ESTIMATING FARE ELASTICITIES OF RAIL DEMAND IN GREAT BRITAIN USING BAYESIAN INFERENCE

by

SAMIRA MARX PINHEIRO

The dissertation is submitted by the Candidate
in Partial Fulfillment of the Requirements for the Degree of

MSc Transport Economics

Submission by the Candidate does not imply
that its content or standard is endorsed by
the Examiners.

Institute for Transporte Studies

University of Leeds

September 2017

Contents

List of Figures

List of Tables

Executive Summary

The purpose of this study is to present an alternative method to estimate coherent and consistent rail fare elasticities in Great Britain, including cross effects of different types of tickets.

Making usage of Bayesian econometrics, the goal is to overcome the elementary issue of previous studies of estimating correct algebraic signs estimates: negative own elasticities and positive cross elasticities, as it is usual to expect from normal competitors goods.

Because of their features, Bayesian methods provide a simple approach to incorporate prior knowledge in the estimation process and restrict the elasticity domain to assure theoretical consistency of estimates' signs. The literature has shown that they have been successfully applied to market research and elasticity estimation in other industries.

In practical terms, this work studies Bayesian regressions for demand models of each ticket type. In these models, the predictors are the fares and two complementary variables representing the other drivers of demand, in accordance with the PDFH - *Passenger Demand Forecasting Handbook*: gross value added (GVA), covering the effects of the external factors, and generalised journey time (GJT), covering the effects of quality variables. This study has applied the *Rail Users and Drivers Dataset* - RUDD, subsetted for non-London long-distance journeys.

Because there are different circumstances of competition among tickets, the data was subsetted in 4 markets, according to the ticket availability. Market 1 covered all routes for which the available fares were *Standard Full* and *Standard Reduced*. In Market 2 the *Standard Advance* fare was included. For Markets 3 and 4 the *First Class* tickets were included: together as a first class effect in Market 3, and segregated in *Full*, *Reduced* and *Advance* in Market 4.

The market categories provided also a complexity scale. Thus, Market 1 is the simplest one and Market 4 is most complex one - indeed, an estimation that has not been covered in previous studies.

The results have shown that the Bayesian regression has successfully estimated correct algebraic signs for all elasticities in all four markets. However, a drawback that must be mentioned is that, even though the estimation process has achieved satisfactory measures of convergence and autocorrelation for the Markov Chains of Monte Carlo, divergent transitions were reported, which might signalise biased estimates. Because this issue is likely to be re-

lated with the domain constraint applied to the estimates, it was not judged as a harm, since it is a necessary evil for the solution adopted.

Additionally, in which regard the interpretation of the coefficients it was noticeable that as complexity increases in the markets, the estimates have lost in precision, but still better than SURE/OLS estimates. Regarding their magnitude, the results have present some unusual values, which must be further investigated. It should be recognised, however, that a proper analysis of magnitude should draw deeper considerations and such complexity was out of the scope of this work.

The overall conclusion is that, as a proof-of-concept study, this work has demonstrated that elasticities estimated by Bayesian methods potentially have practical application for rail demand forecasting. Further developments can bring it closer to reality with the adoption of dynamic effects - short and long-run, and restrictions on the supply side, particularly for *Advance* tickets.

Beyond that, another development that could bring more precision for the train operating companies forecast is the estimation of elasticities applying hierarchical models, which might allow for the estimation route by route, even when few data are available.

1 Introduction

According to economic theory, for a normal good, the own price-elasticity of demand is expected to be a negative value, and the cross-elasticities of its competitive goods are expected to assume positive values. The estimation of these values with the aid of regression models using sales data and the conventional sampling theory approach, however, often produces frustrating wrong algebraic signs outputs (Liu et al. 2009). The problem of estimation of own and cross price elasticities of demand is already acknowledged in the literature (Griffiths et al. (1988), Geweke (1986), Montgomery & Rossi (1999), Liu et al. (2009)).

In the rail industry, this problem emerges in the estimation of fare elasticities of demand. The freedom to commercially explore fares, establishing market-based prices, and the traditional practice of price discrimination in the rail industry enabled the existence of a wide range of different tickets types. However, identifying how these tickets compete with one another and how their price change affects theirs each other demand has been a challenging task. As reported in previous studies, some works based on ticket sales data have "failed to specify cross elasticities whilst those that did often found them to be either wrong sign or statistically insignificant" (Wardman & Toner 2003, p. 6).

The main root of this problem is being considered as the "high degree of correlation between the fares of different tickets" (Wardman & Toner 2003, p. 6), due to the practice of annual fare revisions, which "compounds the already difficult task of estimating what are relatively small effects" (Wardman & Toner 2003, p. 6).

Different approaches have been tried to overcome the data correlation issue. Theoretical constraints and estimation procedures mixing revealed preference and stated preference data can be named as the main alternatives (Wardman & Toner (2003) and ITS & Systra (2016)). Nevertheless, there still is a lack of a methodology to be widely satisfactory applicable across different markets to bring coherent and consistent elasticity estimates.

Because of those difficulties, fare elasticities are considered to "probably represent the most important area of disagreement on rail demand forecasting" (ITS & Systra 2015), embodying a relevant problem of research. This study aims to expand the spectrum of approaches to estimate fare elasticities in the Great Britain introducing Bayesian inference.

Bayesian econometrics bring considerable advantages. In addition to providing “a more natural interpretation of the results of a statistical investigation than does the sampling theory”, it offers a “formal framework for incorporating prior information” (Griffiths et al. 1988, p. 36) available from economic theory. Another appealing feature regards the straightforward method to apply inequality constraints, contrasting to the quadratic programming alternative (Geweke 1986).

