CO907: Assignment II

Alex Bishop

December 4, 2013

2.1 a) Fitting the data with a cubic polynomial

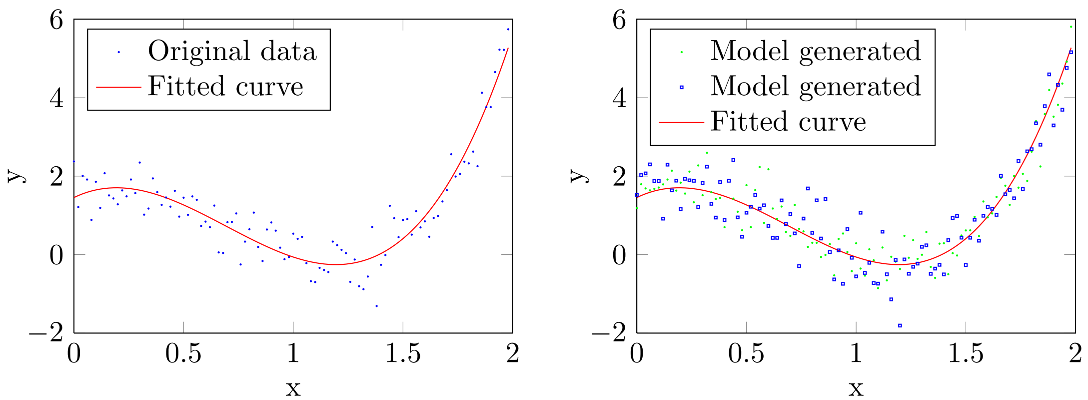


Figure 1: Fit for model $y = ax^3 + bx^2 + cx + d + \xi_t$ with $\widehat{a} = 3.92$, $\widehat{b} = -8.15$, $\widehat{c} = 2.72$, $\widehat{d} = 1.45$, $\xi_t \sim N(0, \widehat{\sigma}^2 = 0.20)$ and two datasets generated from the model demonstrating that the fit is plausible.

2.1 b) Model selection

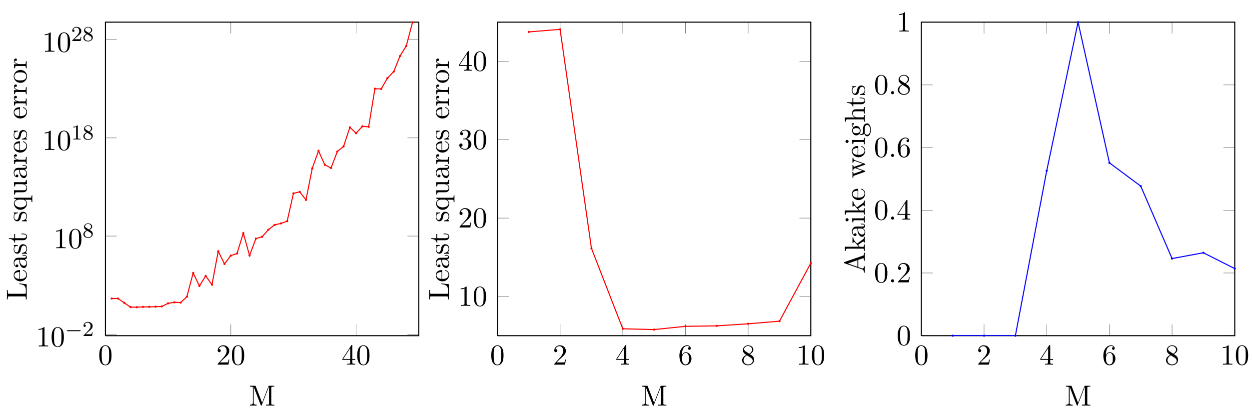


Figure 2: Plots of least squares error for cross validation with a train:test ratio of 50:50 for 1,000 uniformly sampled train/test sets. Also plotted are the Akaike Information Criterion weights for the dataset. Both model selection techniques indicate that the optimal value of M is 5, a quartic equation.

