```
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
----------
      name:  <unnamed>
  log type:  text
 opened on:  18 Sep 2019, 15:54:19

. /* ===================================================================
>                                                    1. Small Firms
>   =================================================================== */
> use cec2004_small.dta, clear ;

. * Rename variables ;
. rename dm firm_id ;

. rename b056 areacode ;

. rename b07 industry ;

. rename b10 type ;

. rename b210 statectl ;

. rename b09 admintier ;

. rename b131 founding_y ;

. rename b132 founding_m ;

. rename b14 status ;

. rename b201 nworkers ;

. rename v210 nbarworkers ;

. rename v207 product ;

. rename v209 sales ;

. rename v213 export ;

. rename f310 capital ;

. rename f313 cur_depre ;

. rename f327 sales_cost ;

. rename f330 sales_tax ;

. rename f331 total_cost ;
```

```
. rename f373 wage ;

. rename f339 nonwage ;

. rename f370 total_rev ;

. rename f325 sales_rev ;

. rename f342 opr_pro ;

. rename f344 total_pro ;

. * Label variables ;
. label variable firm_id "Firm ID";

. label variable areacode "Area code up to the level of subdistrict";

. label variable industry "4-digits Industrial Sector Code (GB2002)";

. label variable type "Ownership Rights of the Firm";

. label variable statectl "Shares in the firm owned by the State";

. label variable admintier ///
>         "Level of Administrative Layer the firm belongs to";

. label variable founding_y "Founding Year";

. label variable founding_m "Founding Month";

. label variable status "Operating Status";

. label variable nworkers "Number of Workers End of Year";

. label variable nbarworkers "Average Annual Number of Workers";

. label variable product "Total value of output (in RMB 1000)";

. label variable sales "Total VALUE of sales (in RMB 1000)";

. label variable export "Total Value of sales from exporting";

. label variable capital "Book Value of Fixed Capital";

. label variable cur_depre "Current Year Depreciation";

. label variable sales_cost "Production Costs for Products Sold";

. label variable sales_tax "Taxes and Fees for Products Sold";
```

```
. label variable total_cost "Miscellaneous Costs of Production";

. label variable wage "Total Wage Compensation";

. label variable nonwage "Total non-wage Compensation";

. label variable total_rev "Total Operating Revenue";

. label variable sales_rev "Total Sales Revenue";

. label variable opr_pro "Operating Profits";

. label variable total_pro "Total Profits";

. destring, replace ;
firm_id contains nonnumeric characters; no replace
areacode has all characters numeric; replaced as double
industry has all characters numeric; replaced as int
type has all characters numeric; replaced as int
statectl has all characters numeric; replaced as byte
(83 missing values generated)
admintier has all characters numeric; replaced as byte
founding_y has all characters numeric; replaced as int
(28498 missing values generated)
founding_m has all characters numeric; replaced as byte
(28498 missing values generated)
status has all characters numeric; replaced as byte
(83 missing values generated)
nworkers already numeric; no replace
b201_02 already numeric; no replace
b201_03 already numeric; no replace
b201_04 already numeric; no replace
b201_05 already numeric; no replace
b201_06 already numeric; no replace
b201_07 already numeric; no replace
b201_08 already numeric; no replace
b201_09 already numeric; no replace
b201_10 already numeric; no replace
b201_11 already numeric; no replace
b201_12 already numeric; no replace
b201_13 already numeric; no replace
b202 already numeric; no replace
b202_02 already numeric; no replace
b202_03 already numeric; no replace
b202_04 already numeric; no replace
b202_05 already numeric; no replace
b202_06 already numeric; no replace
b202_07 already numeric; no replace
b202_08 already numeric; no replace
```

```
b202_09 already numeric; no replace
b202_10 already numeric; no replace
b202_11 already numeric; no replace
b202_12 already numeric; no replace
b202_13 already numeric; no replace
product already numeric; no replace
sales already numeric; no replace
export already numeric; no replace
f300 already numeric; no replace
f305 already numeric; no replace
capital already numeric; no replace
cur_depre already numeric; no replace
f318 already numeric; no replace
f322 already numeric; no replace
f301 already numeric; no replace
