Time-Series Regression and Generalized Least Squares in R

An Appendix to *An R Companion to Applied Regression, Second Edition*

John Fox & Sanford Weisberg

last revision: 11 November 2010

Abstract

Generalized least-squares (*GLS*) regression extends ordinary least-squares (OLS) estimation of the normal linear model by providing for possibly unequal error variances and for correlations between different errors. A common application of GLS estimation is to time-series regression, in which it is generally implausible to assume that errors are independent. This appendix to Fox and Weisberg (2011) briefly reviews GLS estimation and demonstrates its application to time-series data using the `gls` function in the **nlme** package, which is part of the standard R distribution.

1 Generalized Least Squares

In the standard linear model (for example, in Chapter 4 of the text),

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is the $n \times 1$ response vector; $\mathbf{X}$ is an $n \times k+1$ model matrix; $\boldsymbol{\beta}$ is a $k+1 \times 1$ vector of regression coefficients to estimate; and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors. Assuming that $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$ leads to the familiar *ordinary-least-squares* (*OLS*) estimator of $\boldsymbol{\beta}$,

$$\mathbf{b}_{\mathrm{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

with covariance matrix

$$\mathrm{Var}(\mathbf{b}_{\mathrm{OLS}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Let us, however, assume more generally that $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, where the error covariance matrix $\boldsymbol{\Sigma}$ is symmetric and positive-definite. Different diagonal entries in $\boldsymbol{\Sigma}$ correspond to non-constant error variances, while nonzero off-diagonal entries correspond to correlated errors.

Suppose, for the time-being, that $\boldsymbol{\Sigma}$ is known. Then, the log-likelihood for the model is

$$\log_e L(\boldsymbol{\beta}) = -\frac{n}{2}\log_e 2\pi - \tfrac{1}{2}\log_e(\det\boldsymbol{\Sigma}) - \tfrac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

which is maximimized by the *generalized-least-squares* (*GLS*) estimator of $\boldsymbol{\beta}$,

$$\mathbf{b}_{\mathrm{GLS}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

with covariance matrix

$$\mathrm{Var}(\mathbf{b}_{\mathrm{GLS}}) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$$

For example, when $\mathbf{\Sigma}$ is a diagonal matrix of (generally) unequal error variances, then $\mathbf{b}_{\text{GLS}}$ is just the *weighted-least-squares* (*WLS*) estimator.

In a real application, of course, the error covariance matrix $\mathbf{\Sigma}$ is not known, and must be estimated from the data along with the regression coefficients $\boldsymbol{\beta}$. There are, however, vastly too many elements in $\mathbf{\Sigma}$ — $n(n+1)/2$ distinct elements — to estimate the model without further restrictions. With a suitably restrictive parametrization of $\mathbf{\Sigma}$, the model can be estimated by maximum likelihood or another appropriate method.

2 Serially Correlated Errors

One common context in which the errors from a regression model are unlikely to be independent is in *time-series* data, where the observations represent different moments or intervals of time, usually equally spaced. We will assume that the process generating the regression errors is *stationary*: That is, all of the errors have the same expectation (already assumed to be 0) and the same variance (σ^2), and the covariance of two errors depends only upon their separation s in time:[1]

$$C(\varepsilon_t, \varepsilon_{t+s}) = C(\varepsilon_t, \varepsilon_{t-s}) = \sigma^2 \rho_s$$

where ρ_s is the error *autocorrelation at lag s*.

In this situation, the error covariance matrix has the following structure:

$$\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} = \sigma^2 \mathbf{P}$$

If we knew the values of σ^2 and the ρs, then we could apply this result to find the GLS estimator of $\boldsymbol{\beta}$ in a time-series regression, but, of course, these are generally unknown parameters. Moreover, while they are many fewer than the number of elements in the unrestricted error covariance matrix $\mathbf{\Sigma}$, the large number $(n-1)$ of different ρs makes their estimation impossible without specifying additional structure for the autocorrelated errors.