2.1 c) Detrending data

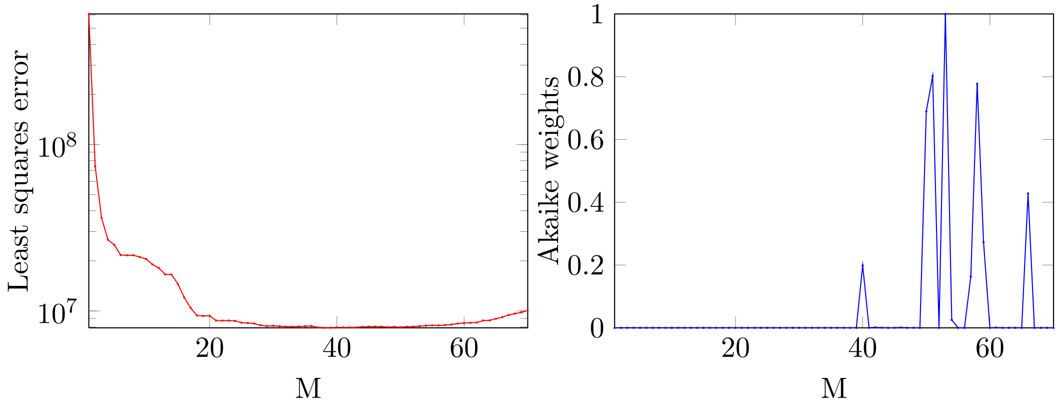


Figure 3: Plots of least squares error on the FTSE data for cross validation with a train:test ratio of 50:50 for 1,000 uniformly sampled train/test sets. Also plotted are the Akaike Information Criterion weights for the dataset. Cross validation suggests the best fit corresponds to M=38 where as the Akaike weights indicate that M=53 is optimal. There is a small peak in the Akaike weights around M=40 therefore assuming M=38 would be reasonable.

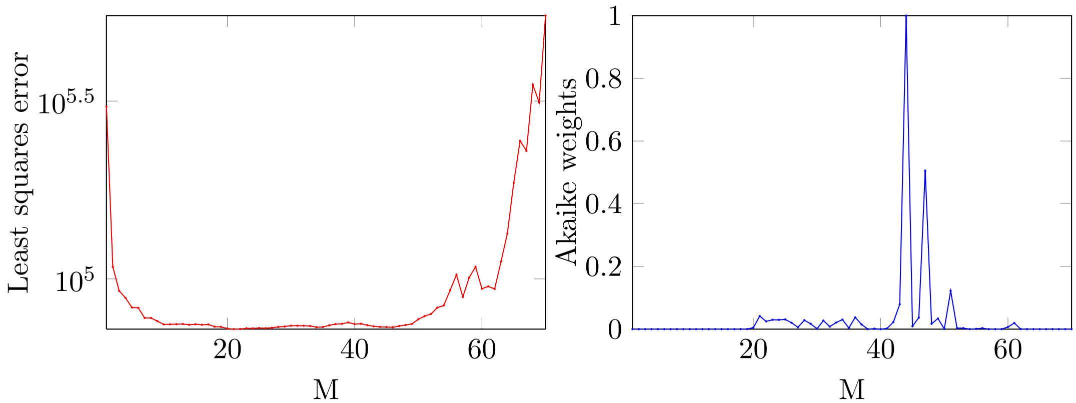


Figure 4: Plots of least squares error on the temperature anomaly data for cross validation with a train:test ratio of 50:50 for 1,000 uniformly sampled train/test sets. Also plotted are the Akaike Information Criterion weights for the dataset. Cross validation suggests the best fit corresponds to M=21 where as the Akaike weights indicate that M=44 is optimal. There are small peaks over a large M range of the Akaike weights, therefore it would be reasonable to assume M=21 to avoid large amounts of overfitting.

One must however consider that we are attempting to extract a general trend from the datasets therefore it can be argued that it is more sensible to take M as a much lower value corresponding to the initial plateaus in the LSE corresponding to M=6 for the FTSE and M=5 for the temperature anomalies.
This is further justified due to the fact that if we take a smaller train:test ratio of the dataset, such as 10:90, then there will be less noise to fit to which will correspond to an earlier minimum in the LSE values.

