Regression with data collected sequentially in time

Data $(y_t, x_{t1}, \dots, x_{tp})$ for $t = 1, \dots, n$ obtained sequentially in time. The assumption of independent ϵ_t in the model:

$$Y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_p x_{tp} + \epsilon_t, \quad \epsilon_t \text{ independent } N(0, \sigma^2)$$

might not hold. For sequential data, it might more reasonable to assume that the ϵ_t are serially correlated over time. If the ϵ_t are positively serially correlated, then SEs of $\hat{\beta}$'s derived under the assumption of independent ϵ_t are generally too small.

For data look at plot and correlation of lagged residual e_{t-1} versus e_t

$$\begin{array}{cc}
\hline
e_1 & e_2 \\
e_2 & e_3 \\
\vdots & \vdots \\
e_{n-1} & e_n \\
\hline
\end{array}$$

The serial correlation of lag 1 (with $\bar{e} = 0$) is

$$\hat{\rho}_1 \overset{\text{def}}{=} \frac{\sum_{t=2}^n e_{t-1} e_t}{\sqrt{\sum_{t=2}^n e_{t-1}^2 \sum_{t=2}^n e_t^2}} \in (-1, 1).$$

The Durbin-Watson statistic is

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

If n is large, then

$$DW = \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2\sum_{t=2}^n e_{t-1} e_t}{\sum_{t=1}^n e_t^2} \approx 2 - 2\hat{\rho}_1.$$

Hence $DW \approx 2$ for serially uncorrelated residuals, $DW \approx 0$ for strong positive serial correlation and $DW \approx 4$ for strong negative serial correlation, since

$$-1 \le \hat{\rho}_1 \le 1 \Rightarrow 0 \le 2 - 2\hat{\rho}_1 \le 4.$$
