Implementing Panel-Corrected Standard Errors in **R**: The **pcse** Package

Delia Bailey
YouGov Polimetrix

Jonathan N. Katz
California Institute of Technology

Abstract

This introduction to the R package **pcse** is a (slightly) modified version of Bailey and Katz (2011), published in the *Journal of Statistical Software*.

Time-series–cross-section (TSCS) data are characterized by having repeated observations over time on some set of units, such as states or nations. TSCS data typically display both contemporaneous correlation across units and unit level heteroskedasity making inference from standard errors produced by ordinary least squares incorrect. Panel-corrected standard errors (PCSE) account for these these deviations from spherical errors and allow for better inference from linear models estimated from TSCS data. In this paper, we discuss an implementation of them in the R system for statistical computing. The key computational issue is how to handle unbalanced data.

Keywords: **pcse**, time-series–cross-section, covariance matrix estimation, contemporaneous correlation, heteroskedasity, R.

1. Introduction

Time-series–cross-section (TSCS) data are characterized by having repeated observations over time on some set of units, such as states or nations. TSCS data have become common in applied studies in the social sciences, particularly in comparative political science applications. These data often show non-spherical errors because of contemporaneous correlation across the units and unit level heteroskedasity. When fitting linear models to TSCS data, it is common to use this non-spherical error structure to improve inference and estimation efficiency by a feasible generalized least squares (FGLS) estimator suggested by Parks (1967) and made popular by Kmenta (1986).

However, Beck and Katz (1995) showed that the Parks (1967) model had poor finite sample properties. In particular, in a simulation study they showed that the estimated standard errors for this model generated confidence intervals that were significantly too small, often underestimating variability by 50% or more, and with only minimal gains in efficiency over a simple linear model that ignored the non-spherical errors. Therefore, Beck and Katz (1995) suggested estimating linear models of TSCS data by ordinary least squares (OLS)[1] and they proposed a sandwich type estimator of the covariance matrix of the estimated parameters, which they called panel-corrected standard errors (PCSE), that is robust to the possibility of

---

[1]OLS is implemented in R's function `lm()`.

non-spherical errors.[2]

Although the PCSE covariance estimator bears some resemblance to heteroskedasity consistent (HC) estimators (see, for example, Huber 1967; White 1980; MacKinnon and White 1985), these other estimators do not explicitly incorporate the known TSCS structure of the data.[3] This leads to important differences in implementation.

This paper describes an implementation of PCSEs in the R system for statistical computing (R Development Core Team 2011). All of the functions described here are available in the package **pcse** that is available from Comprehensive R Archive Network (CRAN) at `http://CRAN. R-project.org/package=pcse`. The key computational issue is how to handle unbalanced data. TSCS data is unbalanced when the number of observations for units vary.

R packages that estimate various models for panel data include **plm** (Croissant and Millo 2008) and **systemfit** (Henningsen and Hamann 2007), that also implement different types of robust standard errors. Some of these are only robust to unit heteroskedasity and possible serial correlation. The **pcse** standard error estimate is robust not only to unit heteroskedacity, but it also robust against possible contemporaneous correlation across the units that is common in TSCS data.[4] Package **plm** also provides an implementation of Beck and Katz (1995) PCSE in the function `vcovBK()`[5] that can be applied to panel models estimated by `plm()`.

The next section fixes notation by briefly reviewing the linear TSCE model and the derivation of PCSEs. Section 3 considers the computational issues with unbalanced panels. Section 4 illustrates the use of the package **pcse**. Finally, Section 5 concludes.

2. TSCS data and estimation

The critical assumption of TSCS models is that of "pooling," that is, all units are characterized by the same regression equation at all points in time. Given this assumption we can write the generic TSCS model as:

$$y_{i,t} = \mathbf{x}_{i,t}\beta + \epsilon_{i,t}; \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{1}$$

where $\mathbf{x}_{i,t}$ is a vector of one or more (k) exogenous variables and observations are indexed by both unit (i) and time (t).