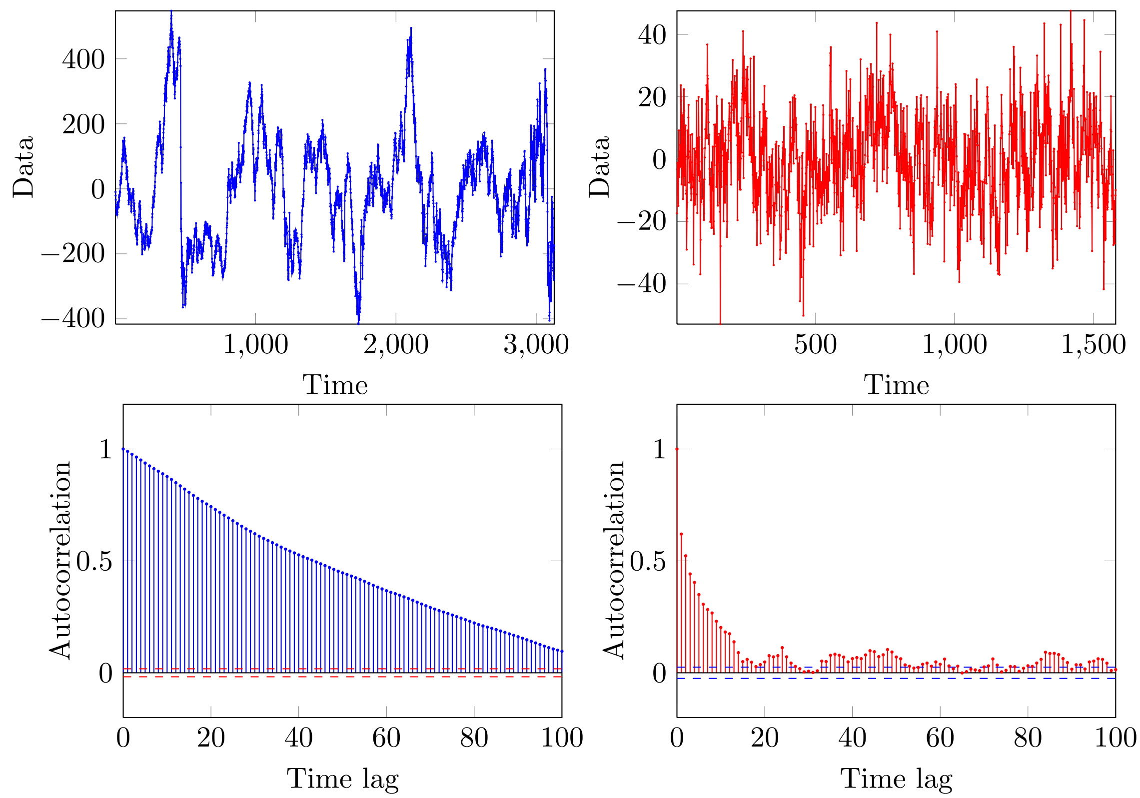
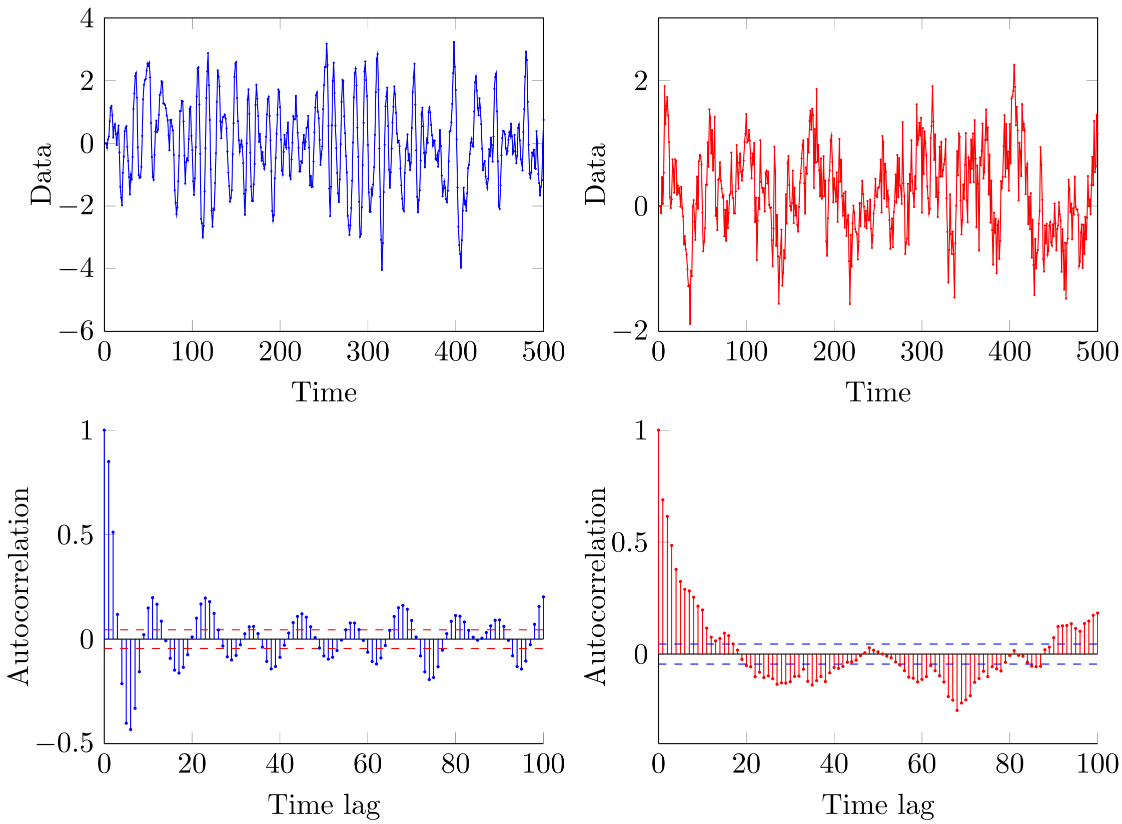