f302 already numeric; no replace
f392 already numeric; no replace
f393 already numeric; no replace
f394 already numeric; no replace
f395 already numeric; no replace
f303 already numeric; no replace
total_rev already numeric; no replace
sales_rev already numeric; no replace
sales_cost already numeric; no replace
sales_tax already numeric; no replace
total_cost already numeric; no replace
opr_pro already numeric; no replace
total_pro already numeric; no replace
wage already numeric; no replace
nonwage already numeric; no replace
nbarworkers already numeric; no replace

. * Aggregate counties to provinces ;
. recode areacode (110000000000/119999999999 = 11)
>         (120000000000/129999999999 = 12) (130000000000/139999999999 = 13)
>         (140000000000/149999999999 = 14) (150000000000/159999999999 = 15)
>         (210000000000/219999999999 = 21) (220000000000/229999999999 = 22)
>         (230000000000/239999999999 = 23) (310000000000/319999999999 = 31)
>         (320000000000/329999999999 = 32) (330000000000/339999999999 = 33)
>         (340000000000/349999999999 = 34) (350000000000/359999999999 = 35)
>         (360000000000/369999999999 = 36) (370000000000/379999999999 = 37)
>         (410000000000/419999999999 = 41) (420000000000/429999999999 = 42)
>         (430000000000/439999999999 = 43) (440000000000/449999999999 = 44)
>         (450000000000/459999999999 = 45) (460000000000/469999999999 = 46)
>         (500000000000/509999999999 = 50) (510000000000/519999999999 = 51)
>         (520000000000/529999999999 = 52) (530000000000/539999999999 = 53)
>         (540000000000/549999999999 = 54) (610000000000/619999999999 = 61)
>         (620000000000/629999999999 = 62) (630000000000/639999999999 = 63)
>         (640000000000/649999999999 = 64) (650000000000/659999999999 = 65),
>         generate(province) ;
```

```
(1098738 differences between areacode and province)

. * Aggregate sectors ;
. recode industry
>         (600/699 = 6) (700/799 = 7) (800/899 = 8) (900/999 = 9)
>         (1000/1099 = 10) (1100/1199 = 11) (1300/1399 = 13) (1400/1499 = 14)
>         (1500/1599 = 15) (1600/1699 = 16) (1700/1799 = 17) (1800/1899 = 18)
>         (1900/1999 = 19) (2000/2099 = 20) (2100/2199 = 21) (2200/2299 = 22)
>         (2300/2399 = 23) (2400/2499 = 24) (2500/2599 = 25) (2600/2699 = 26)
>         (2700/2799 = 27) (2800/2899 = 28) (2900/2999 = 29) (3000/3099 = 30)
>         (3100/3199 = 31) (3200/3299 = 32) (3300/3399 = 33) (3400/3499 = 34)
>         (3500/3599 = 35) (3600/3699 = 36) (3700/3799 = 37) (3900/3999 = 39)
>         (4000/4099 = 40) (4100/4199 = 41) (4200/4299 = 42) (4300/4399 = 43)
>         (4400/4499 = 44) (4500/4599 = 45) (4600/4699 = 46),
>         generate(industry_a) ;
(1098738 differences between industry and industry_a)

. * Aggregate ownership rights ;
. * Data dictionary for type_a:
> *        -> 1-State, 2-Hybrid/collective, 3-Private, 4-HMT, 5-Foreign ;
. recode statectl (1/2 = 0) (9 = 1), generate(stmp) ;
(1098655 differences between statectl and stmp)

. generate type_a = 1 if (type == 110 | type == 141 | type == 151
>                 | (type == 120 & stmp == 0) | (type == 130 & stmp == 0)
>                 | (type == 142 & stmp == 0) | (type == 143 & stmp == 0)
>                 | (type == 149 & stmp == 0) | (type == 159 & stmp == 0)
>                 | (type == 160 & stmp == 0) | (type == 190 & stmp == 0)) ;
(1087344 missing values generated)

. replace type_a = 2 if ((type == 120 & stmp == 1)
>                 | (type == 130 & stmp == 1) | (type == 142 & stmp == 1)
>                 | (type == 143 & stmp == 1) | (type == 149 & stmp == 1)
>                 | (type == 159 & stmp == 1) | (type == 160 & stmp == 1)) ;
(232828 real changes made)

. replace type_a = 3 if (type >= 171 & type <= 174 | (type == 190 & stmp == 1)) ;
(805467 real changes made)

. replace type_a = 4 if (type >= 210 & type <= 240) ;
(26509 real changes made)

. replace type_a = 5 if (type >= 310 & type <= 340) ;
(22489 real changes made)

. drop stmp ;

. label variable type_a "Aggregated Type" ;

. label variable industry_a "2-digits Industry code" ;
```

```
. label variable province "Province" ;

. save cec2004_small.dta, replace ;
file cec2004_small.dta saved

. /* ==================================================================
>                                                     2. Large Firms
>  ================================================================== */
> use cec2004_large_full.dta, clear ;

