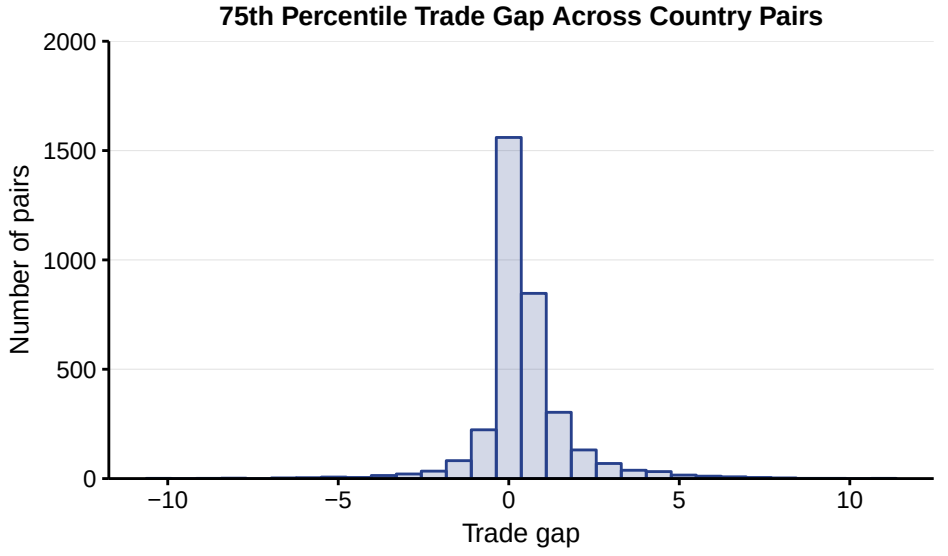



```
ggplot(data=countries, aes(p25)) +
  geom_histogram(col="royalblue4",
```

```
                fill="royalblue4",
                alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 1500),  minor_breaks = NULL) +
  labs(title="25th Percentile Trade Gap Across Country Pairs") +
  labs(x="Trade gap", y="Number of pairs")
```



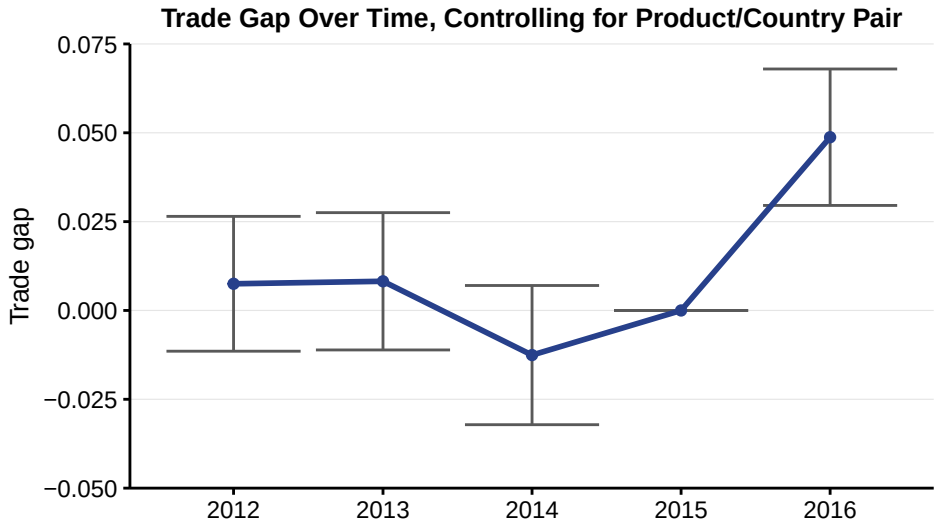

```
ggplot(data=countries, aes(p75)) +
  geom_histogram(col="royalblue4",
                fill="royalblue4",
                alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 2000),  minor_breaks = NULL) +
  labs(title="75th Percentile Trade Gap Across Country Pairs") +
  labs(x="Trade gap", y="Number of pairs")
```




```
rm(periods, products, countries)
```

