The impact of the European Development Fund and European Bank as financing sources on traffic estimation biases

Abstract

This paper aims to model and understand the traffic estimation biases in European transportation infrastructure projects in the light of data from 55 projects developed between 1992 and 2012 in 17 European Union countries. Specifically, we investigate the correlation between overestimated traffic volumes and the presence of either the European Regional Development Policies or the European Investment Bank among the financing sources. We argue that such a correlation may indicate an undesirable selection bias favouring projects that use overestimated traffic volumes, leading to misled strategic decisions with long-term social and economic impacts. We innovate concerning previous studies by introducing a more general model based on classification trees, which accounts for non-linear relationships among the explanatory variables. The findings demonstrate that, in addition to the traditional influencing factors cited in the literature, e.g., contract year and investment value, both the European Regional Development Policies Funds and the European Investment Bank investments are also associated with the traffic estimation bias, albeit with distinct trends.

Keywords:

Transportation Infrastructure; Financing Policies; Project Finance; Traffic Forecasts; Regional Development

JEL codes:

R420; H540; R510; R580

1. Introduction

Traffic forecasting failures have been a central research topic in financial sustainability studies of transport infrastructure investment. According to Cruz and Sarmento (2019), state-of-the-art research presents several reasons for such failures, from political motivations and technical shortcomings to economic dynamics. However, ten years after the Eurozone public debt crisis, studies about the role of the financing source in the decision to invest in transport infrastructure projects are still scarce (Cruz & Sarmento, 2019).

This paper examines the relationship between the participation of two of the primary financing sources for transport investment in Europe (namely, the European Regional Development Policies and the European Investment Bank) and the occurrence of traffic forecasting overestimation errors in transport projects. Given the growing number of projects financed under the Project Financed Contract (PFC) model and the history of EU investment packages throughout the 1980s, 1990s, and 2000s (Overbeek, 2012; Wagenvoort et al. 2012), this article explores whether the investment model of the European Investment Bank (EIB) or the European Regional Development Policies (ERDP) tends to privilege investment in projects with overestimated forecasts.

While discussing renegotiations of transportation infrastructure contracts, Domingues and Zlatkovic (2015) argue that cost and traffic deviations are frequent in such contracts and generally related to opportunistic behaviour, albeit with distinct trends. Van Wee (2007) observes that inaccurate cost forecasts are associated with moral hazard-related differences between the project design and execution phases, while traffic forecasts are associated with adverse selection issues related to methods, data, and project uncertainty. Hence, one can argue that demand forecasts are the most relevant factor concerning the outcome of financing decisions in transport infrastructure projects, as they inform the analysis before the contractual execution. Bearing that in mind and considering the limited availability of data regarding the ex-ante and ex-post data on individual projects (Cruz & Sarmento, 2019), this paper investigates the possible influence of traffic estimation biases on project financing decisions, as these can have enduring adverse consequences.

Comparisons between ERDP and EIB focusing on the effect of the EU's public financing policies on regional development and infrastructure are relatively common in the literature (Pinder et al. 1995; Honohan 1995; Molle 2007; Robinson 2009; European Commission 2010). Such comparison is justified by their supranational responsibilities as redistributive triggers to promote infrastructure development among the EU member states (Molle 2007; Robinson 2009). In addition to their prominence as financing mechanisms, the focus on ERDP and EIB is also justified by the potential influence of political factors on the execution of these funds, as previously highlighted in some studies (Bouvet & Dall'erba, 2010; Dellmuth, 2011; Albalate et al. 2019; Liebe & Howarth, 2020). Specifically, projects with greater demands may be assessed by political entities as more conducive to local GDP growth, as they can potentially generate higher economic benefits for the affected region (Domingues and Zlatkovic, 2015). Therefore, this work evaluates a potential preference for projects with overestimated traffic demands, with a view that such analysis can contribute to a better understanding of the eventual failures of these financing mechanisms and the posterior reparation of such shortcomings.

Despite a similar emphasis on infrastructure development, the EIB's and ERDP's financing decision dynamics differ strongly (Molle 2007; Robinson 2009). In short, local government guarantees are at the forefront of the EIB's selection process, whereas the Member States Policy Committee defines ERDP's financing allocation considering the EU's investment programmes (Molle 2007; Robinson 2009). Hence, we may expect a possible variation of the effect of overestimated forecasted demand between ERDP and EIB due to the distinct decision dynamics. This paper will investigate whether such a variation can be detected in a European transport infrastructure contracts cohort.

We utilised a database of 55 transport infrastructure concessions signed between 1992 and 2012 in 17 European countries to conduct the study. The collection of such a database is a byproduct of the BENEFIT EU research agreement carried out by 14 partners around Europe, which managed to capture the essence of the decision making context in the period (e.g., Roumboutsos et al., 2013; Roumboutsos et al. 2014, Cruz & Sarmento, 2019), despite the contextual barriers to accessing information about each projects' conception, design, decision and forecasting processes (Cruz & Sarmento, 2019). Compared with similar studies, it is a powerful depiction of the characteristics of the decision-making context in the period (Roumboutsos et al. 2013; Roumboutsos et al. 2014, Cruz & Sarmento, 2019).

We employ a decision tree methodology to test the hypothesis that ERDP and the EIB may have a selection bias for financing projects with overestimated traffic volumes. More specifically, we investigate whether the estimation error is correlated with the financing sources, as well as three other technical variables used in previous related studies (namely, contract signature date; investment amount; and infrastructure mode - Road, Rail or Other), see for example, (Bain, 2009; Flyvbjerg, 2005; Flyvbjerg et al. 2005). The decision tree technique adopted is a classic and robust method capable of inferring non-linear relationships between the attributes of the project, including their sources of financing and the deviations in traffic forecasts (Hernandez, Monzon, & de Oña, 2016). Based on the premise that studying the conditions for the occurrence of overestimation is more critical than precisely inferring its size, we choose a classification tree model instead of a regression tree. Considering the limited size of the sample, the former allows a better assessment of the conditions for overestimation.

The results demonstrate that in addition to the traditional influencing factors commonly cited in the literature, e.g., contract year and investment value, both the European Regional Development Policies and the European Investment Bank loans are also associated with the traffic estimation bias. Our results are in line with the literature that suggests that the selection profile regarding projects with overestimated traffic is different between the studied instruments due to the differences in decision-making processes regarding the economic convergence policy in the European Union.