It worths highlighting that Bayesian methods have been a useful tool in marketing, which can be noticed from the discussion in the literature, and commercially, with practical application in retail price optimization (Liu et al. 2009). Leading American retailers - as Target, Walmart, Safeway and Giant Eagle - already make use of Bayesian methods to optimize profits by product category (Liu et al. 2009).

Therefore, the application of Bayesian methods to estimate rail fare elasticities may contribute to the body of knowledge that has been built so far and also contributes to the improvement of the rail industry. Even though this work does not reach promising techniques of Bayesian econometrics - as hierarchical models, it will open the debate performing Bayesian regression models with the aim of achieving coherent algebraic sign estimates - the most elementary issue of estimating estimates.

This work will be structured in five chapters. Following this introduction, chapter two covers the literature review in which the problem of estimation of price effects is debated more deeply, both the general aspects and the specific issues in the rail industry. Additionally, previous studies are revisited to present a holistic view of the evidence so far. Still, a section is dedicated to exploring the Bayesian rationale, which is the basis of this study, and how its application is related to the research problem.

Chapter three comprises the methods applied in this work. It presents information regarding the dataset and the definition of the econometric model. For being an inherent characteristic of the Bayesian inference, sections are dedicated to justifying the choice of prior densities of the model's parameters and the likelihood function.

Chapter four regards the results. It will be assessed the consistency of the estimates' signs and compared with the ordinary least square estimates. Additionally, the uncertainty of the estimation is also appraised, discussed in terms of the range of credible values they may assume - the *high density intervals*. Lastly, an initial debate about the magnitude of estimates is attempted. It will be more descriptive and superficial since a proper analysis of the 95 estimated coefficients holds a complexity that can not be comprised in this work.

In the last chapter, conclusions are discussed and further developments are suggested.

2 Literature Review

2.1 The rail passenger demand

The Passenger Demand Forecasting Council - PDFC - is an association of the main stakeholders of the rail industry in Great Britain: train operating companies, Network Rail, Department for Transport, Transport Scotland, the Office of Rail Regulation, Transport for London, the Urban Transport Group, RSSB, HS1, HS2, Rail North, and Welsh Government. The purpose that brings these institutions together is to improve their knowledge about passenger demand (Rail Delivery Group 2015). The motivation is very natural since increasing knowledge about demand allows to create more business opportunities, raise efficiency and profits.

The PDFC's body of knowledge is the Passenger Demand Forecast Handbook - PDFH, which "summarises over twenty years of research on rail demand forecasting" (Rail Delivery Group 2015). It compiles dozens of studies regarding rail passenger demand and is "recognised within the industry as the key source of evidence in the area" (Rail Delivery Group 2015).

Therefore, it worths taking a step back to portrait a holistic perspective of the rail demand phenomenon according to the PDFH concepts. The broad elements that affect rail demand, also known as drivers of demand, are presented in Figure 2.1.

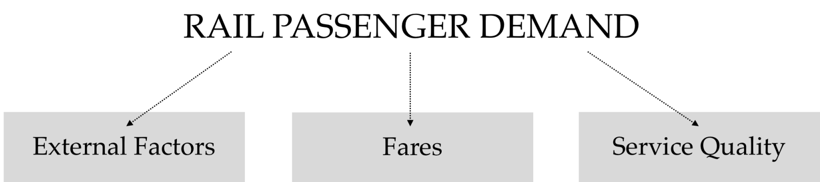


Figure 2.1: Drivers of rail passenger demand.

Source: PDFC (2005), Chaper B0, page 1 [adapted]

The first group, as the name says, comprises aspects that are not under control of the train companies. In broad terms, they are related to economic, socio and demographic aspects of the region. Additionally, the availability, quality and cost of competitive modes - car, bus, coaches and air services, are considered.

The second group regards the fares. Despite the evident impact of fares in the demand, the PDFC has been struggling in quantify this impact when it regards how different type of tickets affect their demand, which is the central issue of this work. It will be convenient to extend the discussion about the different types of fares to better interpret the issues around the estimation of the relationship between demand and fares. It will be done in the next section.

The third group involves the service quality and includes the biggest range of sub elements. The PDFH subsets it as four aspects:

(i) Timetable and related features, which covers journey time, frequency and interchange and is usually summarized in the *generalised journey time* measure.

(ii) Non-timetable attributes, which in turn comprises overcrowding, rolling stock quality, onboard facilities, station facilities, provision of passenger information, staff and security, cleanliness, advertising and promotion.

(iii) Reliability, which covers punctuality and cancellation.

(iv) New services and access, which regards the potential demand hold by new services in new or existing routes, new station and also the effects of improving accessibility to stations.

It is noticeable, thus, that the rail passenger demand is considered to be implicated by several factors. For the fare elasticity estimation, however, the demand model is usually simplified to hold fewer elements that are deemed to represent the mass of the covariance of the others drivers of demand.

Equation 2.1 presents the short version of demand model.

$$V_i = a \ \mathrm{GVA}_i^g \ P_i^f \ \mathrm{GJT}_i^\gamma \tag{2.1}$$

where
V_i is volume of journeys;
a is a constant;
GVA_i is the gross added value;
P_i is the price of the ticket;
GJT_i is the generalised journey time;
g is the GVA elasticity of demand;
f is the price elasticity of demand;
γ is the GJT elasticity of demand.

This simplification is relevant because for the estimation of individual fare elasticities it is necessary to expand the component price for as many fares as exist, which can make the model inherently complicated. Thus, to keep it simple without losing much information, previous studies have established this simple version (ITS & Systra 2016), which will also be adopted in this study.

Additionally, as one may notice, the replacement of GDP by GVA is due to the effort to regionalize the income effect across routes (PDFC 2011). Therefore, instead of applying a national GDP, GVA data is more suitable, since it is available in the NUTS [1] 3 level.

