Advice on approach to assessing forecast uncertainty

16th December 2013

pwc

Disclaimer

This document has been prepared for the Airports Commission in accordance with the terms of the Airports Commission Analysis and Strategy Support framework and the Contract Reference PPRO 04/08/72dated 2nd May 2013 and solely for the purpose and on the terms agreed with the Airports Commission. We accept no liability (including for negligence) to anyone else in connection with this document. This document contains information obtained or derived from a variety of sources as indicated within the document. PwC has not sought to establish the reliability of those sources or verified the information so provided. Accordingly no representation or warranty of any kind (whether express or implied) is given by PwC to any person (except to the Airports Commission under the relevant terms of our engagement) as to the accuracy or completeness of the document.

Scope

As part of PwC's support on analysis and strategy to the Airports Commission, we were asked to review the DfT's approach to assessing forecast uncertainty and advise on any areas for improvement. The AC, in partnership with the DfT, concluded that a Monte Carlo approach offered the best potential and has developed this solution to supplement the DfT's forecasting model.

To focus the review, the AC requested that PwC consider the following questions:

- is the methodology for identifying input distributions appropriate?
- is the stationarity test used robust?
- what are appropriate convergence criteria?
- Is the @Risk approach to handling correlation appropriate?

PwC was not asked to review alternative approaches for handling uncertainty, nor were we asked to comment on the credibility of the central inputs of the forecast model.

This report sets out the findings of PwC's review of the model "*FcModel-v14b_APF02_00_C50*" presented to PwC on 23rd May 2013, and the final methodology as set out in DfT's note "Developing Monte Carlo approach for National Passenger Demand Model" dated June 2013. The conclusions relate only to this model and documentation.

Table of contents

1. Executive summary

1.1. Background and aim of the report

The Airports Commission (AC) outlined in Discussion Paper 01: *Aviation Demand Forecasting* a number of areas that they would like to explore to enhance and supplement the existing DfT aviation forecasting model. One of these areas was the need for a more sophisticated approach to the treatment of uncertainty around a central forecast. The AC, in partnership with the DfT, concluded that a Monte Carlo approach offered the best potential and has developed this solution to supplement the DfT's forecasting model.

The AC have asked PwC to review their approach and advise on any areas for improvement.

To focus the review, the AC requested that PwC consider the following questions:

- is the methodology for identifying input distributions appropriate?
- is the stationarity test used robust?
- what are appropriate convergence criteria?
- Is the @Risk approach to handling correlation appropriate?

PwC was not asked to review alternative approaches for handling uncertainty, nor were we asked to comment on the credibility of the central inputs of the forecast model.

This report sets out the findings of PwC's review of the model "*FcModel-v14b_APF02_00_C50*" presented to PwC on 23rd May 2013, and the final methodology as set out in DfT's note "Developing Monte Carlo approach for National Passenger Demand Model" dated June 2013. The conclusions relate only to this model and documentation.

1.2. The DfT's approach for including uncertainty in the forecasting model

The DfT has chosen to adopt a Monte Carlo style simulation approach using an Excel add-on application called @Risk made by Palisade Corporation. The simulation works by repeatedly drawing observations for inputs from pre-defined distributions, and by inputting these observations into the model, in order to build up an implied distribution of passenger demand forecasts around the central forecast.

Prior to running the Monte Carlo simulation, the DfT conducted a number of preliminary statistical tests to assess:

(i) the stationarity of input variables;

(ii) the distribution of input variables; and

(iii) the correlation between the dependent variable, passenger demand, and the explanatory variables.

The Monte Carlo simulation was then run using @Risk with the automatic assessment set at a 95% confidence that the 75th percentile is within 3% of that estimated. We found that the model "*FcModel-v14b_APF02_00_C50*" converged within 600 iterations, i.e. additional iterations do not subtantially change the mean estimates.

1.3. The key issues and our conclusions

We have identified a number of issues that affect the robustness of the Monte Carlo simulation exercise. These elements and a summary of our conclusions are set out in Figure 1 below. Our conclusions are set out in more detail in the remainder of this report.

