EC2203 QUANTITATIVE METHODS IN ECONOMICS II (1998-1999)

Lecture 15. Multiple regression analysis

15.1. Introduction

In this lecture, we extend the analysis to consider a regression equation with more than one explanatory variable, but retain the reduced form assumption, so that the dependent variable is the only *current* endogenous variable in the equation. The following topics are covered:

Section 15.2. Estimation of a reduced form multivariate equation
Section 15.3. Properties of the OLS estimator and significance tests
Section 15.4. The concept of multicollinearity
Section 15.5. Consequences of misspecification of the equation
Section 15.6. Testing a set of linear restrictions on the parameters

From an exam point of view, the assignment for this lecture is extremely important, and you should make sure that you have got to grips with it by next week.

Some basic results from matrix algebra and matrix differential calculus are employed in Lecture 15. These include the following:

(R1). The transpose of a product of matrices equals the product of the transposes 'in reverse order': e.g. $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$.

(R2). The *rank* of a matrix equals the largest number of linearly independent rows (= the largest number of linearly independent columns) in the matrix.

(R3). If $\mathbf{Ax} = \mathbf{b}$, and $\mathbf{A}$ is invertible, then we can solve for $\mathbf{x}$ as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

(R4). If A is invertible, then $\mathbf{A}^{-1} = (1/\det\mathbf{A})\cdot\text{adj}\mathbf{A}$.

(R5). If $\mathbf{a}$ and $\mathbf{x}$ are two column vectors of the same dimension, then

$$\frac{\partial(\mathbf{a}'\mathbf{x})}{\partial\mathbf{a}} = \mathbf{x}$$

(R6). If $\mathbf{Q} = \mathbf{x}'\mathbf{Ax}$ is a quadratic form with symmetric coefficient matrix, then

$$\frac{\partial(\mathbf{x}'\mathbf{Ax})}{\partial\mathbf{x}} = 2\mathbf{Ax}$$

15.2. Estimation of a reduced form multivariate equation

Suppose we are interested in analysing k variables: $Y, X_2, X_3, \ldots, X_k$. For example, Y might denote *wages*, X_2 might denote *age*, X_3 might denote *years of schooling*, etc.. We can think of the variables as the elements of a k-dimensional 'variable vector' $(Y, X_2, X_3, \ldots, X_k)$. Suppose we have n observations of this vector:

$$\begin{array}{c}(Y_1, X_{21}, X_{31}, \ldots, X_{k1})\\(Y_2, X_{22}, X_{32}, \ldots, X_{k2})\\(Y_3, X_{23}, X_{33}, \ldots, X_{k3})\\\vdots\\(Y_n, X_{2n}, X_{3n}, \ldots, X_{kn})\end{array}$$

Each of the observed X's has two subscripts: the first subscript indicates the variable, and the second subscript indicates the observation number. For example, X_{21} is the 1st observation of variable X_2, X_{34} is the 4th observation of variable X_3, and X_{kn} is the nth observation of variable X_k. In a natural extension of the bivariate regression model of Lecture 14, we assume that the Y and the X's for each t are related according to the following linear equation:

$$Y_t = \beta_1 + \beta_2X_{2t} + \beta_3X_{3t} + \cdots + \beta X_{kt} + u_t \qquad t = 1, 2, \ldots, n$$

This can also be written as

$$Y_t = \sum\nolimits_{i=1}^{k} \beta_i X_{it} + u_t \qquad\qquad t = 1, 2, \ldots, n$$

where $X_{1t} \equiv 1$ for all t so that β_1 is the constant term (introducing the symbol X_{1t} yields a more general formulation of the multiple regression model, since it also covers the case where a constant term is not included and X_1 is a genuine variable). We can write the equations for all n observations in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

which we denote by

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

(**y** and **u** are (n×1) column vectors, **X** is a (n×k) matrix, and β is a (k×1) column vector of coefficients). This is the conventional formulation in econometrics. ***Notice that the notation is different from the usual notation in matrix algebra, where the first subscript denotes the row and the second denotes the column. Here, X_{it} (the tth observation on the ith X-variable) is the element in the ith column and the tth row of the matrix* X.**

In a natural extension of the notation we used in Lecture 14, $\hat{\beta}$ is the (k×1) column vector of estimated coefficients, $\hat{\beta}_i$ being the estimate of β_i (i = 1, 2, . . ., k), and **e** is the (n×1) column vector of residuals: $\mathbf{e} \equiv \mathbf{y} - \mathbf{X}\hat{\beta} \equiv \mathbf{y} - \hat{\mathbf{y}}$, where $\hat{\mathbf{y}} \equiv \mathbf{X}\hat{\beta}$ is the (n×1) column vector of <u>fitted</u> Y-values. ***It is assumed that n > k i.e. that there are more observations than variables.***