Year coefficients controlling for product codes and country pairs

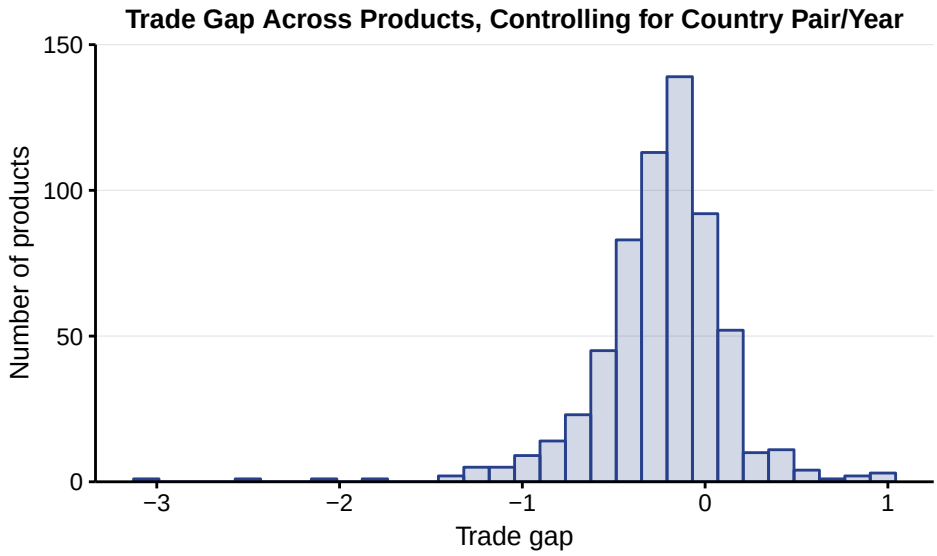
```r
hs12$Period <- as.Date(hs12$Period, "%Y")
hs12$Period <- floor_date(hs12$Period,"year")

hs12$Period.f <- factor(hs12$Period)
hs12$Products.f <- factor(hs12$`Commodity Code`)

hs12$Importer.f <- factor(hs12$`Reporter Code`)
hs12$Exporter.f <- factor(hs12$`Partner Code`)
hs12$Pairs.f <- with(hs12, interaction(Importer.f, Exporter.f))

reg <- felm(Log_gap ~ 1 | Period.f + Products.f + Pairs.f,
            data = hs12,
            exactDOF = FALSE,
            keepX = FALSE,
            keepCX = FALSE)

fes <- getfe(reg,
             se=TRUE,
             bN = 50
)

periodfes <- subset(fes,fe == "Period.f")

periodfes$ci_ub <- periodfes$effect + (1.96 * periodfes$se)
periodfes$ci_lb <- periodfes$effect - (1.96 * periodfes$se)
periodfes <- merge(periodfes,unique(hs12[,list(Period,Period.f)]),by.x = "idx",by.y="Period.f")
periodfes <- rename(periodfes, period = Period)

ggplot(data = periodfes, aes(period,effect)) +
  geom_errorbar(aes(ymin = ci_lb, ymax = ci_ub), color = "grey35") +
  geom_line(color = "royalblue4", size = 1) +
  geom_point(color = "royalblue4") +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(-.050,.075), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Trade gap") +
  labs(title = "Trade Gap Over Time, Controlling for Product/Country Pair")
```



Why the 2016 bump? Could it be something to do with how they revise value estimates when they get more data or convert to most recent HS classification?

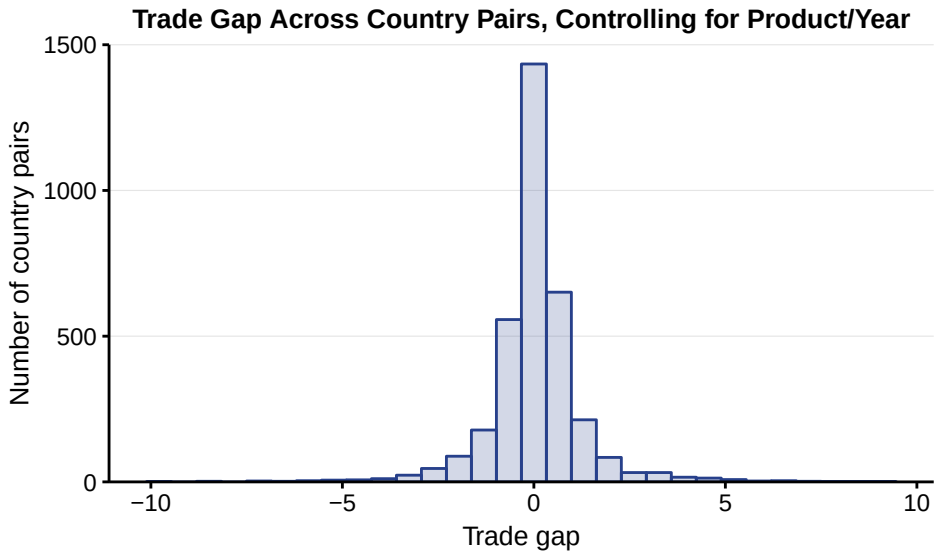
Product code coefficients controlling for country pairs and years

```
productfes <- subset(fes,fe == "Products.f")
productfes <- productfes[,c("effect","idx")]

ggplot(data=productfes, aes(effect)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,150), minor_breaks = NULL) +
  labs(title="Trade Gap Across Products, Controlling for Country Pair/Year") +
  labs(x="Trade gap", y="Number of products")
```



Country pair coefficients controlling for years and product codes

```
pairfes <- subset(fes,fe == "Pairs.f")
pairfes <- pairfes[,c("effect","idx")]

ggplot(data=pairfes, aes(effect)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,1500), minor_breaks = NULL) +
  labs(title="Trade Gap Across Country Pairs, Controlling for Product/Year") +
  labs(x="Trade gap", y="Number of country pairs")
```




```
rm(fes, hs12, pairfes, periodfes, productfes, reg)
```

Quantity Trade Gap

The difference between what the exporting country reports and what the importing country reports in netweight (kg).

