Data Analysis Exercises for Chapter 9: *Applied Regression Analysis, Generalized Linear Models, and Related Methods, Third Edition* (Sage, 2016)

John Fox

Last modified: 2015-03-12

Exercise D9.1 Using a calculator or a computer program that conveniently performs matrix computations, and working with regression that you performed in Exercise D5.5:[1]

(a) Compute the least-squares regression coefficients, $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

(b) Verify that the least-squares slope coefficients $\mathbf{b}_1 = [B_1, B_2, \ldots, B_k]'$ can be computed as $\mathbf{b}_1 = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$ where $\mathbf{X}^*$ and $\mathbf{y}^*$ contain mean deviations for the X's and Y, respectively.

Exercise D9.2 Using a calculator or a computer program that performs matrix computations, and working with the Canadian occupational prestige data (continuing Exercise D9.1):

(a) Calculate the estimated error variance, $S_E^2 = \mathbf{e}'\mathbf{e}/(n-k-1)$ (where $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$), and the estimated covariance matrix of the coefficients, $\widehat{V}(\mathbf{b}) = S_E^2(\mathbf{X}'\mathbf{X})^{-1}$.

(b) Verify that the estimated covariance matrix for the slope coefficients $\mathbf{b}_1 = [B_1, B_2, \ldots, B_k]'$ in this regression can be calculated as $\widehat{V}(\mathbf{b}_1) = S_E^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}$, where $\mathbf{X}^*$ is the mean-deviation matrix for the X's.

Exercise D9.3 A common application of 2SLS estimation and instrumental-variables estimation more generally is to l*inear simultaneous-equation models*, also called *structural-equation models*. The following model ("Model I") was proposed by the econometrician Klein (1950), and is described by Greene (1993, Example 20.5), from which I have taken Klein's data, given in Table 1 — annual data for the U.S. economy from 1921 to 1941, including the following variables:[2]

t	Year, 1921–1941
C_t	Consumption
P_t	Private profits
$W_t^{(p)}$	Private wages
I_t	Investment
K_{t-1}	Capital stock (lagged 1 year)
X_t	Equilibrium demand
$W_t^{(g)}$	Government wages
G_t	Government non-wage spending
T_t	Indirect business taxes and net exports

[1] Many computer programs (for example, APL, Gauss, Lisp-Stat, Mathematica, R, S-PLUS, SAS/IML, and Stata) include convenient facilities for matrix calculations.

[2] L. Klein, *Econometric Fluctuations in the United States, 1921-1941* (Wiley, 1950); W. H. Greene, *Econometric Analysis, Second Edition* (Macmillan, 1993).

With the exception of year (t), the variables are apparently all measured in billions of constant U.S. dollars (variously adjusted). Klein's Model I is defined by the following three regression (or structural) equations:

$$\begin{array}{ll} \text{consumption:} & C_t = \alpha_1 + \beta_{11} P_t + \beta_{12} P_{t-1} + \beta_{13} \left(W_t^{(p)} + W_t^{(g)} \right) + \varepsilon_{1t} \\ \text{investment:} & I_t = \alpha_2 + \beta_{21} P_t + \beta_{22} P_{t-1} + \beta_{23} K_{t-1} + \varepsilon_{2t} \\ \text{private wages:} & W_t^{(p)} = \alpha_3 + \beta_{31} X_t + \beta_{32} X_{t-1} + \beta_{33}(t - 1931) + \varepsilon_{3t} \end{array}$$

There are, in addition, three *identities* — equations that hold definitionally, with no regression parameters to estimate and no errors:

$$\begin{array}{ll} \text{equilibrium demand:} & X_t = C_t + I_t + G_t \\ \text{private profits:} & P_t = X_t - T_t - W_t^{(p)} \\ \text{capital stock:} & K_t = K_{t-1} + I_t \end{array}$$

The three variables on the left-hand side of the regression equations, C_t, I_t, and $W_t^{(p)}$, are *endogenous variables* — determined within the model; each is associated with an *error variable*, respectively ε_{1t}, ε_{2t}, and ε_{3t}, also called a *structural disturbance*. The variables G_t, T_t, $W_t^{(g)}$, $t - 1931$, and the constant regressor associated with the intercepts are *exogenous variables* — determined outside of the model and taken to be independent of the errors. The three *predetermined variables* P_{t-1}, K_{t-1}, and X_{t-1} are also assumed to be independent of the errors at year t.

(a) What instrumental variables are available to estimate the three structural equations of the model? Confirm that there at least as many instrumental variables as regression coefficients to estimate in each structural equation.

(b) How would you estimate each of the structural equations of the model? That is, would you use 2SLS or direct instrumental-variables estimation?

(c) Estimate the structural equations. Depending on the software you use, you will first have to perform some data-management tasks, such as lagging some of the variables in the data set one year. One consequence will be that the data for 1920 are unavailable for the regressions because the necessary data for the previous year, 1919, are not in the data set.

(d) Is it reasonable to suppose that the values of the error variable for each structural equation — e.g., ε_{1t} for the first equation — are independently sampled across years?[3]

[3] Cf., the discussion of time-series regression in Chapter 16.

Table 1: Klein's Data on the U.S. Economy, 1921-1941

Year t	C_t	P_t	$W_t^{(p)}$	I_t	K_{t-1}	$W_t^{(g)}$	X_t	G_t	T_t
1920	39.8	12.7	28.8	2.7	180.1	44.9	2.2	2.4	3.4
1921	41.9	12.4	25.5	−0.2	182.8	45.6	2.7	3.9	7.7
1922	45.0	16.9	29.3	1.9	182.6	50.1	2.9	3.2	3.9
1923	49.2	18.4	34.1	5.2	184.5	57.2	2.9	2.8	4.7
1924	50.6	19.4	33.9	3.0	189.7	57.1	3.1	3.5	3.8
1925	52.6	20.1	35.4	5.1	192.7	61.0	3.2	3.3	5.5
1926	55.1	19.6	37.4	5.6	197.8	64.0	3.3	3.3	7.0
1927	56.2	19.8	37.9	4.2	203.4	64.4	3.6	4.0	6.7
1928	57.3	21.1	39.2	3.0	207.6	64.5	3.7	4.2	4.2
1929	57.8	21.7	41.3	5.1	210.6	67.0	4.0	4.1	4.0
1930	55.0	15.6	37.9	1.0	215.7	61.2	4.2	5.2	7.7
1931	50.9	11.4	34.5	−3.4	216.7	53.4	4.8	5.9	7.5
1932	45.6	7.0	29.0	−6.2	213.3	44.3	5.3	4.9	8.3
1933	46.5	11.2	28.5	−5.1	207.1	45.1	5.6	3.7	5.4
1934	48.7	12.3	30.6	−3.0	202.0	49.7	6.0	4.0	6.8
1935	51.3	14.0	33.2	−1.3	199.0	54.4	6.1	4.4	7.2
1936	57.7	17.6	36.8	2.1	197.7	62.7	7.4	2.9	8.3
1937	58.7	17.3	41.0	2.0	199.8	65.0	6.7	4.3	6.7
1938	57.5	15.3	38.2	−1.9	201.8	60.9	7.7	5.3	7.4
1939	61.6	19.0	41.6	1.3	199.9	69.5	7.8	6.6	8.9
1940	65.0	21.1	45.0	3.3	201.2	75.7	8.0	7.4	9.6
1941	69.7	23.5	53.3	4.9	204.5	88.4	8.5	13.8	11.6