The package ShiftShareSE implements confidence intervals proposed by Adão et al. [2019] for inference in shift-share least squares and instrumental variables regressions, in which the regressor of interest (or the instrument) has a shift-share structure, as in Bartik [1991]. A shift-share variable has the structure $X_i = \sum_{s=1}^{S} w_{is} X_s$, where $i$ indexes regions, $s$ indexes sectors, $X_s$ are sectoral shifters (or shocks), and $w_{is}$ are shares, such as initial share of region $i$’s employment in sector $s$.

This vignette illustrates the use of the package using a dataset from Autor et al. [2013] (ADH hereafter). The dataset is included in the package as the list ADH. The first element of the list, ADH$reg$ is a data-frame with regional variables, the second element, ADH$sic$ is a vector of SIC codes for the sectors, and ADH$W$ is a matrix of shares. See ?ADH for a description of the dataset.

Examples

We now replicate column (1) of Table V in Adão et al. [2019]. First we load the package, define the vector of controls, and define a vector of 3-digit SIC codes:

```r
library("ShiftShareSE")
crls <- paste("t2 + l_shind_manuf_cbp + l_sh_popedu_c +",
 "l_sh_popfborn + l_sh_empl_f + l_sh_routine33 + l_task_outsource",
 "+ division")
sic <- floor(ADH$sic/10)
```

We cluster the standard errors at the 3-digit SIC code (using the option sector_cvar), and, following ADH, weight the data using the weights ADH$reg$weights. See ?reg_ss and ?ivreg_ss for full description of the options.

The first-stage regression:

```r
reg_ss(as.formula(paste("shock ~ ", crlsl)), W = ADH$W,
 X = IV, data = ADH$reg, weights = weights, region_cvar = statefip,
 sector_cvar = sic, method = "all")
```

#> Estimate: 0.6310409

#> Inference:

#> Std. Error    p-value  Lower CI  Upper CI
#> Homoscedastic 0.02732516 0.000000e+00 0.5774846 0.6845973
#> EHW 0.08700719 4.083400e-13 0.4605100 0.8015719
#> Reg. cluster 0.09142372 5.113909e-12 0.4518537 0.8102281
Note that for "AKM0", "Std. Error" corresponds to the normalized standard error, i.e. the length of the confidence interval divided by $2z_{1-\alpha/2}$.

The reduced-form and IV regressions:

\begin{verbatim}
reg_ss(as.formula(paste("d_sh_empl ~", ctrls)), W = ADH$W,
       X = IV, data = ADH$reg, region_cvar = statefip, weights = weights,
       sector_cvar = sic, method = "all")
#> Estimate: -0.4885687
#>
#> Inference:
#> Std. Error p-value Lower CI Upper CI
#> Homoscedastic 0.06332778 1.221245e-14 -0.6126889 -0.3644485
#> EHW 0.11244360 1.392685e-05 -0.7089541 -0.2681833
#> Reg. cluster 0.07578147 1.140306e-10 -0.6370977 -0.3400398
#> AKM 0.16419445 2.924641e-03 -0.8103839 -0.1667535
#> AKM0 0.25437489 4.218033e-04 -1.2368853 -0.2397541
ivreg_ss(as.formula(paste("d_sh_empl ~", ctrls, "| shock")),
         W = ADH$W, X = IV, data = ADH$reg, region_cvar = statefip,
         weights = weights, sector_cvar = sic, method = "all")
#> Estimate: -0.7742267
#>
#> Inference:
#> Std. Error p-value Lower CI Upper CI
#> Homoscedastic 0.1069532 4.523049e-13 -0.9838511 -0.5646022
#> EHW 0.1647892 2.623532e-06 -1.0972075 -0.4512459
#> Reg. cluster 0.1758096 1.096809e-05 -1.1188071 -0.4296462
#> AKM 0.2403730 1.277818e-03 -1.2453492 -0.3031041
#> AKM0 0.3318966 4.218033e-04 -1.6903240 -0.3893132
\end{verbatim}

### Note on collinear sectors

Let $W$ denote the share matrix with the $(i,s)$ element given by $w_{is}$. Suppose that columns of $W$ are collinear, so it that it has rank $S_0 < S$. Without loss of generality, suppose that the first $S_0$ columns of the matrix are full rank, so that the collinearity is caused by the last $S - S_0$ sectors. In this case, it is not possible to recover, $\tilde{X}_s$, the sectoral shifters with the controls partialled out, and the `reg_ss` and `ivreg_ss` functions will return an error message "Share matrix is collinear".

The researcher can either (i) drop the collinear sectors, defining $X_i = \sum_{s=1}^{S_0} w_{is}X_s$, (ii) aggregate the sectors, or (iii) if the only controls are those with shift-share structure, and we have data on $Z_s$, we can estimate $\tilde{X}_s$ by running a sector-level regression of $X_s$ onto $Z_s$, and taking the residual. This third option is not currently implemented in this package. Note that options (i) and (ii) change the definition of the estimand. Since they involve changing the shock vector $X_i$, this has to be done before using the `reg_ss` and `ivreg_ss` functions.
References