Figure 5: Detrended data and autocorrelation functions for the FTSE data (blue, M=6) & temperature anomaly data (red, M=5). At first glance both datasets do not look stationary, this is further supported by their autocorrelation plots, particularly for the FTSE data which descreases in a manner slightly indicative of an AR(q) model.

2.2 a) Generating AR(2) data



$c = 0,\ \phi_1 = 3/2,\ \phi_2 = -3/4,\ \sigma = 1/2$ $\qquad$ $c = 0,\ \phi_1 = 1/2,\ \phi_2 = 1/3,\ \sigma = 1/2$

Figure 6: Data generated from two AR(2) models and their corresponding autocorrelation functions. Both autocorrelation functions display some oscillatory behaviour (particularly the first dataset); however these oscillations seem to have a longer period for the second model, possibly because of its lower ϕ values.

2.2 b) Linear regression of AR(2) models

We can assume that the MLE takes the form of a gaussian due to the fact that the first two terms are fixed to zero and the fact that the stochastic term is gaussian, which implies that all terms are linear combinations of gaussian noise.

The LSE is given and calculated as follows

$$\begin{aligned} E(c, \underline{\phi}) = &\ \tfrac{1}{2} \textstyle\sum_{t=3}^{N} \left(X_t - c - \phi_1 X_{t-1} - \phi_2 X_{t-2}\right)^2 \\ = &\ \tfrac{1}{2} \textstyle\sum_{t=3}^{N} \left(X_t - \langle c,\, \phi_1,\, \phi_2 | \psi_0(t),\, \psi_1(t),\, \psi_2(t) \rangle\right)^2 \\ \langle \bigtriangledown |\, E(c, \underline{\phi}) = &\ \langle c, \underline{\phi} |\, \Phi^T \Phi - \langle \underline{X} |\, \Phi = \langle 0 | \end{aligned}$$

where

$$\begin{aligned} \psi_0(t) = &\ \ \psi_0 = \underline{1} \\ \psi_1(t) = &\ \ (X_2, ..., X_{N-1}) \\ \psi_2(t) = &\ \ (X_1, ..., X_{N-2}) \end{aligned}$$

Giving us the design matrix

$$\implies \Phi = \begin{pmatrix} 1 & X_2 & X_1 \\ \vdots & \vdots & \vdots \\ 1 & X_{N-1} & X_{N-2} \end{pmatrix}$$

$$\implies \Phi^T\Phi = \begin{pmatrix} N-2 & \sum_{t=1}^{N-2} X_{t+1} & \sum_{t=1}^{N-2} X_t \\ \sum_{t=1}^{N-2} X_{t+1} & \sum_{t=1}^{N-2} X_{t+1}^2 & \sum_{t=1}^{N-2} X_t X_{t+1} \\ \sum_{t=1}^{N-2} X_t & \sum_{t=1}^{N-2} X_t X_{t+1} & \sum_{t=1}^{N-2} X_t^2 \end{pmatrix}$$

