An Assessment of Virtual Integration for Passenger Rail Services in Great Britain

Sultan Alsaedi [a], John M. Preston [a]
[a] Transportation Research Group, Faculty of Engineering and Physical Sciences
University of Southampton, Southampton Boldrewood Innovation Campus, University of Southampton, Burgess Road, Southampton SO16 7QF, UK

Abstract

Vertical integration was introduced in the British railway system in the form of a virtual alliance between Network Rail (NR) and South West Trains (SWT). The introduction of this alliance in 2012 was due to the Rail Value for Money Study that was published by McNulty in 2011. However, this alliance was ended in 2015, which was two years earlier than initially agreed by the Department for Transport (DfT). This paper aims to investigate whether the performance quality, in terms of punctuality and reliability, was a reason to end this alliance. The investigation is based on a comparison of the performance quality of SWT with other comparable Train Operating Companies (TOCs), which are Govia Thameslink Railway (GTR) and Southeastern (SE). Furthermore, the measurements of the Public Performance Measures (PPM) and Cancellation and Significant Lateness (CaSL) of these TOCs were used to deliver the comparisons. As a result, the investigation indicated that punctuality and reliability are not influenced by whether the organisation is vertically separated or virtually integrated. Overall, the virtual integration in this case does not seem to have had an impact, on the overall performance quality of passenger rail services provided by SWT.

Keywords
Capacity Utilisation, Punctuality, Reliability, Vertical Separation, Virtual Integration.

Glossary
CaSL: Cancellation and Significant Lateness.
GTR: Govia Thameslink Railway.
IM: Infrastructure Manager.
NR: Network Rail.
ORR: Office of Rail and Road.
PPM: Public Performance Measure.
RU: Railway Undertaking
SE: Southeastern.
SWT: South West Trains.
TOC: Train Operating Company.

1 Introduction

Punctuality and reliability are important factors to measure the performance of the railway system in terms of quality of service and passenger satisfaction (Carey, 1999; Goverde, 2005; Yuan, 2006). These indicators may deteriorate when the railway network is more extensively utilised to accommodate the growth in demand rather than extending or upgrading the track network (Yuan, 2006; Yuan and Hansen, 2007). There are several methods that are used to optimise the performance by designing a robust timetable or control strategy (Carey, 1998; Parbo et al., 2016). Optimal scheduling techniques, for example, can be implemented in the railway system in order to plan, operate and manage passenger train services, and these techniques, for example, can help to maintain the conflict between trains that operate on a single track (Ferreira and Higgins, 1996). Despite the development of scheduling techniques, there is a lack of these methods to mitigate the impact of delay on the performance (Carey and Carville, 2000).

The British rail system was vertically separated as a result of the 1993 Railways Act, with the key distinction being between the Infrastructure Manager (IM), since 2002 Network Rail (NR), and the Railway Undertakings (RUs), known as Train Operating Companies (TOCs). However, there have been long standing concerns about the weak alignment of incentives between the RUs and the IM caused by this vertical separation (Preston, 2002). This issue was revisited by the Rail Value for Money Study chaired by McNulty (2011). Partly as a result, an experiment was conducted between the Wessex Route of NR and the dominant TOC on the route, South West Trains (SWT), in which a form of virtual integration was introduced. The key features of this deep alliance were a single senior management team responsible for trains and track and the joint operations of the Waterloo control centre. This alliance was approved by the Department for Transport (DfT) and the then Office of Rail Regulation (ORR), as regulatory bodies, for a period of five years starting in 2012. In 2015, the DfT announced the end of the virtual alliance two years earlier than scheduled. Given this background, this paper aims to investigate whether changes in punctuality and reliability were reasons to end the virtual alliance between NR and SWT. Therefore, the paper is structured as follows. Section 2 provides a brief description of punctuality and reliability and their causes. In addition, the change in the railway organisation due to the reform is outlined briefly in this section. Section 3 illustrates the research methods that are used to achieve the aim. Section 4 contains a discussion of the results obtained by the three methods. Section 5 draws the final conclusion and key findings.

2 Literature Review

2.1 Punctuality and Reliability: definitions and causes

Punctuality and reliability have various definitions according to different literature. Punctuality is usually related to the running time with respect to an acceptable deviation from the designed timetable, which means a train is considered as punctual if this train runs within the accepted deviation (Olsson and Haugland, 2004; Preston et al., 2009). Punctuality is often described as the proportion of the trains arriving at, passing or departing from a point with a delay lower than a particular time, usually in minutes (Veiseth et al., 2007; Yuan, 2008). The deviation to determine the punctuality of trains varies between railway systems. In Great Britain, the amount of deviation to determine the punctuality relies on the journey length of the service; a train is described as punctual if it arrives at its final destination within five minutes of the timetabled arrival, but the deviation is increased to ten minutes for long distance services (ORR, 2016; Preston et al., 2009). For Switzerland and the Netherlands, a train can be described as punctual if it arrives within four minutes and three minutes respectively (Yuan, 2008). Reliability, on the other hand, is often implemented to illustrate the ratio of trains that have been cancelled (Preston et al., 2009). By contrast, Barron et al. (2013) define the reliability as the predictability of given travel time being experienced by a passenger and the degree of variation around the average travel time. According to Vromans

(2005), the reliability of the railway system depends on whether the trains are operated according to the scheduled timetable.

There is a substantial correlation between the punctuality and reliability indices with the overall delay in train operations. According to Yuan (2008), train delays are classified into three categories. Firstly, the initial delay is recorded when a train crosses the boundary of the investigated network later than timetabled (Yuan, 2008). Secondly, the original delay is the delay caused in the network due to operating trains at a lower speed compared to the scheduled speed, technical faults in the network, excess passenger boarding time and weather conditions (Yuan, 2008). The third category is knock-on delays, and this term is used to describe the delay that is transmitted between trains in the network (Yuan, 2008). When a train is delayed, the other trains that operate on the same route will be delayed (Parbo et al., 2016; Yaghini et al., 2013; Yuan, 2008). A related classification of delays is primary and secondary delays. The primary delay is the direct impact of several factors on the train itself, while the transmitted delays between trains are called secondary delays (Preston et al., 2009; Veiseth et al., 2007; Yuan and Hansen, 2007). According to Preston et al. (2009), the primary delays contributes to 40% of performance delay in the UK, while the remaining 60% is caused by secondary causes. This indicates that the initial and original delay, as stated above, are considered as primary delays, whereas the knock-on delay is defined as a secondary delay. According to Xia et al. (2013), there is a significant impact of bad weather conditions, such as high levels of wind, temperature, humidity and rainfall, on the rail performance, and this impact can lead to a significant delay in train operations. These reasons could have an impact on the train operation such as running the trains at lower speed or derailment.

To ensure punctuality, some aspects should be taken into consideration to design a railway timetable, as stated by Goverde and Hansen (2013). Infrastructure occupation should be considered with respect to three factors that can have an impact on capacity consumption, which are average train speed, number of trains and heterogeneity. These factors have an impact on the headway between trains. For example, the headway is influenced when trains run on the network at different speed levels; fast trains require larger headway due to longer breaking distance. Other aspects are that the timetable should be feasible and robust. Timetable feasibility means the ability of all planned trains to adhere to their scheduled routes. This can eliminate the conflicts between trains, which allow trains to run smoothly without braking. Conflict-free routes can be achieved when the process time of a train exceeds the scheduled time. On the other hand, a timetable can achieve robustness when it is capable of resisting design errors, parameter variations and changing operational conditions. For example, a process time for a scheduled train is calculated with basic parameters based on an estimation made by experts or determined by different methods. The robustness can absorb the design errors of this estimation when the estimated values are slightly different compared to the real values.

