Package 'BartikSE'

August 7, 2019

Title Inference in regressions with shift-share structure

Version 0.1.5

Description Provides confidence intervals in least-squares regressions when the variable of interest has a shift-share structure, and in instrumental variables regressions when the instrument has a shift-share structure.

Depends R (>= 3.4.0)

License GPL-3

Encoding UTF-8

LazyData true

Suggests spelling, testthat, knitr, rmarkdown, ggplot2, AER

Imports Formula

VignetteBuilder knitr

RoxygenNote 6.1.1

URL https://github.com/kolesarm/BartikSE

Language en-US

BugReports https://github.com/kolesarm/BartikSE/issues

R topics documented:

ADH *Dataset from Autor, Dorn and Hanson (2013)*

Description

Subset of data from Autor, Dorn and Hanson (2013, ADH) that is used to illustrate the confidence intervals implemented in this package.

Usage

```
ADH
```

Format

A list, consisting of a data frame, a vector, and a matrix. The first data frame, `ADH$reg`, has 1,444 rows and 16 variables. The rows correspond to 722 commuting zones (CZ) over 2 time periods (1990-1999 and 2000-2007), and the variables are as follows:

d_sh_empl Change in the share of working-age population

d_sh_empl_mfg Change in the share of working-age population employed in manufacturing.

d_sh_empl_nmfg Change in the share of working-age population employed in non-manufacturing.

shock Change in sectoral U.S. imports from China normalized by U.S. total employment in the corresponding sector, aggregated to regional level. This is the variable of interest in ADH.

IV Change in sectoral imports from China by rest of the world, aggregated to regional level. This is the variable used to instrument for `shock`, called `d_tradeotch_pw_lag` in ADH.

weights Regression weights corresponding to start of period CZ share of national populations

statefip State FIPS code

czone CZ number

t2 Indicator for 2000-2007

l_shind_manuf_cbp Employment share of manufacturing

l_sh_popedu_c percent population college-educated

l_sh_popfborn percent population foreign-born

l_sh_empl_f percent employment among women

l_sh_routine33 percent employment in routine occupations

l_task_outsource Offshorability index of occupations in CZ

division US Census division of CZ

The second list component, the vector `ADH$sic` is a vector of length 770 that gives 4-digit SIC industry codes for the sectors used to construct the shift-share IV `ADH$reg$IV`. Finally, `ADH$W` is a 1444-by-700 matrix of shares that correspond to the CZ employment shares in 4-digit SIC sectors.

Source

We thank David Dorn for helping us with the construction of the share matrix. The remaining data was obtained from David Dorn's website, http://ddorn.net/data.htm.

References

Autor, David H., David Dorn, and Gordon H. Hanson, "The China syndrome: Local labor market effects of import competition in the United States," American Economic Review, 2013, 103 (6), 2121–68.

Adão, Rodrigo, Kolesár, Michal, and Morales, Eduardo, "Shift-Share Designs: Theory and Inference", 2018, arXiv:1806.07928

`ivBartik` *Inference in an IV regression with a shift-share instrument*

Description

Computes confidence intervals and p-values in an instrumental variables regression in which the instrument has a shift-share structure, as in Bartik (1991). Several different inference methods can computed, as specified by `method`.

Usage

```
ivBartik(formula, X, data, W, subset, weights, method, beta0 = 0,
  alpha = 0.05, region_cvar = NULL, sector_cvar = NULL)
```

Arguments

`formula`	object of class `"formula"` (or one that can be coerced to that class) of the form `outcome ~ controls | endogenous_regressor`. For a regression with no controls (only an intercept), it takes the form `outcome ~ 1 | endogenous_regressor`
`X`	Shift-share vector with length `N` of sectoral shocks, aggregated to regional level using the share matrix `W`.
`data`	optional data frame, list or environment (or object coercible by `as.data.frame` to a data frame) containing the outcome and running variables in the model. If not found in `data`, the variables are taken from `environment(formula)`, typically the environment from which the function is called.
`W`	A matrix of sector shares, so that `W[i, s]` corresponds to share of sector `s` in region `i`
`subset`	optional vector specifying a subset of observations to be used in the fitting process.
`weights`	an optional vector of weights to be used in the fitting process. Should be `NULL` or a numeric vector. If non-`NULL`, for computing the first stage and the reduced form, weighted least squares is used with weights `weights` (that is, minimizing `sum(weights*residuals^2)`); otherwise ordinary least squares is used.