. * Rename variables ;
. rename dm firm_id ;

. rename b056 areacode ;

. rename b07 industry ;

. rename b10 type ;

. rename b210 statectl ;

. rename b09 admintier ;

. rename b131 founding_y ;

. rename b132 founding_m ;

. rename b14 status ;

. rename b201 nworkers ;

. rename v210 nbarworkers ;

. rename v207 product ;

. rename v209 sales ;

. rename v213 export ;

. * Capital ;
. rename f309 totcapital ;

. rename f310 bookcapital ;

. rename f311 bookcap_prod ;

. rename f312 cumdepr ;

. rename f313 curdepr ;
```

```
. rename f314 nbarcapital ;

. * wage ;
. rename f339 ui ;

. rename f3 pension ;

. rename f11 housing ;

. rename f349 wage ;

. rename f350 wagemajor ;

. rename f351 nonwage ;

. rename f352 nonwagemajor ;

. * value-added ;
. rename f353 vatax ;

. rename f354 itax ;

. rename f355 stax ;

. rename f356 inter ;

. rename f357 intermat ;

. rename f358 manuinter ;

. rename f359 manainter ;

. rename f360 oprinter ;

. * misc information ;
. rename f325 revmajor ;

. rename f327 costmajor ;

. rename f330 taxmajor ;

. rename f334 revminor ;

. rename f335 profminor ;

. rename f328 oprcost ;

. rename f336 manacost ;
```

```
. rename f337 manac_tax ;

. rename f338 manac_insu ;

. rename f8 manac_travel ;

. rename f9 manac_union ;

. rename f4 manac_busi ;

. rename f6 manac_edu ;

. rename f10 manac_emsn ;

. rename f340 fincost ;

. rename f341 fincostint ;

. rename f342 profopr ;

. rename f7 invreturn ;

. rename f343 subsidy ;

. rename f371 revother ;

. rename f372 expenother ;

. rename f344 proftot ;

. rename f345 taxprof ;

. rename f2 ads ;

. * Label variables ;
. label variable firm_id "Firm's Identifier";

. label variable areacode "Firm's Administrative Region";

. label variable industry "4 Digits GB2002 Sector";

. label variable type "Firm's Ownership Rights Type";

. label variable statectl "State Control of Firm";

. label variable admintier ///
>          "Level of government firm registers under";

. label variable founding_y "Founding Year";
```

```
.  label variable founding_m "Founding Month";

.  label variable status "Operating Status of Firm";

.  label variable nworkers "# of Workers END of year";

.  label variable nbarworkers "Annual average # of workers";

.  label variable product "Total Industrial Output Value";

.  label variable sales "Total Sales Value";

.  label variable export "Total Value of sales to abroad";

. * Capital ;
.  label variable totcapital "Book Value of Capital";

.  label variable bookcapital "Original Book Value of Capital";

.  label variable bookcap_prod ///
>          "Book Value of Capital Used in Production";

.  label variable cumdepr "Cumulative Depreciation";

.  label variable curdepr "Current Year Depreciation";

.  label variable nbarcapital "Average Net Value of Capital";

. * wage ;
.  label variable ui "Labor and Unemployment Insurance";

.  label variable pension "Pension and Medicare";

.  label variable housing "Housing Accumulation Fund";

.  label variable wage "Total wage bill";

.  label variable wagemajor "Total non-wage bill";

.  label variable nonwage "Total wage bill associated with major business";

.  label variable nonwagemajor ///
>          "Total non-wage bill associated with majo business";

. * value-added ;
.  label variable vatax "Value-added Tax";

.  label variable itax "Imported Related Tax (Expenditure)";

.  label variable stax "Sales Related Tax (Revenue)";
```

```
. label variable inter "Total value of intermediate input";

. label variable intermat ///
>          "Total value of intermediate production material";

. label variable manuinter ///
>          "Intermediate Input associated with Manufacturing";

. label variable manainter ///
>          "Intermediate Input associated with Management";

. label variable oprinter ///
>          "Intermediate Input associated with Firm Operating";

. * misc information ;
. label variable revmajor "Revenue from Major Business";

. label variable costmajor "Costs from Major Business";

. label variable taxmajor "Taxes from Major Business";

. label variable revminor "Revenue from Minor Business";

. label variable profminor "Profits from Minor Business";

. label variable oprcost "Operating Costs";

. label variable manacost "Management Costs";

. label variable manac_tax "Taxes in Management Costs";

. label variable manac_insu ///
>          "Assets Insurance in Management Costs";

. label variable manac_travel ///