There are several standard models for stationary time-series; the most common for autocorrelated regression errors is the *first-order auto-regressive process*, AR(1):

$$\varepsilon_t = \phi \varepsilon_{t-1} + \nu_t$$

where the *random shocks* ν_t are assumed to be *Gaussian white noise*, $\text{NID}(0, \sigma_\nu^2)$. Under this model, $\rho_1 = \phi$, $\rho_s = \phi^s$, and $\sigma^2 = \sigma_\nu^2/(1-\phi^2)$. Because, as a correlation, $|\phi| < 1$, the error autocorrelations ρ_s decay exponentially towards 0 as s increases.[2]

Higher-order autoregressive models are a direct generalization of the first-order model; for example, the second-order autoregressive model, denoted AR(2), is

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \nu_t$$

[1] Adjacent observations are taken by convention to be separated by 1 unit of time—e.g., 1 year in annual time-series data.

[2] For the AR(1) process to be stationary, $|\phi|$ cannot be equal to 1.

In contrast, in the first-order *moving-average process*, MA(1), the current error depends upon the random shock from the current and previous periods (rather than upon the previous *regression error*),

$$\varepsilon_t = \nu_t + \psi\nu_{t-1}$$

and higher-order MA(q) processes are similarly defined. Finally, AR and MA terms are combined in ARMA(p, q) processes; for example, ARMA(1, 1) errors follow the process

$$\varepsilon_t = \phi\varepsilon_{t-1} + \nu_t + \psi\nu_{t-1}$$

Examining the residual autocorrelations from a preliminary OLS regression can suggest a reasonable form for the error-generating process.[3] The lag-s residual autocorrelation is

$$r_s = \frac{\sum_{t=s+1}^{n} e_t e_{t-s}}{\sum_{t=1}^{n} e_t^2}$$

If the residuals were independently distributed (which they are not), the standard error of each r_s would be approximately $1/\sqrt{n}$, a quantity that can be used as a rough guide to the statistical significance of the residual autocorrelations. A more accurate approach is to calculate the *Dubin-Watson statistics*,

$$D_s = \frac{\sum_{t=s+1}^{n} (e_t - e_{t-s})^2}{\sum_{t=1}^{n} e_t^2}$$

which have a known, if complex, sampling distribution that depends upon the model matrix $\mathbf{X}$. When the sample size is large, $D_s \approx 2(1 - r_s)$, and so Durbin-Watson statistics near 2 are indicative of small residual autocorrelation, those below 2 of positive autocorrelation, and those above 2 of negative autocorrelation.

3 Using The `gls` Function in R

The `gls` function in the **nlme** package (Pinheiro et al., 2010), which is part of the standard R distribution, fits regression models with a variety of correlated-error and non-constant error-variance structures.[4] To illustrate the use of `gls`, let us examine time-series data on women's crime rates in Canada, analyzed by Fox and Hartnagel (1979). The data are in the data frame `Hartnagel` in the `car` package (Fox and Weisberg, 2011):

```
> library(car)
```

. . .

```
> Hartnagel
```

[3]In identifying an ARMA process, it helps to look as well at the *partial autocorrelations* of the residuals. For example, an AR(1) process has an exponentially decaying autocorrelation function, and a partial autocorrelation function with a single nonzero spike at lag 1. Conversely, an MA(1) process has an exponentially decaying *partial autocorrelation* function, and an *autocorrelation* function with a single nonzero spike at lag 1. Of course, these neat theoretical patterns are subject to sampling error.

[4]The **nlme** package also has functions for fitting linear and nonlinear mixed models, as described in the Appendix on mixed-effects models.

```
   year  tfr partic degrees fconvict ftheft mconvict mtheft
1  1931 3200    234    12.4     77.1     NA    778.7     NA
2  1932 3084    234    12.9     92.9     NA    745.7     NA
3  1933 2864    235    13.9     98.3     NA    768.3     NA
4  1934 2803    237    13.6     88.1     NA    733.6     NA
5  1935 2755    238    13.2     79.4   20.4    765.7  247.1
6  1936 2696    240    13.2     91.0   22.1    816.5  254.9
. . .
37 1967 2586    339    80.4    115.2   70.6    781.1  272.0
38 1968 2441    338    90.4    122.9   73.0    849.7  274.7
```

The variables in the data set are as follows:

- `year`, 1931–1968.
- `tfr`, the total fertility rate, births per 1000 women.
- `partic`, women's labor-force participation rate, per 1000.
- `degrees`, women's post-secondary degree rate, per 10,000.
- `fconvict`, women's indictable-offense conviction rate, per 100,000.
- `ftheft`, women's theft conviction rate, per 100,000.
- `mconvict`, men's indictable-offense conviction rate, per 100,000.
- `mtheft`, men's theft conviction rate, per 100,000.