The OLS estimates of the k elements of $\hat{\beta}$ are obtained, as in Lecture 14, by minimising the sum of squared residuals (SSR):

$$\begin{aligned} \text{SSR} &\equiv \sum\nolimits_{t=1}^{n} e_t^2 = \mathbf{e}'\mathbf{e} \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \end{aligned}$$

(Notice that the term $-2\hat{\beta}'\mathbf{X}'\mathbf{y}$ in the last line replaces $-\hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta}$ in the line above it. It is correct to do this because $\hat{\beta}'\mathbf{X}'\mathbf{y}$ is a <u>scalar</u> (i.e. a number), so it is equal to its transpose: $(\hat{\beta}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\hat{\beta}$. Thus, we can simplify $-\hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta}$ as $-2\hat{\beta}'\mathbf{X}'\mathbf{y}$). We can differentiate this expression for SSR with respect to $\hat{\beta}$ (using the elementary rules of matrix differential calculus given in the introduction) to get

$$\frac{\partial \text{SSR}}{\partial \hat{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta}$$

Equating this to zero and rearranging gives the OLS <u>normal equations</u>

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

In order to proceed, we must introduce the multivariate version of assumption (A5) in Lecture 14. This is:

***(A5). X has rank k** (i.e. the <u>columns</u> of* **X** *are <u>linearly independent</u>).*

It can be shown that if this condition holds, then the matrix $\mathbf{X}'\mathbf{X}$ has an inverse $(\mathbf{X}'\mathbf{X})^{-1}$, so we can use elementary matrix algebra to solve for $\hat{\beta}$ from the normal equations as

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y}$$

This is the famous OLS estimator of the coefficient vector in multiple regression analysis.

Notice that we can write the normal equations as $\mathbf{0} = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{X}'\mathbf{e}$. This yields results which are analogous to the 'implications' of OLS discussed in Lecture 14 (see the handout for Lecture 14, pages 4 and 5). As in the bivariate regression model, $\mathbf{X}'\mathbf{e} = \mathbf{0}$ is a direct implication of the first-order conditions for minimisation of the SSRs, and tells us that the OLS estimate $\hat{\beta}$ is always such that it makes the sample covariance between the residuals and the explanatory variables equal to zero. Also notice that $\hat{\mathbf{y}}'\mathbf{e} = \hat{\beta}'\mathbf{X}'\mathbf{e} = \mathbf{0}$, so that the OLS estimate $\hat{\beta}$ is such that the fitted values of the dependent variables and the residuals are also uncorrelated. This can be interpreted intuitively by saying that OLS estimation decomposes the dependent variable vector $\mathbf{y}$ into a fitted-value vector $\hat{\mathbf{y}}$ and a residual vector $\mathbf{e}$: $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$. The OLS estimate $\hat{\beta}$ will always be such that it sets the sample covariance between these two components equal to zero.

15.3. <u>Properties of the OLS estimator and significance tests</u>

In this section, we consider the extension to the multiple regression case of Sections 14.4 to 14.6 in the handout for Lecture 14. Firstly, the unbiasedness and consistency properties of $\hat{\beta}$ carry straight over from the two-variable case (Section 14.4). As an illustration of the method of proof, we consider unbiasedness in the case where all the X-variables (i.e. all the columns of **X**) are <u>non-stochastic</u>. We have

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y}\\ &= (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'(\mathbf{X}\beta + \mathbf{u})\\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\beta + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{u}\\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{u}\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{E}[\hat{\beta}] &= \beta + \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{u}]\\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathrm{E}[\mathbf{u}] && \text{(by assumption (A4) in Lecture 14, page 6)}\\ &= \beta && \text{(by assumption (A1) in Lecture 14: } \mathrm{E}[\mathbf{u}] = \mathbf{0}\text{)}\end{aligned}$$

Next, you will recall that the following expression was derived for the <u>variance</u> of the bivariate regression coefficient in Section 14.4.2 of Lecture 14:

$$V[\hat{\beta}] = E[(\hat{\beta} - \beta)^2] = \frac{\sigma_u^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

The concept of the *variance* of $\hat{\beta}$ generalises in the multiple regression case to the concept of the *variance-covariance matrix* of the vector $\hat{\beta}$, defined as the (k×k) matrix $E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)']$ (Appendix A of Kennedy's book, *A Guide to Econometrics*, is entitled 'All About Variance', and gathers together most of the basic formulas used to compute variances in econometrics). This matrix is symmetric, and contains the variances of the $\hat{\beta}_i$'s on the main diagonal, and the covariances between the $\hat{\beta}_i$'s in the off-diagonal positions:

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)']$$

$$\equiv \begin{bmatrix} E[(\hat{\beta}_1-\beta_1)^2] & E[(\hat{\beta}_1-\beta_1)(\hat{\beta}_2-\beta_2)] & \cdots & E[(\hat{\beta}_1-\beta_1)(\hat{\beta}_k-\beta_k)] \\ E[(\hat{\beta}_2-\beta_2)(\hat{\beta}_1-\beta_1)] & E[(\hat{\beta}_2-\beta_2)^2] & \cdots & E[(\hat{\beta}_2-\beta_2)(\hat{\beta}_k-\beta_k)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(\hat{\beta}_k-\beta_k)(\hat{\beta}_1-\beta_1)] & E[(\hat{\beta}_k-\beta_k)(\hat{\beta}_2-\beta_2)] & \cdots & E[(\hat{\beta}_k-\beta_k)^2] \end{bmatrix}$$

$$= \begin{bmatrix} V[\hat{\beta}_1] & Cov[\hat{\beta}_1,\hat{\beta}_2] & \cdots & Cov[\hat{\beta}_1,\hat{\beta}_k] \\ Cov[\hat{\beta}_1,\hat{\beta}_2] & V[\hat{\beta}_2] & \cdots & Cov[\hat{\beta}_2,\hat{\beta}_k] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[\hat{\beta}_1,\hat{\beta}_k] & Cov[\hat{\beta}_2,\hat{\beta}_k] & \cdots & V[\hat{\beta}_k] \end{bmatrix}$$

When deriving the expression for the variance of the bivariate regression coefficient in Section 14.4.2 of Lecture 14, we needed to assume that the disturbances are not autocorrelated. This assumption is also required here in order to obtain a neat formula for $E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)']$. The assumption can be conveniently expressed in terms of the *variance-covariance* matrix of the u's:

$$E[\mathbf{uu'}] = \begin{bmatrix} E[u_1^2] & E[u_1u_2] & \cdots & E[u_1u_n] \\ E[u_1u_2] & E[u_2^2] & \cdots & E[u_2u_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[u_1u_n] & E[u_2u_n] & \cdots & E[u_n^2] \end{bmatrix}$$

The assumption that the disturbances are not autocorrelated means that $E[u_tu_s] = 0$ for all t and s. This implies that all the off-diagonal elements of the variance-covariance matrix $E[\mathbf{uu'}]$ are zero. The homoskedasticity (constant variance) assumption implies that the diagonal elements of $E[\mathbf{uu'}]$ are all σ_u^2. Thus, assumptions (A2) and (A3) on page 6 of the handout for Lecture 14 can be neatly summarised in the case of multiple regression as

$$E[\mathbf{uu}'] = \begin{bmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & \sigma_u^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_u^2 \end{bmatrix}$$

This can be written more concisely using the (n×n) identity matrix **I** as follows:

$$E[\mathbf{uu}'] = \sigma_u^2\mathbf{I}$$

Above, we derived the following expression for $\hat{\beta}$:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

Therefore we can write

$$\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

Substituting this into the definition of the *variance-covariance* matrix of $\hat{\beta}$ we get

$$\begin{aligned} E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] &= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})'] \\ &= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{uu}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{uu}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} && \text{(assuming that } \mathbf{X} \text{ is non-stochastic)} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma_u^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{I}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} && \text{(since } \sigma_u^2 \text{ is a scalar, we can `take it out')} \\ &= \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

This is the famous formula for the *variance-covariance* matrix of $\hat{\beta}$, assuming that the Classical Assumptions hold. ***The variances of the regression coefficients are given by the diagonal elements of this matrix***. Under the Classical Assumptions (assumptions (A1)-(A5) on page 6 of the handout for Lecture 14), these variances are the smallest of any linear unbiased estimator. Thus the OLS estimates are called the *best linear unbiased estimates* (BLUE). As in the bivariate regression case, we do not know the true σ_u^2, so we replace it in the formula $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$ above by an unbiased estimate. In the bivariate regression model, an unbiased estimate of σ_u^2 is given by $\hat{\sigma}_u^2 = \dfrac{SSR}{n-2}$ (see the handout for Lecture 14, page 12). In general, an unbiased estimate of σ_u^2 can be obtained by dividing the sum of squared residuals by the degrees of freedom remaining after estimation. ***The degrees of freedom are calculated as the number of observations minus the number of parameters estimated***. In the bivariate regression model, only two parameters are estimated (α and β), so the degrees of freedom are n-2. In the multiple regression case, we are estimating k parameters ($\beta_1, \beta_2, \ldots, \beta_k$), so the degrees of freedom are n-k. Thus, the unbiased estimate of σ_u^2 in the multiple regression case is given by

$$\hat{\sigma}_u^2 = \frac{SSR}{n-k}$$

The standard error of $\hat{\beta}_i$ (which, you will recall, is the estimate of the standard deviation of the sampling distribution of $\hat{\beta}_i$) is given by

$$\mathrm{SE}(\hat{\beta}_i) = \hat{\sigma}_u \sqrt{(\text{the ith diagonal element of } (\mathbf{X'X})^{-1})}$$

