Seemingly Unrelated Equations

- Book second edition Chapter 5 pages 155 - 160

- Application:
  - Continuous dependent variables are a “group” which share unobserved effects, but do not directly influence

- Examples:
  - Speed and standard deviation of speed in some time interval
  - A group of continuous variables that share unobserved effects

- Example: A group that shares unobserved effects
  - Consider three equations that model a driver's normal driving speed relative to the speed limit under various speed limits;

    1. Number of miles per hour normally driven above the speed limit in on an interstate with a 70mi/h speed limit and little traffic (will be negative if normally drive below the speed limit under these conditions)
    2. Number of miles per hour normally driven above the speed limit in on an interstate with a 65mi/h speed limit and little traffic

3. Number of miles per hour normally driven above the speed limit in on an interstate with a 55mi/h speed limit and little traffic.

- Estimate as separate equations using OLS?
  - OLS derivation assumes all of the information relating to the regression equation and variables is present to get unbiased and efficient estimators
  - In this speed-limit case, the disturbance term in one of these three regression equations is correlated with the disturbance term in the others. OLS does not account for this.

- When studying the number of miles per hour normally driven above the speed limit at various speed limits, the following equation system can be written (omitting subscripting for observation number):

$$S_{70} = \boldsymbol{\beta}_{70}\mathbf{Z} + \boldsymbol{\alpha}_{70}\mathbf{X} + \varepsilon_{70}$$

$$S_{65} = \boldsymbol{\beta}_{65}\mathbf{Z} + \boldsymbol{\alpha}_{65}\mathbf{X} + \varepsilon_{65}$$

$$S_{55} = \boldsymbol{\beta}_{55}\mathbf{Z} + \boldsymbol{\alpha}_{55}\mathbf{X} + \varepsilon_{55}$$

Where:

- S_{70}, S_{65} and S_{55} are the number of miles per hour drivers normally drive above the speed limit (with little traffic) for 70, 65, and 55 mi/h speed limits, respectively
- **Z** is a vector of driver and driver-household characteristics,
- **X** is a vector of vector of driver preferences and opinions
- β, α, are vectors of estimable parameters, and ε are disturbance terms.

- To obtain efficient estimates, the contemporaneous correlation of disturbances ε_1, ε_2 and ε_3 must be considered and the common approach for doing this is referred to seemingly unrelated regression estimation (SURE) as developed by Zellner (1962).

- Estimation of seemingly unrelated equations is accomplished by using generalized least squares as is done in the third stage of three-stage least squares.

- **Generalized least squares estimation**
- Ordinary least squares (OLS) assumptions are that disturbance terms have equal variances and are not correlated. Generalized least squares (GLS) is used to relax these OLS assumptions. Under OLS assumptions, in matrix notation,

$$\mathrm{E}\left(\varepsilon\varepsilon^{T}\right) = \boldsymbol{\sigma}^2\mathbf{I}$$

Where:

- $E(.)$ denotes expected value,

- ε is an $n \times 1$ column vector of equation disturbance terms (where n is the total number of observations in the data),
- ε^T is the $1 \times n$ transpose of ε,
- σ^2 is the disturbance term variance, and
- $\mathbf{I}$ is the $n \times n$ identity matrix,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & . & 0 \\ 0 & 1 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & 1 \end{bmatrix}$$

- For disturbance-term correlation, $E\left(\varepsilon\varepsilon^T\right)=\sigma^2\mathbf{\Omega}$, where

$$\mathbf{\Omega}=\begin{bmatrix} 1 & \rho & . & \rho^{N-1} \\ \rho & 1 & . & \rho^{N-2} \\ . & . & . & . \\ \rho^{N-1} & \rho^{N-2} & . & 1 \end{bmatrix}$$

- Recall that under ordinary least squares assumptions the disturbance terms have equal variances and are not correlated, resulting in parameters being estimated as,

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$

Where:

- $\hat{\boldsymbol{\beta}}$ is a $p \times 1$ column vector (where p is the number of parameters),
- $\mathbf{X}$ is an $n \times p$ matrix of data (where n is the number of observations),
- $\mathbf{X}^T$ is the transpose of $\mathbf{X}$, and $\mathbf{Y}$ is an $n \times 1$ column vector.

Generalized least squares generalizes this expression by using a matrix that accounts for correlation among equation error terms ($\boldsymbol{\Omega}$), so,

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{Y}$$

- The most difficult aspect of generalized least squares estimation is obtaining an estimate of the $\mathbf{\Omega}$ matrix.