2.2 Rail fares in the UK

2.2.1 Rail fares taxonomy

Standard, first class, saver, super saver, single, return, season, peak, off-peak, advance. Despite one may judge the taxonomy of rail fares to be self-explanatory at a first glance, it can be confusing to structure them into a cohesive system because of the diversity of names and classification used in different types of documents.

After gathering different nomenclatures and definitions, they were compiled in Table 2.1 as a proposed taxonomy to be adopted in this study to interpret the RUDD's dataset classification. Because season fares will not be part of the scope of this is study they were not considered.

The three types of fares *Full*, *Reduced* and *Advance* exist in both *First* and *Standard Class*. It will be common to abbreviated these names to their respective initial - F, R and A, and represent the class as 1 for *First* and 2 for *Standard Class*. Thus, concatenating numbers and letters it is possible to make short reference to the fares - $1F$, $1R$, $1A$, $2F$, $2R$, $2A$. An exception for this taxonomy will be the abbreviation $1N$ (*First Class Non-season*) used to refer to the three first class fares together (*Full, Reduced* and *Advance*).

The relevance given to taxonomy is due to the importance of correctly understanding the differences and particularities of tickets to properly interpret the market segmentation and the consumer behaviour. In the same sense, the next section approaches regulatory affairs since they may be useful to contextualise the results.

[1] Nomenclature of territorial units for statistics.

Table 2.1: Fares categories and conditions

Type	Coverage	Conditions
Full	*full* or *anytime* or *open*, includes single and return, day single and day return.	Passengers can take any train.
Reduced	*reduced* or *off-peak* or *saver/super saver*, includes single and return, day single and day return for off-peak and supper off-peak.	Passengers can take any off-peak train. Peak time may vary from route to route.
Advance	*advance* or *apex*, sold only as single tickets.	Passengers can only take the specific train of the ticket. Must be bought in advance and has limited availability

Source: Own work

2.2.2 Regulated and non-regulated fares

The rail fares in the Great Britain can be regulated, in which case they are set by the franchising authority, or unregulated, in which case they are set by the train operating companies on a commercial basis (Butcher 2017).

According to the last review, run by the Strategic Rail Authority in 2003, there are two groups of regulated fares: the *protected fares* and the *commuter fares* (Butcher 2017). The regulated fares are set in a price-cap mechanism, where the X may vary according to the objectives of the regulatory policy. Since 2004 it became RPI+1% (Butcher 2017).

The *protected fares* seems to serve the broad guidances of regulation policy as they assure reasonable and affordable fares (Butcher 2017), whilst allows train companies to commercially exploit the route through market segmentation and unregulated fares. According to the taxonomy adopted in the SRA's report (SRA 2003), this group of fares comprises:

- *Saver returns*, which regards off-peak services, available for most long-distance journeys. In the taxonomy shown on Table 2.1 it would be classified as *Reduced*.
- *Standard returns*, which regards the full-fare and allows the passenger to travel anytime, but only for journeys without the *saver* option. In that taxonomy shown on Table 2.1 it would be classified as *Full*.

- *Weekly season*, others than the ones included in the *commuters fares*. In the taxonomy shown on Table 2.1 they would not appear since they regard season fares and others not covered in this study.

It is interesting to notice that the regulated fares cover at least one type of ticket that must be available, so a passenger without time restrictions can access the basic service at affordable prices. There is left commercial spots to be price discriminated, targeting users with more strict preferences.

The regulation of *commuter fares* affects mostly season tickets, as well other types as standard single and return within the London Travelcard area or from a pre-defined area of London's suburbs (SRA 2003). Because the London area and commuters tickets are out of the scope of this work, they were mentioned just for completeness, but no light will be shed on them.

2.3 Fare elasticities of demand

2.3.1 Definition

The fare elasticities of demand represent the relationship between the demand for a type of fare and the prices of the fares available in the market - how much the quantity demanded changes when these prices change. It is calculated as the percentage change in quantity demanded by the percentage change in prices, as shown in Equation 2.2.

$$\eta = \frac{\%\Delta Q}{\%\Delta P} \tag{2.2}$$

where
η is the price elasticity of demand
Q is the quantity demanded
P is the price

The fare elasticities of demand may regard the change in the demand of a good A given a change in the price of the good itself, which is called the own-elasticity, shown in Equation 2.3; or it may regard the change in the demand of a good A given a change in the price of good B, which is called cross elasticity, shown in Equation 2.4.

$$\eta_{AA} = \frac{\%\Delta Q_A}{\%\Delta P_A} \tag{2.3}$$

where

η is the price elasticity of demand of good A with respect to changes in the price of good A;

Q is the quantity demanded of good A;

P is the price of good A.

$$\eta_{AB} = \frac{\%\Delta Q_A}{\%\Delta P_B} \tag{2.4}$$

where

η is the price elasticity of demand of good A with respect to changes in the price of good B;

Q is the quantity demanded of good A;

P is the price of good B.

According to economic theory, the own-elasticity of normal goods, say good A, is expected to be negative indicating that the more the price of good A rises, the lower its demand will be. Conversely, the cross-elasticity of a substitute of a normal good A, say good B, is expected to be positive indicating that the more the price of good B rises, the higher the demand of the good A will be because consumers tend to trade-off to the cheaper good.

The degree of competition between two goods depends upon how close substitutes they are. If two goods are close substitutes, switching from one to another is likely to happen so they become very price sensitive. In other words, if two goods are close substitutes they should have high cross price elasticities.

According to the PDFH, there is a belief that the range of fares covered in this study ($1F$, $1R$, $1A$, $2F$, $2R$ and $2A$) are substitute goods, since they assume that for these fares "the own elasticity will be negative and the cross elasticities positive" (PDFC 2005, p. 8, Chapter B2).

In fact, they could be regarded as substitutes because ultimately they provide a transport service from one place to another. However, they might differ in the degree of competition, since particularities and specific conditions of each fare are supposed to segregate these markets.