Coverage

```
options(digits=2)

load(paste(DataPath,"Analysis Data/hs12.Rda", sep = "/"))
hs12 <- as.data.table(hs12)

#Remove observations where one or more countries do not report quantities. 748,839 rows deleted.
hs12 <- hs12[!is.na(Qty_log_gap)]

#There are 411 instances where `qty_log_gap` = inf.
#I removed them -- something else we should do?
hs12 <- hs12[!is.infinite(Qty_log_gap)]

#For each year, how many product*country pairs / all possible product*country pairs?

product <- hs12[, uniqueN(`Commodity Code`)]
product_year <- hs12[, uniqueN(`Commodity Code`), by=Period]
product_year <- rename(product_year, Products = V1)

pair <- unique(setDT(hs12), by = c("Importer", "Exporter"))
pair <- pair[, .N]
pair_year <- unique(setDT(hs12), by = c("Importer", "Exporter", "Period"))
pair_year <- pair_year[, .N, by=Period]
pair_year <- rename(pair_year, Pairs = N)

year_coverage <- merge(product_year, pair_year)
year_coverage$Total_products <- product
year_coverage$Total_pairs <- pair
```

```
year_coverage$Coverage <- (year_coverage$Products*year_coverage$Pairs)/
  (year_coverage$Total_products*year_coverage$Total_pair)

year_coverage
```

```
##    Period Products Pairs Total_products Total_pairs Coverage
## 1:   2012      604  1652            608        3292     0.50
## 2:   2013      604  2153            608        3292     0.65
## 3:   2014      603  2441            608        3292     0.74
## 4:   2015      601  2519            608        3292     0.76
## 5:   2016      605  1778            608        3292     0.54
```

```
rm(pair_year, product_year, year_coverage)

#For each product, how many year*country pairs / all possible year*country pairs?

year <- hs12[, uniqueN(`Period`)]
year_product <- hs12[, uniqueN(`Period`), by=`Commodity Code`]
year_product <- rename(year_product, Years = V1)

pair_product <- unique(setDT(hs12), by = c("Importer", "Exporter", "Commodity Code"))
pair_product <- pair_product[, .N, by= .(`Commodity Code`)]
pair_product <- rename(pair_product, Pairs = N)

product_coverage <- merge(year_product, pair_product)
product_coverage$Total_years <- year
product_coverage$Total_pairs <- pair

product_coverage$Coverage <- (product_coverage$Years*product_coverage$Pairs)/
  (product_coverage$Total_years*product_coverage$Total_pairs)

#Note: Quantity is not reported at the two-digit level
product_coverage[order(Coverage)][1:10]
```

```
##     Commodity Code Years Pairs Total_years Total_pairs Coverage
##  1:         010612     1     2           5        3292  0.00012
##  2:         010231     2     2           5        3292  0.00024
##  3:         010633     2     2           5        3292  0.00024
##  4:         030356     2     2           5        3292  0.00024
##  5:         010239     3     3           5        3292  0.00055
##  6:         020830     4     3           5        3292  0.00073
##  7:         030446     3     4           5        3292  0.00073
##  8:         030455     3     4           5        3292  0.00073
##  9:         010613     4     5           5        3292  0.00122
## 10:         030564     4     5           5        3292  0.00122
```

```
product_coverage[order(-Coverage)][1:10]
```

```
##     Commodity Code Years Pairs Total_years Total_pairs Coverage
##  1:           0901     5  1231           5        3292     0.37
##  2:           0902     5  1022           5        3292     0.31
##  3:           0713     5   986           5        3292     0.30
##  4:           0910     5   982           5        3292     0.30
##  5:           0303     5   981           5        3292     0.30
##  6:           0406     5   940           5        3292     0.29
##  7:           0602     5   863           5        3292     0.26
##  8:           0904     5   863           5        3292     0.26
##  9:           0712     5   853           5        3292     0.26
```

```
## 10:           0402     5   847           5        3292     0.26
```

```
rm(year_product, pair_product, product_coverage)

#For each country pair, how many year*product / all possible year*product?

product_pair <- hs12[, uniqueN(`Commodity Code`), by = c("Importer", "Exporter")]
product_pair <- rename(product_pair, Products = V1)

year_pair <- hs12[, uniqueN(`Period`), by = c("Importer", "Exporter")]
year_pair <- rename(year_pair, Years = V1)

pair_coverage <- merge(product_pair, year_pair, by = c("Importer", "Exporter"))
pair_coverage$T_products <- product
pair_coverage$T_years <- year

pair_coverage$Coverage <- (pair_coverage$Products*pair_coverage$Years)/
  (pair_coverage$T_products*pair_coverage$T_years)

pair_coverage$Exporter <- strtrim(pair_coverage$Exporter, 15)
pair_coverage[order(Coverage)][1:10]
```

```
##       Importer   Exporter Products Years T_products T_years Coverage
##  1:    Albania Madagascar        1     1        608       5  0.00033
##  2:    Algeria Cabo Verde        1     1        608       5  0.00033
##  3:    Algeria     Guinea        1     1        608       5  0.00033
##  4:     Angola Costa Rica        1     1        608       5  0.00033
##  5:     Angola   Honduras        1     1        608       5  0.00033
##  6: Argentina    Hungary        1     1        608       5  0.00033
##  7: Argentina    Lebanon        1     1        608       5  0.00033
##  8:    Armenia     Cyprus        1     1        608       5  0.00033
##  9:    Armenia    Estonia        1     1        608       5  0.00033
## 10:    Austria   Cambodia        1     1        608       5  0.00033
```

```
pair_coverage[order(-Coverage)][1:10]
```