With regards to the railway system in Britain, monitoring the performance quality is highly dependent on two indicators, the Public Performance Measures (PPM) and Cancellation and Significant Lateness (CaSL). PPM is an indicator to measure the performance of train operations for passenger services in order to evaluate both the reliability and punctuality of the service (ORR, 2016). This indicator has two categories based on the journey length of the service to describe the status of the train if it is late or not (ORR, 2016). The first category is designed for regional operators, including London and South East operators; if a train arrives at its destination within five minutes compared to the timetable, this train will be considered as on time (ORR, 2016). The second category is designed for long distance services; a train will be defined as on time when it arrives at its destination within ten minutes compared with the designed timetable (ORR, 2016). CaSL, on the other hand, is the proportion of passenger trains that have been cancelled or arrived at the last destination more than 30 minutes late compared with the designed timetable (ORR, 2016).

A National Task Force (NTF) sub-group proposed new performance metrics to replace PPM and CaSL (NR and ORR, 2017), and a brief description of these metrics is given here. Firstly, Total Passenger Lateness is an indicator to measure the total of time lost for passengers in million hours. This metric focuses on passenger rail serviced by TOCs. Secondly, 'Reliability – cancellations and severe disruption' is a metric to describe the proportion of planned trains that did not serve the full journey or skipped some planned station stops. Moreover, there is a cancellation weight for each train depending on if a train is cancelled fully or partially. This indicator aims to describe a pure reliability of rail services by excluding significant lateness compared to CaSL. Thirdly, 'On Time and Time to 15' metrics are used to describe planned trains that arrive at all recorded stations less than one minute (within 59 seconds) and 15 minutes (within 14 minutes and 59 seconds) respectively. These metrics aim to provide a better explanation of punctuality of rail services.

2.1.1 Previous Studies on Punctuality and Reliability

There are numerous studies to investigate the factors that influence the punctuality and reliability in the railway performance. Each study attempts to determine the factors that have a significant impact on passenger train services in terms of punctuality and reliability, and various methods and models were used for analyses based on the characteristics of variables and collected data (Vromans, 2005).

The first study to consider is the research led by Harris (1992). The purpose of this study was to study the punctuality of railway performance in the UK and Netherlands by selecting different factors. The factors that were considered are the number of previous stops, the length of the train, distance covered, the age of motive power unit and track occupation. The methodology that was used by Harris for analysing was least-squares multiple linear regression. As a result of Harris's analysis, the factors that influenced the determination of the punctuality were the train length and the covered distance.

The second study to consider is the research led by Olsson and Haugland (2004). The purpose of the study was to determine the factors that influence the punctuality on passenger train services in Norway. The factors considered in this study were passenger number, train capacity rate (passenger per seats), the usage of infrastructure capacity, cancellations, the construction work of the network, a temporary decrease in speed, the punctuality of departure and arrival and operational priority rules. As a result, it was found that the punctuality was influenced significantly by the determination of the usage of infrastructure capacity based on the timetable.

The Swedish National Audit Office led a study to investigate the factors influencing the punctuality and reliability between 1976 and 1986. The study found about 50% of the delayed trains were caused by rainfall, temperature and patronage levels (Olsson and Haugland, 2004; Preston et al., 2009).With respect to the latter, if the level of patronage increased by 10% for a month, the punctuality dropped by about 6% on Sundays and about 10% on weekdays, especially on Fridays up to 14% (Olsson and Haugland, 2004; Preston et al., 2009). With respect to weather, the punctuality declined by around 5% as the average temperature decreased by one centigrade below -5C in one month (Olsson and Haugland, 2004; Preston et al., 2009).

The relationship between capacity utilisation (CUI) and congestion-related reactionary delay (CRRD) has been investigated. Armstrong and Preston (2017), for example, delivered research aimed to assess the relationships between capacity utilisation and rail performance, particularly at junctions and stations. The key finding is that there some consistency between CUI and CRRD. The amount of delay escalates due to the increase in the level of capacity utilisation.

2.2 Railway Reform

Privatising the railway system was implemented in order to achieve certain aims. Preston (1996) listed the goals and aims of privatising the railway system, which are to maximise the use the of the railway system; to provide a better satisfaction to the rail users; to improve the performance quality of the railway system; and maximise the net economic benefits of the railway system. Another aim of the rail privatisation that was mentioned by Knowles (2013) is to provide a competitive market for the private sector by limiting the role of the governments in order to improve the performance efficiency and to provide better benefits to the rail users. As reliability and punctuality are used to measure the performance quality of the rail services in terms of the customer satisfaction (Goverde and Meng, 2011), privatising the railway system could provide more reliable and punctual rail services to the rail users.

Railway organisation has changed as a result of liberalising reforms. For instance, Amaral and Thiebaud (2015) illustrated the four types of organisation that have emerged in Europe. The first type is a fully vertically separated organisation, which means that the IM is separated fully from the RUs. The second type is vertically separated organisation with a delegation, which means that the IM and RU are separated, but the RU is responsible for at least some of the IM tasks. The third type is a vertically separated organisation within a holding company, which means that the IM and RU are separated, but both are owned by one holding company. The fourth type is a vertically integrated organisation, which means the IM and RU are managed and operated by one company. However, a new form of railway organisation that has been experimented with in Britain is virtual integration, which retains separation of the IM and RU but encourages joint working, particularly at the operational level.

According to Mizutani et al. (2015), the purpose of the variety of organisation in the railway system is to provide a competitive market for all parties that are involved in the rail market. In Europe, for example, the successive legislations originating with Directive 91/440 require at least an accounting separation between the IM and RU in order to provide a competitive rail market. Furthermore, the separated organisations generate two forms of competition in the rail market, which are open access competition for freight services and competitive tendering for domestic passenger services. However, there is a concern about performance efficiency when the organisation of the railway system is vertically separated or integrated. The concern is that the performance efficiency can deteriorate due to the transaction costs between the infrastructure and train operators and reduced incentives for efficiency and for appropriate investment by the IM (Drew and Nash, 2011).

2.2.1 Previous Studies on Railway Organisation

There are several studies that have attempted to investigate the impact of the organisation forms on the railway system. For example, research published by Merkert et al. (2010) shows that the impact of the vertical separation was not significant on the performance measurements. This research was based on the assessment of the performance efficiency measurements for a cross-section of countries, but the vertical separation may not be the main factor to measure performance efficiency. Similarly, Wetzel (2008) concluded that the performance efficiency is not influenced significantly by vertical separation. However, the cost of rail systems does seem to vary between vertically separated or integrated organisations with regard to train density. Research by Mizutani et al. (2015) concluded that vertically integrated organisations are more beneficial in terms of unit costs at high levels of train density, while the vertically separated organisations are more beneficial at low levels of train density, as shown in Figure 1.