The extent of this segregation is valuable knowledge. Investigating the fare elasticities of different types of tickets may allow identifying how close these markets really are.

Additionally, it may reveal different dynamics with regard to the others drivers of demand. For instance, customers of a given type of fare may be more sensitive to service quality than others.

2.3.2 Difficulties and issues

As introduced in Chapter 1, estimating fare elasticities is not straightforward. Indeed, it is being considered the most important area of disagreement on rail demand forecasting" (ITS & Systra 2015, p. 6). This is so because the traditional estimation of elasticities with the aid of ordinary least squares (OLS) models is hindered by the high correlation of the fares.

Even though it does not cause bias, the high correlation is a known problem in OLS estimation. When the independent variables are highly correlated it is difficult to isolate individual effects of each variable since most of the variation will be common to both. It leaves the OLS with little information to estimate and because of it, the output might be poor estimates.

Another way to interpret the high correlation problem regards the fact that when regressors are highly correlated, it is difficult to identify and address the explanatory power among the variables. This fluctuation causes high variances in the estimators causing them not to be precise. It is like their effects in the dependent variable are blurred together and there is low certainty about which one holds which part of it.

In Bayesian inference, however, correlation gains a new perspective. When two predictors are correlated, their coefficients tend to be anti-correlated, which means the bigger one is, the smaller the other will be. In other terms, one can say that "correlation of predictors causes estimates of their regression coefficients to trade-off" (Kruschke 2014, p. 513).

That feature can be very useful depending on how the prior distributions are defined in the Bayesian model. When non-informative priors are adopted the blurred effect - the large variance, usual to the OLS estimation - is reflected in both posterior density of the estimated coefficients. This means that both coefficients will be estimated with a large range of credible values. However, when a strong prior distribution is defined for one coefficient, which means that one is applying a strong belief to constraint its value, it simultaneously constraints the value of the correlated variable, because of their anti-correlation.

Therefore, when variables are strongly correlated, the constraint applied to one coefficient may "propagate to the estimates of regression coefficients on other predictors that are correlated with the first" (Kruschke 2014, p. 525) and correlation may not be a problem anymore.

There is also the risk of the predictor variables being correlated with omitted variables. In this case, for both methods, it may mislead the interpretation of the coefficients. It worths recovering that an omitted variable, when correlated with a predictor, impacts the predictor estimation to the extent of their correlation (Studenmund 2011). Because the remedies to avoid omitted variable bias are related to the theoretical consistency of the model, which was adopted from previous studies, it will not be covered here.

2.3.3 Previous studies

Several attempts have been tried to solve the issue of estimating fare elasticities applying ticket differentiation. Wardman & Toner (2003) were the first to estimate consistent own and cross-elasticities, with the correct sign and statistically significant. The successful model applied two theoretical constraints known as "Slutsky Symmetry" (Wardman & Toner 2003) and "Dodgson Relationship" (Wardman & Toner 2003). However, the weakness of this study was the absence of *Advance* ticket, "although they were very much in their infancy" (ITS & Systra 2016).

A later study of ITS & Systra (2016) have recovered this method to update the evidence and introduce the *Advance* fare in the estimation. However the conclusion was that "econometric by itself was not able to estimate definitive fare elasticities, but market research would be required in addition" (ITS & Systra 2016, p. 93).

The mix of revealed preference data with stated preference data turns out to be the more consistent approach. Nevertheless, the authors have concluded that this method has presented a weak performance in some markets segments - especially short distance, which brought vulnerability to it.

As one may notice what all these methods have in common is that they were based on the sampling theory approach. However, "unrestricted least-squares estimates of own- and cross-price elasticities are often of incorrect sign and unreasonable magnitude" (Montgomery & Rossi 1999, p. 413) and even the usage of contrivances have not achieved satisfactory estimates that can be generalised. Because of the struggle of the methods attempted so far, it was considered reasonable the exercise to estimate the fare elasticities with the aid of Bayesian regressions.

2.4 The bayesian approach

The Bayesian promise for elasticity estimation is primarily a solution "to get rid of the wrong signs" (Griffiths et al. 1988, p. 36). The advantage of the Bayesian econometrics is the simple way to incorporate prior information from economic theory (Griffiths et al. 1988), which helps to achieve reasonable estimates.

Beyond that, hierarchical Bayes model has been reported as successful solutions for more complex elasticities estimation (Liu et al. 2009). They are useful to problems when it is possible to identify "meaningful hierarchical structure" (Kruschke 2014, p. 221), in which the parameters can be estimated with the aid of layers of information funnelling from general to specific. For instance, Montgomery & Rossi (1999) have estimated price elasticities of multiple brands and stores with the aid hierarchical model. In this work, the authors are aware of the problem of customizing elasticities estimation from a national level to a regional market, or even store to store, which is the description itself of the hierarchical nature of the problem. The results were reported by the authors as "reasonable cross-price elasticities estimates while retaining much of the interesting and potentially valuable store-to-store variation" (Montgomery & Rossi 1999, p. 414).

As seen, the applications of Bayesian econometrics to elasticity estimation are full of potential. However, because this is an introductory work on the subject, hierarchical models will be out of scope, and the focus will be strict to the simplest application o Bayesian regression. The next section work on basic concepts of it.

2.4.1 Bayes' theorem

The Bayesian approach is based on updating the credibility of a prior belief in the light of relevant information, which is the data observed. This abstract rationale becomes a statistical tool through the Bayes' theorem, which is presented in Equation 2.5.

$$\underbrace{p(\theta \mid D)}_{\text{posterior}} = \frac{\overbrace{p(\theta)}^{\text{prior}} \cdot \overbrace{p(D \mid \theta)}^{\text{likelihood}}}{\underbrace{p(D)}_{\text{evidence}}} \tag{2.5}$$

where

θ is the parameter under study; and

D is tha data observed. It can be thought as a random variable Y with observations y_i

The $p(\theta|\text{Data})$ is called the posterior probability density function and represents the updated belief regarding the parameter θ, given an observed data. It is the answer to questions like "Given the data, what do we know about θ?" (Koop 2003).