The remainder of the article is organised as follows. Section 2 presents a review of the role of traffic forecasts in assessing and deciding the investment in transport infrastructure since the emergence of PFCs in Europe. The following section discusses the political landscape underpinning ERDP's and EIB's structural financing mechanisms, highlighting the role of the EU's economic convergence policy. Section 4 introduces the materials and methods. It is followed by the experimental results and discussions presented in Section 5. Finally, Section 6 concludes the paper and features recommendations for future works.

2. Literature review on project financed transportation investments and traffic forecasts biases

Project financed contracts (PFC) were first introduced in continental Europe after the success of Project Financed Initiatives in England in the late 1990s (Clifton et al. 2014). PFCs are long-term contracts developed to support private investment in public services or public infrastructures through a limited financial structure, under which each project's debt is to be financed from cash flows generated by the project itself (Yescombe, 2002).

The cash flow limitations of the transport infrastructure projects place the traffic forecasts in a position of great importance in the decision-making process regarding the financial sustainability of these investments. The entire construction and operation planning becomes a function of the revenues expected by the project, which, in turn, are primarily derived from the use of the infrastructure (Oliveira et al., 2016).

Several studies have examined the influence of traffic forecasting on the concession and management of public transport infrastructure contracts (Cruz & Sarmento, 2019). Many of these studies point out numerous cases of overestimation of future traffic across the globe. They highlight an apparent stagnation of the predictive power of traffic forecasting methods over the years, despite advances in demand forecasting and data collection techniques (Baeza & Vassallo, 2010; Bain, 2009; Cruz & Sarmento, 2019; Flyvbjerg et al., 2005; Næss, 2016; Odeck & Welde, 2017; Vassallo & Baeza, 2007; Wachs & Ortner, 1979).

These two findings combined have motivated research in the field. Since 1978, at least 44 studies have addressed traffic forecast deviations in transport infrastructure projects (Cruz & Sarmento, 2019). Utilising different databases and methodologies, they found distinct explanations for the phenomenon. The influence of opportunistic strategies and an optimistic bias on the reliability of the adopted demand forecasting models are two of the most frequent explanations (Domingues and Zlatkovic, 2015; Cruz & Sarmento, 2019).

Traffic deviation models have historically evolved from the early simplified treatments of technical and macroeconomic variables to incorporate project elements, track records of the forecasts, location, contracting process, and legal context (Cruz & Sarmento, 2019). Nevertheless, none of the 44 studies evaluated in Cruz and Sarmento's (2019) literature review systematically addressed the role of financing. In most cases, the issue of financing appears only in the form of case studies (Baeza & Vassallo, 2010; Albalate et al., 2019; Wachs & Ortner, 1979).

It does not mean that the subject is not relevant. Several authors have addressed the negative effect of the excessive use of public guarantees on public-private partnership projects. Under the umbrella of opportunistic behaviour, case studies have identified the harmful effects of national financing policies on the investment in economically nonviable transport infrastructure projects (Domingues & Zlatkovic, 2015). Authors also discussed the excessive use of public guarantees on highway projects in the United States and Spain as an incentive to develop projects with greater demand risk (Baeza & Vassallo, 2010; Wachs & Ortner, 1979). More recently, it has been observed that the participation of private entities in highly subsidised public-private-partnership (PPP) type contracts tends to support the emergence of lobbying behaviour (Albalate et al., 2019). Therefore, it is not difficult to infer that such behaviour can give rise to financing biases, and it is crucial to verify whether there is evidence of such biases.

3. Defining the EU cohesion policy instruments as the study object

From the outset, the European Commission has the mandate to reduce the economic gaps between the member states to promote a balanced common market. Since the 1950s, the ERDP's and EIB's structural funds have been the primary financing sources for intra-block economic divergence mitigation policies, mainly via infrastructure projects (Bouvet & Dall'erba, 2010; Dellmuth, 2011; Liebe & Howarth, 2020).

Although their purpose is similar, the ERDP and the EIB have distinct decision-making mechanisms. Albeit aligned to the EU's directives, the EIB is organised as a politically autonomous bank. The ERDP's financing allocation decisions, on the other hand, are performed by member states' representatives via public financing calls (Pinder et al., 1995, Bouvet & Dall'erba, 2010; Dellmuth, 2011; Liebe & Howarth, 2020). Hence, one may expect distinct impacts of traffic estimates on these two institutions, given that they are subject to different levels of political influence.

Considering Albalate et al.'s (2019) discussion about the use of political lobby for public guarantees for investment in white elephants, this paper aims to test the hypothesis that the ERDP's and/or the EIB's financing decisions may be biased towards projects with inflated demand forecasts. Such a hypothesis originates from the perception that a financing decision within the European convergence framework is based, at least in part, on the comparative projected GDP growth among the block's regions, and part of this projected growth derives from the projected demand for a given project's infrastructure (Crescenzi & Rodríguez-Pose, 2012; Del Bo et al. 2010). Given such a causal loop between a potential decision to finance the project and the local GDP growth projection, one may be tempted to see the demand forecast as a mechanism to promote the project when a financing decision is pending.

This study evaluates the potential bias caused by the EU's economic convergence policy in both the ERPD's and the EIB's financing decisions concerning projects with overestimated traffic demand. By doing so, one can observe if and how the formal structure of an autonomous bank helped the EIB's financing decisions when faced with unbeknownst inflated demand forecasts. We can also observe if and how EIB's decisions differed from the ERDP under the same circumstances.

4. Methodology

To further evaluate the component of financial effects on the opportunistic behaviour of agents, this study aims to assess the effect of (i) the European Regional Development Policies (ERDP) funds and (ii) the European Investment Bank (EIB) loans on demand forecasting biases, focusing on the overestimation bias. Unlike previous research, this article breaks with the trend to present an analysis focused on case studies (Baeza & Vassallo, 2010; Albalate et al., 2019; Wachs & Ortner, 1979). We propose a meta-analysis, which can compare the influence of the selection process of different agents on a larger project basis than that observed in previous studies (Sanko et al., 2013). Given the importance of the financing structure in the viability of the new projects, this approach assesses the existence of selection biases of the studied financing sources. In other words, the method evaluates whether the projects that a given fund financed tend to overestimate demand forecasts (Button & Chen, 2014; Odeck & Welde, 2017).

