Regression Analysis I

Linear Models for Autocorrelated Data

Hossein Moradi Rekabdarkolaee

South Dakota State University

email: *hossein.moradirekabdarkolaee@sdstate.edu*.

Office: Arch, Math, and Engineering Building, *Room: 254*.

Multiple Linear Regression

- A multiple linear regression model can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p] = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Multiple Linear Regression

- The basic assumption in linear regression model: the error are identically and independently distributed.
- Thus in the model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

it is assumed that

$$E(\epsilon_t, \epsilon_{t-s}) = \begin{cases} \sigma^2 & s = 0 \\ 0 & s \neq 0 \end{cases}$$

- In other words, the correlation between the successive disturbances is zero.
- If this assumption meets, alongside with other assumptions, then OLS can be used for estimating the regression parameters.

Serially Correlated Data

- The identically and independently assumption can be violated in two ways:
  - Not identical: the variance of error term does not remains constant:
    - Problem: heteroskedasticity.
    - Solution: Using Weighted Least Square approach.
  - Not Independent: $E(\epsilon_t, \epsilon_{t-s}) \neq 0$ when $s \neq 0$.
    - Problem: Autocorrelation.
- When autocorrelation exists, some or all off diagonal elements in $E(\epsilon\epsilon^T)$ are nonzero.

Serially Correlated Data

- Sometimes the study and explanatory variables have a natural sequence order over time, i.e., the data is collected with respect to time. Such data is termed as time series data.
- The error terms in time series data are serially correlated.
- The autocovariance at lag s is defined as

$$\gamma_s = E(\epsilon_t, \epsilon_{t-s}); \quad s = 0, \ \pm 1, \ \pm 2, \ldots$$

- At $s = 0$, we have the constant variance.
- The autocorrelation coefficient at lag s is defined as

$$\rho_s = \frac{E(\epsilon_t, \epsilon_{t-s})}{\sqrt{var(\epsilon_t) var(\epsilon_{t-s})}} = \frac{\gamma_s}{\gamma_0}; \quad s = 0, \ \pm 1, \ \pm 2, \ldots$$

Serially Correlated Data

- Let ρ_s and γ_s be constant over time and depend only on length of lag s.
- The autocorrelation between the successive terms (ϵ_2 and ϵ_1), (ϵ_3 and ϵ_2), . . ., (ϵ_n and ϵ_{n-1}), gives the autocorrelation of order one, i. e., ρ_1.
- Similarly, the autocorrelation between the successive terms (ϵ_3 and ϵ_1), (ϵ_4 and ϵ_2), . . ., (ϵ_n and ϵ_{n-2}), gives the autocorrelation of order two, i. e., ρ_2.

Source of Autocorrelation

- Some of the possible reasons for the introduction of autocorrelation in the data are as follows:
  - Carryover of effect, at least in part, is an important source of autocorrelation. For example, the monthly data on expenditure on household is influenced by the expenditure of preceding month.
  - Another source of autocorrelation is the effect of deletion of some variables.
  - The misspecification of the form of relationship can also introduce autocorrelation in the data.
  - The difference between the observed and true values of variable is called measurement error or errors-in-variable.

Structure of Error Term

- Consider the situation where the disturbances are autocorrelated:

$$
\begin{aligned}
E(\epsilon\epsilon^T) &= \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix} \\
&= \gamma_0 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}
\end{aligned}
$$

Serially Correlated Data

- Observe that now there are $(n + p)$ parameters- $\beta_1, \beta_2, \ldots, \beta_p, \sigma^2, \rho_1, \rho_2, \ldots, \rho_{n-1}$.
- These parameters are to be estimated on the basis of available n observations.
- Since the number of parameters are more than the number of observations, so the situation is not good from the statistical point of view.
- In order to handle the situation, some special form and the structure of the disturbance term is needed to be assumed so that the number of parameters in the covariance matrix of disturbance term can be reduced.

Serially Correlated Data

- Popular autocorrelation structures are:
  - Autoregressive (AR) process.
  - Moving average (MA) process.
  - Joint autoregression moving average (ARMA) process.

Estimation under the first order autoregressive process:

- Consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Assume ϵ_i's follow a first order AR.

Serially Correlated Data

- First order AR is defined as

$$\epsilon_i = \rho\epsilon_{i-1} + \nu_i$$

- where $|\rho| < 1$ and $E(\nu_i) = 0$. Thus the error structure will be

$$E(\nu_i, \nu_{i-s}) = \begin{cases} \sigma_\nu^2 & s = 0 \\ 0 & s \neq 0 \end{cases}$$

for all $i = 1, 2, \ldots, n$ and ρ is the first order autocorrelation between ϵ_i and ϵ_{i-1}.