The $p(\theta)$ is called the prior probability density function. It represents the pre-existent belief regarding θ, unconditional to the observations. The prior belief may be defined due to a theoretical background or to any kind of accumulated knowledge of the subject.

That is often a critique of the Bayesian approach, being considered a subjective element in the analysis which may pollute the final results. However, a good effect of it is that it demands the researcher to have a prior interpretation of the phenomenon under study and consider reasonable values that it may assume. Indeed, the practice of interpreting a problem before looking at the estimated parameters of a model is highly recommended in the frequentist inference (Gujarati & Porter (2009), Kennedy (2003), Studenmund (2011)).

Also, it is important to highlight that the establishment of a prior probability density might not be a mere researcher's opinion. On the contrary, it must be justified to a "sceptical and scientific audience" (Kruschke et al. 2012). Even whether sceptics disagree about the distribution of a prior, there is still the advantage of measure the impact of the disagreement testing the impact of different priors on the final estimate.

The $p(D|\theta)$ represents the likelihood function, which is a function of the parameter θ for a given data Y assuming that θ follows a given probability distribution. Gujarati & Porter (2009) teaches that one may think the likelihood and the probability density function as related functions that are both composed of three parts: i) the parameter under study, θ; the data Y, whose observations are y_i; and the probability distribution of the parameter under study - how it is expected to behave. It may be useful thinking the function as the analogy of a machine with inputs and outputs. When the input is a possible value of the parameter under study θ, it is the likelihood function, which uses the other two pieces of information to generate the output - the likelihood. Whether the input is the observed data, y_i, it is a probability function, which uses the other two pieces of information to generate the output - the (joint) probability of observing y_i.

The last element in the Bayes' theorem is the $p(D)$, which represents the observed data. It may be interpreted as the unconditional probability of observing a given Y, in the whole universe of possibilities. However, because the interest regards in learning about the parameter under study θ, and $p(D)$ is not related to it, it is usually not considered in the analysis.

Because of that, it is usually said that the posterior probability density is *proportional* to the prior probability density and the likelihood, as shown in Equation 2.6 (Koop 2003).

$$p(\theta|D) \propto p(\theta) * p(D|\theta) \tag{2.6}$$

2.4.2 Running a bayesian regression model

Model definition

Expanding the Bayesian theorem to a practical application in a regression model does not change the Bayes' rationale. However, it may be useful making some explicit considerations since, instead of a unidimensional problem - as presented before, this study will regard a multi dimensional problem because the interest is to estimate several parameters together.

To easily illustrate, consider a regression model with two parameters of interest β_0 and β_1, as given by Equation 2.7.

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} \tag{2.7}$$

Assuming that β_0 and β_1 are normally distributed with mean M_0 and M_1 and standard deviation S_0 and S_1, respectively, as shown by Figure 2.2, then

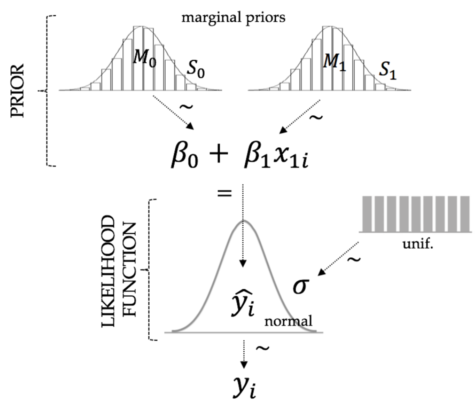


Figure 2.2: Bayesian regression scheme.

Source: Kruschke et al. (2012) p.727 [adapted]

the characterization of the prior probability density would be a three-dimensional space, as shown by Figure 2.3. One may notice that the prior is a joint probability of β_0's and β_1's individual - or marginal - prior probability densities, which integrates to 1.

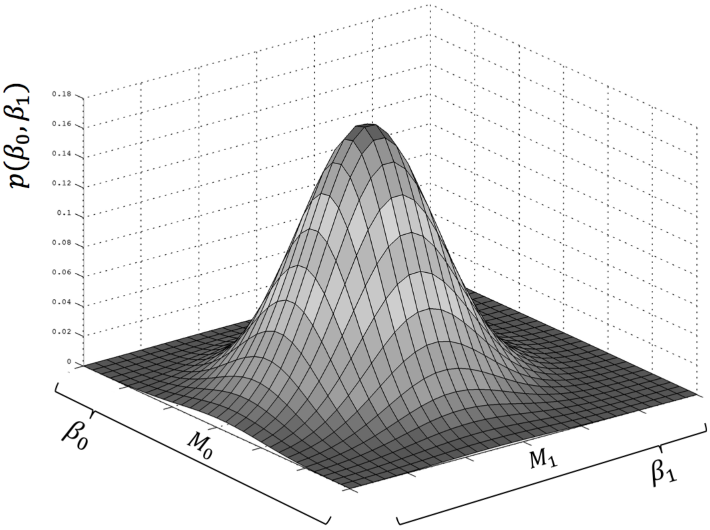


Figure 2.3: Prior probability density for parameters β_0 and β_1.

Source: Own work

The second element to define a Bayesian regression model is the likelihood function, also illustrated in Figure 2.2. In this example, it was assumed that the predicted variable Y, which has y_i realizations, is normally distributed with mean $\hat{y}_i$ - which can also be interpreted as $\hat{y}_i = \beta_0 + \beta_1 x_{1i}$, and standard deviation σ. The arrow linking the βs with the $\hat{y}_i$ shows that there is an equality between it, so y_i is a linear function of the explanatory variables.