We will estimate the regression of `fconvict` on `tfr`, `partic`, `degrees`, and `mconvict`. The rationale for including the last predictor is to control for omitted variables that affect the crime rate in general. Let us begin by examining the time series for the women's conviction rate (Figure 1):

```
> plot(fconvict ~ year, type="o", pch=16, data=Hartnagel,
+     ylab="Convictions per 100,000 Women")
```

Including the argument `type="o"` to `plot` overplots points and lines, as is traditional for a time-series graph, and `pch=16` specifies filled dots as the plotting characters.[5] We can see that the women's conviction rate fluctuated substantially but gradually during this historical period, with no apparent overall trend.

A preliminary OLS regression produces the following fit to the data:

```
> mod.ols <- lm(fconvict ~ tfr + partic + degrees + mconvict, data=Hartnagel)
> summary(mod.ols)

Call:
lm(formula = fconvict ~ tfr + partic + degrees + mconvict, data = Hartnagel)
```

[5]For more information on drawing graphs in R, see Chapter 7 of the text. There is a `ts.plot` function in the **stats** package in R for graphing time-series data. Though we will not bother to do so, it is also possible to define special time-series data objects in R. For more information, consult `?ts`.

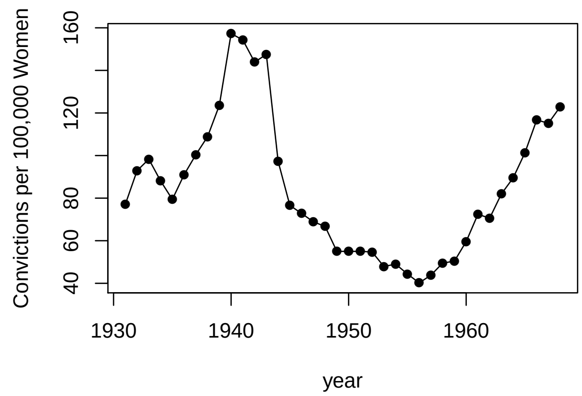

Figure 1: Time series of Canadian women's indictable-offense conviction rate, 1931–1968.

```
Residuals:
   Min     1Q Median     3Q    Max 
-42.96  -9.20  -3.57   6.15  48.38 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 127.64000   59.95704    2.13    0.041
tfr          -0.04657    0.00803   -5.80  1.8e-06
partic        0.25342    0.11513    2.20    0.035
degrees      -0.21205    0.21145   -1.00    0.323
mconvict      0.05910    0.04515    1.31    0.200

Residual standard error: 19.2 on 33 degrees of freedom
Multiple R-squared: 0.695,	Adjusted R-squared: 0.658 
F-statistic: 18.8 on 4 and 33 DF,  p-value: 3.91e-08
```

The women's crime rate, therefore, appears to decline with fertility and increase with labor-force participation; the other two predictors have nonsignificant coefficients. A graph of the residuals from the OLS regression (Figure 2), however, suggests that they may be substantially autocorrelated:[6]

```
> plot(Hartnagel$year, residuals(mod.ols), type="o", pch=16,
+     xlab="Year", ylab="OLS Residuals")
> abline(h=0, lty=2)
```

The `acf` function in the R **stats** package computes and plots the *autocorrelation* and *partial-autocorrelation functions* of a time series, here for the OLS residuals (Figure 3):

```
> acf(residuals(mod.ols))
```

```
> acf(residuals(mod.ols), type="partial")
```

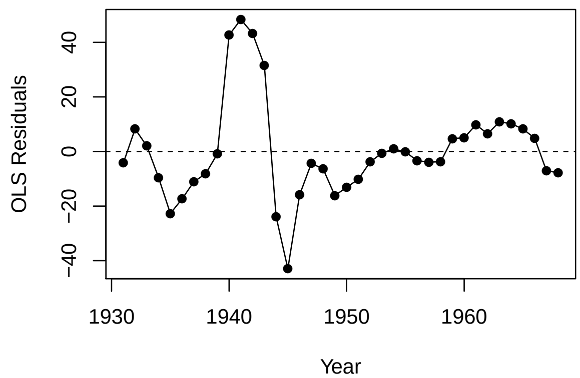

Figure 2: Residuals from the OLS regression of women's conviction rate on several predictors.