- Thus, we have

$$\begin{aligned} \epsilon_i &= \rho\epsilon_{i-1} + \nu_i \\ &= \rho(\epsilon_{i-2} + \nu_{i-1}) + \nu_i \\ &\vdots \\ &= \nu_i + \rho\nu_{i-1} + \rho^2\nu_{i-2} + \ldots = \sum_{j=0}^{\infty} \rho^j \epsilon_{i-j} \end{aligned}$$

Serially Correlated Data

- Thus, for all i, we have

$$\begin{aligned} E(\epsilon_i) &= 0 \\ E(\epsilon_i^2) &= E(\epsilon_i^2) + \rho^2 E(\epsilon_{i-1}^2) + \rho^4 E(\epsilon_{i-2}^2) + \ldots = (1 + \rho^2 + \rho^4)\sigma_\nu^2 \\ E(\epsilon_i^2) &= \sigma_\epsilon^2 = \frac{\sigma_\nu^2}{1 - \rho^2} \end{aligned}$$

- In addition

$$\begin{aligned} E(\epsilon_i \epsilon_{i-1}) &= E\left[(\epsilon_i + \rho\epsilon_{i-1} + \rho^2\epsilon_{i-2} + \ldots) \times (\epsilon_{i-1} + \rho\epsilon_{i-2} + \rho^2\epsilon_{i-3} + \ldots)\right] \\ &= E\left[(\epsilon_i + \rho(\epsilon_{i-1} + \rho\epsilon_{i-2} + \ldots) \times (\epsilon_{i-1} + \rho\epsilon_{i-2} + \rho^2\epsilon_{i-3} + \ldots)\right] \\ &= \rho E\left[(\epsilon_{i-1} + \rho\epsilon_{i-2} + \ldots)^2\right] = \rho\sigma_\epsilon^2. \end{aligned}$$

Similarly,

$$E(\epsilon_i \epsilon_{i-2}) = \rho^2 \sigma_\epsilon^2$$

And in general,

$$E(\epsilon_i \epsilon_{i-s}) = \rho^s \sigma_\epsilon^2$$

Serially Correlated Data

- The general form of the covariance function is

$$E(\epsilon\epsilon^T) = \mathbf{\Omega} = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}$$

- The variance of the errors are not $\sigma_\epsilon^2\mathbf{I}$.

Consequences of autocorrelated error

- Consider the model with first order autoregressive disturbances

$$y = \mathbf{X}\beta + \epsilon; \quad \epsilon_t = \rho\epsilon_{t-1} + \nu_t, \quad t = 1, 2, \ldots, n$$

- With the following assumptions

$$\begin{aligned} E(\epsilon) &= 0; \quad E(\epsilon\epsilon^T) = \mathbf{\Omega} \\ E(\nu_t) &= 0; \quad E(\nu_t\nu_{t-s}) = \begin{cases} \sigma_\nu^2 & s = 0 \\ 0 & s \neq 0 \end{cases} \end{aligned}$$

  where $\mathbf{\Omega}$ is positive definite matrix.

- The OLS estimate of β is

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\beta + \epsilon) \\ &\Rightarrow E(\hat{\beta} - \beta) = E\left[(\mathbf{X}^T\mathbf{X})^{-1}\epsilon\right] = 0 \end{aligned}$$

- Thus the OLS remains unbiased.

Consequences of autocorrelated error

- The variance of β is

$$var(\beta) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Omega}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}.$$

- The residual vector is

$$\begin{aligned} e &= y - \mathbf{X}\beta = \bar{H}y = \bar{H}\epsilon \\ e^Te &= y^T\bar{H}y = \epsilon^T\bar{H}\epsilon \\ E(e^Te) &= E(\epsilon^T\bar{H}\epsilon) = E(\epsilon^T\epsilon) - E(\epsilon^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\epsilon) \\ &= n\sigma_\epsilon^2 - tr((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Omega}\mathbf{X}). \end{aligned}$$

- Since $s^2 = \frac{e^Te}{n-1}$, thus

$$E(s^2) = \frac{\sigma_\epsilon^2}{n-1} - \frac{1}{n-1}tr((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Omega}\mathbf{X})$$

- Thus, s^2 is biased estimator of σ^2.