- In seemingly unrelated regression estimation, $\mathbf{\Omega}$ is estimated from initial ordinary least squares estimates of individual equations.

- In 3SLS, it is estimated using the initial 2SLS parameter estimates.

Mary Martchouk's Dissertation:

- Using Bluetooth detection, collect real-time travel time information on I-69 in Indianapolis

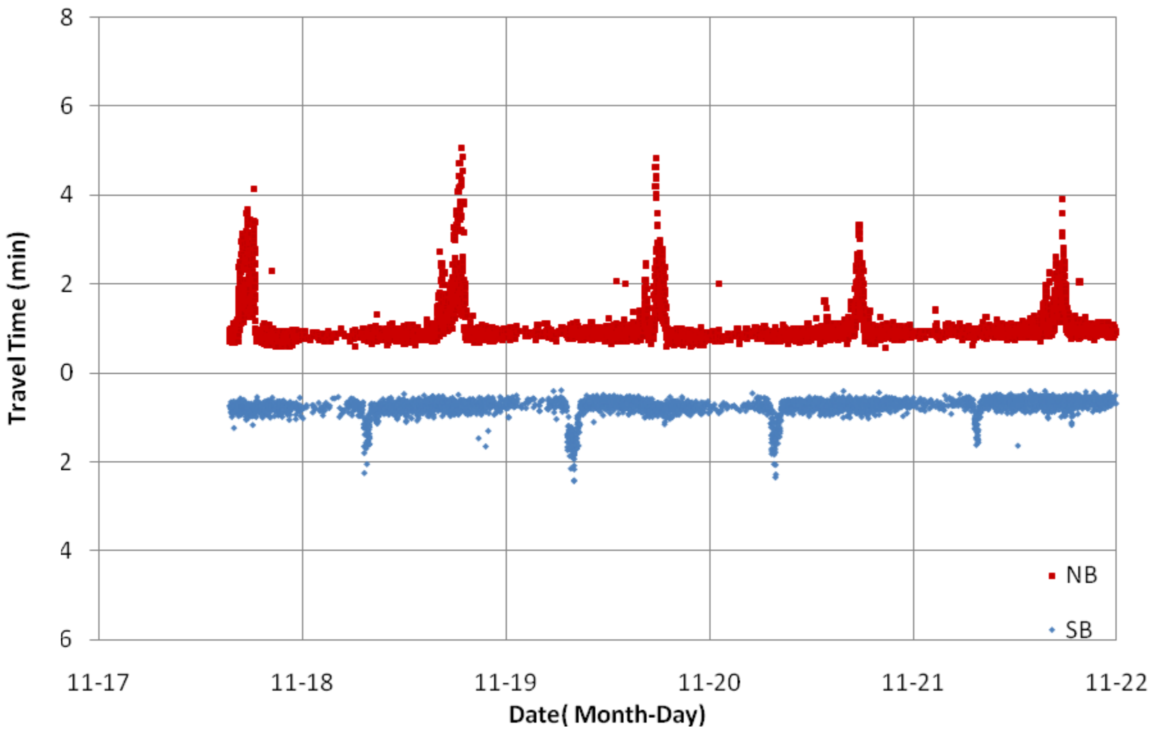

Figure 1. Travel Time on Segment 1 (between mile marker 2.9 and mile marker 3.7) from 11/17 through 11/22 (NB=northbound, SB=southbound).

- The average travel time of vehicles traversing the same segment of the road during a specified time period (for example, a 15-minute period) can vary significantly from one time period to the next.

- Standard deviation of individual vehicle travel times during 15-minute travel-time intervals ranges from 5% to 15% of the average travel time during off-peak hours.

- During the peak hours, this standard deviation can rise up to 50% of the average travel time or more.

- This inherent individual vehicle travel-time variability within the traffic stream is likely the result of individuals' lane and speed choice.