Figure 1

Key issues and why it is important	DfT's approach	Our conclusions
Inputs selection – to mitigate the issue of omitted variable bias	Use a correlation matrix in variable selection	This approach reduces subjectivity in variable selection and ensures that key variables are not left out of the model
Input stationarity – to mitigate problems of spurious regressions	Use Augmented Dickey-Fuller (ADF) test and considers drifts and trends under the null hypothesis of a unit root	The chosen stationarity test is robust and is a standard unit root test
Inputs distribution identification – @Risk requires the user to provide the distribution of the variables used in each model	Use Minitab to identify each variable's statistical distribution based on Anderson-Darling statistics (AD)	The AD test is one of the most powerful normality tests available
Convergence of Monte Carlo simulation – to ensure the stability of the simulation estimates	Number of iterations set to automatic	This is in line with @Risk best modeling practice. The latter ensures that the simulation converges to a stable output distribution
Choice of sampling technique; Monte Carlo or Latin Hypercube – to improve speed of convergence	Use Latin Hypercube sampling	Latin Hypercube sampling generates more efficient estimates of desired parameters than Monte Carlo sampling
Oil price floor and ceiling – to avoid using absurd oil prices in the simulation	Oil price floor based on long run marginal cost of extraction curve. Price ceiling informed by the price at which electric vehicles become viable	In our view, a simpler and relatively more objective alternative for the ceiling price would be to construct a price using the mean of the growth rate of oil prices plus two standard deviations as an approximate upper bound

2. Review of the DfT's approach to forecasting uncertainty

The DfT has adopted the Monte Carlo approach to analyse the uncertainty around the central forecasts provided by its National Aviation Passenger Demand Model (NAPDM). The Monte Carlo simulation is run using an Excel add-on application called @Risk made by Palisade Corporation.

Prior to running the Monte Carlo simulation, DfT conducted a number of preliminary statistical tests to assess: (i) the stationarity of input variables; (ii) the distribution of input variables; and (iii) the correlation of input variables with passenger demand. The Monte Carlo simulation was then run with @Risk automatic assessment set at 95% confidence that the 75^{th} percentile is within 3% of that estimated.

To reflect AC's principal concerns, our review of this approach covers four elements: variable identification, stationarity and distribution; defining an appropriate convergence criterion; deciding on whether to use the Latin Hypercube or Monte Carlo sampling approach; and oil prices considerations.

For each of these elements, we provide a brief description of the issues, the approach adopted by the DfT and conclude with our assessment of the robustness and suitability of the method used.

2.1. Variable identification, stationarity and distribution assessment

Figure 2 lists the explanatory variables used by the DfT to model and forecast air passenger demand. These variables are incorporated in the NAPDM and form the basis for the Monte Carlo simulation. These variables are selected using a correlation matrix between passenger demand (the dependent variable) and potential explanatory or input variables. This technique reduces the subjectivity of input variable selection and tracks the strength of the relationship between variables over time.

Figure 2 - NAPDM inputs

	Description of distribution
UK GDP	Distribution of historic growth rates
UK Consumer Expenditure	Distribution of historic growth rates
Foreign Region GDP	Distribution of historic growth rates
Oil Prices	Distribution of historic growth rates adjusted to reflect DECC's assessment of ceilings and floors for oil price levels
Market Maturity	Uniform distribution across existing judgement-based range
Econometric coefficients	Normal distribution with parameters informed by t-stats outputs
Carbon Prices	Assumption (normal distribution, assume DECC's projections cover the 90% confidence interval)
Video conferencing factor	Uniform distribution

@Risk requires users to specify a correlation matrix between the input variables as part of the Monte Carlo simulation. To ensure that this correlation matrix is robust, the DfT has adopted PwC's recommendation to override economically meaningless correlation values with zero, and to keep economically sensible varaibles intact even when they are statistically insignificant. The benefit of this approach is that it ensures internal consistency of the model through the exclusion of meaningless correlations and helps convergence of the simulation towards more plausible values.

Once the variables have been identified, the next step is to assess variable stationarity by conducting unit root tests. When testing for unit roots, it is important to specify the null and alternative hypotheses to account for the trend properties of the data.

The DfT has used the Augmented Dickey-Fuller (ADF) test to assess the stationarity of the key input variables in the NAPDM. Unit root tests tend to over-reject the null, increasing the risk of using a spurious regression.[1] To mitigate this problem, conventional practice is to perform the test under three null hypotheses: first, the process is assumed to be a random walk without drift, second a drift is included, and third a trend is included with or without a drift (see figure 3).[2] DfT's methodology is consistent with this approach and hence, in our view, is robust.