Consequences of autocorrelated error

- Application of OLS fails in case of autocorrelation in the data and leads to serious consequences as
  - Overly optimistic view from R^2.
  - narrow confidence interval.
  - usual t-ratio and F-ratio tests provide misleading results.
  - prediction may have large variances.
- Generalized Least Squares estimate of β is more efficient than OLS.
- The Durbin-Watson (DW) test is used for testing the hypothesis of lack of first order autocorrelation in the disturbance term.
- The DW test statistic is

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

Consequences of autocorrelated error

- For large n:

$$d \approx 1 + 1 - 2r = 2(1 - r).$$

  where r is the sample autocorrelation coefficient from residuals based on OLS and can be regarded as the regression coefficient of e_t on e_{t-1}.
- Here
  - Positive Autocorrelation of ϵ_t's $\Rightarrow d < 2$,
  - Negative Autocorrelation of ϵ_t's $\Rightarrow d > 2$,
  - Zeor Autocorrelation of ϵ_ts $\Rightarrow d \approx 2$
- Since $-1 \leq r \leq 1$, thus we have
  - $2 < d < 4$ for $-1 < r < 0$,
  - $0 < d < 2$ for $0 < r < 1$,

  Thus d is between 0 and 4.

Consequences of autocorrelated error

- Exact critical values for DW are difficult to obtain.
- However, there are tables of critical values for lower and upper bounds d_L and d_U
- the decision rule is

Table: DW test procedure for $H_0 : \rho = 0$

Nature of H_1	Reject H_0	Fail to Reject H_0	Inconclusive
$\rho > 0$	$d < d_L$	$d > d_U$	$d_L < d < d_U$
$\rho < 0$	$d > (4 - d_L)$	$d < (4 - d_L)$	$(4 - d_U) < d < (4 - d_L)$
$\rho \neq 0$	$d < d_L$ or $d > (4 - d_L)$	$d_U < d < (4 - d_U)$	$d_L < d < d_U$ or $(4 - d_U) < d < (4 - d_L)$

Limitations of DW test

- If d falls in the inconclusive zone, then no conclusive inference can be drawn.
- This zone becomes fairly larger for low degrees of freedom.
- One solution is to reject H_0 if the test is inconclusive.
- A better solutions is to modify the test as
  - Reject the H_0 when $d < d_U$.
  - Fail to reject the H_0 when $d \geq d_U$.
- This test gives satisfactory solution when values of regressors change slowly.
- The DW test is not applicable when intercept term is absent in the model.
- In such a case, one can use another critical values, say d_M in place of d_L. The tables for critical values d_M are available.
- The test is not valid when lagged dependent variables appear as explanatory variables.

Estimation procedures with autocorrelated errors when autocorrelation coefficient is known

- Consider the estimation of regression coefficient under first order autoregressive

$$\begin{aligned} y &= \mathbf{X}\boldsymbol{\beta} + \epsilon, \\ \epsilon_t &= \rho\epsilon_{t-1} + \nu_t \end{aligned}$$

- with $E(\epsilon) = 0$, $E(\epsilon\epsilon^T) = \Omega$, $E(\nu) = 0$, and $E(\nu\nu^T) = \sigma^2\mathbf{I}$.
- Estimate the regression parameters with following steps:
  - Find the matrix P such that $P^TP = \Omega^{-1}$.
  - Transform the variables as $y^* = Py, \mathbf{X}^* = P\mathbf{X}, \epsilon^* = P\epsilon$.
  - Now, you can use OLS to obtain $\hat{\beta}$.

Estimation procedures with autocorrelated errors when autocorrelation coefficient is unknown

- Unlike the case that autocorrelation coefficient is known, there multiple different procedures for unknown case:
  - Cochran Orcutt procedure.
  - Hildreth-Lu procedure or Grid-search procedure.
  - First difference procedure.

Cochran Orcutt procedure

- Cochran Orcutt procedure is an iterative procedure:
  - Use OLS to $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$ and obtain residual verctor e.
  - Estimate ρ using
$$r = \frac{\sum_{t=2}^{n} e_{t-1} e_t}{\sum_{t=1}^{n} e_t^2},$$
  - Fit the transformed model by
$$y_t - r y_{t-1} = \beta_0(1 - r) + \beta_1(x_t - r x_{t-1}) + \nu_t$$
using OLS.
  - Test if iteration is needed using DW test.

    If the test indicates that the residuals are uncorrelated, the procedure terminates.

    If the test indicates that the residuals are correlated, reestimate the ρ from the new residuals for the fitted model and redo previous steps.