We analysed 55 transport infrastructure concessions signed between 1992 and 2012 using data from the BENEFIT project, the report of the European Court of Auditors 2014, and the European Investment Bank's project database (European Court of Auditors, 2014; European Investment Bank, 2015; Roumboutsos et al. 2013; Roumboutsos et al. 2014). With data from 17 countries with different contextual variables over 20 years, this sample is comprehensive but scattered, given the large standard deviations for all observed continuous variables.

The technique used for the analysis of the 55 projects was decision trees. Decision trees are considered a classic and robust technique for classification. They enable us to infer non-linear relationships between the project's features, including their sources of financing, and the deviations in demand forecasts (Breiman et al., 1984; Loh, 2011). We opt to use decision trees because linear regression models have little explanatory power on the variability of traffic forecasts using a database of projects from different countries (Odeck & Welde, 2017; Parthasarathi & Levinson, 2010; Welde & Odeck, 2011). In addition, linear regression does not capture non-linear relationships amongst the variables, as do decision trees. Finally, decision trees provide a non-parametric analysis tool that considers the possibility that the effects of some variables may not remain homogeneous throughout the sample (Loh, 2011).

Initially, the study intended to use the regression decision tree technique. However, it was observed that the analysis of the magnitude of the observed traffic diversion did not present reliable results due to the limitations of the sample size. Therefore, we decided to evaluate only the occurrence of demand overestimation, since this information would already be significant to assess the impact of the dependent variables under analysis.

In transport research, decision tree techniques have been used to improve the estimation of non-parametric models in studies about the perception of transport quality. They also helped identify factors that cause accidents and predict the modal choice of public transport passengers (Hernandez et al., 2016; Tsami et al., 2018). Since they provide an estimation of the importance of the explanatory variables on the observed outcome, Classification Trees are often used to study non-linear relationships between dependent and independent variables (Breiman et al., 1984; Loh, 2011).

4.1 Classification Decision Trees

Decision tree models are based upon the construction of binary trees (Ghiasi et al., 2020). The model produces decision nodes to help predict a dependent variable *Y* as a function of an independent variable *X* with *p* attributes. Each node splits the tree into two branches as a function of a single attribute of the independent variable X. The underpinning technique estimates a function d(x) to map each point in the X domain to a corresponding point in the Y domain. To that end, we need a set of n observations of input-output pairs given by L = {(x1, y1), . . . , (xn, yn)}.

A classification tree defines each node Aj as a union of subsets, where each subset is obtained from the dataset by recursive partitioning of the sample space X. The tree partitions Y into *J* disjoint sets such that Aj = {x : d(x) = j} | X = $\cup^{J}_{j=1}$ Aj. The partitioning procedure may be based on different criteria, such as entropy-based metrics, information gain, Gini index, normalised impurity, etc. In this work, the partitioning is based on the entropy criterion (Lee & Yang, 2010).

Entropy is a measure of the dispersion of the observed sample with respect to the independent variable *Y*. It is given by (Lee & Yang, 2010):

$$Entropy = -\sum_j \quad p_j \cdot log_2 \cdot p_j \ ,$$

Where p_j is the probability that an observation belongs to class *j*. Maximum entropy is attained when the sample is evenly distributed, i.e. all classes have the same probability. On the other hand, a pure node is observed when the entropy attains the minimum value of 0, representing a single group, i.e.:

$$Entropy_{min} = -1 \cdot log_2(1) = 0$$

$$Entropy_{max} = -0.5 \cdot log_2(0.5) - 0.5 \cdot log_2(0.5) = 1$$

The classification algorithm will seek an optimal split at each node by choosing the attribute with the least entropy within the sample space *L*. Typically, the validation of the model's results uses cross-validation, where a segment of the dataset is used to learn the tree, and the rest is used to test the accuracy of the prediction against observed values. Specifically, we used a K-fold validation technique, which utilises a set of K validation subsets in turn, considering that the elements not currently used for validation will be used to train the model. Such a technique helps avoid biases and reach better classification results (Anguita et al., 2012).

4.2 Modelling the financing effect on traffic forecast biases

Based on the model developed by Flyvbjerg et al. (2005), we accounted for the traffic deviations according to the following equation:

- $trafInac = \frac{(Ta - Tf)}{Tf}$ (1)

where Ta is the *actual traffic,* and Tf is the *forecasted traffic.*

By computing these deviations, the proposed model classifies the observations into two groups of traffic performance: below or inline/above expected. Thus:

- $trafInac$ = Inline or above if > - 0,05
- $trafInac$ = below if $\leq$ - 0,05

This classification considers that underestimated, and within-expected forecasts have the same effect on the project's financial sustainability. It assumes that the deviation magnitude is not significant for bias identification when studying the role of financing instruments in selecting overestimated projects. In other words, we focused on understanding whether the funds' selection processes favoured projects with overestimated forecasts, regardless of the extent of the deviation.

The observed model was built to assess the role of the financial dimension by comparing the effect of the participation of the two European financing instruments on the traffic forecast overestimation outcome. This study was controlled by three variables considered relevant in previous studies (Bain, 2009; Flyvbjerg et al., 2005). Table 1 presents the selected variables for each group:

Table 1 - Description of variables used in decision tree models

Variable's name	Description	Descriptive statistics
Dependent Variable		
TrafInac	trafInac = Inline or above if > - 0,05 trafInac = below if ≤ - 0,05	Inline = 25 (45.45%) Below = 30 (54.54%) avg: -0.127 sd: 0.3907
Exposure variable		
EIBpar	EIB finance investment allocated in the project / Total investment (%)	$\overline{x}$: 0.21; sd: 0.22
ERDPpar	ERDP finance investment allocated in the project / Project investment size (%)	$\overline{x}$: 0.08; sd: 0.14
Control variables		
Total.inv	Project investment size in millions of euros	$\overline{x}$: 675.39; sd: 704.61
Road	Dummy for Road Projects: 1- Road	55% = 0

		45% = 1
Rail	Dummy for Rail Projects: 1- Rail	61% = 0 29% = 1
Other	Dummy for Airports, Ports and other infrastructure projects Projects: 1-Others	64% = 0 26% = 1
Contract.date	Signature date of the project contract	1992 (1); 1994 (2); 1995 (3); 1996 (2); 1997 (3); 1999 (5); 2000 (4); 2001 (1); 2002 (3); 2003 (3); 2004 (2); 2005 (3); 2006 (3); 2007 (4); 2008 (5); 2009 (5); 2010 (2); 2011 (3); 2012 (1)

Table 1 shows 30 projects with demands greater than 5% below the forecasted and 25 projects with demands within or above the predicted. With an average traffic deviation of -12.07% from the expected and a standard deviation of 39.07%, this sample is within the expected values compared to previous studies. Specifically, the sample is similar to the cases observed by Baeza and Vassalo (2012; 2010), who also studied European projects.