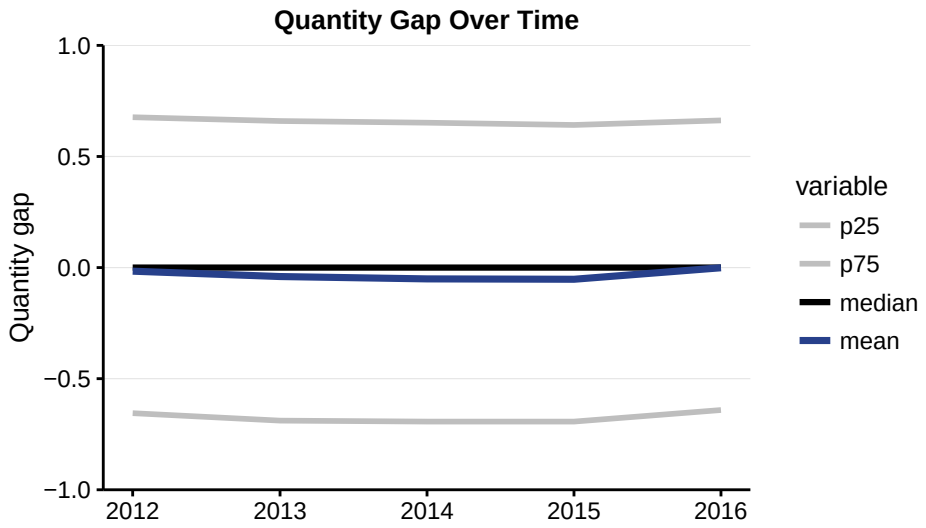
```
##       Importer Exporter Products Years T_products T_years Coverage
##  1:     Belgium   France      564     5        608       5     0.93
##  2:       Italy   France      542     5        608       5     0.89
##  3:     Germany   France      537     5        608       5     0.88
##  4: Luxembourg   France      536     5        608       5     0.88
##  5:      France  Belgium      526     5        608       5     0.87
##  6:      France    Italy      525     5        608       5     0.86
##  7: Luxembourg  Belgium      523     5        608       5     0.86
##  8:      France  Germany      519     5        608       5     0.85
##  9:       Italy  Germany      507     5        608       5     0.83
## 10:     Germany    Italy      505     5        608       5     0.83
```

```
rm(product_pair, year_pair, pair_coverage, pair, product, year)
```

Quantity trade gap over time

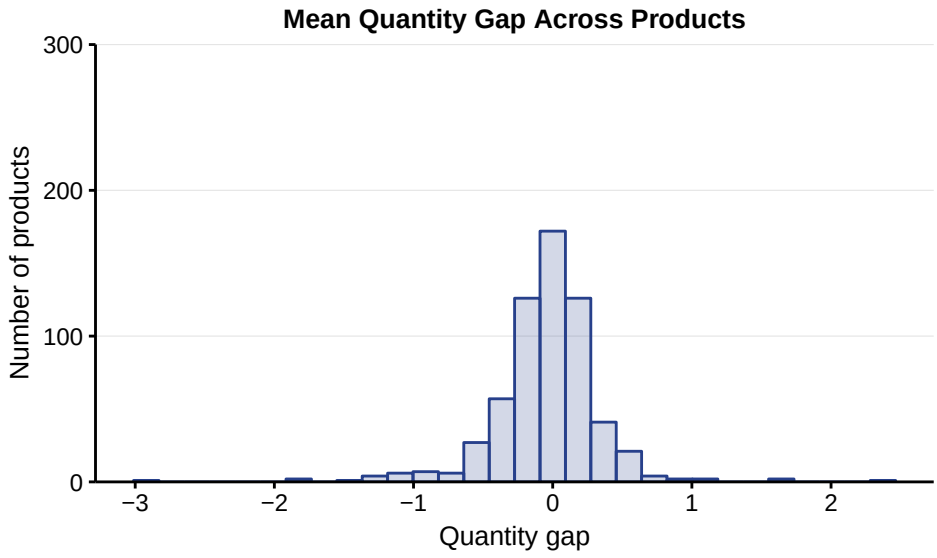
```
hs12$Period <- as.Date(hs12$Period, "%Y")
hs12$Period <- floor_date(hs12$Period,"year")


periods <- hs12[, .(mean = as.double(mean(Qty_log_gap)),
                    median = as.double(median(Qty_log_gap)),
                    p25 = as.double(quantile(Qty_log_gap,.25)),
                    p75 = as.double(quantile(Qty_log_gap,.75))
```

```
),
by=Period]

periods <- melt(periods, id = 'Period')
periods$variable <- factor(periods$variable, levels = c("p25","p75","median","mean"))

ggplot(data=periods ) +
  geom_line(data=periods, aes(x = Period, y = value, colour = variable, size=variable)) +
  background_grid(major = 'y', minor = "none") +
  scale_colour_manual(values=c("grey","grey","black","royalblue4")) +
  scale_size_manual(values = c(1,1,1.1,1.25)) +
  scale_y_continuous(expand = c(0, 0), limits = c(-1,1), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Quantity gap") +
  labs(title="Quantity Gap Over Time")
```