TSCS analysts typically put some structure on the assumed error process. In particular, they usually assume that for any given unit, the error variance is constant, so that the only source of heteroskedasticity is differing error variances across units. Analysts also assume that all spatial correlation is both contemporary and does not vary with time. The temporal dependence exhibited by the errors is also assumed to be time invariant, and may also be invariant across units. We, however, will be ignoring temporal dependence for the remainder of this paper by assuming that the analyst has controlled for it either by including the lagged dependent variable, $y_{i,t-1}$, in the set of regressors, $\mathbf{x}_{i,t}$, or using some sort of differencing.

---

[2]The estimator is actually rather poorly named as it really used for TSCS data, in which the time dimension is large enough for serious averaging within units, as opposed to panel data, which typically have short time dimensions. However, this is the nomenclature used in the literature.

[3] These heteroskedastic constituent covariance estimators are available in the R in the **sandwich** package (Zeileis 2004)

[4]For a discussion of the differences between TSCS and panel data see Beck and Katz (2011).

[5]The function `vcovBK()` was not yet part of **plm** when the first version of **pcse** was developed.

Since these assumptions are all based on the panel nature of the data, we call them the "panel error assumptions."

As is well known, the correct formula for the sampling variability of the OLS estimates from Equation 1 is given by the square roots of the diagonal terms of

$$\text{Cov}(\hat{\beta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \{ \mathbf{X}^\top \mathbf{\Omega} \mathbf{X} \} (\mathbf{X}^\top \mathbf{X})^{-1}. \tag{2}$$

If the errors obey the spherical error assumption — i.e., $\mathbf{\Omega} = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is an $NT \times NT$ identity matrix — this simplifies to the usual OLS formula, where the OLS standard errors are the square roots of the diagonal terms of

$$\widehat{\sigma^2} (\mathbf{X}^\top \mathbf{X})^{-1}$$

where $\widehat{\sigma^2}$ is the usual OLS estimator of the common error variance, σ^2. If the errors obey the panel structure, then this provides incorrect standard errors. Equation 2, however, can still be used, in combination with that panel structure of the errors to provide accurate PCSEs. For panel models with contemporaneously correlated and panel heteroskedastic errors, $\mathbf{\Omega}$ is an $NT \times NT$ block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances, $\mathbf{\Sigma}$, along the diagonal. To estimate Equation 2 we need an estimate of $\mathbf{\Sigma}$. Since the OLS estimates of Equation 1 are consistent, we can use the OLS residuals from that estimation to provide a consistent estimate of $\mathbf{\Sigma}$. Let $e_{i,t}$ be the OLS residual for unit i at time t. We can estimate a typical element of $\mathbf{\Sigma}$ by

$$\hat{\mathbf{\Sigma}}_{i,j} = \frac{\sum_{t=1}^{T_{i,j}} e_{i,t} e_{j,t}}{T_{i,j}}, \tag{3}$$

with the estimate $\hat{\mathbf{\Sigma}}$ being comprised of all these elements. We then use this to form the estimator $\hat{\mathbf{\Omega}}$ by creating a block diagonal matrix with the $\hat{\mathbf{\Sigma}}$ matrices along the diagonal. With balanced data where $T_{i,j} = T, \ \forall i = 1, \ldots, N$, we can simplify this to

$$\hat{\mathbf{\Sigma}} = \frac{(\mathbf{E}^\top \mathbf{E})}{T} \tag{4}$$

where $\mathbf{E}$ is the $T \times N$ matrix of residuals and hence estimate $\mathbf{\Omega}$ by

$$\hat{\mathbf{\Omega}} = \hat{\mathbf{\Sigma}} \otimes I_T \tag{5}$$

where $\otimes$ is the Kronecker product. PCSEs are thus computed by taking the square root of the diagonal elements of

$$\text{PCSE} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{\Omega}} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}. \tag{6}$$

3. Computational issues

3.1. Balanced data

The computational issues are fairly straightforward for balanced data. We need only the vector of residuals from a linear fit, the model matrix ($\mathbf{X}$), and indicators for group and time.