Regarding the choice of exposure variables, the model observes not only the participation of the instrument in the total investment of the project but also its absolute value. Of the 55 projects in the sample, 22 did not receive investments from EIB, 34 did not receive funds from ERDP, 14 did not receive from either source, and 13 received from both sources.

Since the decision tree techniques independently analyse the effect of each explanatory variable on the dependent variable, the overlap between the effects of these two variables does not undermine the reliability of the model.

As mentioned earlier in this section, we considered variables related to project size, mode of transport, and year of contract signature because they are essential for explaining traffic deviations in previous articles about the topic. Flyvbjerg et al. (2005) and Bain (2009) also observed these three variables as possible explanatory factors for the emergence of traffic forecast deviations. Therefore, they were introduced in the model as control variables for assessing the role of financing structure in the discrepancies.

4.3 Data

Flybjerg et al. (2005) reported that information that allows the estimation of traffic deviations in transport infrastructure projects is relatively rare. In most cases, this information is not even produced, and, when generated, there are several difficulties in comparing the initial estimates with the observations made after the infrastructure started operating (Cruz & Sarmento, 2019).

The database used in this study started to be consolidated from case studies reported in the European project BENEFIT (Roumboutsos et al. 2013, 2014). The BENEFIT project studied 86 cases of investment in transport infrastructure to develop a business modelling methodology for the development of transport infrastructure projects, according to their expected results.

The expected and realised demand indicators are among the variables collected in the case studies.

We selected all 47 cases from the initial sample that included information regarding traffic forecasts, observed traffic, and information on the project's financial structure. Eight projects from a European Court of Auditors report on the effectiveness of investments financed by the European Union (European Court of Auditors, 2014) were added to this sample, totalling 55 projects. Compared with previous studies, this sample seems to have good representativeness regarding the period observed and the countries covered.

Finally, the study regards the "ramp-up effect" problem on the traffic forecast inaccuracy analysis described in Flyvbjerg et al. (2005). All traffic forecast data were collected considering the demand level in 2013, 2014 and 2015. The exceptions are renegotiated projects due to errors in the traffic forecast before 2014. In these cases, the information presented considers the forecasting error in the last year before the renegotiation.

5. Results and discussion

A decision tree model development consists of three stages: growth, pruning, and selection (Breiman et al., 1984; Loh, 2011). The decision tree growth process starts with the inclusion of all independent variables (i.e., exposure and control) in an unrestricted model. Then, the algorithm estimates the entropy of each split to evaluate the effects of distinct groupings of independent variables on the traffic forecast accuracy. Figure 1 presents the decision tree generated after applying the growth step.

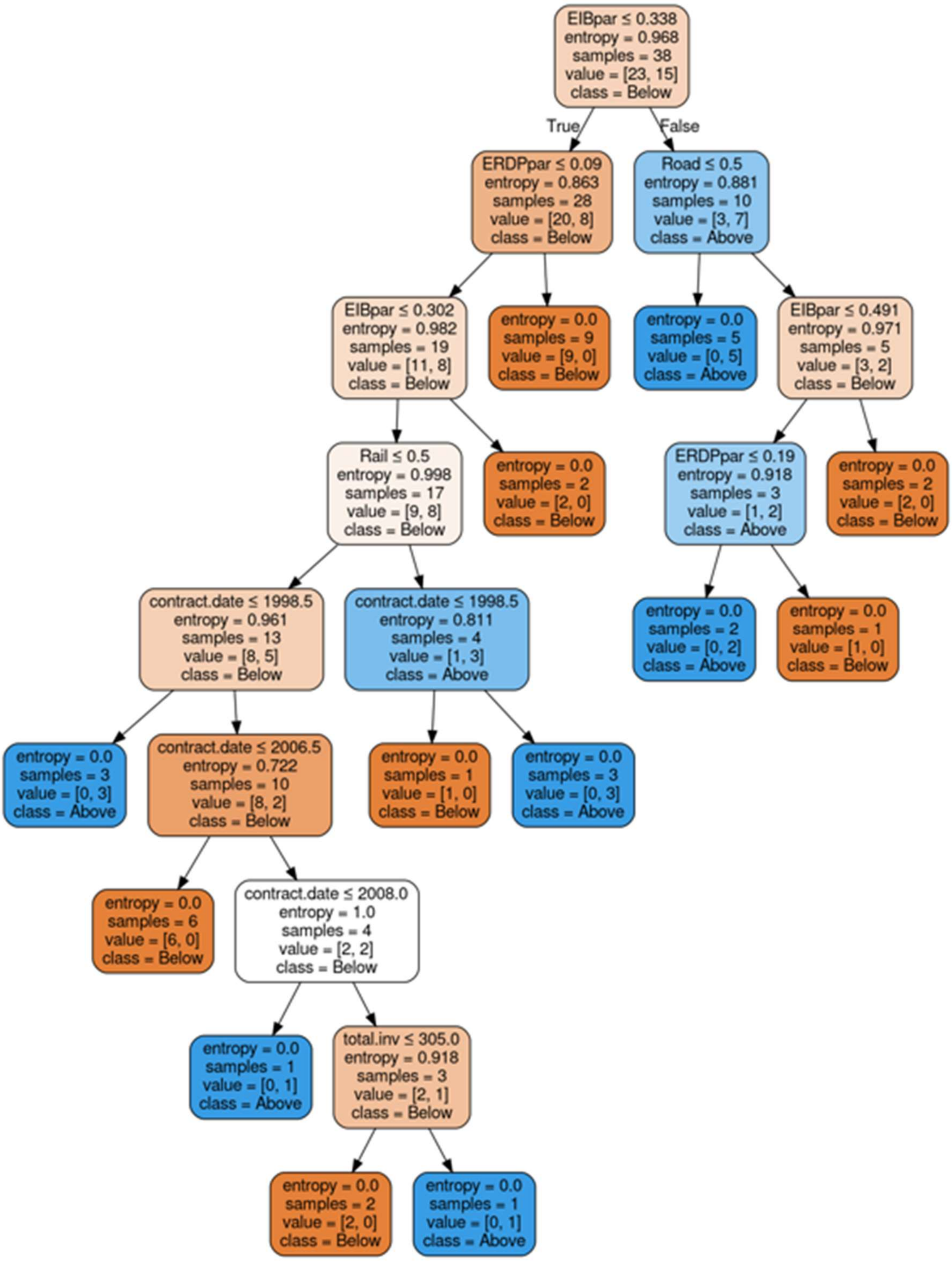

Figure 1 - Estimated results of the unrestricted model

Since the number of observations in the sample was reduced, the study evaluated the model's predictive ability by applying the K-fold technique in five groups of training and test observations, which corresponded to 70% and 30% of the sample, respectively.