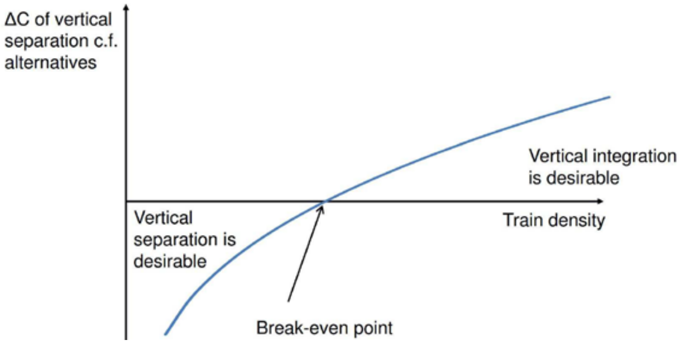

Figure 1: The effect of train density on the cost of different organisations (Source: (Mizutani et al., 2015))

2.3 The British situation

Since the British railway was reformed and privatised, demand levels and rail performance quality have changed dramatically. Figure 2 shows that passenger rail demand as measured by passenger-kilometres has increased substantially since the British railway was reformed (Merkert, 2005). This indicates that the rail performance required more attention to accommodate the increase of the demand, especially for the efficiency of the performance quality.

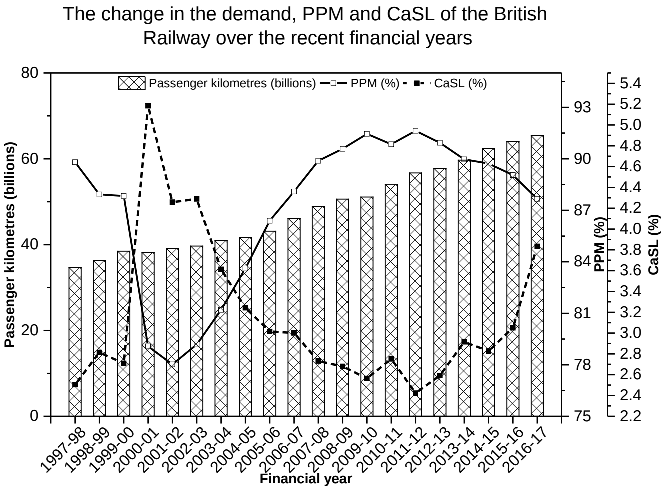

Figure 2: The change in the demand on the British Rail (Source: (ORR, 2017))

With regard to the performance quality, the change in the PPM and CaSL indicators is shown in Figure 2. According to this figure, there was a substantial adverse change in the PPM and CaSL measurements following the Hatfield accident between 2000-01 and 2002-03. This was exacerbated due to the sharp increase of the rail traffic in the rail network and the failure of Railtrack to maintain the track sufficiently (Drew and Ludewig, 2011). As a result, the railway industry in Britain was reformed by replacing Railtrack by NR in 2002 (Drew and Ludewig, 2011) and there was some recovery in performance.

However, the Rail Value for Money Study (McNulty, 2011) remained concerned about the misalignment of incentives. As a result a new scheme was proposed in the form of deep-alliances NR and the TOCs (Thompson, 2013). This would include the joint operations of control centres, with the purpose being to enhance rail performance and minimise the cost, which should improve customer satisfaction.

3 Methodology

The methodology of this research is based on data analysis techniques to investigate the impact of virtual integration on performance quality. The investigation is a comparison between the performance of SWT with similar operators, SE (Southeastern) and GTR (Govia Thameslink Railway). The process of the investigation is divided into three approaches. The first approach is an assessment of the change in the performance quality of SWT to evaluate the changes in the measurements of PPM and CaSL. The second approach is the organisation effect analysis to assess the effect of virtual integration on performance quality. The last approach is a prediction assessment of the performance quality of SWT to deliver a comparison between the actual measurements of PPM and CaSL with the predicted measurements. The data that is used for this investigation is collected from ORR.

3.1 Change Assessment

In this analysis, different approaches of comparisons are considered. The first approach is an individual comparison for SWT, which means that the PPM and CaSL measurements are used to assess the change in the performance quality performance prior, during and post the alliance for the mainline and suburban routes. For example, when the PPM measurements are considered, the proportion of trains arrived on-time of each route during the prior-alliance period is compared with the proportion during the alliance period and with the post-alliance period, and Table 1 outlines the start and end of prior, during and post-alliance periods as financial years and periods. The process is repeated similarly for the CaSL measurements.

Table 1: The allocation of the investigated periods in the financial years and periods.

Prior-alliance period		**Alliance period**		**Post-alliance period**	
Start	**End**	**Start**	**End**	**Start**	**End**
2010/11 Period 01	2011/12 Period 13	2012/13 Period 01	2015/16 Period 05	2015/16 Period 06	2018/19 Period 06

The second approach is ‘the cross comparison’, which means that the change in the performance quality of SWT is compared to the change of GTR and SE for each comparable route, and this comparison consists of two scenarios. Firstly, the change in the measurements of the PPM and CaSL is considered to compare the change of the SWT performance quality with the other TOCs. For example, the proportion of trains that arrived on-time for SWT during the prior-alliance period is compared with the same proportion of each comparable route of GTR and SE during the same period. The process is repeated similarly for the remaining periods and for the CaSL measurements.

Secondly, the amount of change in the quality performance is assessed between the TOCs through prior, during and post the alliance. Moreover, the average change in the measurements of the PPM and CaSL indicators for SWT from the prior-alliance period to the alliance period and from the alliance period to post-alliance period are examined. In addition, these changes are compared with the average changes in the measurements of GTR and with SE. However, the average value of the PPM and CaSL indicators will be calculated based on the periods in the three investigated periods separately for each TOC. The change of each indicator then is obtained based on the following equations:

$$\mathrm{PI}_{\mathrm{C}_i} = \mathrm{PI}_{\mathrm{P}_i} - \mathrm{PI}_{\mathrm{A}_i} \tag{1}$$

Where:

$\mathrm{PI}_{\mathrm{Ci}}$ = The change of the average value of the performance indicator, whether PPM or CaSL, from the prior to the alliance period.
$\mathrm{PI}_{\mathrm{Pi}}$ = The average value of the performance indicator in the prior-alliance period.
$\mathrm{PI}_{\mathrm{Ai}}$ = The average value of the performance indicator in the alliance period.
i = Period.

$$\mathrm{PI}_{\mathrm{D}_i} = \mathrm{PI}_{\mathrm{A}_i} - \mathrm{PI}_{\mathrm{E}_i} \tag{2}$$

Where:

$\mathrm{PI}_{\mathrm{Di}}$ = The change of the average value of the performance indicator, whether PPM or CaSL, from the alliance to post-alliance period.
$\mathrm{PI}_{\mathrm{Ai}}$ = The average value of the performance indicator in the alliance period.
$\mathrm{PI}_{\mathrm{Ei}}$ = The average value of the performance indicator in the post-alliance period.
i = Period.

The test that is used for these comparisons is the hypothesis test of two sample proportions as the provided data is expressed in proportions (Johnson, 2001).