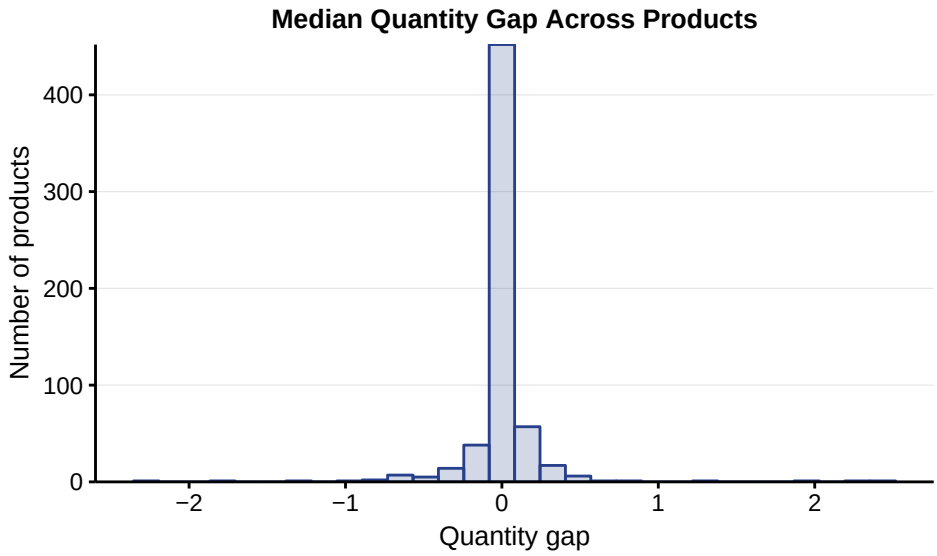


Quantity trade gap across products

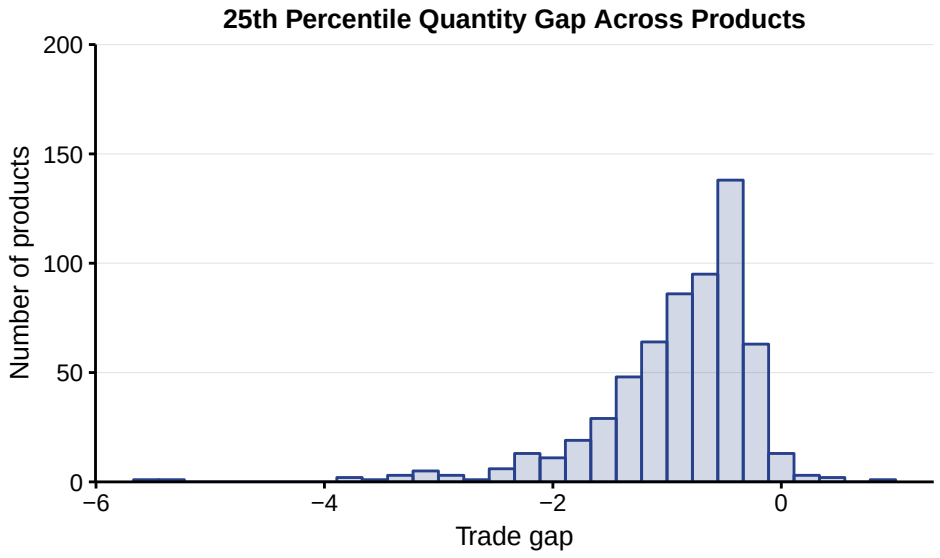
```
products <- hs12[, .(mean = as.double(mean(Qty_log_gap)),
                     median = as.double(median(Qty_log_gap)),
                     p25 = as.double(quantile(Qty_log_gap,.25)),
                     p75 = as.double(quantile(Qty_log_gap,.75))
),
by= `Commodity Code`]

ggplot(data=products, aes(mean)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,300), minor_breaks = NULL) +
  labs(title="Mean Quantity Gap Across Products") +
  labs(x="Quantity gap", y="Number of products")
```



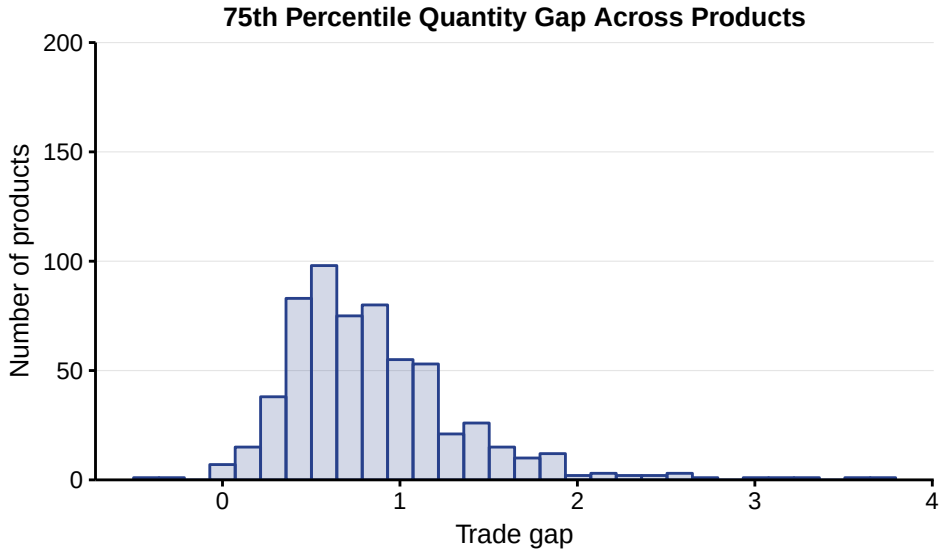

```
ggplot(data=products, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0)) +
  labs(title="Median Quantity Gap Across Products") +
  labs(x="Quantity gap", y="Number of products")
```




```
ggplot(data=products, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,200), minor_breaks = NULL) +
  labs(title="25th Percentile Quantity Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```



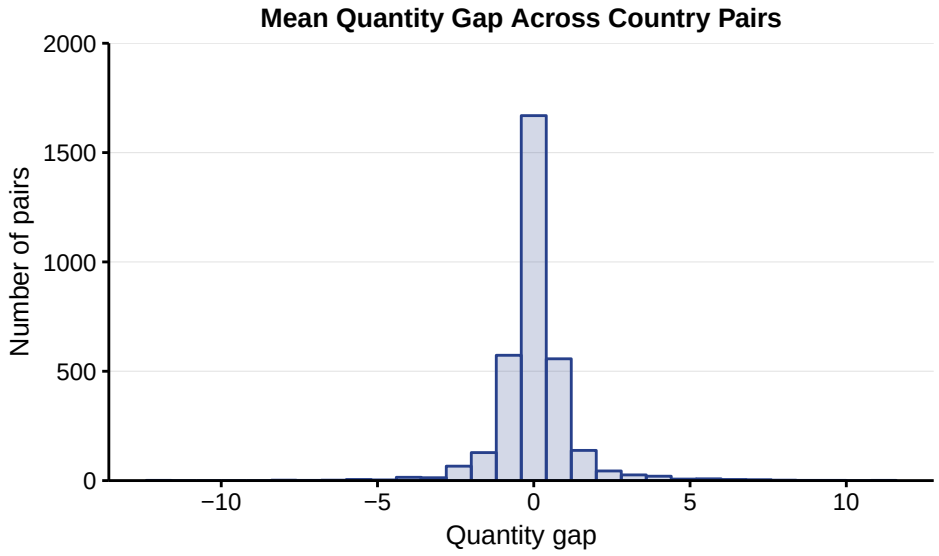