>          "Travel Costs in Management Costs";

. label variable manac_union ///
>          "Union Activity Costs in Management Costs";

. label variable manac_busi ///
>          "Business Operating Costs in Management Costs";

. label variable manac_edu ///
>          "Employees Education in Management Costs";

. label variable manac_emsn "Pollution Emission Fee";

. label variable fincost "Financial Costs";
```

```
. label variable fincostint ///
>         "Interest rates in Financial Costs";

. label variable profopr "Operating Profits";

. label variable invreturn "Investment Returns";

. label variable subsidy "Subsidy";

. label variable revother "Revenue from Other Activities";

. label variable expenother ///
>         "Expenditure from Other Activities";

. label variable proftot "Total Profits";

. label variable taxprof "Total Profit Taxes";

. label variable ads "Advertisement Expenditures";

. destring, replace ;
firm_id contains nonnumeric characters; no replace
areacode has all characters numeric; replaced as double
industry has all characters numeric; replaced as int
type has all characters numeric; replaced as int
statectl has all characters numeric; replaced as byte
(108 missing values generated)
admintier has all characters numeric; replaced as byte
founding_y has all characters numeric; replaced as int
(236 missing values generated)
founding_m has all characters numeric; replaced as byte
(236 missing values generated)
status has all characters numeric; replaced as byte
(107 missing values generated)
nworkers already numeric; no replace
b201_02 already numeric; no replace
b201_03 already numeric; no replace
b201_04 already numeric; no replace
b201_05 already numeric; no replace
b201_06 already numeric; no replace
b201_07 already numeric; no replace
b201_08 already numeric; no replace
b201_09 already numeric; no replace
b201_10 already numeric; no replace
b201_11 already numeric; no replace
b201_12 already numeric; no replace
b201_13 already numeric; no replace
b202 already numeric; no replace
b202_02 already numeric; no replace
```

```
b202_03 already numeric; no replace
b202_04 already numeric; no replace
b202_05 already numeric; no replace
b202_06 already numeric; no replace
b202_07 already numeric; no replace
b202_08 already numeric; no replace
b202_09 already numeric; no replace
b202_10 already numeric; no replace
b202_11 already numeric; no replace
b202_12 already numeric; no replace
b202_13 already numeric; no replace
product already numeric; no replace
sales already numeric; no replace
export already numeric; no replace
f300 already numeric; no replace
f304 already numeric; no replace
f1 already numeric; no replace
f390 already numeric; no replace
f305 already numeric; no replace
f306 already numeric; no replace
f308 already numeric; no replace
f391 already numeric; no replace
totcapital already numeric; no replace
bookcapital already numeric; no replace
bookcap_prod already numeric; no replace
cumdepr already numeric; no replace
curdepr already numeric; no replace
nbarcapital already numeric; no replace
f317 already numeric; no replace
f318 already numeric; no replace
f319 already numeric; no replace
f389 already numeric; no replace
f321 already numeric; no replace
f322 already numeric; no replace
f323 already numeric; no replace
f301 already numeric; no replace
f302 already numeric; no replace
f392 already numeric; no replace
f393 already numeric; no replace
f394 already numeric; no replace
f395 already numeric; no replace
f303 already numeric; no replace
revmajor already numeric; no replace
costmajor already numeric; no replace
taxmajor already numeric; no replace
revminor already numeric; no replace
profminor already numeric; no replace
oprcost already numeric; no replace
manacost already numeric; no replace
manac_tax already numeric; no replace
```

```
manac_insu already numeric; no replace
manac_travel already numeric; no replace
manac_union already numeric; no replace
manac_busi already numeric; no replace
manac_edu already numeric; no replace
manac_emsn already numeric; no replace
fincost already numeric; no replace
fincostint already numeric; no replace
profopr already numeric; no replace
invreturn already numeric; no replace
subsidy already numeric; no replace
revother already numeric; no replace
expenother already numeric; no replace
proftot already numeric; no replace
taxprof already numeric; no replace
ads already numeric; no replace