3.2 Effect Analysis

The effect of organisation forms on performance quality is assessed based on regression analysis. Moreover, different regression models are created for each performance quality indicator. These models contain numerical and categorical variables. The numerical variable is the average rolling stock age (RSA). The categorical variables are the financial years (FY), periods (P), route types (RT) and organisation forms (OF). As these variables cannot be implemented directly into the regression analysis, each group will be recoded as 1 and 0 through dummy coding, which is an approach to recode the categorical data to be applicable to use in the regression model (Fox, 2015, p. 118). The process of the dummy coding involves excluding one variable of each category to be set as a reference of this category and recoding the remaining variables. The dummy coding for the four categories are the following:

$$\mathrm{TOC} = \begin{cases} \mathrm{D_{GTR}} = 1, \text{if the TOC is GTR, otherwise } \mathrm{D_{GTR}} = 0 \\ \mathrm{D_{SE}} = 1, \text{if the TOC is SE, otherwise } \mathrm{D_{SE}} = 0 \\ \text{SWT is the reference for this category} \end{cases}$$

$$\mathrm{FY} = \begin{cases} \mathrm{D_{2011-12}} = 1, \text{if } 2011-12 \text{ is the financial year, otherwise } \mathrm{D_{2011-12}} = 0 \\ \mathrm{D_{2012-13}} = 1, \text{if } 2012-13 \text{ is the financial year, otherwise } \mathrm{D_{2012-13}} = 0 \\ \mathrm{D_{2013-14}} = 1, \text{if } 2013-14 \text{ is the financial year, otherwise } \mathrm{D_{2013-14}} = 0 \\ \mathrm{D_{2014-15}} = 1, \text{if } 2014-15 \text{ is the financial year, otherwise } \mathrm{D_{2014-15}} = 0 \\ \mathrm{D_{2015-16}} = 1, \text{if } 2015-16 \text{ is the financial year, otherwise } \mathrm{D_{2015-16}} = 0 \\ \mathrm{D_{2016-17}} = 1, \text{if } 2016-17 \text{ is the financial year, otherwise } \mathrm{D_{2016-17}} = 0 \\ \mathrm{D_{2017-18}} = 1, \text{if } 2017-18 \text{ is the financial year, otherwise } \mathrm{D_{2017-18}} = 0 \\ 2010-11 \text{ is the reference for this category} \end{cases}$$

$$\mathrm{RT} = \begin{cases} \mathrm{D_{Suburban}} = 1, \text{if the route is suburban, otherwise } \mathrm{D_{Suburban}} = 0 \\ \text{The mainline route is the reference for this category} \end{cases}$$

$$\mathrm{OF} = \begin{cases} \mathrm{D_{VS}} = 1, \text{if the organisation is vertically separated, otherwise } \mathrm{D_{VI}} = 0 \\ \text{The virtually integrated organisation (VI) is the reference for this category} \end{cases}$$

$$\mathrm{P} = \begin{cases} \mathrm{D_{P02}} = 1, \text{if period is the second period, otherwise } \mathrm{D_{P02}} = 0 \\ \mathrm{D_{P03}} = 1, \text{if period is the third period, otherwise } \mathrm{D_{P03}} = 0 \\ \mathrm{D_{P04}} = 1, \text{if period is the fourth period, otherwise } \mathrm{D_{P04}} = 0 \\ \mathrm{D_{P05}} = 1, \text{if period is the fifth period, otherwise } \mathrm{D_{P05}} = 0 \\ \mathrm{D_{P06}} = 1, \text{if period is the sixth period, otherwise } \mathrm{D_{P06}} = 0 \\ \mathrm{D_{P07}} = 1, \text{if period is the seventh period, otherwise } \mathrm{D_{P07}} = 0 \\ \mathrm{D_{P08}} = 1, \text{if period is the eighth period, otherwise } \mathrm{D_{P08}} = 0 \\ \mathrm{D_{P09}} = 1, \text{if period is the ninth period, otherwise } \mathrm{D_{P09}} = 0 \\ \mathrm{D_{P10}} = 1, \text{if period is the tenth period, otherwise } \mathrm{D_{P10}} = 0 \\ \mathrm{D_{P11}} = 1, \text{if period is the eleventh period, otherwise } \mathrm{D_{P11}} = 0 \\ \mathrm{D_{P12}} = 1, \text{if period is the twelveth period, otherwise } \mathrm{D_{P12}} = 0 \\ \mathrm{D_{P13}} = 1, \text{if period is the thirteenthperiod, otherwise } \mathrm{D_{P13}} = 0 \\ \text{The first period is the reference for this category} \end{cases}$$

The regression model that will be used to estimate the coefficients for the PPM and CaSL indicators is the following:

$$\begin{aligned} \mathrm{PI} = \beta_0 &+ \beta_1 \mathrm{RSA} + \beta_2 \mathrm{D_{GTR}} + \beta_3 \mathrm{D_{SE}} + \beta_4 \mathrm{D_{2011-12}} + \beta_5 \mathrm{D_{2012-13}} + \beta_6 \mathrm{D_{2013-14}} \\ &+ \beta_7 \mathrm{D_{2014-15}} + \beta_8 \mathrm{D_{2015-16}} + \beta_9 \mathrm{D_{2016-17}} + \beta_{10} \mathrm{D_{2017-18}} + \beta_{11} \mathrm{D_{P02}} \\ &+ \beta_{12} \mathrm{D_{P03}} + \beta_{13} \mathrm{D_{P04}} + \beta_{14} \mathrm{D_{P05}} + \beta_{15} \mathrm{D_{P06}} + \beta_{16} \mathrm{D_{P07}} + \beta_{17} \mathrm{D_{P08}} \\ &+ \beta_{18} \mathrm{D_{P09}} + \beta_{19} \mathrm{D_{P10}} + \beta_{20} \mathrm{D_{P11}} + \beta_{21} \mathrm{D_{P12}} + \beta_{22} \mathrm{D_{P13}} \\ &+ \beta_{23} \mathrm{D_{Suburban}} + \beta_{24} \mathrm{D_{VS}} + \varepsilon \end{aligned} \tag{3}$$

As the measurements of the PPM and CaSL indicators are limited between 0 and 1, these measurements are required to be transformed on the logit scale because the data is bounded between 0 and 1 (Fox, 2015, p. 72). The formula that will be used for transforming is the following:

$$Logit(PI_i) = log\frac{PI_i}{1-PI_i} \tag{4}$$

The regression analysis relies on disaggregated data published by ORR. This means that other explanatory variables are not included in the analysis, such as train length, service frequency and passenger rail demand, which could contribute to a better explanation of changes in performance quality.

3.3 Performance Quality Prediction

The objective of predicting the performance quality is to analyse and compare the actual measurements of the performance quality of the SWT with the predicted measurements, and this is delivered by two stages. Firstly, the actual measurements of PPM and CaSL of the prior-alliance period are used to forecast the measurements during the alliance. Secondly, the actual measurements of PPM and CaSL of the alliance period are used to forecast the measurements post the alliance.

The process of creating the forecast models is based on the Autoregressive Integrated Moving Average (ARIMA) model process. The ARIMA models contain the Autoregressive (AR) and Moving Average (MA) as the parameters of the ARIMA model are ARIMA(p,d,q), where p, d and q are related to the autoregressive order, the moving average order and the required difference of the model to achieve stationarity respectively (Washington et al., 2003, p. 180). The process of the ARIMA models according to Washington et al. (2003, p. 183) is as follows:

1. Plot the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF).
2. Estimate the parameters of the ARIMA model.
3. Check the accuracy of the model.

Plotting Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) is significant in creating the ARIMA model. As the ARIMA model is a combination of Autoregressive (AR) and Moving Average (MA), ACF and PACF are used to determine the order of these combinations. Moreover, Table 2 contains the process of selecting the ARIMA model and estimating the orders of these models. In some cases, ACF plot shows repeated significant lags in the same period. This indicates that there is a seasonal effect on the data, and this requires an upgrade of the ARIMA model to include the seasonal effect, which means that the seasonal ARIMA model (SARIMA) will be more appropriate for forecasting. The order of the SARIMA is SARIMA (p,d,q)x(P,D,Q)t, where p, d and q are stated above, and P, D and Q are the order of the autoregressive order, the moving average order and the required difference of the model to achieve stationarity respectively for the seasonal effect at period t. However, ARIMA models can be developed to accommodate the effect of periods as regressors, and this method is known as ARIMAX models.