$$\Phi^T \left|\underline{X}_N\right\rangle = \begin{pmatrix} 1 & \dots & 1 \\ X_2 & \dots & X_{N-1} \\ X_1 & \dots & X_{N-2} \end{pmatrix} \begin{pmatrix} X_3 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} \sum_{t=3}^{N} X_t \\ \sum_{t=3}^{N} X_t X_{t-1} \\ \sum_{t=3}^{N} X_t X_{t-2} \end{pmatrix}$$

$$\therefore \left|\widehat{c},\, \widehat{\phi}_1,\, \widehat{\phi}_2\right\rangle = (\Phi^T\Phi)^{-1}\Phi^T \left|\underline{X}_N\right\rangle = \begin{pmatrix} N-2 & \sum_{t=1}^{N-2} X_{t+1} & \sum_{t=1}^{N-2} X_t \\ \sum_{t=1}^{N-2} X_{t+1} & \sum_{t=1}^{N-2} X_{t+1}^2 & \sum_{t=1}^{N-2} X_t X_{t+1} \\ \sum_{t=1}^{N-2} X_t & \sum_{t=1}^{N-2} X_t X_{t+1} & \sum_{t=1}^{N-2} X_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=3}^{N} X_t \\ \sum_{t=3}^{N} X_t X_{t-1} \\ \sum_{t=3}^{N} X_t X_{t-2} \end{pmatrix}$$

2.2 c) Fitting parameters to the data

Numerically solving the above equations, and repeating over 1,000 generated datasets yields the following parameters for the two AR(2) models previously mentioned:
$\widehat{c} = 0.000 \pm 0.007,\ \widehat{\phi}_1 = 1.496 \pm 0.009,\ \widehat{\phi}_2 = -0.747 \pm 0.009,\ \widehat{\sigma} = 0.499 \pm 0.000$
$\widehat{c} = 0.000 \pm 0.001,\ \widehat{\phi}_1 = 0.496 \pm 0.001,\ \widehat{\phi}_2 = 0.327 \pm 0.001,\ \widehat{\sigma} = 0.498 \pm 0.000$

However, this would not be possible given we only had one dataset. An alternative method would be to use cross-validation with contiguous chunks of data (as uniformly sampling would destroy the correlations).

2.2 d) Real-world time-series as AR(2) models

It is highly unlikely that any of the real-world time-series studied could be reasonably modeled with an AR(2) process as this would be much too simple for modelling something as complicated as the FTSE.

Fitting an AR(2) process to the FTSE data (original and detrended with M=6 respectively) yields:
$\widehat{c} = 0.18,\ \widehat{\phi}_1 = 1.08,\ \widehat{\phi}_2 = -0.08,\ \widehat{\sigma} = 24.47$
$\widehat{c} = 0.04,\ \widehat{\phi}_1 = 1.08,\ \widehat{\phi}_2 = -0.09,\ \widehat{\sigma} = 24.38$

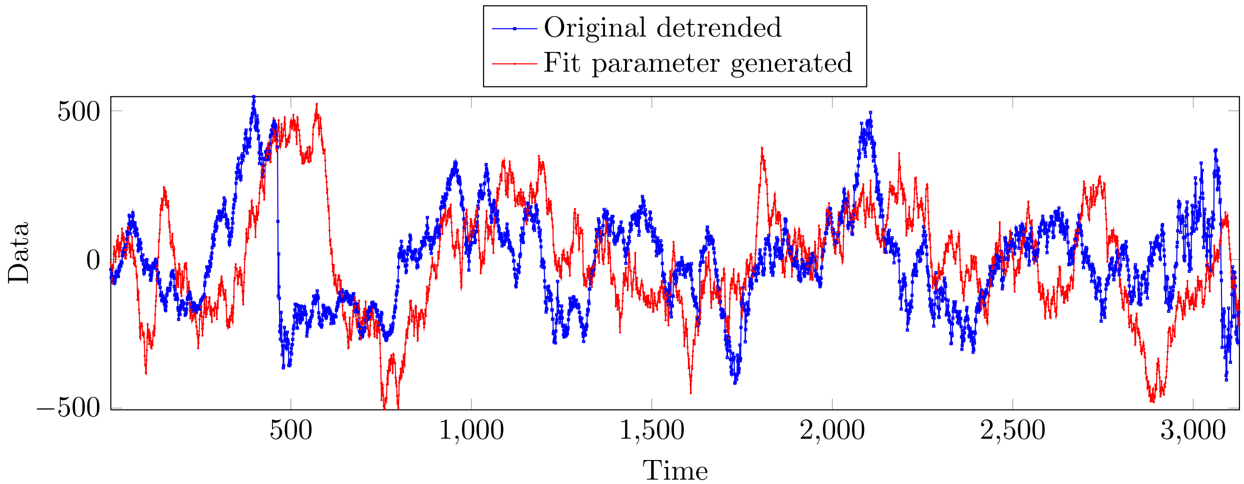


Figure 7: Detrended (with M=6) FTSE data plotted with AR(2) data generated with the linear regression fit parameters. The data looks qualitatively the same; however it is not quantitatively close enough to be able to model this time series with an AR(2) model.