In this unrestricted model, we observed that the main variable for the definition of the nodes is the date of signature of the contract, followed by the participation of the ERDP and the EIB loans. The model reaches 100% accuracy for the training segments but presents only an average of

45% precision among the 5 test segments defined by the K-fold analysis, thus characterising overfitting: a minute explanation of the training set but with reduced predictive value.

The overfitting was already expected due to the sample size and the number of independent variables. This result means that the estimated model is not generalisable because it is overfitting the training samples. To make the model more generalisable to both training and test samples, we proceeded to prune the tree using the max depth pruning procedure. This procedure seeks the longest path between the root node and the tree leaves that maintains the predictive ability to concentrate the explanatory power in the most critical levels and prevent the use of nodes with poor generalisation capability (Ghiasi et al. 2020).

Ghiasi et al. (2020) suggest a trial and error approach to estimate the optimal tree length. We searched models limiting the paths to lengths varying from 1 to 24 nodes. Using again the five training and test segments estimated by the K-fold technique, we observed the variability of the model accuracy over the training and test populations as a function of the estimated tree depth. The results are presented in Figure 2.

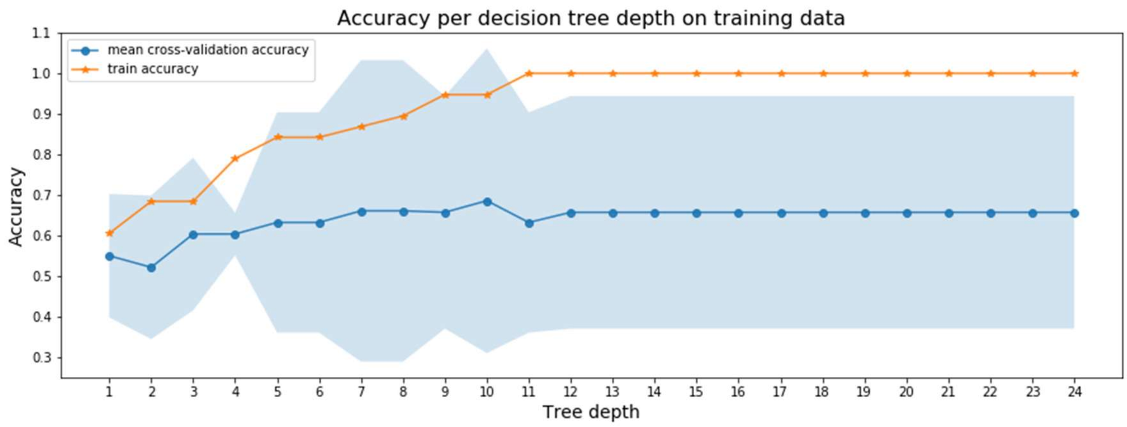

Figure 2 - Estimated tree accuracy per tree depth for train and test samples

Figure 2 shows that the model's accuracy for the training population grows as the depth of the tree increases until it reaches 100%, while the accuracy for the test population fluctuates. In figure 2, the maximum point of the estimated model accuracy for the mean K-fold test segments is at level 10 of the decision tree. At that point, the mean cross-validation accuracy of the test segment reaches almost 70%.

These results prove satisfactory compared to previous studies that model the effects of contract models on deviations in traffic estimates (Athias & Nuñez, 2008; Button & Chen, 2014; Horowitz & Emslie, 1978; Odeck & Welde, 2017; Parthasarathi & Levinson, 2010). The following section discusses the results and the decision thresholds estimated by the model for each node.

5.1 Understanding the occurrence of traffic deviations

Supported by the decision tree presented in Figure 3, we observed the effects of the exposure and control variables on overestimating traffic forecasts in the sample. The method of Derived Importance or Variable Importance Measure (VIM) was used to hierarchise the importance of the explanatory variables. The VIM is calculated based on the average entropy reductions over all tree nodes where a division was made from this variable (Breiman et al. 1984).

