Appendices
Appendix A

ARMA modeling details

Steps for fitting the appropriate statistical model to given data series

The steps to obtain a relation between dependent (eg. PDR) and independent (eg. speed) variable is as follows,

1. Create an object in R and enter the data series values.

2. Make that series as time series by setting the frequency = 1/value of variable on x-axis (speed)

3. Calculate the ACF (Autocorrelation function) of the series and plot it. The autocorrelation can be used for the following two purposes:
   
   - To detect non-randomness in data (it measures the linear predictability of the series)
   - To identify an appropriate time series model if the data are not random.

When the autocorrelation is used to identify an appropriate time series model, the autocorrelations are usually plotted for many lags

4. Calculate the PACF (partial autocorrelation function) and plot it. Partial autocorrelations are useful in identifying the order of an autoregressive model. The partial autocorrelation of an AR (p) process is zero at lag p+1 and greater. If the sample autocorrelation plot indicates that an AR model may be appropriate, then the sample partial autocorrelation plot is examined to help identify the order.

5. Comparing the results of the ACF and PACF of the series with the table 6.1, one can select order of the model. If it is difficult to get correct order and model type from ACF and PACF values then, model is to be selected using suitable criteria like Akaike’s information
criterion (AIC). To finalize the model for available data, it is desirable to have minimum value of AIC.

6. Use proper fitting model to data series and examine the results. At time of using any model give proper entries like object name (data series name), order of the model that best fits to given series and method of Estimation (use for estimating the coefficients).

7. Fitting summary gives Coefficients of the model and standard error.

8. Final relation (e.g PDF vs SPEED) can be obtained by putting the coefficient values in the model equation.

9. To predict the value at the future instance following options are possible-
   
   • ‘Predict’ function in R can be used or
   
   • Obtained relation between the parameters can also be used.

The prominent feature of R is its flexibility. R stores the analyzed result as ‘object’ so the user can extract only the part of the results which is of interest. R can display the estimated coefficients. R software can perform statistical or data analysis.
Appendix B

Terms used in multiple regression technique

Terms used in multiple regression techniques There are certain terms which allow us to understand the results of this statistical technique [78].

1. B-value (standardized regression coefficients)
   The B value is a measure of how strongly each predictor variable influences the criterion variable. The beta is measured in units of standard deviation. Higher the B value the greater the impact of the predictor variable on the criterion variable. The B regression coefficient is used to assess the strength of the relationship between each predictor variable to the criterion variable.

2. Standard Error of Estimate (SEE)
   The standard error of the estimate is a measure of the accuracy of predictions. It is the standard deviation of observed values, Y, around predicted values, \( \hat{Y} \). Recall that the regression line is the line that minimizes the sum of squared deviations of prediction (also called the sum of squares error). The standard error of the estimate is closely related to this quantity and is defined below:

\[
SEE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n - k - 1}} = \sqrt{\frac{SSE}{n - k - 1}} \quad (B.1)
\]

The terms used in the equation B.1 are,

- \( y_i \) = the sample value of the dependent variable,
- \( \hat{y}_i \) = corresponding value estimated from regression equation,
n = no. of observations,

k = no. of predictors or independent variables,

n-k-1 = degree of freedom (DF) for Standard error

3. Sum of Squares due to Error (SSE) is given by,

\[ SSE = \sum_{i=1}^{n} (e_i)^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \]  (B.2)

\[ = \sum_{i=1}^{n} (Y_i - b_0 - b_1X_{1i} - b_2X_{2i} - b_3X_{3i} - b_kX_{ki})^2 \]  (B.3)

4. Sum of Squares due to Regression (SSR) represented by,

\[ SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y}_i)^2 \]  (B.4)

5. Total Sum of Squares (SST) is given by equation B.5 and equation B.6.

\[ SST = \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2 \]  (B.5)

\[ SST = SSR + SSE \]  (B.6)

6. R is a measure of the correlation between the observed value and the predicted value of the criterion variable.

7. R Square, \( R^2 \), is the square of the measure of correlation and indicates the proportion of the variance in the criterion variable which is accounted for by our model. A high value of \( R^2 \), suggesting that the regression model explains the variation in the dependent variable well, is obviously important if one wishes to use the model for predictive or forecasting Purposes. The multiple coefficient of determination \( R^2 \) is a measure of how well the multiple regression equation fits the sample data.

\[ R^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2} = \frac{explained SS}{total SS} = \frac{SSR}{SST} \]  (B.7)

To compute \( R^2 \) as given in equation B.7, as ratio of SSR and SST, where, SS means sum of squared deviations.
In essence, this is a measure of how good a prediction of the criterion variable we can make by knowing the predictor variables.