Given the indicators for group and time, we can appropriately reshape the vector of residuals into an $N \times T$ matrix. We can then directly calculate $\hat{\Sigma}$ from Equation 4 and, therefore, the PCSE for the fit. Within R, the function `lm` returns a `lm` object that contains, among other items, the residuals and the model matrix. It does not, however, include indicators for unit and time and these must be supplied by the user. The package **pcse** implements the estimation described above in the function `pcse`, which takes the following arguments:

```
pcse(lmobj, groupN, groupT, ...)
```

The first argument `lmobj` is a fitted linear model object as returned by `lm`. The argument `groupN` is a vector indication which cross-sectional unit an observation is from and `groupT` indicates which time period.

3.2. Unbalanced data

The only interesting computational issue is how to handle unbalanced data sets. With an unbalanced dataset, Equation 4 is no longer valid. There have been two alternatives estimation procedures suggested for unbalanced data. The first is to estimate Σ using a balanced subset of the data. The second alternative is to calculate the elements of Σ pairwise. We will consider each in turn.

The advantage of using the balanced subset approach is its computation ease. The largest balanced subset of the data can be found using the following simple R code:

```
units <- unique(groupN)
N     <- length(units)
time  <- unique(groupT)
T     <- length(time)
brows <- c()
for (i in 1:T) {
  br <- which(groupT == time[i])
  check <- length(br) == N
  if (check) {
    brows <- c(brows, br)
  }
}
```

It first computes the `unit` and `time` identifiers and their respective number N and T. The index `brows` gives all of the balanced rows. We can restrict the calculations of Σ to this balanced subset of data. This allows us to once again use Equations 4. The downside to this approach is that we are not using all of the available data to estimate Σ.

Recall that Σ is the contemporaneous correlation between every pair of units in our sample. The alternative approach then is to use Equation 3 for each pair $i, j \in N$ to construct our estimate $\hat{\Sigma}$. That is, for each pair of units we determine with temporal observations overlap between the two. We use this pairwise balanced sample to estimate $\hat{\Sigma}_{i,j}$.

We could do this directly by looping over all possible pairs and using Equation 3. However, for large N this can be a large set to loop over. We can improve on this by instead filling in the residual vector with zeros for the missing observations needed to balance out the data.

Clearly, these filled in observation do not alter the sum of the product of the residuals, since they contribute zero if either i or j have been filled in. As long as we divide by the appropriate $T_{i,j} = \min(T_i, T_j)$, we will appropriately calculate the correlation between i and j. This is approach we use in `pcse()` when the option `pairwise = TRUE` is used.

4. Example

In this section we demonstrate the use of the package **pcse**. The data we will use is from Alvarez, Garrett, and Lange (1991), hereafter AGL, and were reanalyzed using a simple linear model in Beck, Katz, Alvarez, Garrett, and Lange (1993). The data set is available in the package as the data frame `agl`. The data cover basic macro-economic and political variables from 16 OECD nations from 1970 to 1984. AGL estimated a model relating political and labor organization variables (and some economic controls) to economic growth, unemployment, and inflation. The argument was that economic performance in advanced industrialized societies was superior when labor was both encompassing and had political power or when labor was weak both in politics and the market. Here we will only look at their model of economic growth.

First both the package and the data need to be loaded into R with

```
R> library("pcse")
R> data("agl")
```

We can then fit their basic model of economic growth with

```
R> agl.lm <- lm(growth ~ lagg1 + opengdp + openex + openimp + central +
+    leftc + inter + as.factor(year), data = agl)
```

The model assumes that growth depends on lagged growth (`lagg1`), vulnerability to OECD demand (`opengdp`), OECD export (`openex`), OECD import (`openimp`), labor organization index (`central`), year fixed effects (`as.factor(year)`), the fraction of cabinet portfolios held by "left" parties (`leftc`), and interaction of `central` and `leftc` (`inter`). The interest focuses on the last three variables, particularly the interaction.