We can proceed from this formula to t-ratios and tests of hypotheses just as in Section 14.5 of the handout for Lecture 14. *If we want to test the null hypothesis* H_0: $\beta_i = 0$, *then we calculate the t-ratio* $\hat{\beta}_i/\mathrm{SE}(\hat{\beta}_i)$, ***and compare it with the tabulated percentage points of the t-distribution with n-k degrees of freedom. These percentage points are roughly the same as those of the standard normal distribution for reasonably large values of n.***

As in Section 14.5 of the handout for Lecture 14, this procedure is only valid if the errors are not autocorrelated. ***If the errors are autocorrelated, then*** $E[u_t u_s] \neq 0$ ***for some*** $t \neq s$, ***and so*** $E[\mathbf{uu'}]$ ***is no longer a diagonal matrix.*** In this case, we must replace the assumption $E[\mathbf{uu'}] = \sigma_u^2\mathbf{I}$ by the more general (n×n) matrix: $E[\mathbf{uu'}] = \mathbf{V}$. Then we would have

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = (\mathbf{X'X})^{-1}\mathbf{X'}E[\mathbf{uu'}]\mathbf{X}(\mathbf{X'X})^{-1} = (\mathbf{X'X})^{-1}\mathbf{X'VX}(\mathbf{X'X})^{-1}$$

These variances are no longer the smallest possible, so the OLS estimates are no longer BLUE. Calculating standard errors by the previous formula will give biased answers, since they are based on the *wrong* covariance matrix. ***The table for validity of our testing procedure is identical to that for the two-variable case given in Section 14.5 of the handout for Lecture 14.*** This is a problem to which we will return when considering autocorrelation in Lecture 16.

Finally, the coefficient of determination R^2 is calculated in exactly the same way as in the bivariate regression case (see Section 14.6 in the handout for Lecture 14):

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = 1 - \frac{\text{unexplained variation}}{\text{total variation}} = 1 - \frac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

It is interpreted in exactly the same way. As before, R^2 should not be used as a measure of goodness of fit when there is no constant term in the equations.

In the context of multiple regression analysis, R^2 has a property that is sometimes considered rather undesirable. Addition of an extra explanatory variable, however useless, to the equation will never reduce R^2. At worst, it will leave it the same. ***It is sometimes thought that inclusion of variables that explain very little should be penalised.*** An alternative measure which is sometimes reported in empirical work is the adjusted-R^2, defined as

$$\text{adjusted-}R^2 = 1 - \frac{n-1}{n-k}(1-R^2)$$

It can be shown that the adjusted-R^2 decreases when an extra variable is added if the new variable's t-ratio is less than 1 in absolute value.

15.4. The concept of multicollinearity

Recall that the extended version of assumption (A5) now requires that **X** be of rank k. ***This will be violated if any two or more columns of* X *are linearly dependent.*** For example, suppose that we are trying to estimate the equation

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, 2, \ldots, n$$

but that $X_{3t} = \lambda X_{2t}$. In this case, every movement in X_2 will be matched by a movement in X_3, and we will not be able to separate the influence of X_2 on Y from that of X_3. Substituting $X_{3t} = \lambda X_{2t}$ into the above equation gives

$$Y_t = \beta_1 + (\beta_2 + \lambda\beta_3)X_{2t} + u_t$$

Thus, we can estimate $(\beta_2 + \lambda\beta_3)$, but we have no way of 'separating out' the estimate to give estimates of β_2 and β_3.

More generally, if rank(X) < k, then X'X will not be invertible. Thus, we will not be able to calculate $\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$. Such a situation is called <u>perfect multicollinearity</u>. It usually arises from a misformulation of the regression equation (in the example above, X_2 and X_3 should not have *both* been in the equation in the first place). However, the problem of multicollinearity comes in all 'degrees'. Usually it is '*less-than-perfect*' multicollinearity that we experience. ***The most common case in which this arises is when two of the explanatory variables are highly, but <u>not</u> perfectly, correlated.*** The effect is to make it difficult to separate their influences with any degree of accuracy.

Two columns of **X** being highly (but not perfectly) correlated implies that the determinant of **X'X** will be close to (but not equal to) zero. Since $(\mathbf{X'X})^{-1} = (1/\det(\mathbf{X'X}))\cdot\text{adj}(\mathbf{X'X})$, it follows that the elements of $(\mathbf{X'X})^{-1}$ will be very large. ***The consequence of this is that the <u>variances</u> of the*** $\hat{\beta}_i$ ***will be large, since we saw earlier that***

$$\text{SE}(\hat{\beta}_i) = \hat{\sigma}_u \sqrt{(\text{the ith diagonal element of } (\mathbf{X'X})^{-1})}$$

This means that the $\hat{\beta}_i$ will not be very accurately determined. The observed t-ratios will be low. ***Thus, the classic symptoms of multicollinearity are "a highish R^2 together with insignificant coefficients".*** The high R^2 means that one or more of the explanatory variables has a systematic influence on the dependent variable, but we cannot tell which ones because all the t-ratios are tiny.