A formal way to state this model could be as presented in Equation 2.8.

$$
\begin{aligned}
\mathrm{y}_i \sim & Normal(\hat{y}_i, \sigma) && \text{[likelihood]} \\
\hat{y}_i = & \beta_0 + \beta_1 x_{1i} && \text{[linear model]} \\
\beta_0 \sim & Normal(M_0, S_0) && \text{[prior]} \\
\beta_1 \sim & Normal(M_1, S_0) && \text{[prior]} \\
\sigma \sim & uniform(a, b) && \text{[prior]}
\end{aligned}
\tag{2.8}
$$

Estimation of the posterior

Once defined these elements in the Bayesian regression framework, the result will be a posterior probability density. Analogously to the prior, the posterior will be multidimensional probability space which integrates to 1, similarly to Figure 2.3 - the exact shape of the posterior will depend on the prior and the likelihood.

Still analogously to the prior probability density, the posterior is a joint probability space compounded by the marginal posterior distribution of each predictor, in this example β_0 and β_1.

The posterior probability density is generated by approximation by collecting from it a large representative sample. This method, called Markov Chain of Monte Carlo - MCMC, is useful when the analytical solution is not possible given the complexity of the posterior distribution (Kruschke 2014). "It is the MCMC algorithms and software, along with faster computer hardware, that allows Bayesian data analysis for realistic applications that would have been effectively impossible 30 years ago"(Kruschke 2014, p. 144).

The core idea of an MCMC lies in the generation of a sequence of samples in the probability space through a random walk process - it may vary among software and algorithms. Each sample is virtually equivalent to a step in the random walk, in which is defined a value for all predictors in the model. For instance, in the example used so far, a step would contain a pair of β_0 and β_1 (β_0, β_1).

The random walk is the way the algorithm explores the probability space of possible values. To decide whether or not to accept a step, the algorithm compares the probability of the current step with the probability of the proposed step and applies a decision rule - which may also vary among software and algorithms. These probabilities used in to decide whether or not a proposed step should be accepted are computed from combinations of the prior and the likelihood - recover Equation 2.6.

From a given position, the next step is accepted as credible value if the combined probability of the likelihood and the prior is bigger than the current step. If the next step has the lower probability, then a specific decision rule determines whether it is accepted or not. The practical effect is that the MCMC always accepts a step - and keep it as a sample - to explore regions where the probability space has high probability and only accepts part of the steps in regions with lower probability. With long MCMC, there would have been enough steps to provide a good approximation of the posterior probability density.

3 Methods

3.1 Data

The available dataset for the current study is a subset of the *Rail Users and Drivers Dataset* - RUDD, for non-London long distance (over twenty miles) journeys.

Th RUDD “includes just over a twenty thousand flows, for twenty-one years (1994/95 to 2013/15), with each flow including more than 900 variables” (DfT 2016, p. 135). In the current application of RUUD, the data regards 6184 bi-directional origin-destination pairs with annual information volume of journeys and revenues, and some drivers of demand. Figure 3.1 illustrates all stations comprised in these OD pairs.

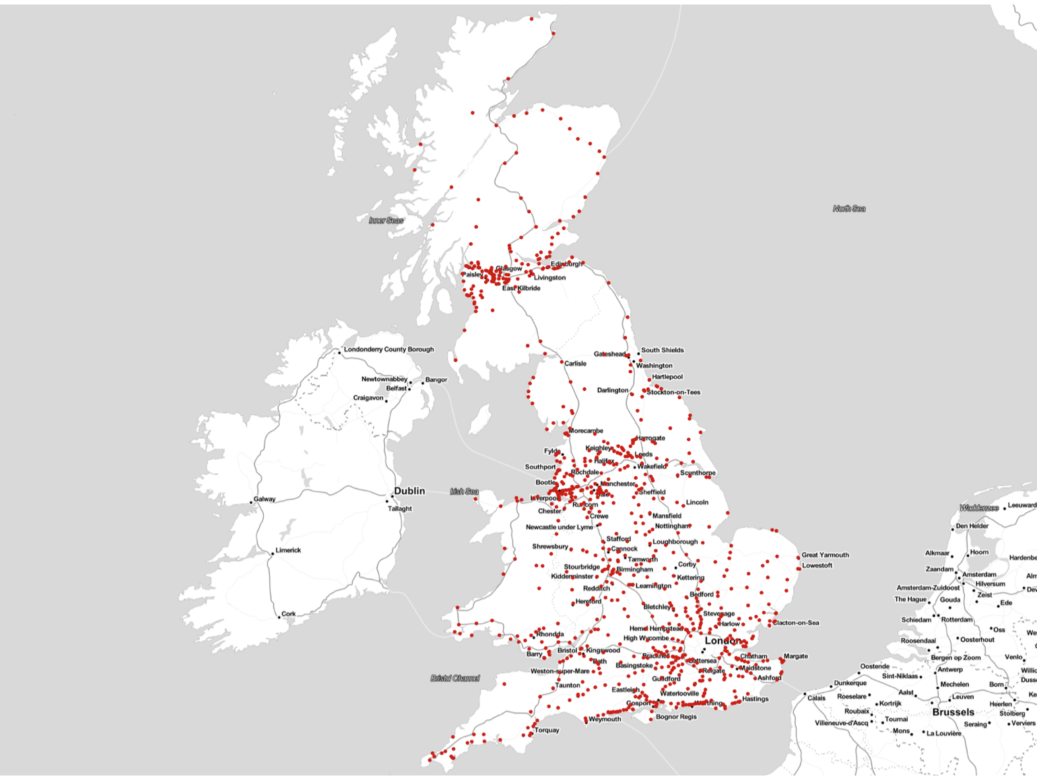


Figure 3.1: Stations covered in the 6.184 OD pairs of the RUDD’s subset

Source: Own work

3.1.1 Variables

As discussed in Chapter 2, the variables of interest for the current work will be journeys, fares, generalised journey time, and gross added value. This section is dedicated to discuss these variables, what they mean and some particularities about what they actually represent.