After selecting the order of the ARIMA model, the parameters of the ARIMA model will be estimated in order to examine and check the selected model accuracy. This means that even if the order of the ARIMA model was selected based on the plots of the ACF and PACF, the order can be modified based on the model accuracy process. The parameters will be estimated by using the application R.

Table 2: The process of selecting the ARIMA model and estimating the orders (Source: (Washington et al., 2003, p. 183)).

	AR(p)	**MA(q)**	**ARMA(p,q)**
ACF	Trails off exponentially	Cuts off after lag q	Trails off exponentially
PACF	Cuts off after lag p	Trails off exponentially	Trails off exponentially

The model check is divided into two approaches. The first approach is to assess the errors of the fitted values of the created model. This assessment is based on the equations that are provided by Washington et al. (2003, p. 190) such as root mean square error (RMSE). After calculating the accuracy measures, the interpretation of the results is as the values become closer to 0 so the ARIMA models are more accurate for forecasting. The second approach is to check the diagnostic of the residuals, and the purpose of this process is to check the estimated parameters (Cryer and Chan, 2008, p. 238). The process contains checking the standardized residuals, the ACF of the residuals and the Ljung-Box test. For the standardized residuals, the residuals will be plotted in order to assess the pattern of these residuals. On the other hand, the ACF of the residuals will be plotted in order to examine if there is any residual that is statistically significant. However, Table 3 contains the estimation of the model coefficients that are used to forecast the performance quality.

Table 3: Parameter estimation for ARIMA family.

	Model 1	**Model 2**	**Model 3**	**Model 4**
$AR_1(p_1)$	-0.54891**	0.10893*	0.09447*	0.88883***
$AR_2(p_2)$	-0.62864***	---	0.43099**	---
$MA_1(q_1)$	---	---	---	-0.60429**
$SAR_1(D_1)$	---	---	-0.70447***	-0.72249***
Intercept	---	0.01479***	0.93483***	0.01975***
D_{P02}	0.00630	-0.00203	-0.00654	0.00036
D_{P03}	-0.01262	0.00719	-0.00658	-0.00202
D_{P04}	-0.01039	0.00356	-0.02731***	0.00401
D_{P05}	-0.00538	0.00004	-0.02427***	0.00830**
D_{P06}	0.00587	-0.00516	0.01065	-0.00485
D_{P07}	0.00044	-0.00424	-0.01366	0.00406
D_{P08}	-0.03821***	0.00049	-0.07141***	0.01820***
D_{P09}	-0.06277***	0.02103***	-0.08180***	0.00938**
D_{P10}	-0.06045***	0.01093*	-0.08476***	0.02972***
D_{P11}	-0.01839*	0.00531	-0.07177***	0.01979***
D_{P12}	-0.02614**	0.00225	-0.05445***	0.01542***
D_{P13}	-0.00167	0.00263	-0.01524*	0.00095
RMSE	0.00788	0.00540	0.01282	0.00650

Model 1: ARIMAX (2,1,0). Model 2: ARIMAX (1,0,0). Model 3: SARIMAX (2,0,0)×(1,0,0). Model 4: SARIMAX (1,0,1) ×(1,0,0).

*, **, *** indicates significance at 0.05, 0.01, and 0.001 respectively.

4 Results and Discussion

4.1 Change Assessment

The investigation of the change in the performance quality of SWT was processed through two assessments. The first assessment focussed on SWT itself, through the three periods divided on the basis of performance indicators, PPM and CaSL, and route types, mainline and suburban routes, as provided in Table 4. The results of this assessment indicated that there was a continuous and statistically significant reduction of the performance quality of both routes of SWT through the three periods. Precisely, the proportion of trains arrived on-time reduced continually through the three periods. Similarly, the proportion of trains cancelled or significantly late escalated through the three periods.

Table 4: A summary of the individual and cross comparisons for both routes.

	Measurement differences			
Comparison status	**Mainline**		**Suburban**	
	PPM	**CaSL**	**PPM**	**CaSL**
Individual comparison (SWT itself)				
Case (1)	0.0268***	-0.0077***	0.0240***	-0.0084***
Case (2)	0.0606***	-0.0112***	0.0377***	-0.0116***
Case (3)	0.0874***	-0.0189***	0.0617***	-0.0200***
Proportion comparison (SWT with GTR)				
Case (4)	0.0295***	-0.0107***	0.0188***	-0.0090***
Case (5)	0.0372***	-0.0133***	0.0450***	-0.0197***
Case (6)	0.0546***	-0.0333***	0.0741***	-0.0375***
Proportion comparison (SWT with SE)				
Case (4)	0.0348***	-0.0101***	0.0188***	-0.0120***
Case (5)	0.0014**	0.0004	0.0037***	-0.0027***
Case (6)	-0.0239***	0.0034***	-0.0126***	0.0001
Change comparison (SWT with GTR)				
Case (7)	-0.0084	0.0025	-0.0285**	0.0112*
Case (8)	-0.0210**	0.0219*	-0.0281**	0.0178*
Change comparison (SWT with SE)				
Case (7)	0.0334***	0.0112	0.0160*	-0.0098*
Case (8)	0.0256	0.0178	0.0153	-0.0027

Cases: Cases (1): Prior-alliance minus alliance periods. Case (2): Alliance minus post-alliance periods. Case (3): Prior-alliance minus post-alliance periods. Case (4): Prior-alliance period for SWT minus prior-alliance period for other TOCs. Case (5): Alliance period for SWT minus alliance period for other TOCs. Case (6): Post-alliance period for SWT minus post-alliance period for other TOCs. Case (7): Prior- to alliance periods for SWT minus similar change for other TOCs. Case (8): Alliance to post-alliance periods for SWT minus similar change for other TOCs.

*, **, *** indicates significance at 0.05, 0.01, and 0.001 respectively.

The second assessment was based on comparing the change in the performance quality of SWT with GTR and SE through the three periods, and the comparisons were divided similarly to the individual assessment in terms of performance indicators and route types, as shown in Table 4. For SWT and GTR comparison, SWT performed more effectively in terms of the PPM and CaSL measurements compared to GTR in the three periods. This means that SWT had a statistically significant higher proportion of trains arrived on-time and a lower proportion of trains cancelled or significantly late compared to GTR. However, when the change in the PPM and CaSL measurements is considered, both TOCs had statistically similar changes in the performance quality from the pre-alliance to the alliance periods for mainline services, whilst there was a statistical significant difference for suburban services. Although there are a decrease in the proportion of trains arrived on-time and an increase in the proportion of trains cancelled or significantly late for SWT, these changes are significantly higher for GTR. In contrast, there is a statistical significant difference in the PPM and CaSL measurements between SWT and GTR from the alliance to post-alliance periods for both routes. Although there is a deterioration in the measurements of the performance quality of SWT, the change in these measurements is significantly higher for GTR.

For SWT and SE comparisons, the performance quality of both routes of SWT was statistically better in terms of the PPM measurements compared with SE in the prior- and during the alliance periods, but SE performed better post the alliance period. For the CaSL indicator, SWT performed more effectively for both routes during the prior-alliance period compared to SE. For the alliance period, both TOCs had a similar performance in the mainline route while SWT performed better for the suburban route. For the post-alliance period, SE had a better performance quality in the mainline route while there is a similarity in the performance quality for both TOCs in the suburban route. With regard to the changes in performance quality, both routes of SWT faced significant adverse changes in the PPM measurements from the prior-alliance to alliance periods compared to the changes in SE’s performance, whilst the change in CaSL was also adverse for suburban services. In contrast, there are no statistically significant differences in the changes in the measurements of PPM and CaSL for both TOCs in the mainline and suburban routes from the alliance to post-alliance periods.