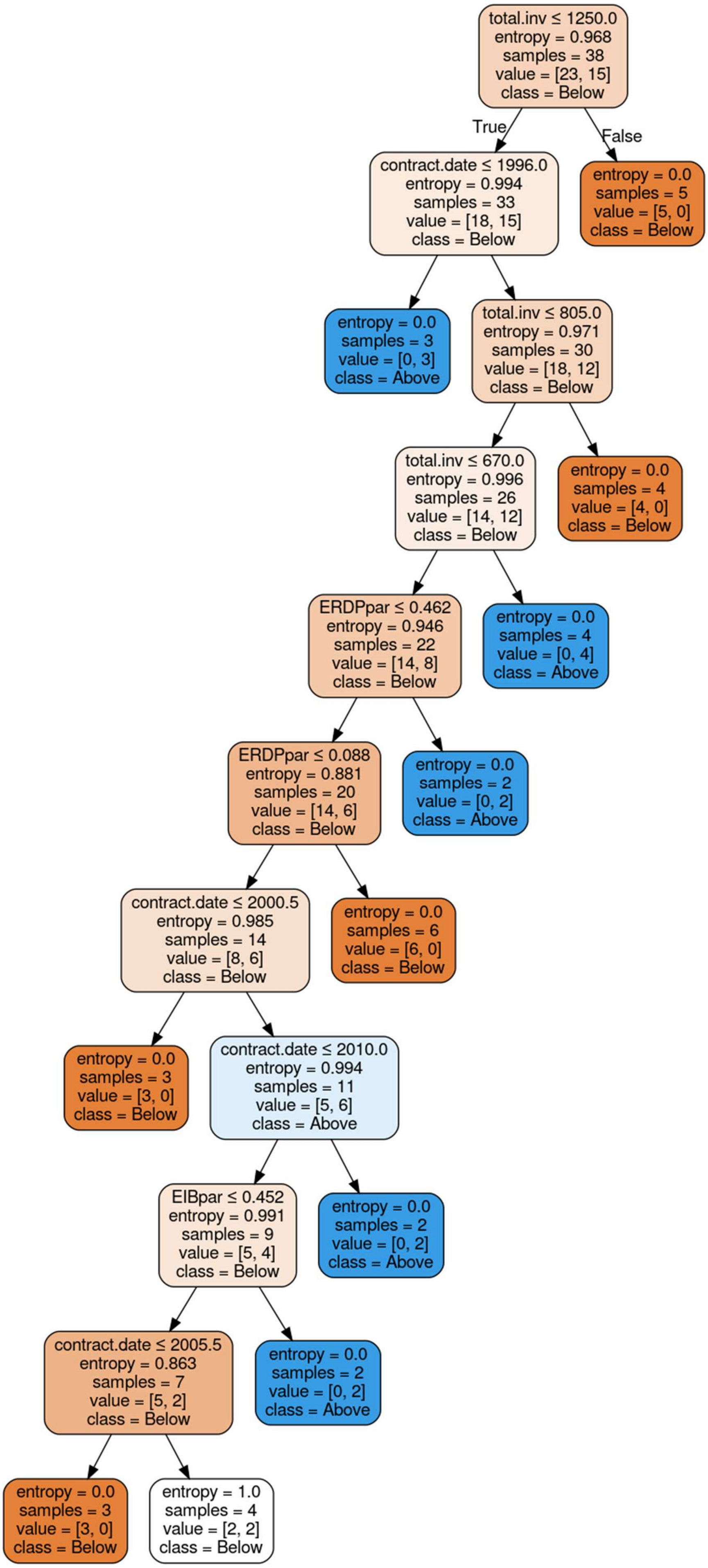

Figure 3 - Estimated results of the Pruned Tree model

According to Hernandez et al. (2016), this method is particularly attractive because it can be calculated without any additional computational expense beyond the standard training procedure. The entropy reductions considered in the calculation were weighted according to node size to avoid the overvaluation of group divisions of a few observations (i.e., not very generalisable). Figure 4 shows the importance index of each of the variables in the estimated model.

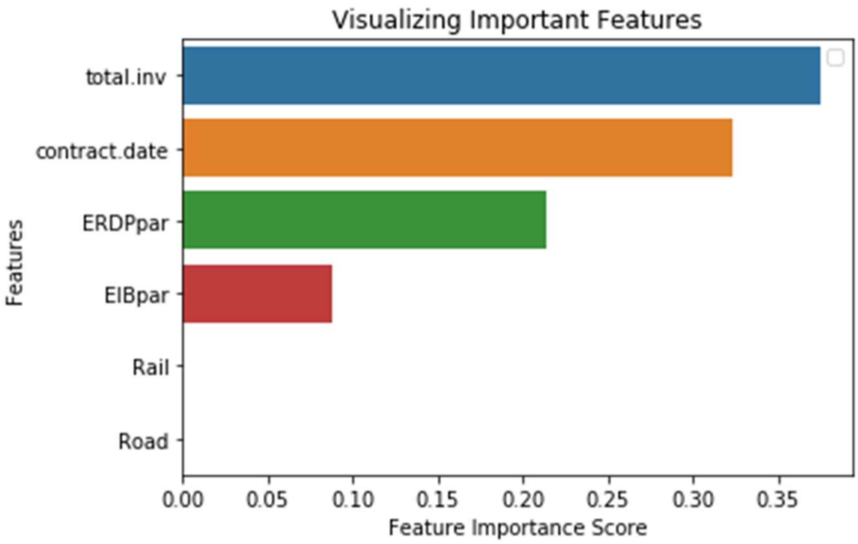

Figure 4 - Hierarchy chart of the importance of the model variables

As presented in Figure 4, the four most important contractual variables for differentiating traffic forecast deviation patterns among the observed projects are the total investment, contract date, ERPD investment and EIB Loans participation, respectively. The following section examines the thresholds of the effect of each of these variables on the model.

5.2 Control Variables: Mode, Year and Investment

The three control variables (i.e., contract.date, Road/Rail, and total.inv) were chosen to compare the findings of this study with the state-of-the-art presented so far. Before discussing the role of exposure to the financing of the two instruments observed, this section compares the findings on the control variables vis-à-vis the observations drawn on the same factors by Flyvbjerg et al. (2005) and Bain (2009).

5.2.1 Project size

The total project investment appears in three nodes in the pruned tree. Table 2 describes the conditioning factors for each node and the distribution results of each group identified.

Table 2 - Nodes determined by the variable investment size in the estimated tree

	Node 1	Node 2	Node 3
Split Variable	total.inv >1250 mi EUR	total.inv > 805 mi EUR	total.inv > 670 mi EUR AND < 805 mi EUR

Upper nodes conditions		Contract.date < 1996	Contract.date < 1996
Outcome	5 below expected traffic	4 cases below expected traffic	4 cases within or above expected traffic

The model shows that the larger projects tend to present overestimated traffic forecasts only after 805 million Euros. Below this threshold, the influence of the investment is not relevant for the differentiation of these projects. This finding is aligned with Flyvbjerg et al.'s (2005) results that find only significance in the logarithmic relation between cost and traffic inaccuracy for railway projects.

5.2.2 Contract date

The model displays three nodes created from the contract signature year variable, which refer to 1996, 2000 and 2010. Table 3 describes the information presented at each node in order of importance.

Table 3 - Nodes determined by the contract signature year variable in the estimated tree

	Node 1	**Node 2**	**Node 3**
Split Variable	contract_date < 1996	contract_date > 1996 and contract_date < 2000	contract_date > 2000 and contract_date <2010
Upper nodes conditions	Total.inv < 1250	ERDPpar < 0.088; Total.inv < 670	ERDPpar < 0.088; Total.inv < 670
Outcome	3 cases within or above expected traffic	3 cases below	2 cases within or above expected traffic

The result of the contract.date factor confirms the findings of previous studies, which showed no consistent improvement in the predictive power of traffic forecasts over the years (Flyvbjerg et al. 2005, Cruz e Sarmento, 2019). Instead, Table 3 shows a non-linear trend for the contract.date effect among projects signed before 1996, between 1996 and 2000, and from 2000 to 2010. This oscillation supports the identification of contextual cycles of project production, quality evaluation and accuracy of traffic forecasts (e.g., pre-1996; between 1996 and 2000; and from 2000 to 2010) (Robinson, 2009, Baeza & Vassallo, 2010). Given the non-linear approach adopted by decision tree models, this finding may be related to the detection of a non-linear behaviour created by other possible latent contextual conditions (e.g., political periods more susceptible to error or even the very behaviour of financing agencies) not expected by the hypothesis of a learning curve evaluated in previous studies.