```
ggplot(data=products, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  scale_y_continuous(expand = c(0, 0), limits = c(0,200), minor_breaks = NULL) +
  background_grid(major = 'y', minor = "none") +
  labs(title="75th Percentile Quantity Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```



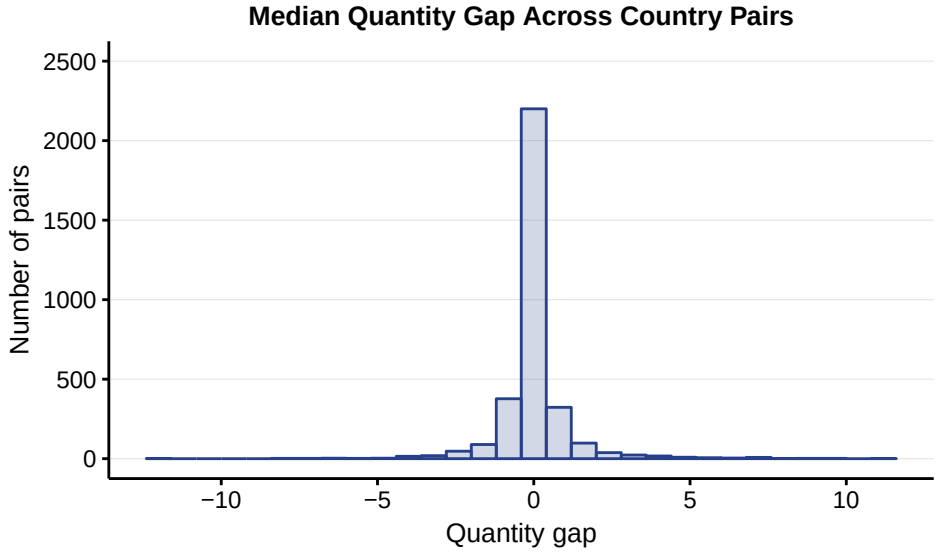
Quantity trade gap across country pairs

```
countries <- hs12[, .(mean = as.double(mean(Qty_log_gap)),
                      median = as.double(median(Qty_log_gap)),
                      p25 = as.double(quantile(Qty_log_gap,.25)),
                      p75 = as.double(quantile(Qty_log_gap,.75))
),
by= c("Importer", "Exporter")]

ggplot(data=countries, aes(mean)) +
  geom_histogram(col="royalblue4",
```

```
                fill="royalblue4",
                alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,2000), minor_breaks = NULL) +
  labs(title="Mean Quantity Gap Across Country Pairs") +
  labs(x="Quantity gap", y="Number of pairs")
```

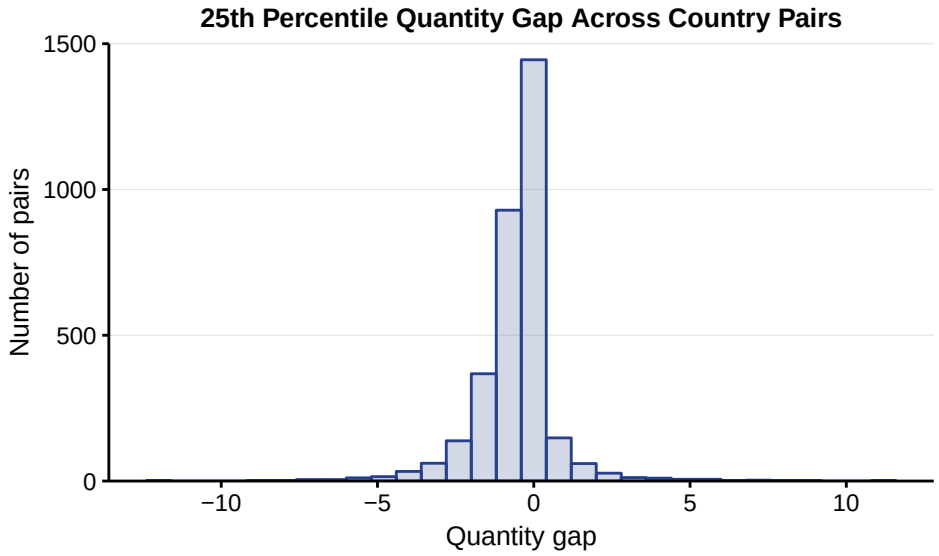



```
ggplot(data=countries, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(limits = c(0,2500), minor_breaks = NULL) +
  labs(title="Median Quantity Gap Across Country Pairs") +
  labs(x="Quantity gap", y="Number of pairs")
```

