```
> # ------------------------------------------------------------------------
> #                       Program Description
> # -------------------- .... [TRUNCATED]

> # Load in packages
> library(foreign)

> library(data.table)

> # ------------------------------------------------------------------------
> #                           Small Firms
> # ------------------------- .... [TRUNCATED]

> # For faster processing, save data as data.table
> DSMALL <- data.table(tmpread, key=c("firm_id"))

> rm(tmpread) # free the temp file

> # Commenting Variables
> comment(DSMALL$firm_id) <- "Firm's Identifier"

> comment(DSMALL$areacode) <- "Firm's Administrative Region"

> comment(DSMALL$industry) <- "4 Digits GB2002 Sector"

> comment(DSMALL$type) <- "Firm's Ownership Rights Type"

> comment(DSMALL$statectl) <- "State Control of Firm"

> comment(DSMALL$admintier) <-
+     "Level of government firm registers under"

> comment(DSMALL$founding_y) <- "Founding Year"

> comment(DSMALL$founding_m) <- "Founding Month"

> comment(DSMALL$status) <- "Operating Status of Firm"

> comment(DSMALL$nworkers) <- "# of Workers END of year"

> comment(DSMALL$product) <- "Total Industrial Output Value"

> comment(DSMALL$sales) <- "Total Sales Value"

> comment(DSMALL$export) <- "Total Value of sales to abroad"

> comment(DSMALL$capital) <- "Book Value of Fixed Capital"

> comment(DSMALL$cur_depre) <- "Current Year Depreciation"
```

```
> comment(DSMALL$sales_cost) <-
+     "Production Costs for Products Sold"

> comment(DSMALL$sales_tax) <- "Taxes and Fees for Products Sold"

> comment(DSMALL$total_cost) <-
+     "Miscellaneous Costs of Production"

> comment(DSMALL$wage) <- "Total Wage Compensation"

> comment(DSMALL$nonwage) <- "Total non-wage Compensation"

> comment(DSMALL$nbarworkers) <- "Average Number of Workers"

> comment(DSMALL$total_rev) <- "Total Operating Revenue"

> comment(DSMALL$sales_rev) <- "Total Sales Revenue"

> comment(DSMALL$opr_pro) <- "Operating Profits"

> comment(DSMALL$total_pro) <- "Total Profits"

> # Save in R internal binary form
> save(DSMALL, file = "./Data/DSMALL_R.RData")

> rm(DSMALL)

> # ------------------------------------------------------------------------
> #                            Large Firms
> # ------------------------- .... [TRUNCATED]

> # For faster processing, save data as data.table
> DLARGE <- data.table(tmpread, key=c("firm_id"))

> rm(tmpread) # free the temp file

> # Commenting Variables
> # Basic Infomation
> comment(DLARGE$firm_id) <- "Firm's Identifier"

> comment(DLARGE$areacode) <- "Firm's Administrative Region"

> comment(DLARGE$industry) <- "4 Digits GB2002 Sector"

> comment(DLARGE$type) <- "Firm's Ownership Rights Type"

> comment(DLARGE$statectl) <- "State Control of Firm"

> comment(DLARGE$admintier) <- "Level of government firm registers under"
```

```
> comment(DLARGE$founding_y) <- "Founding Year"

> comment(DLARGE$founding_m) <- "Founding Month"

> comment(DLARGE$status) <- "Operating Status of Firm"

> comment(DLARGE$nworkers) <- "# of Workers END of year"

> comment(DLARGE$nbarworkers) <- "Annual Average # of Workers"

> # Production and Financial
> comment(DLARGE$product) <- "Total Industrial Output Value"

> comment(DLARGE$sales) <- "Total Sales Value"

> comment(DLARGE$export) <- "Total Value of sales to abroad"

> comment(DLARGE$totcapital) <- "Current Book Value of Capital"

> comment(DLARGE$bookcapital) <- "Original Book Value of Capital"

> comment(DLARGE$bookcap_prod) <-
+     "Book Value of Capital Used in Production"

> comment(DLARGE$cumdepr) <- "Cumulative Depreciation"

> comment(DLARGE$curdepr) <- "Current Year Depreciation"