4.2 The Effect Analysis

The effect analysis aims to assess the change in the PPM and CaSL indicators with respect to several factors. With regards to the PPM indicator, Table 5 contains summary results of the regression model of the PPM indicator. According to these results, the impact of the average age of the rolling stock on the PPM measurements is not statistically significant. This is expected, if the rolling stock is well maintained, its age should not affect the measurements of PPM. For the comparison between TOCs, the difference in the PPM measurements between SWT and SE is not statistically significant while the difference between SWT and GTR is statistically significant. Moreover, the odds ratio of the PPM indicator for GTR is 36% lower than SWT. This could be explained as the rail infrastructure of GTR is not reliable, and the timetable is not suitable for the peak services, and these reasons could affect train operation (Gibb, 2016). For the route type comparison, the difference in the PPM measurements between the mainline and suburban routes is statistically significant. Precisely, the odds ratio of the PPM indicator for the suburban route is higher by 20.70% compared to the mainline route. For the comparison between different organisation forms, the difference in the PPM measurements between the virtually integrated and separated organisation is not statistically significant. This means that there is no change in the performance quality measured by the PPM indicator between the virtually integrated and vertically separated organisation. For the financial year comparison, the difference in the PPM measurements between the financial years is statistically significant, except for 2011-12 and 2012-13 where the difference is not statistically significant. The results indicates that the odds ratios of 2013-14, 2014-15, 2015-16, 2016-17 and 2017-18 are lower by 22.29%, 30.38%, 40.36%, 54.50% and 48.66% respectively compared to 2010-11. For the comparison between the financial periods, there is a statistical difference between the first period and the other periods, except for the second period where the difference is not statistically significant. The worst financial periods in terms of the PPM measurements are the eighth, ninth, tenth, eleventh, twelfth periods (broadly October to February) where the odds ratios are 54.48%, 62.29%, 56.46% 46.98% and 41.97% lower respectively. This could be explained by the change in weather conditions in autumn (leaf fall) and winter (snow and frost) seasons when bad weather is most likely to affect performance.

Table 5: Regression results for the PPM and CaSL indicators.

Term	**PPM model**			**CaSL model**		
	β	**exp(β)**	**Odds ratio**	**β**	**exp(β)**	**Odds ratio**
Constant	2.6984***			-4.0655***		
Rolling stock age	0.0042	1.0043	0.42	-0.0089	0.9911	0.88
TOC						
GTR	-0.4524***	0.6361	36.39	0.6109***	1.8422	84.22
SE	-0.0625	0.9394	6.05	0.1309*	1.1399	13.98
Route type						
Suburban	0.1881***	1.2070	20.69	-0.0185	0.9816	-1.84
Organisation form						
VS	0.0643	1.0665	6.65	-0.1062	0.8992	-10.07
Financial year						
2011-12	0.0102	1.0104	1.03	0.02795	1.0283	2.83
2012-13	-0.0768	0.9260	7.39	0.08692	1.0908	9.08
2013-14	-0.2522***	0.7771	22.29	0.3011***	1.3514	35.13
2014-15	-0.3620***	0.6962	30.37	0.3434***	1.4098	40.97
2015-16	-0.5168***	0.5964	40.36	0.5445***	1.7238	72.38
2016-17	-0.7875***	0.4550	54.50	0.8383***	2.3126	131.26
2017-18	-0.6666***	0.5134	48.65	0.7011***	2.0160	101.59
Period						
P02	-0.0583	0.9433	5.66	0.0224	1.0227	2.27
P03	-0.1600**	0.8521	14.79	0.2303**	1.2590	25.90
P04	-0.2205***	0.8021	19.78	0.1809*	1.1983	19.83
P05	-0.1848**	0.8312	16.88	0.2010*	1.2227	22.27
P06	-0.1267*	0.8810	11.90	0.0917	1.0961	9.60
P07	-0.2565***	0.7737	22.62	0.1444	1.1554	15.53
P08	-0.7869***	0.4552	54.47	0.4311***	1.5391	53.90
P09	-0.9923***	0.3707	62.92	0.7120***	2.0381	103.81
P10	-0.8315***	0.4354	56.46	0.7489***	2.1147	111.47
P11	-0.6344***	0.5302	46.97	0.5236***	1.6882	68.82
P12	-0.5442***	0.5803	41.96	0.5000***	1.6488	64.88
P13	-0.2659***	0.7665	23.35	0.2486**	1.2823	28.22
R-squared		0.7237			0.5673	
Adj. R-squared		0.7126			0.5499	

*, **, *** indicates significance at 0.05, 0.01, and 0.001 respectively.

Besides the analysis of the PPM indicator, Table 5 contains an estimation of the coefficients of the regression model for the CaSL indicator. According to this table, the impact of the average age of rolling stock on the CaSL indicator is also not statistically significant. Regarding the comparison between TOCs, the difference in CaSL measurements between SWT and GTR is statistically significant and the odds ratio of the CaSL indicator for GTR is 84.22% higher than SWT, and this can be linked with the reasons that are stated for the PPM indicator. Similarly, the difference in the CaSL measurements between SWT and SE is statistically significant and SE has a higher odds ratio by almost 14% compared to SWT. This means that SWT performed more efficiently in terms of the CaSL indicator compared to GTR and SE from 2010-11 to 2017-18 financial years. For the route type comparison, there is no statistical difference in the CaSL measurements between the mainline and suburban routes. This means that there is no difference in the proportion of trains cancelled or significantly late between the mainline and suburban routes. With regard to the impact of different organisation forms on the CaSL indicator, the results show no statistical difference between the vertical separation and virtual integration. This means that there is no difference in the CaSL measurements due to the change in the railway organisation. For the financial years, the difference in the CaSL measurements between 2011-12 and 2012-13 with the reference financial year, 2010-11, is not statistically significant. This indicates that the change in the proportion of trains cancelled or significantly late in 2011-12 and 2012-13 is not statistically significant compared to 2010-11. In contrast, the results show that there is a statistical difference in the CaSL measurements in 2013-14, 2014-15, 2015-16, 2016-17 and 2017-18 compared to the reference year. This means that the odds ratios of the CaSL indicator for 2013-14, 2014-15, 2015-16, 2016-17 and 2017-18 are higher by 35.13%, 40.97%, 72.38%, 131.26% and 101.59% respectively than the reference year. In addition, 2016-17 can be observed as the worst financial year in terms of the CaSL indicator. In a similar way to the financial year comparison, the results have two indications for the period comparison. The first indication is that there is no statistical difference in the comparison of the second, sixth and seventh periods with the first periods, which is treated as a reference. The second indication is that the difference in the CaSL measurements in the third, fourth, fifth, seventh, eighth, ninth, tenth eleventh, twelfth and thirteenth periods compared to the first period is statistically significant. This indicates that the odds ratios of the CaSL indicator for the third, fourth, fifth, seventh, eighth, ninth, tenth eleventh, twelfth and thirteenth periods are higher by 25.9%, 19.83%, 22.27%, 53.90%, 103.81% 111.47% 68.82% 64.88% and 28.22% respectively compared to the reference period. Additionally, the tenth period can be considered as the worst period in terms of the CaSL indicator. The explanation of this effect follow the same reasons for bad weathers as described above.