Figure 3 - Power of unit root test

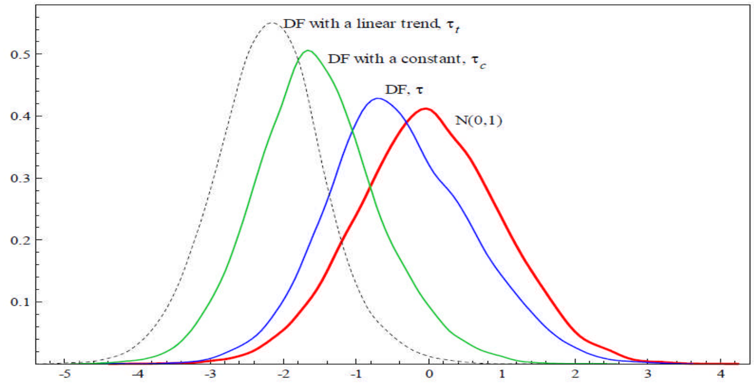


Figure 4 sets out the conclusions obtained by the DfT from these tests:

Figure 4 - Stationarity test results

Variable	Stationary?
Annual growth of UK GDP	Yes
Oil price	No
Annual growth of oil price	Yes
Growth of oil price 10 yr rolling average	No
Growth of oil price 5 yr rolling average	No
Growth of oil price 20 yr rolling average	No
Annual growth of foreign GDP WE	Yes
Annual growth of foreign GDP OECD	Yes
Annual growth of foreign GDP NIC	Yes

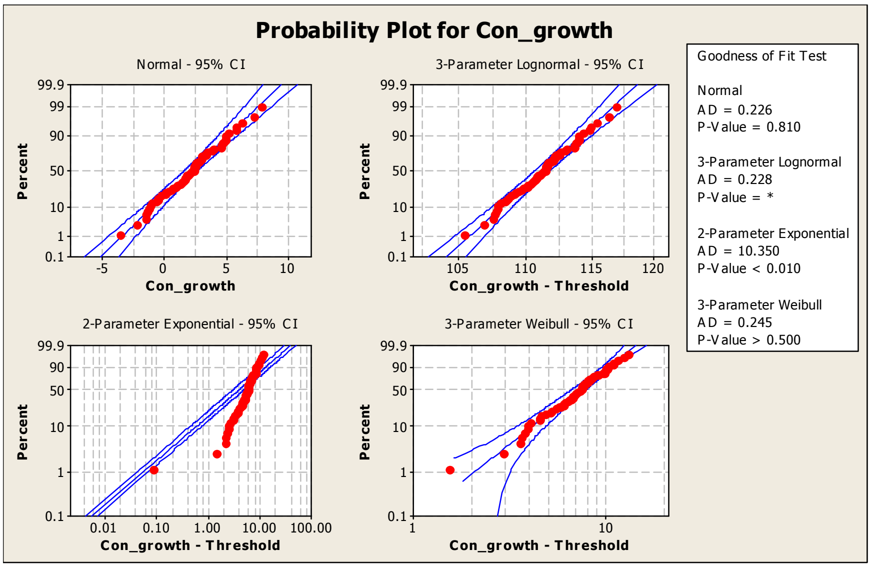
[1] Cochrane, John, H. (1991) *A critique of the application of Unit Root tests,* Journal of Economic Dynamics and Control 275 – 284. North-Holland

[2] See for example Rudebusch, G. D., (1993): *The Uncertain Unit Root in Real GDP,* American Economic Review 83, pp. 264- 272.

Annual growth of foreign GDP LDC	Yes
Annual growth of UK consumption	Yes

The final step before starting the Monte Carlo simulation is to identify the distribution of the input variables series that are used in the NAPDM. This is done by using MiniTab's distribution fitting function for each of the stationary input variables. Figure 5 provides an example of the probability plots for annual growth of UK consumer expenditure produced by the DfT using Minitab. In our view, this is a suitable approach.

Figure 5 - Distribution of consumer expenditure growth



The AD is a statistical test on whether a given sample of data is drawn from a given probability distribution, in our case, a normal distribution. The AD is also known as a Goodness of Fit (GoF) test as it assesses how well a given sample of n observations, $x_1, \ldots, x_n$, fits a specified distribution. Other well-known statistical goodness-of-fit tests include the chi-square (CS) test, Kolmogorov-Smirnov (KS) test, Cramer von Mises (CVM) test, and the probability plot correlation coefficient (PPCC) test. In his study on Normality Testing, Islam (2008) finds the AD to be the most powerful test when compared to other normality tests such as; Jarque-Bera, Shapiro-Francia, D'Agostino & Pearson and Anderson-Darling & Lilliefors. Desmoulins-Lebeault (2004) finds that the modified AD test (modified by Stephens, 1974) and the Shapiro-Wilks as the best omnibus test. Furthermore, the AD is found to perform well in small samples. Based on these evidence, we believe that choosing the AD test to assess the distribution of the the DfT model is appropriate.