Here are the fit and standard errors without correcting for the panel structure of the data (note to save space the estimates for the year effects have been excluded from the printout.):

```
R> summary(agl.lm)


Call:
lm(formula = growth ~ lagg1 + opengdp + openex + openimp + central +
    leftc + inter + as.factor(year), data = agl)

Residuals:
   Min     1Q Median     3Q    Max
-5.837 -1.197  0.067  1.170  4.529

Coefficients:
```

```
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          5.968890   0.916402    6.51  5.0e-10 ***
lagg1                0.050315   0.139204    0.36  0.71812
opengdp             -0.002330   0.001867   -1.25  0.21342
openex               0.002008   0.001208    1.66  0.09786 .
openimp             -0.000609   0.001679   -0.36  0.71720
central             -0.763563   0.216258   -3.53  0.00051 ***
leftc               -0.024712   0.009276   -2.66  0.00829 **
inter                0.012868   0.003614    3.56  0.00045 ***
as.factor(year)1971 -1.205114   0.658282   -1.83  0.06851 .
as.factor(year)1972  0.530824   0.699475    0.76  0.44874
as.factor(year)1973  0.572342   0.702811    0.81  0.41633
as.factor(year)1974 -5.085447   1.100673   -4.62  6.6e-06 ***
as.factor(year)1975 -5.280737   0.953471   -5.54  8.7e-08 ***
as.factor(year)1976 -0.210344   0.885918   -0.24  0.81255
as.factor(year)1977 -2.267336   0.676865   -3.35  0.00095 ***
as.factor(year)1978 -1.258657   0.742950   -1.69  0.09167 .
as.factor(year)1979 -1.632229   0.702948   -2.32  0.02116 *
as.factor(year)1980 -3.902415   0.775032   -5.04  1.0e-06 ***
as.factor(year)1981 -4.652619   0.824901   -5.64  5.2e-08 ***
as.factor(year)1982 -4.532462   0.926328   -4.89  1.9e-06 ***
as.factor(year)1983 -2.775442   0.744784   -3.73  0.00025 ***
as.factor(year)1984 -0.770129   0.807864   -0.95  0.34150
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.82 on 218 degrees of freedom
Multiple R-squared:  0.502,        Adjusted R-squared:  0.454
F-statistic: 10.5 on 21 and 218 DF,  p-value: <2e-16
```

We can correct the standard errors by using:

```
R> agl.pcse <- pcse(agl.lm, groupN = agl$country, groupT = agl$year)
```

Included in the package is a summary function, `summary.pcse`, that can redisplay the estimates with the PCSE used for inference:

```
R> summary(agl.pcse)
```

```
 Results:

                     Estimate     PCSE t value Pr(>|t|)
(Intercept)          5.968890  0.89298   6.684 1.91e-10
lagg1                0.050315  0.15188   0.331 7.41e-01
opengdp             -0.002330  0.00179  -1.301 1.94e-01
openex               0.002008  0.00114   1.753 8.09e-02
openimp             -0.000609  0.00166  -0.368 7.13e-01
```

```
central              -0.763563 0.26569   -2.874 4.46e-03
leftc                -0.024712 0.00668   -3.698 2.75e-04
inter                 0.012868 0.00295    4.367 1.95e-05
as.factor(year)1971  -1.205114 0.14328   -8.411 5.34e-15
as.factor(year)1972   0.530824 0.27687    1.917 5.65e-02
as.factor(year)1973   0.572342 0.28987    1.974 4.96e-02
as.factor(year)1974  -5.085447 0.83220   -6.111 4.51e-09
as.factor(year)1975  -5.280737 0.67528   -7.820 2.24e-13
as.factor(year)1976  -0.210344 0.67377   -0.312 7.55e-01
as.factor(year)1977  -2.267336 0.22281  -10.176 3.76e-20
as.factor(year)1978  -1.258657 0.36798   -3.420 7.46e-04
as.factor(year)1979  -1.632229 0.31179   -5.235 3.87e-07
as.factor(year)1980  -3.902415 0.42860   -9.105 5.63e-17
as.factor(year)1981  -4.652619 0.52777   -8.816 3.83e-16
as.factor(year)1982  -4.532462 0.64582   -7.018 2.81e-11
as.factor(year)1983  -2.775442 0.39866   -6.962 3.89e-11
as.factor(year)1984  -0.770129 0.53750   -1.433 1.53e-01


  ---------------------------------------------


# Valid Obs = 240; # Missing Obs = 0; Degrees of Freedom = 218.
```

We note that the standard error on `central` has increased a bit, but the standard errors of the other two variables of interest, `leftc` and `inter` have actually decreased.