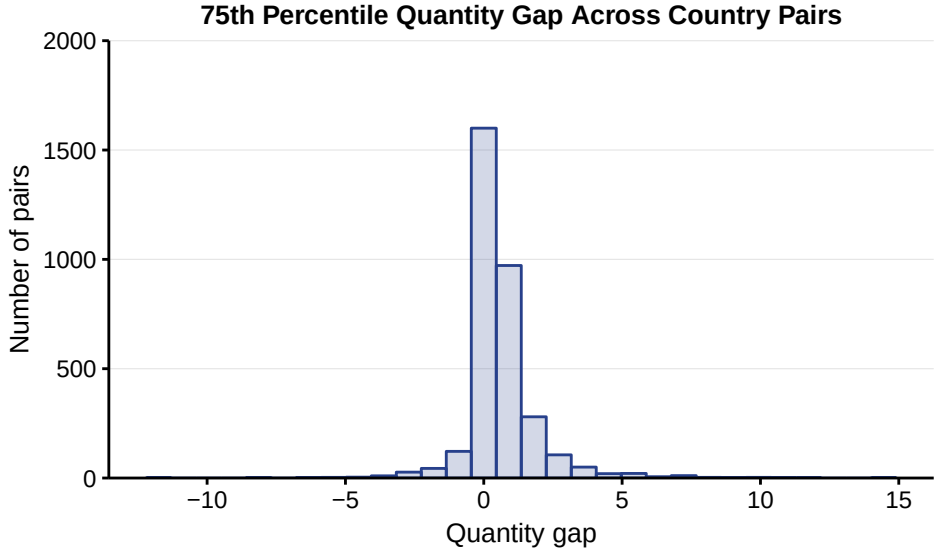



```
ggplot(data=countries, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
```

```
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,1500), minor_breaks = NULL) +
  labs(title="25th Percentile Quantity Gap Across Country Pairs") +
  labs(x="Quantity gap", y="Number of pairs")
```




```
ggplot(data=countries, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,2000), minor_breaks = NULL) +
  labs(title="75th Percentile Quantity Gap Across Country Pairs") +
  labs(x="Quantity gap", y="Number of pairs")
```

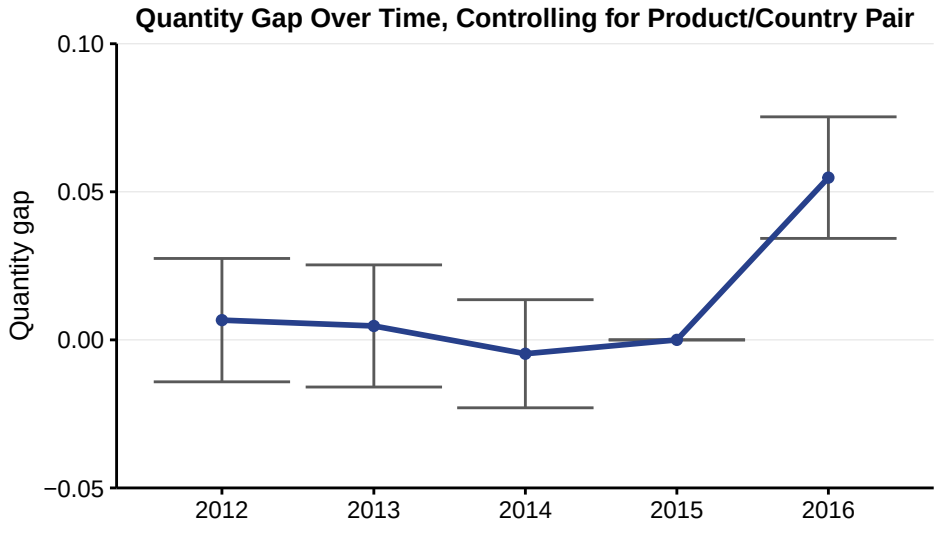



```
rm(periods, products, countries)
```