An equally robust alternative to the DfT current input distribution identification is to use the distribution fitting companion of @Risk, BestFit. Using BestFit in this situation is not essential, but it would have some benefits:

- the number and form of distributions listed as options in Minitab (14 distributions) do not always match with those available in @Risk (38 distributions);
- Minitab can sometimes fail to return a distribution that may be available in @Risk; and
- BestFit allows the user to have control over the selection of the fit statistics.

2.2. Appropriate convergence criterion

The concept of convergence is used to test sampling methods. The sample is said to have converged when the output distribution is considered stable. An output distribution isconsidered stable when additional iterations do not markedly change the shape or statistics of the sampled distribution. The sample mean versus the true mean is the main measure of convergence, but skewness, percentile probabilities and other statistics can also be used.[3]

@Risk gives the user the option to set the number of iterations *or* to set the number of iterations to 'automatic' and instead specify a convergence tolerance and a confidence level. In this option, the simulation stops when the conditions have been met.

The DfT has used this latter option, with the the criteria set at *95% confidence that the 75th percentile is within 3% of that estimated.* We believe that this is the most appropriate option and is in line with @Risk modelling best practice.

2.3. Latin Hypercube or Monte Carlo?

@Risk provides the user with two sampling techniques: Monte Carlo and Latin Hypercube sampling.

Latin Hypercube sampling is designed to recreate the input distribution with fewer iterations than the Monte Carlo method. Latin Hypercube sampling is "sampling without replacement" where the number of stratifications of the cumulative distribution is equal to the number of iterations performed. With Latin Hypercube, the samples more accurately reflect the distribution of values of the input probability distribution. Latin Hypercube sampling converges faster on the true distributions than does Monte Carlo sampling.[4]

Latin Hypercube also aids the analysis of situations where low probability outcomes are represented in input probability distributions. By forcing the sampling of the simulation to include the outlying events, Latin Hypercube sampling assures that they are accurately represented in the simulation outputs.

Figure 6 - Monte Carlo

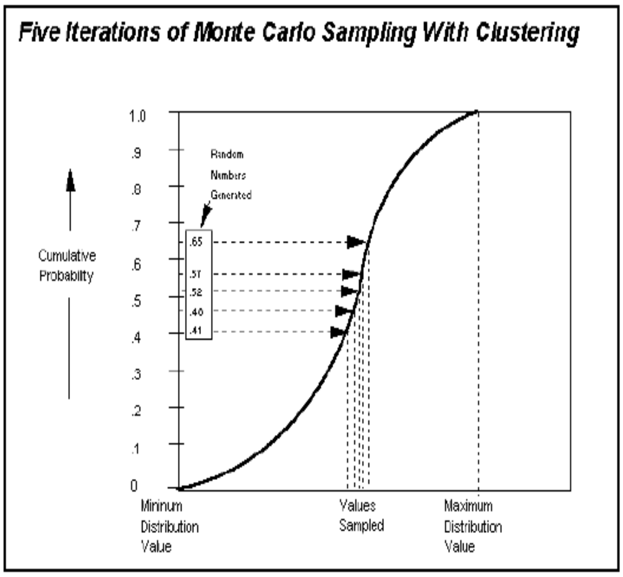
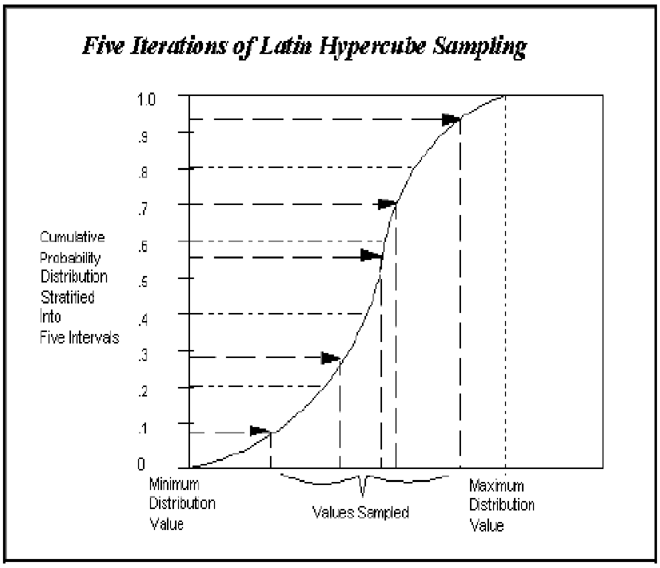