We have also included an unbalanced version of the AGL data set, `aglUn`, that was created by randomly deleting some observations. This was only done to demonstrate how estimates vary by casewise and pairwise estimation of the covariance matrix and is not a recommended modeling strategy. As before, we can estimate the same model by:

```
R> data("aglUn")
R> aglUn.lm <- lm(growth ~ lagg1 + opengdp + openex + openimp + central +
+    leftc + inter + as.factor(year), data = aglUn)
R> aglUn.pcse1 <- pcse(aglUn.lm, groupN = aglUn$country,
+    groupT = aglUn$year, pairwise = TRUE)
R> summary(aglUn.pcse1)

 Results:


               Estimate     PCSE  t value Pr(>|t|)
(Intercept)    6.198165 0.87255   7.1035 1.90e-11
lagg1         -0.007295 0.15069  -0.0484 9.61e-01
opengdp       -0.001962 0.00181  -1.0811 2.81e-01
openex         0.002215 0.00115   1.9319 5.47e-02
openimp       -0.000914 0.00166  -0.5516 5.82e-01
central       -0.842650 0.24450  -3.4464 6.87e-04
leftc         -0.028387 0.00702  -4.0451 7.37e-05
inter          0.014521 0.00307   4.7286 4.17e-06
```

```
as.factor(year)1971 -0.817961 0.16592   -4.9299 1.68e-06
as.factor(year)1972  0.577742 0.26045    2.2182 2.76e-02
as.factor(year)1973  0.628091 0.26656    2.3562 1.94e-02
as.factor(year)1974 -4.988305 0.81934   -6.0882 5.43e-09
as.factor(year)1975 -5.072186 0.66781   -7.5952 1.03e-12
as.factor(year)1976 -0.523931 0.66352   -0.7896 4.31e-01
as.factor(year)1977 -2.115323 0.18978  -11.1464 6.31e-23
as.factor(year)1978 -1.173037 0.34402   -3.4098 7.81e-04
as.factor(year)1979 -1.867747 0.31185   -5.9893 9.15e-09
as.factor(year)1980 -4.083112 0.43107   -9.4721 6.32e-18
as.factor(year)1981 -4.448396 0.50744   -8.7663 6.61e-16
as.factor(year)1982 -4.456624 0.62940   -7.0807 2.17e-11
as.factor(year)1983 -2.831709 0.42371   -6.6831 2.11e-10
as.factor(year)1984 -0.984174 0.57233   -1.7196 8.70e-02


    ---------------------------------------------


# Valid Obs = 230; # Missing Obs = 10; Degrees of Freedom = 208.
```

Here we see the estimates of the pairwise version of the PCSE, since the option `pairwise = TRUE` was given. The results are close to the original results for the balanced data.

If we preferred the casewise estimate that uses the largest balanced subset to estimate the contemporaneous correlation matrix, we do that by:

```
R> aglUn.pcse2 <- pcse(aglUn.lm, groupN = aglUn$country,
+    groupT = aglUn$year, pairwise = FALSE)


Warning message:
In pcse(aglUn.lm, groupN = aglUn$country, groupT = aglUn$year, ...
  Caution! The number of CS observations per panel, 7, used to compute
  the vcov matrix is less than half theaverage number of obs per panel
  in the original data.You should consider using pairwise selection.


R> summary(aglUn.pcse2)


 Results:


               Estimate       PCSE  t value Pr(>|t|)
(Intercept)    6.198165  0.721172   8.5946 2.00e-15
lagg1         -0.007295  0.123454  -0.0591 9.53e-01
opengdp       -0.001962  0.001243  -1.5784 1.16e-01
openex         0.002215  0.000782   2.8327 5.07e-03
openimp       -0.000914  0.001191  -0.7674 4.44e-01
central       -0.842650  0.264484  -3.1860 1.66e-03
leftc         -0.028387  0.006387  -4.4444 1.43e-05
inter          0.014521  0.002829   5.1324 6.55e-07
```

```
as.factor(year)1971 -0.817961 0.203861  -4.0124 8.39e-05
as.factor(year)1972  0.577742 0.230489   2.5066 1.30e-02
as.factor(year)1973  0.628091 0.237070   2.6494 8.68e-03
as.factor(year)1974 -4.988305 0.583396  -8.5505 2.66e-15
as.factor(year)1975 -5.072186 0.482912 -10.5033 5.62e-21
as.factor(year)1976 -0.523931 0.546879  -0.9580 3.39e-01
as.factor(year)1977 -2.115323 0.194109 -10.8976 3.61e-22
as.factor(year)1978 -1.173037 0.298877  -3.9248 1.18e-04
as.factor(year)1979 -1.867747 0.283784  -6.5816 3.72e-10
as.factor(year)1980 -4.083112 0.361778 -11.2862 2.36e-23
as.factor(year)1981 -4.448396 0.411130 -10.8199 6.22e-22
as.factor(year)1982 -4.456624 0.504551  -8.8329 4.29e-16
as.factor(year)1983 -2.831709 0.386623  -7.3242 5.19e-12
as.factor(year)1984 -0.984174 0.456290  -2.1569 3.22e-02


---------------------------------------------

# Valid Obs = 230; # Missing Obs = 10; Degrees of Freedom = 208.
```