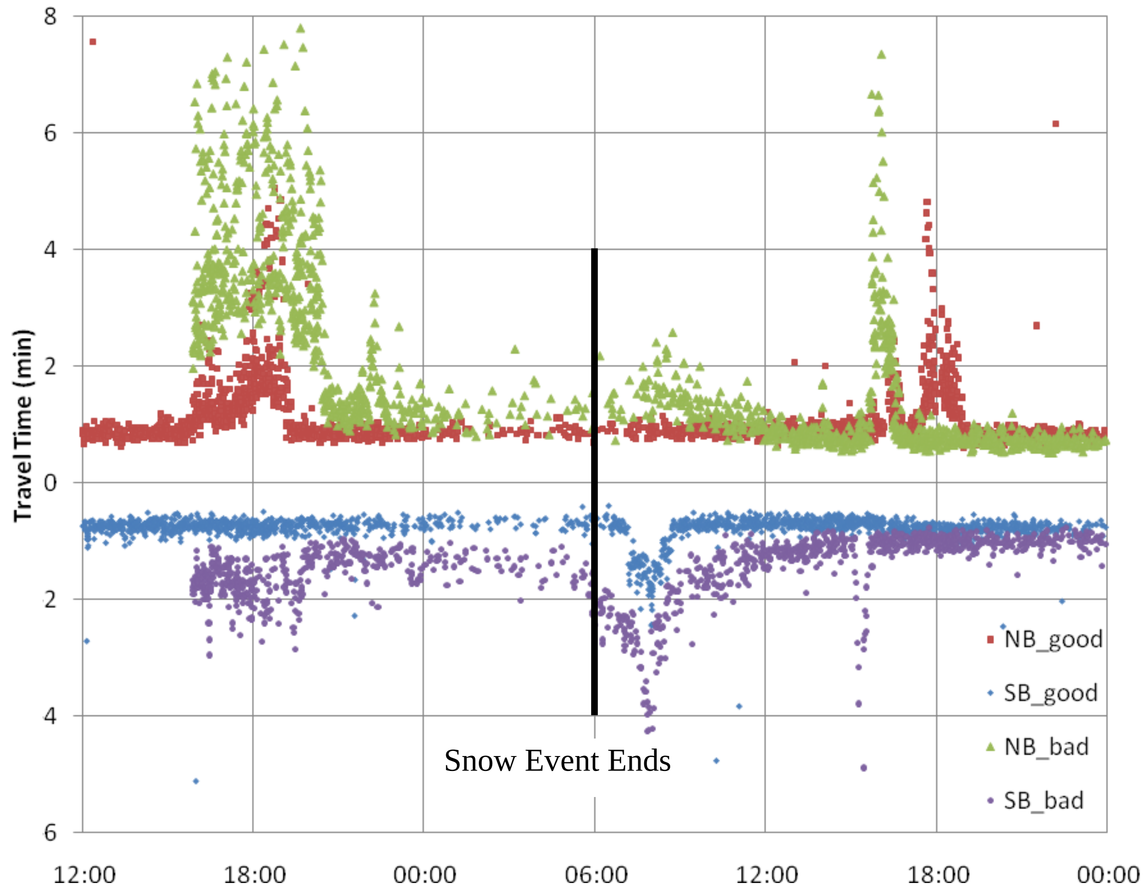

Figure 2. Travel Times for Snow Conditions (bad – January 27 and 28) and Sunny Day (good – November 18 and 19) on segment 1 (NB=northbound, SB=southbound).

To study average travel time and standard deviation of travel time in 15-minute intervals, consider the equation system:

$$TT = \boldsymbol{\beta}_1 \mathbf{X}_1 + \varepsilon_1$$

$$STDEV = \boldsymbol{\beta}_2 \mathbf{X}_2 + \varepsilon_2$$

Where:

- *TT* represents the average travel time during the 15-minute interval,
- *STDEV* is the standard deviation of individual vehicle travel times during the same 15-minute interval,
- $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of estimated parameters,

- $\mathbf{X}_1$ and $\mathbf{X}_2$ are vectors of independent variables, and
- ε_1 and ε_2 are error terms.

Note that *TT* and *STDEV* do not have a direct effect on each other and are related only indirectly via contemporaneous correlation in error terms (thus they need to be estimated by seemingly unrelated regression).

Table 1. Seemingly Unrelated Regression Model of Average Travel Time (in minutes).

Variable	**Parameter Estimate**	**Standard Error**	**t-statistic**
Constant	0.275777	0.029261	9.43
15-minute average volume to capacity ratio (raised to the 6.87 power) during the 15-minute time period one hour prior to observation	0.686212	0.179953	3.81
15-minute average travel time during the previous 15-minute time period (min)	0.796879	0.029788	26.75
Night-time indicator (1 if time between 21:00 and 6:00, 0 otherwise)	–0.09069	0.012721	–7.13
Peak-hour indicator (1 if time between 17:00 and 18:30, 0 otherwise)	0.31528	0.023262	13.55
Number of Observations	958		
Adjusted *R*-squared	0.63		

Table 2. Seemingly Unrelated Regression Model of Travel Time Standard Deviation (in minutes).