> comment(DLARGE$nbarcapital) <- "Average Net Value of Capital"

> # Wage
> comment(DLARGE$ui) <- "Labor and Unemployment Insurance"

> comment(DLARGE$pension) <- "Pension and Medicare"

> comment(DLARGE$housing) <- "Housing Accumulation Fund"

> comment(DLARGE$wage) <- "Total wage bill"

> comment(DLARGE$nonwage) <- "Total non-wage bill"

> comment(DLARGE$wagemajor) <-
+     "Total wage bill associated with major business"

> comment(DLARGE$nonwagemajor) <-
+     "Total non-wage bill associated with majo business"

> # Value-added
> comment(DLARGE$vatax) <- "Value-added Tax"
```

```
> comment(DLARGE$itax) <- "Imported Related Tax (Expenditure)"

> comment(DLARGE$stax) <- "Sales Related Tax (Revenue)"

> comment(DLARGE$inter) <- "Total value of intermediate input"

> comment(DLARGE$intermat) <-
+     "Total value of intermediate production material"

> comment(DLARGE$manuinter) <-
+     "Intermediate Input associated with Manufacturing"

> comment(DLARGE$manainter) <-
+     "Intermediate Input associated with Management"

> comment(DLARGE$oprinter) <-
+     "Intermediate Input associated with Firm Operating"

> # Updated Information
> comment(DLARGE$revmajor) <- "Revenue from Major Business"

> comment(DLARGE$costmajor) <- "Costs from Major Business"

> comment(DLARGE$taxmajor) <- "Taxes from Major Business"

> comment(DLARGE$revminor) <- "Revenue from Minor Business"

> comment(DLARGE$profminor) <- "Profits from Minor Business"

> comment(DLARGE$oprcost) <- "Operating Costs"

> comment(DLARGE$manacost) <- "Management Costs"

> comment(DLARGE$manac_tax) <- "Taxes in Management Costs"

> comment(DLARGE$manac_insu) <- "Assets Insurance in Management Costs"

> comment(DLARGE$manac_travel) <- "Travel Costs in Management Costs"

> comment(DLARGE$manac_union) <-
+     "Union Activity Costs in Management Costs"

> comment(DLARGE$manac_busi) <-
+     "Business Operating Costs in Management Costs"

> comment(DLARGE$manac_edu) <-
+     "Employees Education in Management Costs"

> comment(DLARGE$manac_emsn) <- "Pollution Emission Fee"
```

```
> comment(DLARGE$fincost) <- "Financial Costs"

> comment(DLARGE$fincostint) <- "Interest rates in Financial Costs"

> comment(DLARGE$profopr) <- "Operating Profits"

> comment(DLARGE$invreturn) <- "Investment Returns"

> comment(DLARGE$subsidy) <- "Subsidy"

> comment(DLARGE$revother) <- "Revenue from Other Activities"

> comment(DLARGE$expenother) <-
+     "Expenditure from Other Activities"

> comment(DLARGE$proftot) <- "Total Profits"

> comment(DLARGE$taxprof) <- "Total Profit Taxes"

> comment(DLARGE$ads) <- "Advertisement Expenditures"

> # Save in R internal binary form
> save(DLARGE, file = "./Data/DLARGE_R.RData")

> rm(DLARGE)

> # ---------------------------------------------------------------------
> #                     Merge the Two Datasets
> # -------------------- .... [TRUNCATED]

> DLARGE <- as.data.table(DLARGE)

> load("./Data/DSMALL_R.RData")  # Small firms

> DSMALL <- as.data.table(DSMALL)

> # Merge the two samples
> # Variables Needed:
> # industry, status, industry_a, type_a,
> #   founding_y, nbarworkers, product, totcapital, wage, n .... [TRUNCATED]

> sjunk <- DSMALL[,list(status,industry, industry_a, type_a, founding_y,
+       province, product, capital, nbarworkers, wage, nonwage)]

> setnames(sjunk,"capital","totcapital")

> CNEC_avgp <- rbind(ljunk,sjunk)

> CNEC_avgp$type_a <- factor(CNEC_avgp$type_a)
```

```
> CNEC_avgp$province <- factor(CNEC_avgp$province)
> CNEC_avgp$industry <- factor(CNEC_avgp$industry)
> CNEC_avgp$industry_a <- factor(CNEC_avgp$industry_a)
> CNEC <- CNEC_avgp
> save(CNEC_avgp, file = "./Data/CNEC_avgp.RData")
> rm(DLARGE,DSMALL,ljunk,sjunk, CNEC_avgp)
```