It is important to emphasise that the 'problem' of multicollinearity is one of <u>degree</u>: it is perfectly normal for explanatory variables to be correlated with each other to some extent. The 'problem' of multicollinearity only arises when explanatory variables are so highly correlated with each other that it becomes difficult to separate their individual effects.

There is no all-purpose 'remedy' for multicollinearity. You might be able to deal with it by dropping variables from the model, or by adding more observations to the data set (if this is possible), but you will have to take care not to cause more problems than you solve by taking such steps.

15.5. <u>Consequences of misspecification of the equation</u>

Correct 'inference' in econometrics (i.e. estimation of population parameters and hypothesis testing) is heavily dependent on the assumption that the deterministic part of our equation is correctly specified. ***This is largely a matter of whether or not our equation contains the correct variables***. Economic theory may suggest, for example, that an important variable has been left out (possibly due to the data not being available). We will also see later (in Lectures 16 and 17) that the presence of autocorrelated or heteroskedastic residuals can be an indication of misspecification of the variables that should be included in the equation. In this section, we will briefly state the general consequences of such misspecification. Algebraic 'proofs' are not provided here. They can be found in most econometrics textbooks.

(1). *A relevant explanatory variable left out*
When a relevant explanatory variable is not included in an equation, the OLS estimate of the coefficient vector will in general be biased and inconsistent. The usual t-tests will not be valid.

(2). *An irrelevant explanatory variable included*
When an irrelevant explanatory variable is included in an equation, the OLS estimate of the regression coefficient will in general be 'inefficient' (i.e. it will not have the smallest possible variance) and will therefore be less likely to be accurate, although it will still be unbiased and conistent, and the usual t-tests will still be valid.

15.6. Testing a set of linear restrictions on the parameters

One of the most important uses of the regresson model in econometrics is in testing the validity of linear restrictions on the parameters of an equation. In this section, we consider the testing of joint exclusion restrictions, but the method applies to more general kinds of linear restrictions as well.

We have seen that we can test whether an individual parameter in a model is equal to zero by computing the t-ratio of its OLS estimate, and then comparing this with appropriate cutoff points from a t-distribution. ***Now suppose that we are interested in testing the null hypothesis that two or more parameters are jointly equal to zero.*** As an illustration, suppose we are estimating the equation

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \gamma_1 Q_{1t} + \gamma_2 Q_{2t} + \gamma_3 Q_{3t} + u_t$$

where Q_1, Q_2 and Q_3 are explanatory variables which have been included to account for any seasonal variation in the dependent variable (for example, Q_1 might indicate 'summer', Q_2 might indicate 'autumn', and Q_3 might indicate 'winter'). We might wish to test the null hypothesis that there is no seasonal variation i.e. that all three variables are irrelevant in explaining Y_t. This would be done by testing the set of restrictions $\gamma_1 = \gamma_2 = \gamma_3 = 0$. The restricted equation is given by

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

The general test procedure is as follows:
(1). Impose the restrictions to be tested on the equation to obtain the *restricted form* of the equation. Estimate this restricted form by OLS and calculate the sum of squared residuals. Call this SSR_R.
(2). Next, estimate the *unrestricted form* of the equation (i.e. the original equation), and calculate here the sum of squared residuals. Call this SSR_U.
(3). Finally, calculate the ratio

$$F = \frac{(SSR_R - SSR_U)/d}{SSR_U/(n-k)}$$

where
d = the number of restrictions (this is often called the 'numerator degrees of freedom');
k = the number of parameters in the original (i.e. unrestricted) equation; and
n = the number of observations used in the estimation of the equations.

This will be our *test statistic.* ***It can be shown that under the null hypothesis that the restrictions are true, this ratio has a probability distribution known as 'Fisher's F-distribution with d and (n - k) degrees of freedom'.*** If the restrictions are satisfied (i.e. if the null hypothesis is true), then we would expect the restricted and unrestricted forms of the equation to give very similar results. So we would expect SSR_R and SSR_U to be very similar. Thus, we would expect our test statistic to take a small positive value i.e. to be 'close' to zero (note that imposing a restriction on an equation can never reduce the sum of squared residuals, so it must always be the case that $SSR_R > SSR_U$; this means that the F-statistic must always be positive). ***So we will not want to reject the null hypothesis when the test statistic gives a small value.*** If, on the other hand, one or more of the restrictions does not hold (i.e. the null hypothesis is not true), then the restricted form of the equation will have had an invalid restriction imposed upon it, and hence we would expect SSR_R to be considerably greater than SSR_U. In this case, our test statistic will be considerably greater than zero. ***Therefore we will want to reject the null hypothesis when the test statistic gives a large value.***

Given the distribution of the test statistic, we will accept the null hypothesis at the 5% significance level if the value of the test statistic is below the 95% threshold. Given the value of d and the value of (n-k), the appropriate threshold is obtained by consulting the table of 95th percentiles of the F-distribution (this table is attached, along with the table of 99th percentiles).