4.3 Performance Quality Prediction

The prediction of the PPM and CaSL indicators of SWT is divided into two approaches. The first approach is predicting the PPM and CaSL measurements during the virtual integration based on the measurements of the prior-integration period. The second approach is that the measurements of the PPM and CaSL indicators during the virtual integration are used to predict the measurements post the integration period. The procedure of predicting the PPM and CaSL measurements is based on ARIMAX models to accommodate the effect of periods on the PPM and CaSL indicators, as discussed in Section 3.3. For the first approach, the models that are used to predict the PPM and CaSL measurements are ARIMAX (2,1,0) and ARIMAX (1,0,0) respectively. Figure 3 and Figure 4 show the observed and predicted measurements of the PPM and CaSL indicators during the virtual integration. According to these figures, it can be seen that there is a reduction in the performance of the PPM and CaSL indicators in the eighth, ninth and tenth periods compared to the predicted values. In addition, as discussed in Section 4.2, the eighth, ninth and tenth periods can be observed as the worst periods that have significant deterioration in the PPM and CaSL indicators for all TOCs, including SWT. Having said that, the virtual integration did not contribute to mitigating the deterioration in the performance quality during that period.

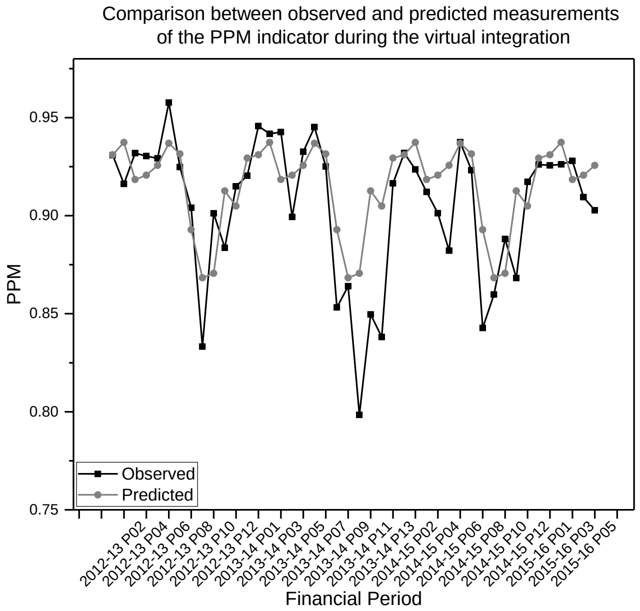

Figure 3: The observed and predicted measurements of the PPM indicator during the virtual integration for SWT.

The second approach is to predict the PPM and CaSL measurements in the post-integration period based on the integration period. The models that are used to predict the PPM and CaSL measurements are SARIMAX (2,0,0)×(1,0,0) and SARIMAX (1,0,1)×(1,0,0) respectively. Figure 5 shows the predicted and observed measurements of the PPM indicator post the integration period. It can be seen that there is a downtrend in the measurements of the PPM indicator in the post-integration period. In addition, several financial periods have a decrease in the proportion of trains arrived on-time compared to the predicted measurements. For the CaSL indicator, the predicted and observed values for this indicator are shown in Figure 6. According to this figure, there is a fluctuation in the observed measurements as predicted, but several periods have a higher proportion of trains cancelled or significantly late than predicted. Having said that, there is a significant deterioration in the performance quality of SWT post the integration period.

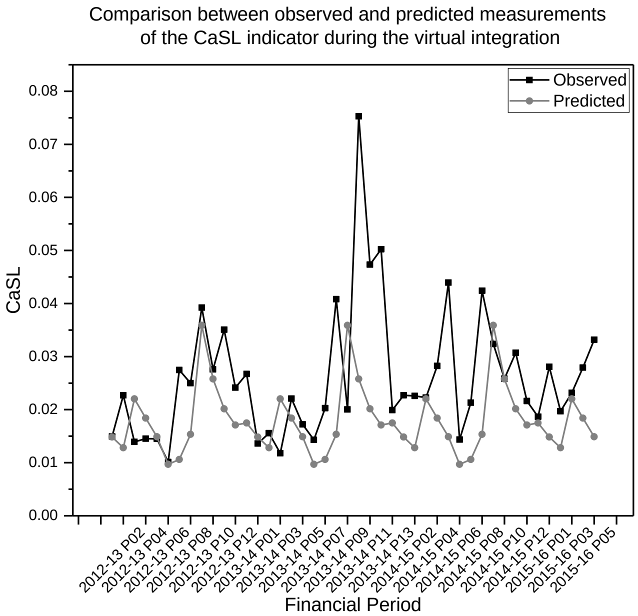
Comparison between observed and predicted measurements
of the CaSL indicator during the virtual integration

Figure 4: The observed and predicted measurements of the CaSL indicator during the virtual integration for SWT.

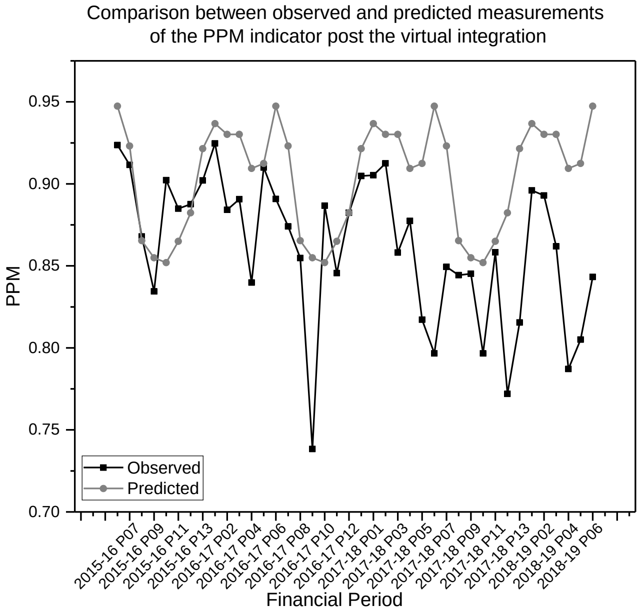
Comparison between observed and predicted measurements
of the PPM indicator post the virtual integration

Figure 5: The observed and predicted measurements of the PPM indicator post the virtual integration for SWT.

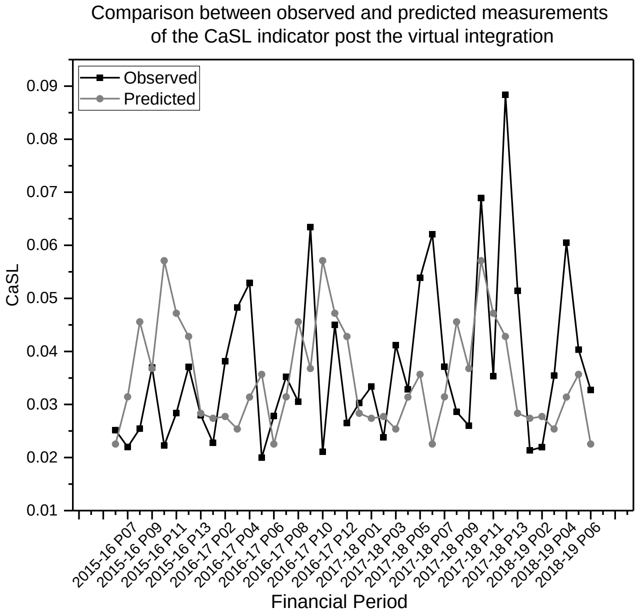

Figure 6: The observed and predicted measurements of the CaSL indicator post the virtual integration for SWT.