method	Vector specifying which inference methods to use. The vector elements have to be one or more of the following strings: "homosk" Assume i.i.d. homoskedastic errors "ehw" Eicker-Huber-White standard errors "region_cluster" Standard errors clustered at regional level "akm" Adão-Kolesár-Morales "akm0" Adão-Kolesár-Morales with null imposed. Note the reported standard error for this method corresponds to the normalized standard error, given by the length of the confidence interval divided by $2z_{1-\alpha/2}$ "all" All of the methods above
beta0	null that is tested (only affects reported p-values)
alpha	Determines confidence level of reported confidence intervals, which will have coverage 1-alpha.
region_cvar	A vector with length N of cluster variables, for method "cluster_region". If the vector 1:N is used, clustering is effectively equivalent to ehw
sector_cvar	A vector with length S of cluster variables, if sectors are to be clustered, for methods "akm" and "akm0". If the vector 1:S is used, this is equivalent to not clustering.

Note

subset is evaluated in the same way as variables in formula, that is first in data and then in the environment of formula.

References

Bartik, Timothy J., Who Benefits from State and Local Economic Development Policies?, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1991.

Examples

```
## Use ADH data from Autor, Dorn, and Hanson (2013)
ivBartik(d_sh_empl ~ 1 | shock, X=IV, data=ADH$reg, W=ADH$W,
         method=c("ehw", "akm", "akm0"))
```

---

ivBartik.fit *Inference in an IV regression with a shift-share instrument*

Description

Basic computing engine to calculate confidence intervals and p-values in an instrumental variables regression with a shift-share instrument, using different inference methods, as specified by method.

Usage

```
ivBartik.fit(y1, y2, X, W, Z, w = NULL, method = c("akm", "akm0"),
  beta0 = 0, alpha = 0.05, region_cvar = NULL, sector_cvar = NULL)
```

Arguments

y1	Outcome variable, vector of length N
y2	Endogenous variable, vector of length N
X	Shift-share vector with length N of sectoral shocks, aggregated to regional level using the share matrix W.
W	A matrix of sector shares, so that W[i, s] corresponds to share of sector s in region i
Z	Matrix of regional controls, matrix with N rows
w	vector of weights (length N) to be used in the fitting process. If not NULL, weighted least squares is used with weights w, i.e., sum(w * residuals^2) is minimized.
method	Vector specifying which inference methods to use. The vector elements have to be one or more of the following strings: "homosk" Assume i.i.d. homoskedastic errors "ehw" Eicker-Huber-White standard errors "region_cluster" Standard errors clustered at regional level "akm" Adão-Kolesár-Morales "akm0" Adão-Kolesár-Morales with null imposed. Note the reported standard error for this method corresponds to the normalized standard error, given by the length of the confidence interval divided by $2z_{1-\alpha/2}$ "all" All of the methods above
beta0	null that is tested (only affects reported p-values)
alpha	Determines confidence level of reported confidence intervals, which will have coverage 1-alpha.
region_cvar	A vector with length N of cluster variables, for method "cluster_region". If the vector 1:N is used, clustering is effectively equivalent to ehw
sector_cvar	A vector with length S of cluster variables, if sectors are to be clustered, for methods "akm" and "akm0". If the vector 1:S is used, this is equivalent to not clustering.

---

lmBartik *Inference in linear regression with a shift-share regressor*

Description

Computes confidence intervals and p-values in a linear regression in which the regressor of interest has a shift-share structure, as the instrument in Bartik (1991). Several different inference methods can computed, as specified by method.