Variable	**Parameter Estimate**	**Standard Error**	**t-statistic**
Intercept	0.406418	0.029367	13.84
15-minute average speed during the previous 15-minute time period (mi/h)	–0.00704	0.000497	–14.16
Night-time indicator (1 if time between 21:00 and 6:00, 0 otherwise)	–0.06607	0.007752	–8.52
Peak-hour indicator (1 if time between 17:00 and 18:30, 0 otherwise)	0.176096	0.013685	12.87
Length of the segment (miles)	0.183697	0.017656	10.40
Number of Observations	958		
Adjusted R-squared	0.46		

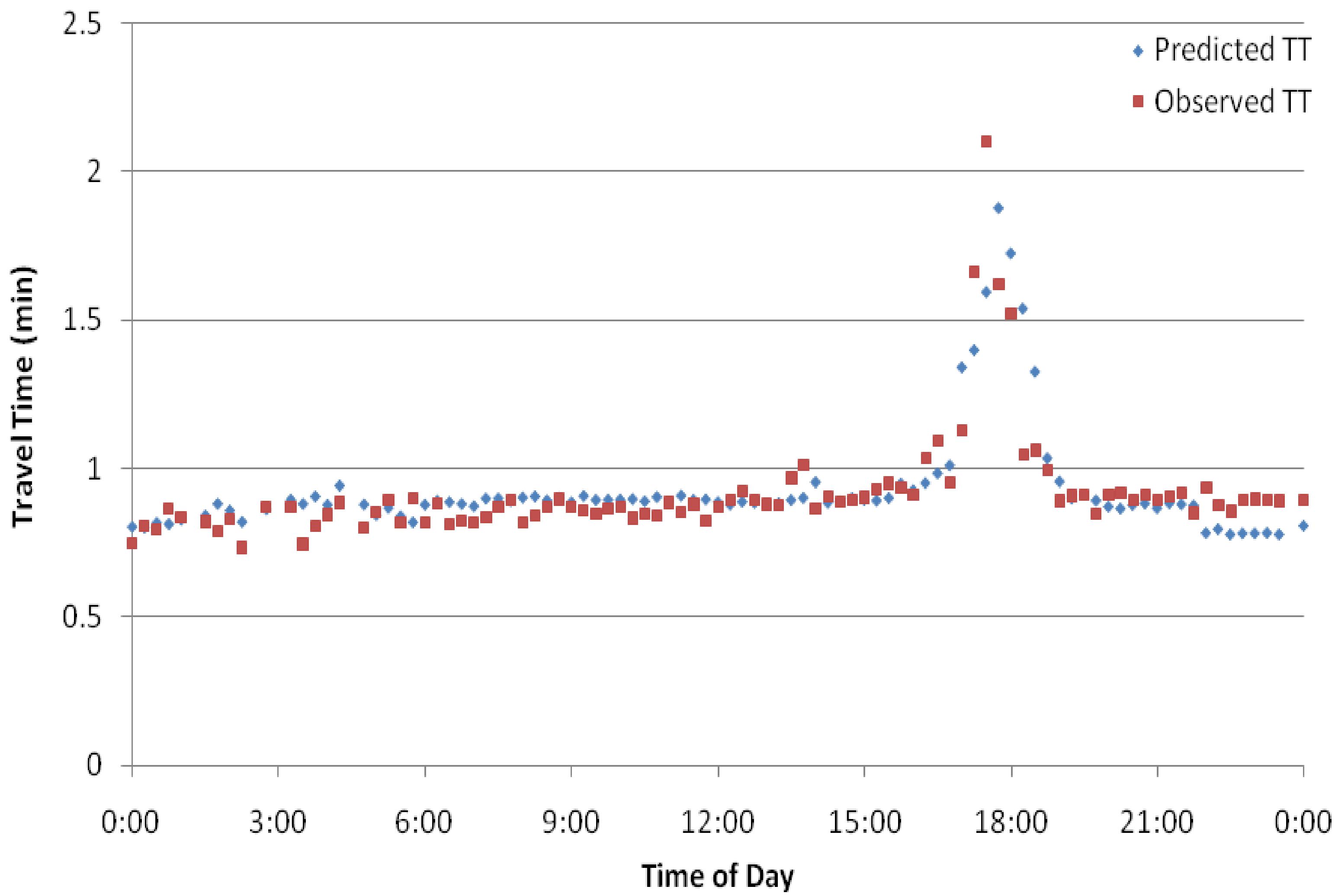

Figure 3. Segment 1 Travel Time on November 20.