5.2.3 Transport Mode

Finally, it is observed that the estimated model does not present nodes using the transport mode variables for the differentiation of the projects. This result differs from Flyvbjerg et al. (2005) concerning the trend of demand overestimation by railway projects. However, it is understood that this result may be associated with the identification of behavioural patterns of the demand prediction of project contracts in more significant variables such as project cost, year of contract signature or the financial variables

In summary, the examination of the three control variables shows that part of the explanatory power of the factors remains relevant with the introduction of financial variables. Furthermore, this novel approach presents more explanatory results than previous studies that disregard both a non-linear relationship and the other effects of distinct variables. Although divergent in some cases, the results found in this section may contribute to the study of the topic.

5.3 Exposure variables: the effect of financing sources

This section looks at the model nodes estimated for the two exposure variables EIBpar and ERDPpar. The model presented in Figure 2 identified three grouping nodes defined by these variables.

Starting with the nodes estimated from the European Regional Development Policies Funds participation (ERDPpar) variable, the model shows a share of participation of this financing source that may be associated with the overestimation of demand. The model shows that within the group of projects with a total investment of less than EUR 670M and signed after 1996, all six projects that had an ERDP fund participation between 46.2% and 8.8% presented traffic estimates below the expected, while the two projects that had participation above 46.2% presented estimates above or within the expected.

Still, regarding the participation of ERDP funds, the model shows that below the threshold of 8.8% participation, all three projects signed between 1996 and 2000 presented overestimated traffic forecasts, while the two projects signed after 2010 had demand results within or above expectations. This result reinforces the hypothesis that different investment cycles affect the behaviour of projects and financing mechanisms.

Finally, among the remaining nine projects with a total investment lower than EUR 670M, signed between 2000 and 2010 and with ERDP financing participation lower than 8.8%, the model finds two projects with demand within or above expected. These had a participation of EIB loans above 45.2%. Unlike the ERDP, the EIB participation has a selection bias towards projects with demand above the expected.

6. Conclusion

The examination of the effects of financing structures on project financial sustainability outcomes presents a relevant dimension for a better understanding of the potential failures of a transport infrastructure investment project. The investigation of the role of the financial agent in the selection process of these projects has the potential to ensure a more precise assessment of the possible reasons for deviations in traffic forecasts.

Given the role that traffic forecasts have on the planning and feasibility of transport infrastructure investments, this study aimed to explore the behaviour of traffic forecasts in two different investment contexts. Following optimism bias and previous studies' strategic forecast errors findings, we argued that the relationship between traffic deviations and the financing structure could be an explanatory component of this bias.

Based on an analysis of the relationship between European regional development policies funds' investments and European investment bank loans, we found that differences in the instrument decision autonomy regarding the EU convergence policy and the practice of putting local government guarantees at the centre of the finance decision are related to different probabilities of deviations in demand forecasts for transport infrastructure investment projects. These results support the hypothesis that the financing structure can act as a regulator or enabler of the optimistic expectations and strategic errors strategies in selecting overestimated project finance transportation investments, as observed in previous studies.

Specifically, regarding the EU's public infrastructure financing policies, the findings suggest that the financing bodies should aim to ensure the accountability of local governments for projects put forward for financing bids. One possibility is to introduce local public guarantees in the financing contract. Regarding the financing decisions of the financing bodies, the study suggests that the EU should raise awareness regarding the possible consequences of the convergence policy in the public financing mechanisms.

References

Albalate, D.; Bel, G. & Gragera, A. (2019). Politics, risk, and white elephants in infrastructure PPPs. Utilities Policy, 58(September 2018), 158–165. https://doi.org/10.1016/j.jup.2019.05.001

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012). The 'K'in K-fold cross validation. In 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) (pp. 441-446). i6doc. com publ.

Athias, L., & Nuñez, A. (2008). Winner's curse in toll road concessions. Economics Letters, 101(3), 172–174. https://doi.org/10.1016/j.econlet.2008.07.017

Baeza, M. A., & Vassallo, J. M. (2012). Traffic uncertainty in toll motorway concessions in Spain. Proceedings of the Institution of Civil Engineers: Transport, 165(2), 97–105. https://doi.org/10.1680/tran.2012.165.2.97

Baeza, M. D. L. Á., & Vassallo, J. M. (2010). Private concession contracts for toll roads in Spain: analysis and recommendations. Public Money & Management, 30(March 2015), 299–304. https://doi.org/10.1080/09540962.2010.509179

Bain, R. (2009). Error and optimism bias in toll road traffic forecasts. Transportation, 36(5), 469–482. https://doi.org/10.1007/s11116-009-9199-7

Bouvet, F., & Dall'erba, S. (2010). European regional structural funds: How large is the influence of politics on the allocation process? Journal of Common Market Studies, 48(3), 501–528. https://doi.org/10.1111/j.1468-5965.2010.02062.x

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.

Button, K., & Chen, Z. (2014). Demand forecasting errors and the ownership of infrastructure. Applied Economics Letters, 21(7), 494–496. https://doi.org/10.1080/13504851.2013.870642

Clifton, J., Díaz-Fuentes, D., & Revuelta, J. (2014). Financing utilities: How the role of the European Investment Bank shifted from regional development to making markets. Utilities Policy, 29, 63–71. https://doi.org/10.1016/j.jup.2013.10.004

Crescenzi, R., & Rodríguez-Pose, A. (2012). Infrastructure and regional growth in the European Union. Papers in regional science, 91(3), 487-513. doi:10.1111/j.1435-5957.2012.00439.x