Cochran Orcutt procedure

- The Cochran Orcutt procedure does not always work properly.

  Major reason: when error terms are positively correlated, r tends to be an underestimate of ρ.

- Under certain circumstances, it may be helpful to construct pseudo-transformed values for period 1 so that the regression for the transformed variables is based on n, rather than $n - 1$, cases.

- The least squares properties of the residuals, such as that the sum of the residuals is zero, apply to the residuals for the fitted regression function with the transformed variables, not to the residuals for the fitted regression function transformed back to the original variables.

Hildreth-Lu procedure or Grid-search procedure

- The Hildreth-Lu procedure has following steps:
  - Apply OLS to

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + \nu_t, \quad t = 2, 3, \ldots, n.$$

    using different values of ρ such as $0, \pm 0.1, \pm 0.2, \ldots, \pm 1$.
  - Calculate SSE for each ρ.
  - Select the ρ that minimizes the SSE.
- Once the value of ρ is chosen, the fitted regression function corresponding to that value of ρ is examined to see if the transformation has successfully eliminated the autocorrelation.

Hildreth-Lu procedure or Grid-search procedure

- The Hildreth-Lu procedure, unlike the Cochrane-Orcutt procedure, does not require any iterations once the estimate of the autocorrelation parameter ρ is obtained.
- Note that SSE as a function of ρ is quite stable in a wide region around the minimum, as is often the case.
- It indicates that the numerical search for finding the best value of ρ need not be too fine unless there is particular interest in the intercept term β_0, since the estimate β_0 is sensitive to the value of r.

First Differences Procedure

- The autocorrelation parameter ρ is usually large.
- In addition, SSE as a function of ρ often is quite flat for large values of ρ up to 1.
- Some economists and statisticians have suggested use of $\rho = 1$ in the transformed model; i.e.

$$y_t^* = \beta_1^* x_t^* + \epsilon_t$$

where $y_t^* = y_t - y_{t-1}$ and $x_t^* = x_t - x_{t-1}$.

- These transformed variables are ordinary first differences.
- The fitted regression function in the transformed variables can be transformed back to the original variables as follows:

$$\hat{\beta}_0 = \bar{y} - \beta_1^* \bar{x}; \quad \hat{\beta}_1 = \beta_1^*.$$

Comparison of Three Methods

- All of the estimates of β_1 are quite close to each other.
- The estimated standard deviations of β_1 based on Hildreth-Lu and first differences transformation methods are quite close to each other; that with the Cochrane-Orcutt procedure is somewhat smaller.
- The estimated standard deviation of β_1 based on OLS with the original variables is still smaller.
- All three transformation methods provide essentially the same estimate of σ_ϵ^2.

Forecasting with Autocorrelated Error Terms

- One important use of autoregressive error regression models is to make forecasts.
- With these models, information about the error term in the most recent period n can be incorporated into the forecast for period $n+1$.
- This provides a more accurate forecast because, when autoregressive error regression models are appropriate, the error terms in successive periods are correlated.
- Suppose we have a simple linear regression with autoregressive error (Extending to the multiple linear regression is direct):

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t; \quad \epsilon_t = \rho \epsilon_{t-1} + \nu_t.$$

- Thus:

$$y_t = \beta_0 + \beta_1 x_t + \rho \epsilon_{t-1} + \nu_t$$

Forecasting with Autocorrelated Error Terms

- Then for $n+1$, we have

$$y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \rho \epsilon_n + \nu_{n+1}$$

which is made up the following components:

  - Expected value of $\beta_0 + \beta_1 x_{n+1}$.
  - A multiplication of ρ and the preceding error term.
  - An independent random error (also called disturbance) with mean 0.

- The forecast process for $n+1$ is denoted by F_{n+1} and it deals with all three components.

Forecasting with Autocorrelated Error Terms

- Given x_{n+1}, we estimate the expected value $\beta_0 + \beta_1 x_{n+1}$ via fitted regression function:

$$\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$$

- ρ is estimated by r and ϵ_n is estimated by $e_n = y_n - \hat{y}_n$. Thus, $\rho\epsilon_n$ is estimated by re_n.
- The expected value of ν_{n+1} is equal to zero.
- Thus,

$$F_{n+1} = \hat{Y}_{n+1} + re_n.$$

- An approximate $1 - \alpha$ prediction interval for $Y_{n+1,\{new\}}$ is

$$F_{n+1} \pm t_{1-\frac{\alpha}{2};n-3} s_{pred}$$

where

$$s_{pred} = MSE\left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$