Summer Seminar: Philosophy of Statistics

Empirical exercise
Consider the simple Normal model in table 1.

Table 1: The simple Normal model		
Statistical GM:		$X_t = \mu + u_t,\ t \in \mathbb{N} := (1, 2, ..., n, ...)$
[1]	Normal:	$X_t \backsim \mathsf{N}(.,.),\ x_t \in \mathbb{R},$
[2]	Constant mean:	$E(X_t) = \mu,\ \mu \in \mathbb{R},\ \forall t \in \mathbb{N},$
[3]	Constant variance:	$Var(X_t) = \sigma^2,\ \forall t \in \mathbb{N},$
[4]	Independence:	$\{X_t,\ t \in \mathbb{N}\}$-independent process.

Using the following data in Yule (1926) over the period 1866-1911:
y_t– the mortality rate in England,
x_t– the ratio of Church of England marriages to all marriage; data file attached.

(a) Estimate the simple Normal model for both data series separately and test the underlying assumptions using the following auxiliary regressions specified in terms of the residuals $\{u_t = x_t - \overline{x}_n,\ t = 1, 2, ..., n\}$ where $\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ is an estimate of μ:

$$\widehat{u}_t = \delta_0 + \overbrace{\delta_1 t + \delta_2 t^2}^{[2]} + \overbrace{\delta_3 x_{t-1}}^{[4]} + \varepsilon_{1t},\ H_0\colon\ \delta_1 = \delta_2 = \delta_3 = 0 \text{ vs. } H_1\colon\ \delta_1 \neq 0 \text{ or } \delta_2 \neq 0 \text{ or } \delta_3 \neq 0.$$

$$\widehat{u}_t^2 = \gamma_0 + \overbrace{\gamma_1 t + \gamma_2 t^2}^{[3]} + \overbrace{\gamma_3 x_{t-1}^2}^{[4]} + \varepsilon_{2t},\ H_0\colon\ \gamma_1 = \gamma_2 = \gamma_3 = 0 \text{ vs. } H_1\colon\ \gamma_1 \neq 0 \text{ or } \gamma_2 \neq 0 \text{ or } \gamma_3 \neq 0.$$

NOTE that the above choices of the various terms for the auxiliary regressions are only indicative of the direction of departure from the model assumptions!

(b) In light of your results in (a) explain which of the model assumptions is invalid for each of the two data series.

(c) In light of your conclusions in (b) respecify the original simple Normal for both data series to account for the statistical information that the original model did not account for.

Hint: explore models of the form:

$$X_t = \alpha_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + u_t.$$

(d) Choosing the best (in terms of statistical adequacy) respecified models from (c) use the residuals to evaluate the contemporaneous correlation coefficient:

$$Crr(X_t, Y_t) = \frac{E\left\{[(X_t - E(X_t|X_{t-1}, X_{t-2})][(Y_t - E(Y_t|Y_{t-1}, Y_{t-2})]\right\}}{\sqrt{[(X_t - E(X_t|X_{t-1}, X_{t-2})]^2 E[(Y_t - E(Y_t|Y_{t-1}, Y_{t-2})]^2}}.$$

(c) Compare and contrast your results in (d) with the estimated correlation coefficient based on:

$$\widehat{Corr(X_t, Y_t)} = \frac{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \overline{Y})(X_t - \overline{X})}{\sqrt{\left[\frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})^2\right]\left[\frac{1}{n}\sum_{t=1}^{n}(Y_t - \overline{Y})^2\right]}}$$

Comment on your results.