Example: Consider the 'seasonal variation' example above. Suppose we use 50 observations of each variable to estimate the equation

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \gamma_1 Q_{1t} + \gamma_2 Q_{2t} + \gamma_3 Q_{3t} + u_t$$

and obtain a sum of squared residuals SSR = 1.6. We will call this the 'unrestricted' sum of squared residuals, and write $SSR_U = 1.6$. Suppose we now impose the set of restrictions $\gamma_1 = \gamma_2 = \gamma_3 = 0$ on the equation to get the restricted form

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

and estimate this restricted form by OLS getting SSR = 5.2. We will call this the 'restricted' sum of squared residuals, and write $SSR_R = 5.2$. We have d = 3, and n-k = 50 - 6 = 44, so the F-statistic is

$$F = \frac{(SSR_R - SSR_U)/d}{SSR_U/(n-k)} = \frac{(5.2-1.6)/3}{1.6/44} = \frac{1.2}{0.0364} = 32.9670$$

From the table of 95th percentiles of the F-distribution (attached), we see that the critical value when d = 3 and (n-k) = 44 is less than 2.84. Since F > 2.84, we reject the null hypothesis that the parameters γ_1, γ_2, and γ_3 are jointly equal to zero. ***Intuitively, imposing this joint restriction leads to a considerable worsening of the model, producing an excessively large sum of squared residuals in the restricted model compared to the unrestricted one.***

(End of Lecture 15)

Assignment for Lecture 15. Multiple regression analysis

Please make sure that you can do the following problems by next week. You will get very similar ones in your exams. Full solutions are attached, but you should not look at them until you have made every effort to answer the questions yourself.

Question 1

The following estimated equation was obtained by Ordinary Least Squares using quarterly data for 1991 to 1996 inclusive (24 observations):

$$\underset{}{Y_t} = \underset{(2.12)}{1.10} - \underset{(0.0034)}{0.0096}\,X_{1t} - \underset{(3.35)}{4.56}\,X_{2t} + \underset{(0.007)}{0.034}\,X_{3t} \qquad SSR = 20.22$$

The figures in brackets are standard errors.

(i). Test the significance of each of the slope coefficients.

(ii). When three variables representing the first three quarters of the year were <u>added</u> to the equation, the sum of squared residuals fell to 19.35. Test the null hypothesis that there is no seasonal variation in the dependent variable, both at the 1% and 5% significance levels.

Question 2

The following equations were estimated on 24 observations, 1918-1941, where D_t is dividends and E_t is earnings in year t. Standard errors are given in parentheses, and SSR is the sum of squared residuals.

$$(A)\quad D_t = \underset{(0.20)}{0.59} + \underset{(0.10)}{0.40}\,E_t \qquad SSR = 2.1849$$

$$(B)\quad D_t = \underset{(0.17)}{-0.14} + \underset{(0.08)}{0.32}\,E_t - \underset{(0.10)}{0.10}\,E_{t-1} + \underset{(0.14)}{0.70}\,D_{t-1} \qquad SSR = 0.84821$$

Test the following hypotheses, giving the relevant critical values at both the 1 per cent level and the 5 per cent level:

(a). that the coefficients of E_{t-1} and D_{t-1} are equal to zero individually.

(b). that the coefficients of E_{t-1} and D_{t-1} are equal to zero jointly, using the F test.

SOLUTIONS

Question 1

(i) $n = 24 \quad k = 4 \quad$ so $\quad (n-k) =$ degrees of freedom $= 20$

Coeff. of X_{1t} : $\quad t\text{-ratio} = \dfrac{-0.0096}{0.0034} = 2.824$

Coeft of X_{2t} : $\quad t\text{-ratio} = \dfrac{-4.56}{3.35} = -1.361$

Coeft of X_{3t} : $\quad t\text{-ratio} = \dfrac{0.034}{0.007} = 4.857$

Critical value for test at 5% level with 20 d.f $= 2.086$

Thus, coefficients of X_{1t} and X_{3t} are significant at 5% level, but coefficient of X_{2t} is insignificantly different from zero.

(ii) $SSR_u = 19.35 \qquad d = 3$

$SSR_R = 20.22 \qquad n-k = 24-7 = 17$

The statistic is

$$F(d=3,\ n-k=17) = \frac{(20.22-19.35)/3}{19.35/17} = 0.255$$

Critical value for 1% test: 5.18

Critical value for 5% test: 3.20

Do <u>not</u> reject the null hypothesis at either significance level.

Question 2

(a) $n = 24$ $k = 4$ so $n - k$ = degrees of freedom = 20.