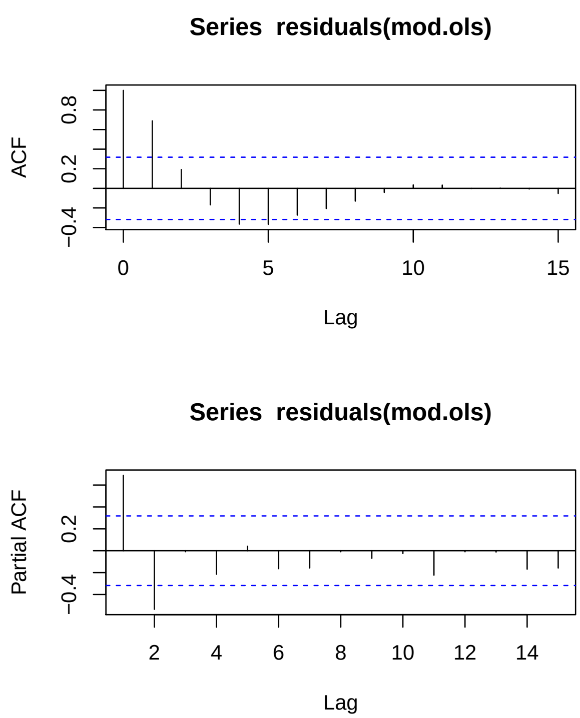

Figure 3: Autocorrelation and partial-autocorrelation functions for the residuals from the OLS regression of women's conviction rate on several predictors.

The broken horizontal lines on the plots correspond to 95-percent confidence limits. The general pattern of the autocorrelation and partial autocorrelation functions — sinusoidal decay in the former; two spikes, one positive, the other negative, in the latter — is suggestive of an AR(2) process with $\phi_1 > 0$ and $\phi_2 < 0$.

We follow up by computing Durbin-Watson statistics for the OLS regression, using the `durbinWatsonTest` function in the **car** package. By default, this function computes bootstrapped p-values for the Durbin-Watson statistics:[7]

```
> durbinWatsonTest(mod.ols, max.lag=5)

 lag Autocorrelation D-W Statistic p-value
   1       0.6883        0.6169   0.000
   2       0.1923        1.5994   0.164
   3      -0.1686        2.3187   0.328
   4      -0.3653        2.6991   0.000
   5      -0.3673        2.6521   0.004
 Alternative hypothesis: rho[lag] != 0
```

Three of the first five Durbin-Watson statistics are statistically significant, including the first. As an alternative, the `dwtest` function in the **lmtest** package (Zeileis and Hothorn, 2002) computes the p-value for the first-order Durbin-Watson statistic analytically:

```
> library(lmtest)
> dwtest(mod.ols, alternative="two.sided")

        Durbin-Watson test

data:  mod.ols
DW = 0.6169, p-value = 1.392e-08
alternative hypothesis: true autocorelation is not 0
```

Many of the arguments for the `gls` function are the same as for `lm` — in particular, `gls` takes `model`, `data`, `subset`, and `na.action` arguments.

- In `gls`, `na.action` defaults to `na.fail`: Missing data in a time-series in any event require special consideration.
- The `weights` argument to `gls` can be used to specify a model for the error variance,
- The `correlation` argument (as we will illustrate presently) can be used to specify a model for error autocorrelation.
- The `method` argument selects the method of estimation — `method="ML"` for maximum-likelihood estimation.[8]

[6]There also seems to be something unusual going on during World War II that is not accounted for by the predictors, a subject that we will not pursue here.

[7]See Section 4.3.7 of the text and the Appendix on bootstrapping.

[8]The default `method` for `glm` is `"REML"` for `RE`stricted `M`aximum-`L`ikelihood, which may be thought of as correcting for degrees of freedom. In the current illustration, REML estimation produces very different results from ML. (Try it!) To see the full range of arguments to `glm`, consult the on-line help.