Cruz, C. O., & Sarmento, J. M. (2019). Traffic forecast inaccuracy in transportation: a literature review of roads and railways projects. In Transportation. https://doi.org/10.1007/s11116-019-09972-y
Cruz, C. O., & Marques, R. C. (2013). Risk-sharing in highway concessions: Contractual diversity in Portugal. Journal of professional issues in engineering education and practice, 139(2), 99-108.
Del Bo, C., Florio, M., & Manzi, G. (2010). Regional infrastructure and convergence: Growth implications in a spatial framework. Transition Studies Review, 17(3), 475–493. https://doi.org/10.1007/s11300-010-0160-4
Dellmuth, L. M. (2011). The cash divide: the allocation of European Union regional grants. Journal of European Public Policy, 18(7), 1016–1033. https://doi.org/10.1080/13501763.2011.599972
Domingues, S., & Zlatkovic, D. (2015). Renegotiating ppp contracts: reinforcing the 'p'in partnership. Transport Reviews, 35(2), 204-225.
European Court of Auditors. (2014). Effectiveness of EU-supported public urban transport projects. https://doi.org/10.2865/30848
European Investment Bank. (2015). EIB Project Financed database. Retrieved from http://www.eib.org/projects/loans/index.htm
Flyvbjerg, B. (2005). Measuring inaccuracy in travel demand forecasting: Methodological considerations regarding ramp up and sampling. Transportation Research Part A: Policy and Practice, 39(6), 522–530. https://doi.org/10.1016/j.tra.2005.02.003
Flyvbjerg, B., Holm, M., & Buhl, S. (2005). How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation. Journal of the American Planning Association, 71(2), 131–146. https://doi.org/10.1080/01944360508976688
Ghiasi, M. M., Zendehboudi, S., & Mohsenipour, A. A. (2020). Decision tree-based diagnosis of coronary artery disease: CART model. Computer Methods and Programs in Biomedicine, 192, 105400. https://doi.org/10.1016/j.cmpb.2020.105400
Hernandez, S., Monzon, A., & de Oña, R. (2016). Urban transport interchanges: A methodology for evaluating perceived quality. Transportation Research Part A: Policy and Practice, 84, 31–43. https://doi.org/10.1016/j.tra.2015.08.008
Horowitz, J., & Emslie, R. (1978). Comparison of measured and forecast traffic volumes on urban interstate highways. Transportation Research, 12(1), 29–32.
Lee, M. Y., & Yang, C. S. (2010). Entropy-based feature extraction and decision tree induction for breast cancer diagnosis with standardised thermograph images. Computer methods and programs in biomedicine, 100(3), 269-282.
Liebe, M., & Howarth, D. (2020). The European Investment Bank as Policy Entrepreneur and the Promotion of Public-Private Partnerships. New Political Economy, 25(2), 195–212. https://doi.org/10.1080/13563467.2019.1586862
Loh, W. Y. (2008). Classification and Regression Tree Methods, In Encyclopaedia of Statistics in Quality and Reliability.
Loh, W. Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 14–23. https://doi.org/10.1002/widm.8
Macário, R., Couchinho, R., Costa, J., & Ribeiro, J. (2013). Understanding pitfalls in application of PPPs in transport infrastructure in Portugal.
Næss, P. (2016). Inaccurate and biased: Cost--benefit analyses of transport infrastructure projects. In Crisis System (pp. 58–78). Routledge.
Odeck, J., & Welde, M. (2017). The accuracy of toll road traffic forecasts: An econometric evaluation. Transportation Research Part A: Policy and Practice, 101, 73–85. https://doi.org/10.1016/j.tra.2017.05.001
Oliveira, M., Ribeiro, J., & Macário, R. (2016). Are we planning investments to fail? Consequences of traffic forecast effects on PPP contracts: Portuguese and Brazilian cases. Research in Transportation Economics, 59, 1–8. https://doi.org/10.1016/j.retrec.2016.04.003
Overbeek, H. (2012). Sovereign Debt Crisis in Euroland: Root Causes and Implications for European Integration. International Spectator, 47(1), 30–48. https://doi.org/10.1080/03932729.2012.655006
Parthasarathi, P., & Levinson, D. (2010). Post-construction evaluation of traffic forecast accuracy. Transport Policy, 17(6), 428–443. https://doi.org/10.1016/j.tranpol.2010.04.010
Robinson, N. (2009). The European investment bank: The EU's neglected institution. Journal of Common Market Studies, 47(3), 651–673. https://doi.org/10.1111/j.1468-5965.2009.01818.x

Roumboutsos, A., Farrell, S., Liyanage, C. L., & Macário, R. (Eds.). (2013). COST Action TU1001 Public Private Partnerships in Transport: Trends & Theory P3T3, 2013 Discussion Papers Part II Case Studies. Brussels: COST Office.

Roumboutsos, A., Farrell, S., & Verhoest, K. (Eds.). (2014). COST Action TU1001 Public Private Partnerships in Transport : Trends & Theory, 2014 Discussion Papers Country Profiles & Case Studies. Brussels: COST Office.

Sanko, N., Morikawa, T., & Nagamatsu, Y. (2013). Post-project evaluation of travel demand forecasts: Implications from the case of a Japanese railway. Transport Policy, 27, 209–218. https://doi.org/10.1016/j.tranpol.2013.02.002

Tsami, M., Adamos, G., Nathanail, E., Budilovich, E., Yatskiv, I., & Magginas, V. (2018). A decision tree approach for achieving high customer satisfaction at urban interchanges. Transport and Telecommunication, 19(3), 194–202. https://doi.org/10.2478/ttj-2018-0016

Van Wee, B. (2007). Large Infrastructure Projects: A Review of the Quality of Demand Forecasts and Cost Estimations. Environment and Planning B: Planning and Design, 34(4), 611–625. doi:10.1068/b32110

Vassallo, J. M., & Baeza, M. (2007). Why traffic forecasts in PPP contracts are often overestimated. EIB University Research Sponsorship Programme, EIB Luxembourg.

Wachs, M., & Ortner, J. (1979). Capital grants and recurrent subsidies: A dilemma in American transportation policy. Transportation, 8, 3–19. Retrieved from http://link.springer.com/article/10.1007/BF00149848

Wagenvoort, R., De Nicola, C., & Kappeler, A. (2012). Infrastructure Finance in Europe: Composition, Evolution and Crisis Impact. SSRN Electronic Journal, 15(1), 16–39. https://doi.org/10.2139/ssrn.1829921

Welde, M., & Odeck, J. (2011). Do planners get it right? The accuracy of travel demand forecasting in Norway. European Journal of Transport and Infrastructure Research, 11(1), 80–95. https://doi.org/10.18757/ejtir.2011.11.1.2913

Yescombe, E. R. (2002). Principles of Project Finance. Elsevier Science.