Figure 7 - Latin Hypercube



[3] Source @Risk

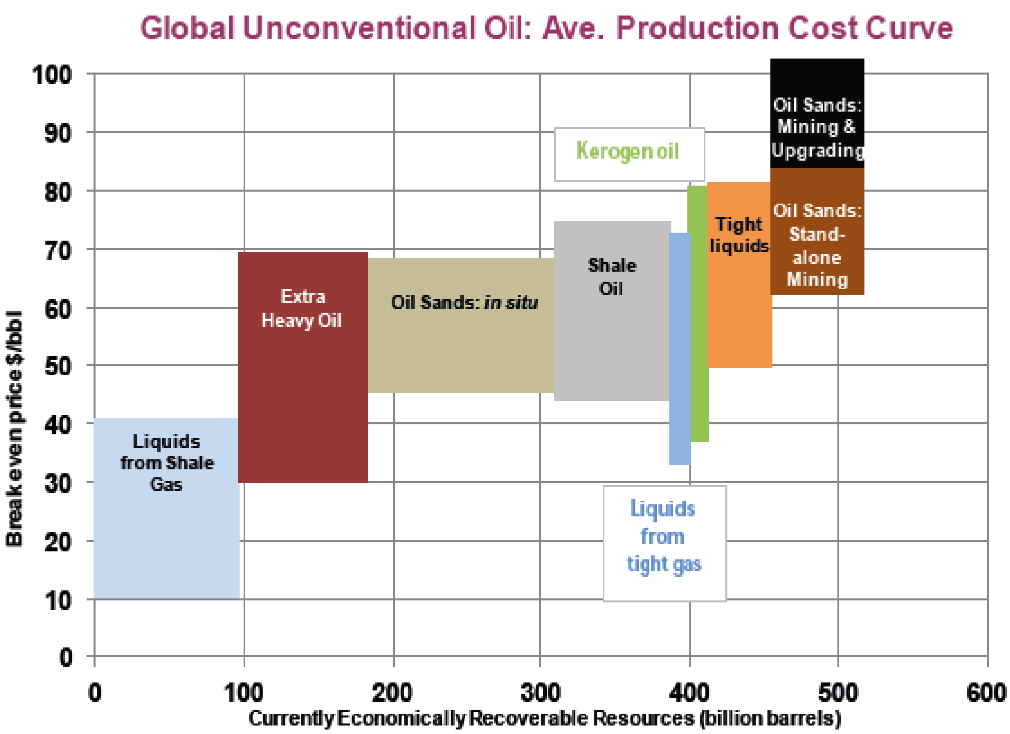
[4] Damblin, M. Couplet and Looss, B, G. D., (2013): *Numerical studies of space filling designs: optimization of Latin Hypercube Samples and subprojection properties,* Submitted to the Journal of Simulation.

Our review of the statistical literature confirms that the Latin Hypercube sampling generates more efficient estimates of desired parameters than simple Monte Carlo sampling (see for example Stocki, R. (2005), Stein, M. (1987), and Carnell, R. (2013)). although experts disagree on which method is more accurate. We concur with the DfT's decision to use Latin Hypercube as its default setting in the NAPDM.

2.4. Oil price floors and ceilings

The DfT wants to impose an oil price floor and ceiling in the sampling process to remove the potential for including absurdly high or low oil prices in the simulation exercise. The DfT has based its oil price floor on the IEA's long run marginal cost of extraction curve (see Figure 8 below), whilst the price ceiling was informed by the price at which electric vehicles become viable.

Figure 8



Source: IEA analysis of Rystad Energy data. Rystad develops estimates based on bottom up analysis of global fields, licenses, and potentially recoverable resources given currently available technology and activity levels. All resource values depicted in the graph are cumulative expected production from 2012 until 2100, excluding already produced oil through 2011. Oil and field condensate only, not natural gas plant liquids. Note that for oil sands development costs CERI, Alberta ERCB, and NEB are used.

Source: IEA MTOMR 2012 (note these are 2012 prices)

The DfT ssumed a uniform distribution between the floor and ceiling prices to account for their uncertainty. The Monte Carlo/LHC simulation is then run using the historic distribution of oil price growth rates with the above constraints.

Whilst in principle we do not think that there are hard floors or ceilings for oil prices, nevertheless it is reasonable to consider what would be plausible high and low oil price scenarios that might encompass some percentage of the probability distribution of possible outcomes over the period of interest.

Whilst using a cost curve allows the DfT to draw objective conclusions regarding the oil price floor, it is more difficult to determine the oil price ceiling objectively, especially given the inelasticity of oil prices to demand. In our view, a simpler and relatively more objective alternative to the price of electric vehicles would be to construct a floor and ceiling price using use the mean of the growth rate of oil prices plus two standard deviations as an approximate upper bound. While this solution is necessarily ad hoc, it has the benefit of having some precedents in addressing this types of issue.