For the Canadian women's crime data:

```
> library(nlme)
> mod.gls <- gls(fconvict ~ tfr + partic + degrees + mconvict,
+     data=Hartnagel, correlation=corARMA(p=2), method="ML")
> summary(mod.gls)

Generalized least squares fit by maximum likelihood
  Model: fconvict ~ tfr + partic + degrees + mconvict
  Data: Hartnagel
    AIC   BIC logLik
  305.4 318.5 -144.7

Correlation Structure: ARMA(2,0)
 Formula: ~1
 Parameter estimate(s):
   Phi1    Phi2
 1.0683 -0.5507

Coefficients:
            Value Std.Error t-value p-value
(Intercept) 83.34     59.47   1.401  0.1704
tfr         -0.04      0.01  -4.309  0.0001
partic       0.29      0.11   2.568  0.0150
degrees     -0.21      0.21  -1.016  0.3171
mconvict     0.08      0.04   2.162  0.0380

 Correlation:
         (Intr) tfr    partic degres
tfr      -0.773
partic   -0.570  0.176
degrees   0.093  0.033 -0.476
mconvict -0.689  0.365  0.047  0.082

Standardized residuals:
    Min      Q1     Med      Q3     Max
-2.4992 -0.3717 -0.1495  0.3372  2.9095

Residual standard error: 17.70
Degrees of freedom: 38 total; 33 residual
```

Specifying the correlation structure as `correlation=corARMA(p=2)` fits an AR(2) process for the errors; that is, the moving-average component is implicitly of order `q=0`, and hence is absent. In this instance, the ML estimates of the regression parameters under the AR(2) error-correlation model are not terribly different from the OLS estimates (although the coefficient for `mconvict` is now statistically significant). The ML estimates of the error-autoregressive parameters are sizable, $\widehat{\phi}_1 = 1.068$ and $\widehat{\phi}_2 = -0.551$.

We can employ likelihood-ratio tests to check whether the parameters of the AR(2) process for the errors are necessary, and whether a second-order autoregressive model is sufficient. We proceed by updating the original `gls` model, respecifying the time-series process for the errors; we then compare nested models using the generic `anova` function, which has a method for `gls` objects:

```
> mod.gls.3 <- update(mod.gls, correlation=corARMA(p=3))
> mod.gls.1 <- update(mod.gls, correlation=corARMA(p=1))
> mod.gls.0 <- update(mod.gls, correlation=NULL)
> anova(mod.gls, mod.gls.1)

          Model df   AIC   BIC logLik   Test L.Ratio p-value
mod.gls       1  8 305.4 318.5 -144.7
mod.gls.1     2  7 312.4 323.9 -149.2 1 vs 2   9.009  0.0027

> anova(mod.gls, mod.gls.0)

          Model df   AIC   BIC logLik   Test L.Ratio p-value
mod.gls       1  8 305.4 318.5 -144.7
mod.gls.0     2  6 339.0 348.8 -163.5 1 vs 2   37.59  <.0001

> anova(mod.gls.3, mod.gls)

          Model df   AIC   BIC logLik   Test L.Ratio p-value
mod.gls.3     1  9 307.4 322.1 -144.7
mod.gls       2  8 305.4 318.5 -144.7 1 vs 2 0.01847  0.8919
```

An AR(3) specification would be unusually complicated, but in any event the tests support the AR(2) specification.

4 Complementary Reading and References

Time-series regression and GLS estimation are covered in Fox (2008, chap. 16). GLS estimation is a standard topic in econometrics texts. There are substantial treatments in Judge et al. (1985) and in Greene (2003), for example. Likewise, ARMA models are a standard topic in the time-series literature; see, for example, Chatfield (1989).

References

Chatfield, C. (1989). *Analysis of Time Series: An Introduction*. Chapman and Hall, London, fourth edition.

Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models*. Sage, Thousand Oaks, CA, second edition.

Fox, J. and Hartnagel, T. F. (1979). Changing social roles and female crime in Canada: a time series analysis. *Canadian Review of Sociology and Anthropology*, 16:96–104.

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA, second edition.

Greene, W. H. (2003). *Econometric Analysis*. Prentice-Hall, Upper Saddle River, NJ, fifth edition.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*. Wiley, New York, second edition.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Development Core Team (2010). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-97.

Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3):7–10.