ui already numeric; no replace
pension already numeric; no replace
housing already numeric; no replace
wage already numeric; no replace
wagemajor already numeric; no replace
nonwage already numeric; no replace
nonwagemajor already numeric; no replace
vatax already numeric; no replace
itax already numeric; no replace
stax already numeric; no replace
inter already numeric; no replace
intermat already numeric; no replace
manuinter already numeric; no replace
manainter already numeric; no replace
oprinter already numeric; no replace
nbarworkers already numeric; no replace
v211 already numeric; no replace

. * Aggregate counties to provinces ;
. recode areacode (110000000000/119999999999 = 11)
>          (120000000000/129999999999 = 12) (130000000000/139999999999 = 13)
>          (140000000000/149999999999 = 14) (150000000000/159999999999 = 15)
>          (210000000000/219999999999 = 21) (220000000000/229999999999 = 22)
>          (230000000000/239999999999 = 23) (310000000000/319999999999 = 31)
>          (320000000000/329999999999 = 32) (330000000000/339999999999 = 33)
>          (340000000000/349999999999 = 34) (350000000000/359999999999 = 35)
>          (360000000000/369999999999 = 36) (370000000000/379999999999 = 37)
>          (410000000000/419999999999 = 41) (420000000000/429999999999 = 42)
>          (430000000000/439999999999 = 43) (440000000000/449999999999 = 44)
>          (450000000000/459999999999 = 45) (460000000000/469999999999 = 46)
>          (500000000000/509999999999 = 50) (510000000000/519999999999 = 51)
>          (520000000000/529999999999 = 52) (530000000000/539999999999 = 53)
>          (540000000000/549999999999 = 54) (610000000000/619999999999 = 61)
>          (620000000000/629999999999 = 62) (630000000000/639999999999 = 63)
```

```
>               (640000000000/649999999999 = 64) (650000000000/659999999999 = 65),
>           generate(province) ;
(276410 differences between areacode and province)

. * Aggregate sectors ;
. recode industry
>           (600/699 = 6) (700/799 = 7) (800/899 = 8) (900/999 = 9)
>           (1000/1099 = 10) (1100/1199 = 11) (1300/1399 = 13) (1400/1499 = 14)
>           (1500/1599 = 15) (1600/1699 = 16) (1700/1799 = 17) (1800/1899 = 18)
>           (1900/1999 = 19) (2000/2099 = 20) (2100/2199 = 21) (2200/2299 = 22)
>           (2300/2399 = 23) (2400/2499 = 24) (2500/2599 = 25) (2600/2699 = 26)
>           (2700/2799 = 27) (2800/2899 = 28) (2900/2999 = 29) (3000/3099 = 30)
>           (3100/3199 = 31) (3200/3299 = 32) (3300/3399 = 33) (3400/3499 = 34)
>           (3500/3599 = 35) (3600/3699 = 36) (3700/3799 = 37) (3900/3999 = 39)
>           (4000/4099 = 40) (4100/4199 = 41) (4200/4299 = 42) (4300/4399 = 43)
>           (4400/4499 = 44) (4500/4599 = 45) (4600/4699 = 46),
>           generate(industry_a) ;
(276410 differences between industry and industry_a)

. * Aggregate ownership rights ;
. * Data dictionary for type_a:
> *          -> 1-State, 2-Hybrid/collective, 3-Private, 4-HMT, 5-Foreign ;
. recode statectl (1/2 = 0) (9 = 1), generate(stmp) ;
(276302 differences between statectl and stmp)

. generate type_a = 1 if (type == 110 | type == 141 | type == 151
>                 | (type == 120 & stmp == 0) | (type == 130 & stmp == 0)
>                 | (type == 142 & stmp == 0) | (type == 143 & stmp == 0)
>                 | (type == 149 & stmp == 0) | (type == 159 & stmp == 0)
>                 | (type == 160 & stmp == 0) | (type == 190 & stmp == 0)) ;
(243245 missing values generated)

. replace type_a = 2 if ((type == 120 & stmp == 1)
>                 | (type == 130 & stmp == 1) | (type == 142 & stmp == 1)
>                 | (type == 143 & stmp == 1) | (type == 149 & stmp == 1)
>                 | (type == 159 & stmp == 1) | (type == 160 & stmp == 1)) ;
(66348 real changes made)

. replace type_a = 3 if (type >= 171 & type <= 174 | (type == 190 & stmp == 1)) ;
(119730 real changes made)

. replace type_a = 4 if (type >= 210 & type <= 240) ;
(28399 real changes made)

. replace type_a = 5 if (type >= 310 & type <= 340) ;
(28765 real changes made)

. drop stmp ;

. label variable type_a "Aggregated Type" ;
```

```
.  label variable industry_a "2-digits Industry code" ;

.  label variable province "Province" ;

.  save cec2004_large_full.dta, replace ;
file cec2004_large_full.dta saved

.  log close ;
       name:  <unnamed>
   log type:  text
  closed on:  18 Sep 2019, 15:55:16
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
----------
```