However, $R^2$ tends to somewhat over-estimate the success of the model when applied to the real world.

8. An Adjusted $R^2$ value is calculated which takes into account the number of variables in the model and the number of observations (participants), our model is based on equation,

$$
Adjusted R^2 = 1 - \left( \frac{n - 1}{n - k - 1} \right) (1 - R^2) = 1 - \frac{SSE/(n - k - 1)}{SST/(n - 1)}
$$

(B.8)

(B.9)

This Adjusted $R^2$ value gives the most useful measure of the success of our model.

If, for example we have an Adjusted $R^2$ value of 0.75 we can say that our model has accounted for 75% of the variance in the criterion variable.

9. F value or F ratio: The F Value or F ratio is the test statistic used to decide whether the model as a whole has statistically significant predictive capability, that is, whether the regression SS is big enough, considering the number of variables needed to achieve it.

F is the ratio of the Model Mean Square (MSR) to the Error Mean Square (MSE) given by,

$$
F = \frac{MSR}{MSE}
$$

(B.10)

If we are dealing with k explanatory variables, X1 through Xk, then there will be k+1 coefficients, b0 through bk, and the degrees of freedom for each part of the partition are calculated as given by set of equations.

Total : $dfT = n - 1$ (no.of observs. − 1)  
Explained : $dfR = k$ (no.of coeff.s. − 1)  
Unexplained : $dfE = n - k - 1$ (no.of obs. − no.of coeff.s)  

(B.11)  
(B.12)  
(B.13)

It is now possible to construct overall F-test to check on the statistical significance of the regression model. Since an F statistic is defined as the ratio of two variables (or “mean
squares” as statisticians often call them), we have to convert “sums of squares” to “mean squares” as represented by,

\[
\text{Total} : \quad MST = \frac{SST}{dfT} \quad \text{(B.14)}
\]

\[
\text{Explained} : \quad MSR = \frac{SSR}{dfR} \quad \text{(B.15)}
\]

\[
\text{Unexplained} : \quad MSE = \frac{SSE}{dfE} \quad \text{(B.16)}
\]

The F statistic that tests the overall significance of the regression model is given by equation B.17.

\[
F = \frac{MSR}{MSE} = \frac{\sum(\hat{Y} - \bar{Y})^2/k}{\sum(Y - \bar{Y})^2/(n - k - 1)} \quad \text{(B.17)}
\]

Note that this F-test is sensitive to the relative strengths of the numerator and denominator. If the unexplained MS is large, then regression model is not doing well, and F becomes smaller. If the explained MS is large relative to the unexplained MS, then F becomes larger. Looking up an F table, we can make decision as to the significance of the regression model. If the ratio is much larger than one, then it is likely that MSR is estimating a larger quantity than is MSE and that the null hypothesis is false. Larger value of F is desirable.

10. P-value: The P-value is a measure of the overall significance of the multiple regression equation. Like the adjusted R square, this P-value is a good measure of how well the equation fits the sample data. The Small value of P indicates that the multiple regression equation has good overall significance and is usable for predictions.
Appendix C

Colored Petri Net details

More about Colored Petri nets Colored Petri Nets (CP-nets or CPNs) is a formal modeling language that is well suited for modeling and analyzing large and complex systems for several reasons: hierarchical models can be constructed, complex information can be represented in the token values and inscriptions of the models, timing information can be included in the models, and mature and well-tested tools exist for creating, simulating, and analyzing CPN models [79]. CP-nets are often used to model and verify the logical correctness of network protocols.

Colored Petri nets provide a framework for the design, specification, validation, and verification of systems. CP-nets have a wide range of application areas like communication protocols, operating systems, hardware designs, embedded systems, software system designs, and business process re-engineering. CPN is a graphical computer tool supporting the practical use of CP-nets [80]. The tool supports the construction, simulation, functional and performance analysis of CPN models. CPN models can be structured into a number of related modules. The module concept of CP-nets is based on a hierarchical structuring mechanism (either top down or bottom up approach).

CP-nets include a time concept which makes it possible to capture the time taken by different activities in the system. Timed CPN models and simulation can be used to analyze the performance of a system, by investigating Quality of Service (QoS) parameters like, delay, and throughput. It is possible to investigate the functional correctness of systems modeled by means of timed CP-nets. Abstract CPN models can be used in an early phase of system development to determine the boundaries of the project and specify requirements [81].

CP-nets have a sound, mathematically well founded execution semantics, are well-proven, and have proper tool support. The design and specification can be supported by modeling and sim-