Coeff. of E_{t-1} : t-ratio = $\frac{-0.10}{0.10} = -1$

coeff. of D_{t-1} : t-ratio = $\frac{0.70}{0.14} = 5$

critical value for 1% test : 2.845

critical value for 5% test : 2.086

Coefficient of D_{t-1} is significant at both levels.

Coefficient of E_{t-1} is insignificant at both levels.

(b) $SSR_u = 0.84821$ $d = 2$

$SSR_R = 2.1849$ $n-k = 20$

The test statistic is

$$F(d=2,\ n-k=20) = \frac{(2.1849 - 0.84821)/2}{2.1849/20} = 6.118$$

Critical value for 1% test : 5.85

Critical value for 5% test : 3.49

Reject the null hypothesis at both significance levels.

TABLE 3. Percentiles of the Student's t Distribution; Table Entry Is x Such That $\text{Prob}(t_n \leq x) = P$

P n	.750	.900	.950	.975	.990	.995
1	1.000	3.078	6.314	12.71	31.82	63.66
2	0.817	1.886	2.920	4.303	6.965	9.925
3	0.766	1.638	2.354	3.182	4.541	5.841
4	0.741	1.533	2.132	2.777	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.708
7	0.711	1.415	1.895	2.365	2.998	3.500
8	0.706	1.397	1.860	2.306	2.897	3.355
9	0.703	1.383	1.833	2.262	2.822	3.250
10	0.700	1.372	1.813	2.228	2.764	3.169
11	0.698	1.363	1.796	2.201	2.718	3.106
12	0.696	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.161	2.650	3.012
14	0.693	1.345	1.762	2.145	2.624	2.977
15	0.691	1.341	1.753	2.132	2.602	2.947
16	0.691	1.337	1.746	2.120	2.584	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.879
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.326	1.725	2.086	2.528	2.845
21	0.687	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.686	1.319	1.714	2.069	2.500	2.807
24	0.686	1.318	1.711	2.064	2.492	2.797
25	0.685	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.704	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.311	1.697	2.042	2.457	2.750
35	0.682	1.307	1.690	2.030	2.438	2.724
40	0.681	1.303	1.684	2.021	2.423	2.705
45	0.680	1.301	1.680	2.014	2.412	2.690
50	0.680	1.299	1.676	2.009	2.403	2.678
60	0.679	1.296	1.671	2.000	2.390	2.660
70	0.679	1.294	1.667	1.994	2.381	2.648
80	0.679	1.292	1.664	1.990	2.374	2.639
90	0.678	1.291	1.662	1.987	2.368	2.632
100	0.677	1.290	1.660	1.984	2.364	2.626
∞	0.674	1.282	1.645	1.960	2.326	2.576

TABLE 5. 95th Percentiles of the F Distribution; Table Entry Is f Such That $\text{Prob}(F_{n_1,n_2} \le f) = 0.95$

	n_1 = Degrees of Freedom for the Numerator																	
n_2	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	50	60	∞
1	161.	200.	216.	225.	230.	234.	237.	239.	241.	242.	244.	246.	248.	250.	251.	251.	252.	254.
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	6.54	9.12	5.89	8.94	5.63	8.85	5.49	8.79	8.74	5.32	8.66	8.62	8.59	8.58	8.57	8.83
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.97	5.91	5.86	5.80	5.75	5.72	5.70	5.69	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.46	4.44	4.43	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.88	3.81	3.78	3.75	3.74	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.58	3.51	3.45	3.38	3.34	3.32	3.31	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	3.04	3.02	3.01	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.83	2.80	2.79	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.78	2.70	2.66	2.64	2.62	2.54
11	4.84	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.57	2.53	2.51	2.49	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.47	2.43	2.40	2.39	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.72	2.67	2.60	2.53	2.46	2.38	2.34	2.31	2.30	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.54	2.46	2.39	2.31	2.27	2.24	2.22	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.21	2.18	2.16	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.43	2.35	2.28	2.19	2.15	2.12	2.11	2.01
17	4.45	3.59	3.20	2.97	2.81	2.70	2.62	2.55	2.50	2.45	2.38	2.31	2.23	2.15	2.10	2.08	2.06	1.96
18	4.41	3.56	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.11	2.06	2.04	2.02	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.24	2.16	2.07	2.03	2.00	1.98	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.39	2.35	2.28	2.20	2.13	2.04	1.99	1.97	1.95	1.84
21	4.32	3.47	3.07	2.84	2.69	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.01	1.97	1.94	1.92	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	1.99	1.94	1.91	1.89	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.38	2.32	2.28	2.20	2.13	2.05	1.96	1.91	1.89	1.87	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.26	2.18	2.11	2.03	1.94	1.89	1.86	1.84	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.17	2.09	2.01	1.92	1.87	1.84	1.82	1.71
26	4.23	3.37	2.98	2.74	2.59	2.48	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.90	1.85	1.82	1.80	1.70
27	4.22	3.36	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.21	2.13	2.06	1.97	1.89	1.84	1.81	1.79	1.68