5 Conclusion

The assessment of the virtual integration started by evaluating the change in the performance quality of SWT. The results of this analysis pointed out the reduction in performance quality, as measured by the PPM and CaSL indicators for SWT. However, the change assessment was inconclusive, with SWT broadly performing better than GTR during the alliance period but performing worse than SE. The effect analysis was implemented in order to assess the effect of the virtual integration on the performance quality. The results of this analysis indicate that there is no evidence to support the effect of vertical integration on performance quality. The final analysis was predicting the performance quality of the SWT. The indication of this analysis is that the actual measurements of the performance quality are almost similar compared with the predicted measurements except for some periods that could be affected by other factors such as adverse weather. Overall, virtual integration does not seem to have had a significant impact on the performance quality of SWT.

Bibliography

Amaral, M. and Thiebaud, J.-C. (2015), Vertical separation in rail transport: How prices influence coordination, in '3rd Florence Conference on the Regulation of Infrastructures: Taking stock of current challenges', p. 50.

Armstrong, J. and Preston, J., (2017). 'Capacity utilisation and performance at railway stations'. *Journal of rail transport planning & management*, 7(3), pp.187-205.

Barron, A., Melo, P., Cohen, J. and Anderson, R., (2013). "Passenger-focused management approach to measurement of train delay impacts”, *Transportation Research Record: Journal of the Transportation Research Board*, 2351, 46–53.

Carey, M. (1998). ‘Optimizing scheduled times, allowing for behavioural response’, *Transportation Research Part B: Methodological*, 32(5), 329–342.

Carey, M. 1999. ‘Ex ante heuristic measures of schedule reliability’, *Transportation Research Part B: Methodological*, 33(7), 473–494.

Carey, M. and Carville, S. 2000) ‘Testing schedule performance and reliability for train stations’, *Journal of the Operational Research Society* 51(6), 666–682.

Cryer, J. D. and Chan, K.-S. (2008), *Time Series Analysis with Applications in R*, 2nd ed., Springer-Verlag. New York.

Drew, J. and Nash, C. (2011), ‘Vertical separation of railway infrastructure: Does it always make sense?, *Institute of Transport Studies, University of Leeds* Working Paper 594.

Ferreira, L. and Higgins, A. (1996), ‘Modeling reliability of train arrival times’, *Journal of transportation engineering*, 122(6), 414–420.

Fox, J. (2015), *Applied regression analysis and generalized linear models*, 3rd ed., Sage Publications, London.

Gibb, C. (2016), ‘Changes to improve the performance of the Southern network and train services, and restore passenger confidence [pdf]’. URL: https://goo.gl/Bv2gpx [Accessed date: 15 September 2017]

Goverde, R. M. (2005), ‘Punctuality of railway operations and timetable stability analysis’, PhD thesis, Technical University Delft.

Goverde, R. M. and Meng, L. (2011), ‘Advanced monitoring and management information of railway operations’, *Journal of Rail Transport Planning & Management* 1(2), 69–79.

Goverde, R. M., & Hansen, I. A. (2013). Performance indicators for railway timetables. In *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings* (pp. 301-306). IEEE.

Harris, N. G. (1992), *Planning Passenger Railways: A Handbook.* Transport Publishing Co., UK.

Knowles, R. D. (2013), ‘Railway franchising in Great Britain and effects of the 2008/09 economic recession’, *Environment and Planning A*, 45(1), 197–216.

McNulty, R. (2011), ‘Realising the potential of GB rail, report of the rail value for money study, summary report’, *Department for Transport and ORR.*

Merkert, R. (2005), The restructuring and future of the British Rail system, Technical Report Working Paper 586, Leeds, UK. Copyright of the Institute of Transport Studies, University Of Leeds. URL: http://eprints.whiterose.ac.uk/2288/

Merkert, R., Smith, A. S. and Nash, C. A. (2010), ‘Benchmarking of train operating firms–a transaction cost efficiency analysis’, *Transportation Planning and Technology*, 33(1), 35–53.

Mizutani, F., Smith, A., Nash, C. and Uranishi, S. (2015), ‘Comparing the costs of vertical separation, integration, and intermediate organisational structures in European and East Asian railways’, *Journal of Transport Economics and Policy (JTEP)*, 49(3), 496–515.

NR and ORR (2017), ‘Mandate L3 AR 004: Review of New Performance Metrics. [pdf]’ URL: https://orr.gov.uk/__data/assets/pdf_file/0005/25376/review-of-new-performance-metrics-2017-07-18.pdf [Accessed: 12 April 2019]

Olsson, N. O. and Haugland, H. (2004), ‘Influencing factors on train punctuality results from some Norwegian studies’, *Transport policy*, 11(4), 387–397.

ORR (2016), ‘Passenger Freight Rail Performance: Quality and Methodology Report7. [pdf]’. URL: https://goo.gl/5oyhu1 [Accessed: 25 March 2017]

ORR (2017), The National Rail Trends (NRT) Data Portal. URL: https://dataportal.orr.gov.uk/ [Accessed: 25 July 2017]

Parbo, J., Nielsen, O. A. and Prato, C. G. (2016), 'Passenger perspectives in railway timetabling: a literature review', *Transport Reviews*, 36(4), 500–526.

Preston, J. (1996), 'The economics of British rail privatization: an assessment', *Transport Reviews*, 16(1), 1–21.

Preston, J. (2002) 'The Transaction Cost Economics of Railways'. *Trasporti Europei*, 20/21, 6 – 15.

Preston, J., Wall, G., Batley, R., Ib´a˜nez, J. and Shires, J. (2009), 'Impact of delays on passenger train services: evidence from Great Britain', *Transportation Research Record: Journal of the Transportation Research Board*, (2117), 14–23.

Thompson, L. S. (2013), 'Recent developments in rail transportation services'. Organisation for Economic Co-operation and Development (OECD). URL: http://www.oecd.org/daf/competition/Rail-transportation-Services-2013.pdf [Accessed: 20 July 2017]

Veiseth, M., Olsson, N. and Saetermo, I. (2007), 'Infrastructures influence on rail punctuality', *WIT Transactions on the Built Environment* 96.

Vromans, M. (2005), Reliability of Railway Systems, PhD thesis, Erasmus University Rotterdam, Rotterdam School of Management. URL: http://hdl.handle.net/1765/6773 [Accessed: 20 July 2017]

Washington, S. P., Karlaftis, M. G. and Mannering, F. (2003), *Statistical and econometric methods for transportation data analysis*, CRC press.

Wetzel, H. (2008), European Railway Deregulation: The Influence of Regulatory and Environmental Conditions on Efficiency, Working Paper Series in Economics 86, University of Lneburg, Institute of Economics. URL: https://ideas.repec.org/p/lue/wpaper/86.html

Xia, Y., Van Ommeren, J. N., Rietveld, P. and Verhagen, W. (2013), 'Railway infrastructure disturbances and train operator performance: The role of weather', *Transportation research part D: transport and environment*, 18, 97–102.

Yaghini, M., Khoshraftar, M. M. and Seyedabadi, M. (2013), 'Railway passenger train delay prediction via neural network model', *Journal of advanced transportation* 47(3), 355–368.

Yuan, J. (2006), Stochastic modelling of train delays and delay propagation in stations, PhD thesis, Technical University Delft.

Yuan, J. and Hansen, I. A. (2007), 'Optimizing capacity utilization of stations by estimating knock-on train delays', *Transportation Research Part B: Methodological*, 41(2), 202–217.

Yuan, J. (2008), Statistical analysis of train delays, in I. A. Hansen and J. Pachl, eds, 'Railway Timetable Traffic: Analysis - Modelling - Simulation', Hamburg: Eurailpress, chapter 10, pp. 170–181.