Quantity year coefficients controlling for product codes and country pairs

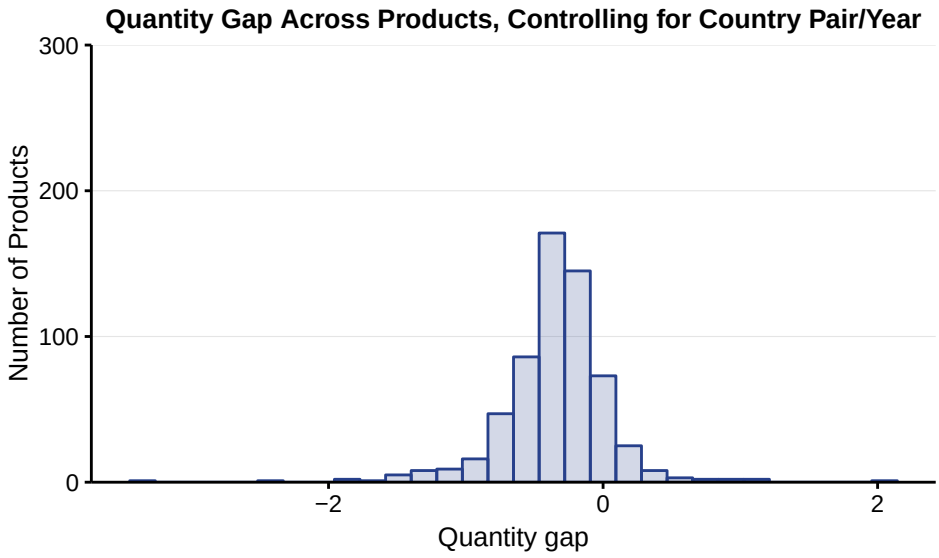
```
hs12$Period <- as.Date(hs12$Period, "%Y")
hs12$Period <- floor_date(hs12$Period,"year")
```

```
hs12$Period.f <- factor(hs12$Period)
hs12$Products.f <- factor(hs12$`Commodity Code`)

hs12$Importer.f <- factor(hs12$`Reporter Code`)
hs12$Exporter.f <- factor(hs12$`Partner Code`)
hs12$Pairs.f <- with(hs12, interaction(Importer.f, Exporter.f))

reg <- felm(Qty_log_gap ~ 1 | Period.f + Products.f + Pairs.f,
            data = hs12,
            exactDOF = FALSE,
            keepX = FALSE,
            keepCX = FALSE)

fes <- getfe(reg,
             se=TRUE,
             bN = 50
)

periodfes <- subset(fes,fe == "Period.f")

periodfes$ci_ub <- periodfes$effect + (1.96 * periodfes$se)
periodfes$ci_lb <- periodfes$effect - (1.96 * periodfes$se)
periodfes <- merge(periodfes,unique(hs12[,list(Period,Period.f)]),by.x = "idx",by.y="Period.f")
periodfes <- rename(periodfes, period = Period)

ggplot(data = periodfes, aes(period,effect)) +
  geom_errorbar(aes(ymin = ci_lb, ymax = ci_ub), color = "grey35") +
  geom_line(color = "royalblue4", size = 1) +
  geom_point(color = "royalblue4") +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(-.05, .10), breaks=seq(-.05, .10, .05)) +
  xlab(label = "") +
  ylab(label = "Quantity gap") +
  labs(title = "Quantity Gap Over Time, Controlling for Product/Country Pair")
```



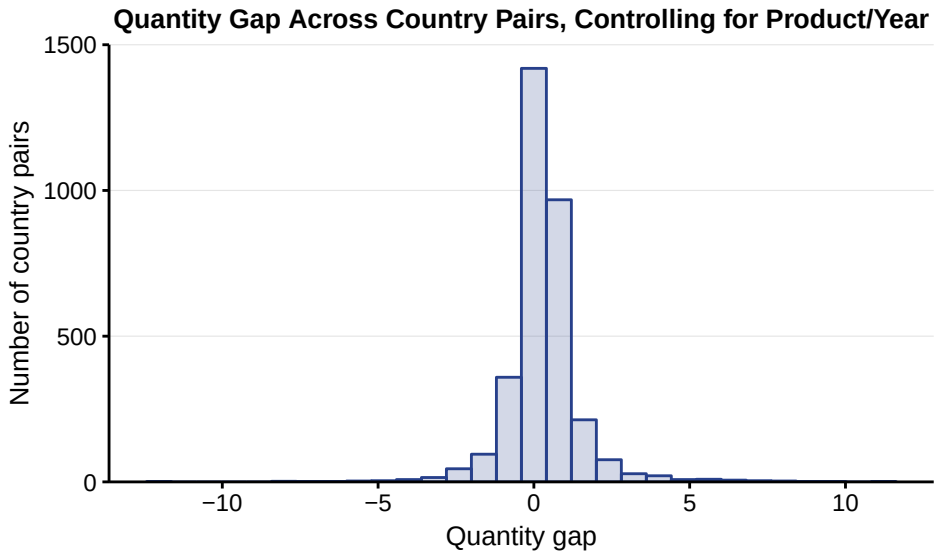
Quantity product code coefficients controlling for country pairs and years

```
productfes <- subset(fes,fe == "Products.f")
productfes <- productfes[,c("effect","idx")]

ggplot(data=productfes, aes(effect)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,300), minor_breaks = NULL) +
  labs(title="Quantity Gap Across Products, Controlling for Country Pair/Year") +
  labs(x="Quantity gap", y="Number of Products")
```



Quantity country pair coefficients controlling for years and product codes

```
pairfes <- subset(fes,fe == "Pairs.f")
pairfes <- pairfes[,c("effect","idx")]

ggplot(data=pairfes, aes(effect)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,1500), minor_breaks = NULL) +
  labs(title="Quantity Gap Across Country Pairs, Controlling for Product/Year") +
  labs(x="Quantity gap", y="Number of country pairs")
```




```
rm(fes, hs12, pairfes, periodfes, productfes, reg)
```