Here we see that the software has issued a warning about the calculation of the standard errors. Although there only 10 missing observations, they are intermingled through out the data. This means that the largest balanced panel only has seven time points, whereas the data runs for 14. In this case, it is not clear that PCSEs will be correctly estimated although in this case they are not that different from the casewise estimate.

5. Summary

This paper briefly reviews estimation of panel-corrected standard errors for time-series–cross-section (TSCS) data. It discusses an implementation of estimating them in the R system for statistical computing in the **pcse** package.

Computational details

The results in this paper were obtained using R 3.4.3 with the package **pcse** 1.9.1.1. R and the **pcse** package are available from CRAN at http://CRAN.R-Project.org/.

References

Alvarez RM, Garrett G, Lange P (1991). "Government Partisanship, Labor Organization, and Macroeconomic Performance." *American Political Science Review*, **85**, 539–556.

Bailey D, Katz JN (2011). "Implementing Panel-Corrected Standard Errors in R: The **pcse** Package." *Journal of Statistical Software, Code Snippets*, **42**(1), 1–11. URL http://www.jstatsoft.org/v42/c01/.

Beck N, Katz JN (1995). "What To Do (and Not To Do) with Times-Series–Cross-Section Data in Comparative Politics." *American Political Science Review*, **89**(3), 634–647.

Beck N, Katz JN (2011). "Modeling Dynamics in Time-Series–Cross-Section Political Economy Data." *Annual Review of Political Science*. doi:10.1146/annurev-polisci-071510-103222. Forthcoming.

Beck N, Katz JN, Alvarez RM, Garrett G, Lange P (1993). "Government Partisanship, Labor Organization, and Macroeconomic Performance: A Corrigendum." *American Political Science Review*, **87**, 945–948.

Croissant Y, Millo G (2008). "Panel Data Econometrics in R: The **plm** Package." *Journal of Statistical Software*, **27**(2), 1–43. URL http://www.jstatsoft.org/v27/i02/.

Henningsen A, Hamann JD (2007). "**systemfit**: A Package for Estimating Systems of Simultaneous Equations in R." *Journal of Statistical Software*, **23**(4), 1–40. URL http://www.jstatsoft.org/v23/i04/.

Huber PJ (1967). "The Behavior of Maximum Likelihood Estimation Under Nonstandard Conditions." In LM LeCam, J Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, CA.

Kmenta J (1986). *Elements of Econometrics*. 2nd edition. Macmillan, New York, NY.

MacKinnon JG, White H (1985). "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics*, **29**, 305–325.

Parks R (1967). "Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated." *Journal of the American Statistical Association*, **62**(318), 500–509.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

White H (1980). "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity." *Econometrica*, **48**, 817–838.

Zeileis A (2004). "Econometric Computing with HC and HAC Covariance Matrix Estimators." *Journal of Statistical Software*, **11**(10), 1–17. URL http://www.jstatsoft.org/v11/i10/.

Affiliation:

Jonathan N. Katz
California Institute of Technology
DHSS 228-77
Pasadena, CA 91125, United States of America
E-mail: jkatz@caltech.edu
URL: http://jkatz.caltech.edu/