14

15

n_1 = Degrees of Freedom for the Numerator

n_2	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	50	60	∞
28	4.20	3.34	2.95	2.72	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.87	1.82	1.79	1.77	1.66
29	4.18	3.33	2.94	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.11	2.03	1.95	1.86	1.81	1.78	1.75	1.65
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.17	2.09	2.02	1.93	1.84	1.79	1.76	1.74	1.62
35	4.12	3.27	2.88	2.64	2.49	2.37	2.29	2.22	2.16	2.12	2.04	1.96	1.88	1.79	1.74	1.70	1.68	1.57
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.13	2.08	2.00	1.92	1.84	1.75	1.69	1.66	1.64	1.51
45	4.06	3.21	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	1.98	1.90	1.81	1.71	1.66	1.63	1.60	1.48
50	4.05	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.87	1.79	1.69	1.63	1.60	1.58	1.45
60	4.00	3.15	2.76	2.53	2.37	2.26	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.65	1.60	1.56	1.54	1.39
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.89	1.81	1.72	1.62	1.57	1.53	1.51	1.36
80	3.96	3.11	2.72	2.49	2.33	2.22	2.13	2.06	2.00	1.95	1.88	1.79	1.70	1.60	1.55	1.51	1.48	1.34
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.86	1.78	1.69	1.59	1.53	1.49	1.47	1.32
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.98	1.93	1.85	1.77	1.68	1.57	1.52	1.48	1.45	1.30
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.39	1.34	1.31	1.00

TABLE 6. 99th Percentiles of the F Distribution; Table Entry Is f Such That $\mathrm{Prob}(F_{n_1,n_2} \leq f) = 0.99$

	n_1 = Degrees of Freedom for the Numerator																	
n_2	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	50	60	∞
1	4052	5000	5403	5625	5724	5859	5928	5982	6023	6056	6106	6157	6209	6261	6287	6302	6313	6366
2	98.5	99.0	99.2	99.3	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	21.3	28.7	19.0	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.5	26.4	26.4	26.3	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.6	14.4	14.2	14.0	13.8	13.8	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.6	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.38	9.29	9.24	9.20	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.88	7.72	7.56	7.40	7.23	7.14	7.09	7.06	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	5.99	5.91	5.86	5.82	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.20	5.12	5.07	5.03	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.65	4.57	4.52	4.48	4.31
10	10.0	7.56	6.55	6.00	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.25	4.17	4.12	4.08	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	3.94	3.86	3.81	3.78	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.70	3.62	3.57	3.54	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.67	3.51	3.43	3.38	3.34	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.35	3.27	3.22	3.18	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.01	3.90	3.81	3.67	3.52	3.37	3.22	3.13	3.08	3.05	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.10	3.02	2.97	2.93	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.00	2.92	2.87	2.84	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.92	2.84	2.79	2.75	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.85	2.76	2.71	2.68	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.57	3.46	3.37	3.23	3.09	2.94	2.78	2.70	2.64	2.61	2.42
21	8.02	5.78	4.88	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.72	2.64	2.58	2.55	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.67	2.58	2.53	2.50	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.08	2.93	2.78	2.62	2.54	2.48	2.45	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.58	2.49	2.44	2.40	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.54	2.45	2.40	2.36	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.10	2.96	2.82	2.66	2.50	2.42	2.36	2.33	2.13
27	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.47	2.38	2.33	2.29	2.10

n_2	n_1 = Degrees of Freedom for the Numerator																	
	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	50	60	∞
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.44	2.35	2.30	2.26	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.01	2.87	2.73	2.58	2.41	2.33	2.27	2.24	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.31	3.17	3.07	2.98	2.84	2.70	2.55	2.39	2.30	2.25	2.21	2.01
35	7.41	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.74	2.60	2.45	2.28	2.19	2.14	2.10	1.91
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.67	2.52	2.37	2.20	2.12	2.06	2.02	1.81
45	7.23	5.11	4.25	3.77	3.46	3.23	3.07	2.94	2.83	2.74	2.61	2.47	2.31	2.15	2.06	2.00	1.96	1.75
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79	2.70	2.56	2.42	2.27	2.10	2.01	1.95	1.91	1.68
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.03	1.94	1.88	1.84	1.60
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.45	2.31	2.15	1.98	1.89	1.83	1.79	1.53
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.42	2.27	2.12	1.94	1.85	1.79	1.75	1.49
90	6.93	4.85	4.01	3.54	3.23	3.01	2.85	2.72	2.61	2.53	2.39	2.25	2.09	1.92	1.82	1.76	1.72	1.48
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.70	2.59	2.50	2.37	2.22	2.07	1.89	1.80	1.74	1.69	1.43
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.70	1.57	1.51	1.46	1.00