To start it should be highlighted that the variable fares are not directly collected. Instead, it will be approximated by the average revenue per journey. The fact that the main variable in the study is not actually known may represent a vulnerability, nevertheless, it would not be further debated since it is the best information at hand.

Another fact is that, because the dataset is bi-directional, it means that, for example, a trip from Leeds to York and from York to Leeds are identified by the same code, and the variables of journeys and revenues regard both directions summed up. Also, the GVA was collapsed into a single measure that reflects the average GVA of the origin and destination regions. It is, therefore, helpful to consider the OD pair as a connexion, or a route, since the variables will always regard the pair, instead of the location of the origin or the destination by itself. Table 3.1 summarises information on the variables.

Table 3.1: Definition of variables

Variable	Measure and Unit	Conexion Equivalent Measure
Journey	annual volume of tickets sold, by fare type.	sum of origin and destination values.
Fares	total annual revenue divided by annual volume of tickets sold, by fare type, 2014 values.	sum of origin and destination values.
GJT	sum of journey time, frequency penalty and interchange penalty, as defined in PDFH, in minutes.	-
GVA	regional GVA, at NUTS 3 level, in millions of pounds, 2014 values.	simple average of origin's and destination's GVA.

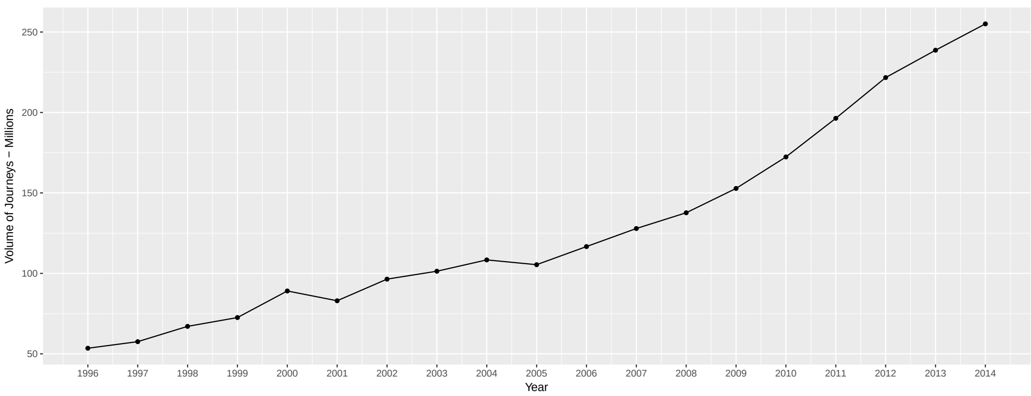
Source: Own work

Journeys

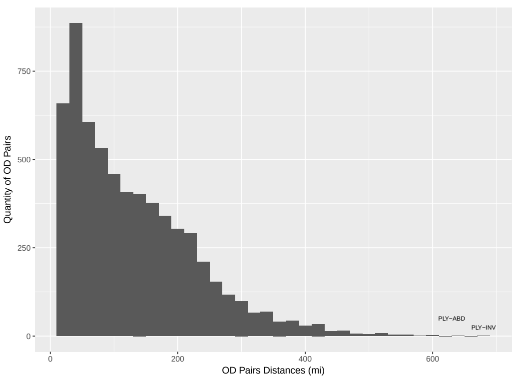
The variable *journeys* is the number of tickets sold per year in each bi-directional route. It does not precisely reflect the number of trips since some tickets allow the passengers to break the journey in a route and make stopovers. Nevertheless, it is a good measure to represent the rail demand.

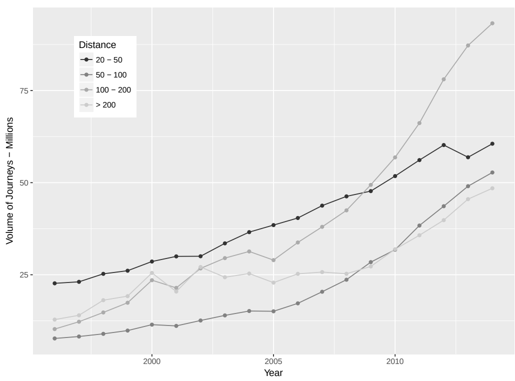
To understand the dynamics of this variable Figure 3.2 brings two dimensions of journeys: its trend across time and a characterization of journeys by distance.

As shown in Figure 3.2a the volume of journeys have consistently increased in the past years. But which kind of journeys are these? Figure 3.2b shows that, despite there are very distant routes, as Plymouth to Inverness or Plymouth to Aberdeen, the mass of routes covered by the study peaks around 50 miles and is decrescent as the distance increases. Anyhow, one can affirm that no matter is the class of their length, the aggregate volume of journeys is increasing over the years for all of them, especially for journeys between 100 and 200 miles, as shown by Figure 3.2c.



(a) Aggregated growth of journeys





(b) Distribution of OD pairs' distance (c) Growth of journeys per distance class

Figure 3.2: Explanatory Analysis on *Journeys*

Source: Own work

Fares

In which regards the variable *fares*, Figure 3.3 illustrates the range of values in universe of fares for all routes (OD pairs) comprised in the dataset. To make them more comparable they were converted in pounds per mile.

The top row covers the first class fares, separated by type of fare - *full*, *reduced* and *advance*. The bottow row refers to the standard class.

From these graphs one may notice that the range of *full* fares is increasing over the time, both for first and standard class. The *reduced* fare, in turn, shows a significantly smaller range in the standard class, whilst for the first class, it does not seem to follow a pattern. The restricted range of the standard reduced fares may be due to the ceiling imposed by regulation, as discussed in Chapter 2.