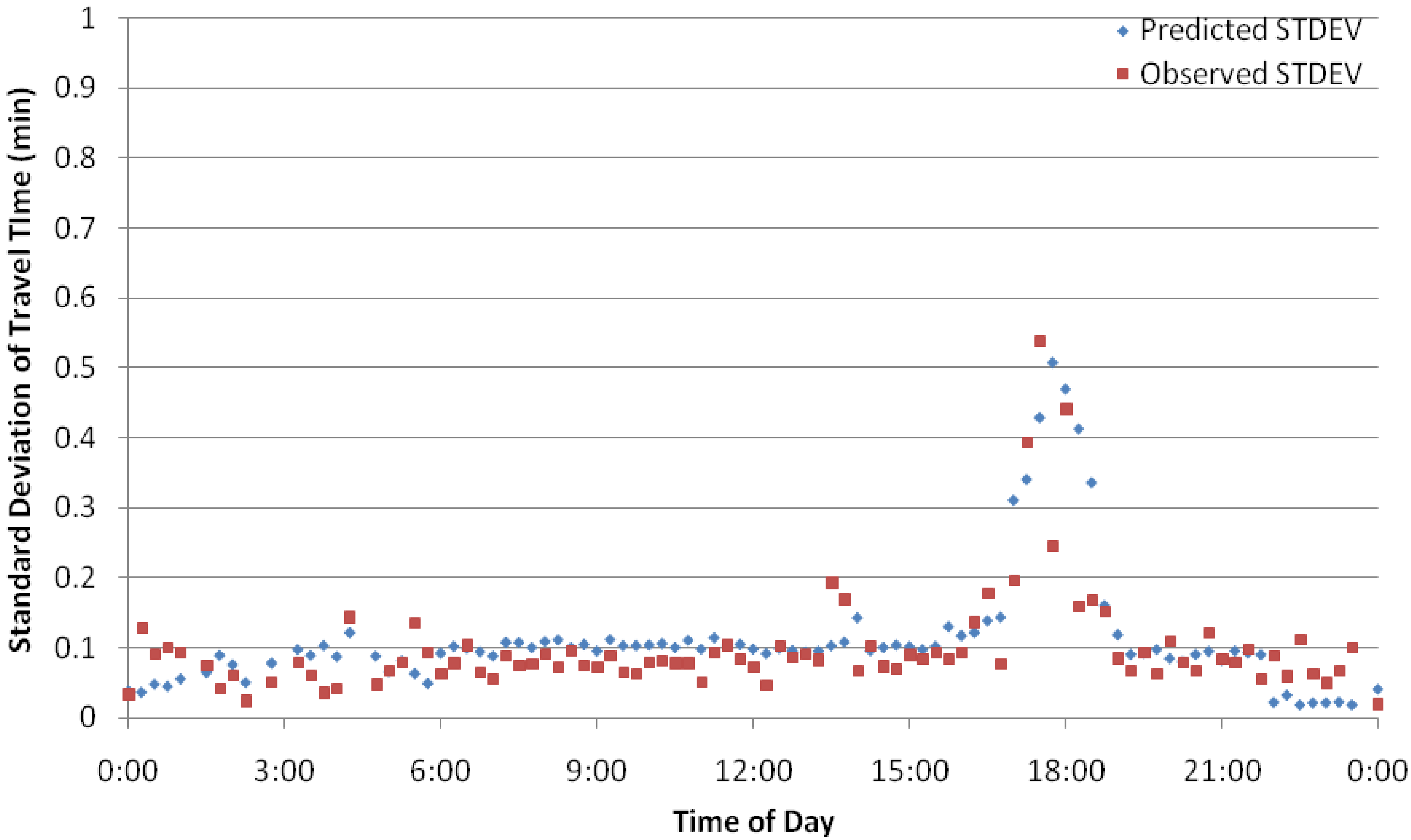

Figure 4. Segment 1 Standard Deviation (STDEV) of Travel Time.