Usage

```
lmBartik(formula, X, data, W, subset, weights, method, beta0 = 0,
  alpha = 0.05, region_cvar = NULL, sector_cvar = NULL)
```

Arguments

formula	object of class "formula" (or one that can be coerced to that class) of the form outcome ~ controls. For a regression with no controls (only an intercept), it takes the form outcome ~ 1
X	Shift-share vector with length N of sectoral shocks, aggregated to regional level using the share matrix W.
data	optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the outcome and running variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which the function is called.
W	A matrix of sector shares, so that W[i, s] corresponds to share of sector s in region i
subset	optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of weights to be used in the fitting process. Should be NULL or a numeric vector. If non-NULL, weighted least squares is used with weights weights (that is, minimizing sum(weights*residuals^2)); otherwise ordinary least squares is used.
method	Vector specifying which inference methods to use. The vector elements have to be one or more of the following strings: "homosk" Assume i.i.d. homoskedastic errors "ehw" Eicker-Huber-White standard errors "region_cluster" Standard errors clustered at regional level "akm" Adão-Kolesár-Morales "akm0" Adão-Kolesár-Morales with null imposed. Note the reported standard error for this method corresponds to the normalized standard error, given by the length of the confidence interval divided by $2z_{1-\alpha/2}$ "all" All of the methods above
beta0	null that is tested (only affects reported p-values)
alpha	Determines confidence level of reported confidence intervals, which will have coverage 1-alpha.
region_cvar	A vector with length N of cluster variables, for method "cluster_region". If the vector 1:N is used, clustering is effectively equivalent to ehw
sector_cvar	A vector with length S of cluster variables, if sectors are to be clustered, for methods "akm" and "akm0". If the vector 1:S is used, this is equivalent to not clustering.

Note

subset is evaluated in the same way as variables in formula, that is first in data and then in the environment of formula.

References

Bartik, Timothy J., Who Benefits from State and Local Economic Development Policies?, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1991.

Examples

```
## Use ADH data from Autor, Dorn, and Hanson (2013)
lmBartik(d_sh_empl ~ 1, X=IV, data=ADH$reg, W=ADH$W,
         method=c("ehw", "akm", "akm0"))
```

`lmBartik.fit` *Inference in a shift-share regression*

Description

Basic computing engine to calculate confidence intervals and p-values in shift-share designs using different inference methods, as specified by `method`.

Usage

```
lmBartik.fit(y, X, W, Z, w = NULL, method = c("akm", "akm0"),
  beta0 = 0, alpha = 0.05, region_cvar = NULL, sector_cvar = NULL)
```

Arguments

`y`	Outcome variable, vector of length `N`
`X`	Shift-share vector with length `N` of sectoral shocks, aggregated to regional level using the share matrix `W`.
`W`	A matrix of sector shares, so that `W[i, s]` corresponds to share of sector `s` in region `i`
`Z`	Matrix of regional controls, matrix with `N` rows corresponding to regions.
`w`	vector of weights (length `N`) to be used in the fitting process. If not `NULL`, weighted least squares is used with weights `w`, i.e., `sum(w * residuals^2)` is minimized.
`method`	Vector specifying which inference methods to use. The vector elements have to be one or more of the following strings: `"homosk"` Assume i.i.d. homoskedastic errors `"ehw"` Eicker-Huber-White standard errors `"region_cluster"` Standard errors clustered at regional level `"akm"` Adão-Kolesár-Morales `"akm0"` Adão-Kolesár-Morales with null imposed. Note the reported standard error for this method corresponds to the normalized standard error, given by the length of the confidence interval divided by $2z_{1-\alpha/2}$ `"all"` All of the methods above
`beta0`	null that is tested (only affects reported p-values)
`alpha`	Determines confidence level of reported confidence intervals, which will have coverage `1-alpha`.

region_cvar A vector with length N of cluster variables, for method "cluster_region". If the vector 1:N is used, clustering is effectively equivalent to ehw

sector_cvar A vector with length S of cluster variables, if sectors are to be clustered, for methods "akm" and "akm0". If the vector 1:S is used, this is equivalent to not clustering.

Index