2.3 a) Regression form

$$\begin{aligned}
A\cos(\omega t+\phi) = &\quad A\cos(\phi)\cos(\omega t) - A\sin(\phi)\sin(\omega t)\\
\implies B_1 = &\quad A\cos(\phi)\\
\implies B_2 = &\quad -A\sin(\phi)\\
\therefore \phi = \arctan(\tfrac{-B_2}{B_1}), &\quad A = \tfrac{B_1}{\cos(\phi)}
\end{aligned}$$

2.3 b) Linear regression on the data

$$\begin{aligned}
& B_1 = -0.63, \quad B_2 = -2.16\\
\implies & \phi = -1.29, \quad A = -2.25
\end{aligned}$$

2.3 c) Plotting the signal

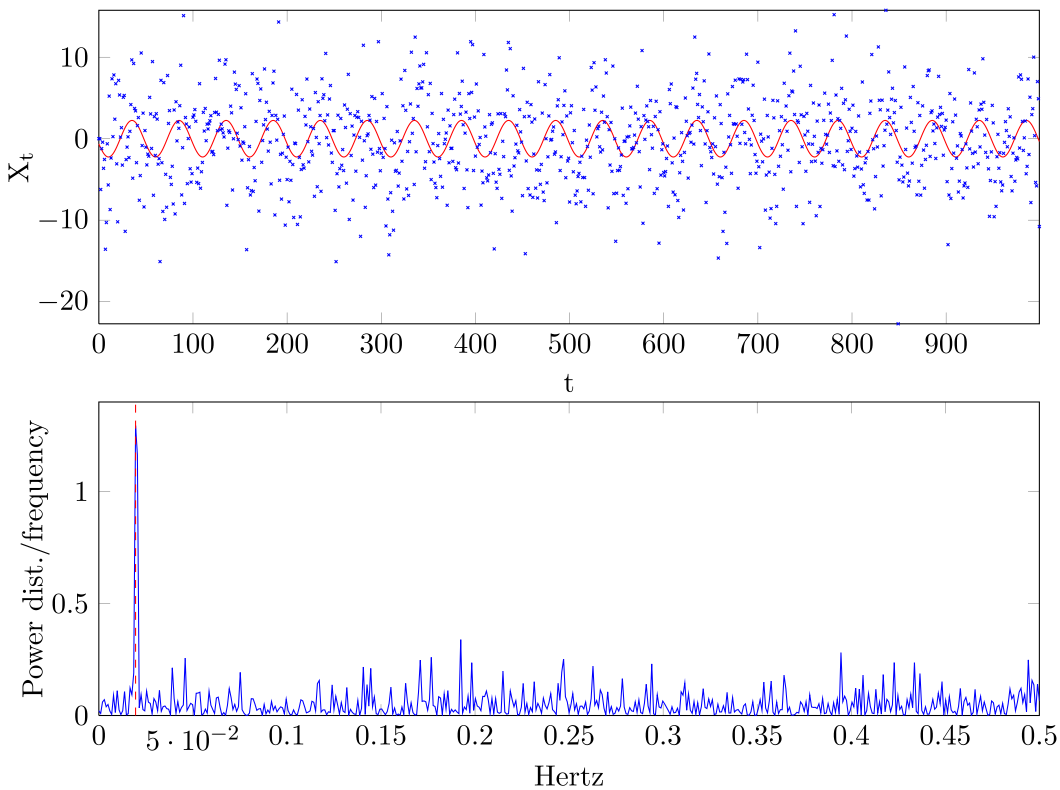


Figure 8: Signal in noise and periodogram showing that the frequency of oscillation is $\approx 0.020Hz$. The periodogram and linear regression method both suggest that the result is plausible and there is a signal in the noise. The phase (ϕ) is given by the value corresponding to the peak frequency.