Lastly, in which regards the *advance* fares, which is a mix of discounted *full* and *reduced*, they present a steady median, slightly slower than the reduced fares but with a wider range in the standard class. For the first class, despite the lack o pattern for the *reduced* fare, the advance also shows a steady median, lower than the *full*, with a step change in the year 2000, which was also observed in the *reduced*.

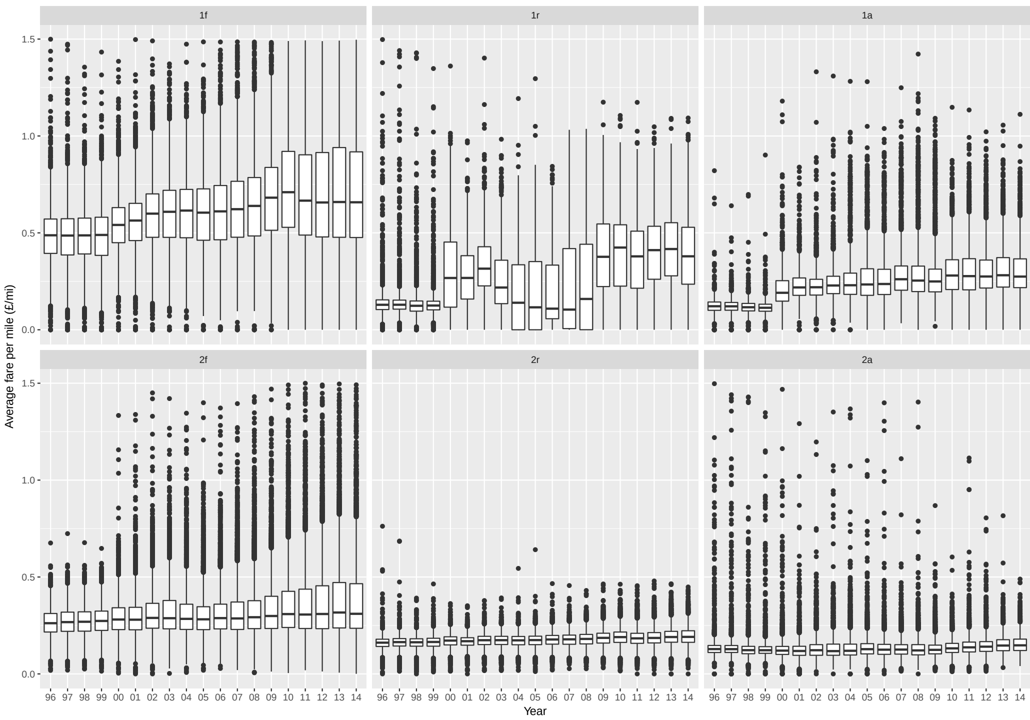


Figure 3.3: Dispersion of fares per mile accross the OD pairs

Source: Own work

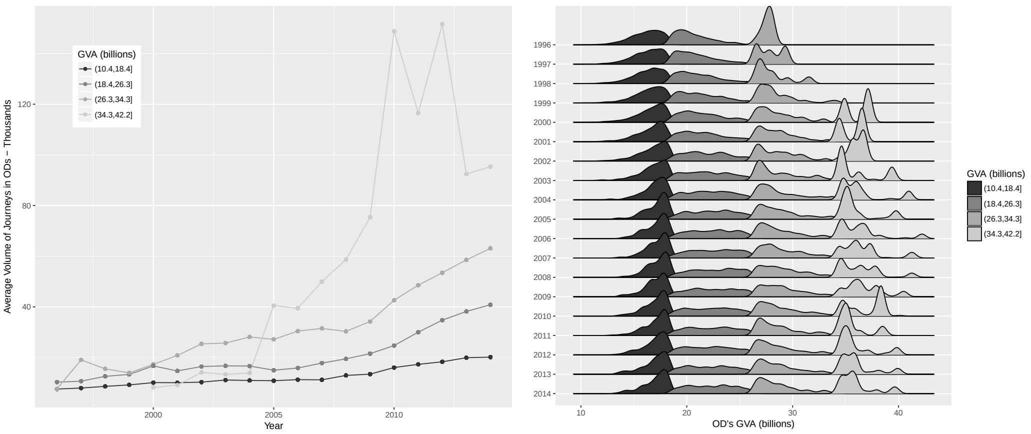
Gross Added Value - GVA

The GVA is a metric for the level of economic activity in a given region. Because rail demand is expected to have a positive polarity with economic activity (PDFC 2005) - the higher GVA, the higher volume of journeys - it is interesting to check this dynamic on the data.

To visualise it, the variable GVA was segmented in four levels. For each level, the average volume of journeys was calculated per year. As shown in Figure 3.4a, the higher the GVA level, the higher the average volume of journeys per route. Also, it is possible to notice an overall trend of growth of journeys for all levels of GVA.

To complement the visualization, Figure 3.4b brings more information about the range of GVA among the ODs pairs and how was the dynamic across the years. Because the OD pairs in the dataset remained constant over the years, with very few exceptions, one can notice that, in general terms, there was a trend towards the right. That means that regions are escalating to levels of higher economic activity.

For the lowest level of GVA, at the left, it seems that have occurred a concentration movement which peaks very close to the boundary of the next level. Eventually, these OD pairs may have changed to next level of GVA. Also, the second and third levels seem to have spread towards the right. Lastly, the highest level of GVA appeared only in the year 2000 and seems to have a mild movement of expansion and contraction.



(a) Growth journeys by level of GVA (b) Distribution of OD pairs' GVA

Figure 3.4: Exploratory Analysis on *GVA*

Source: Own work

In short, eyeballing the *GVA* variable it is possible to identify consistency between the theoretical relation of economic activity and rail journeys in the data.

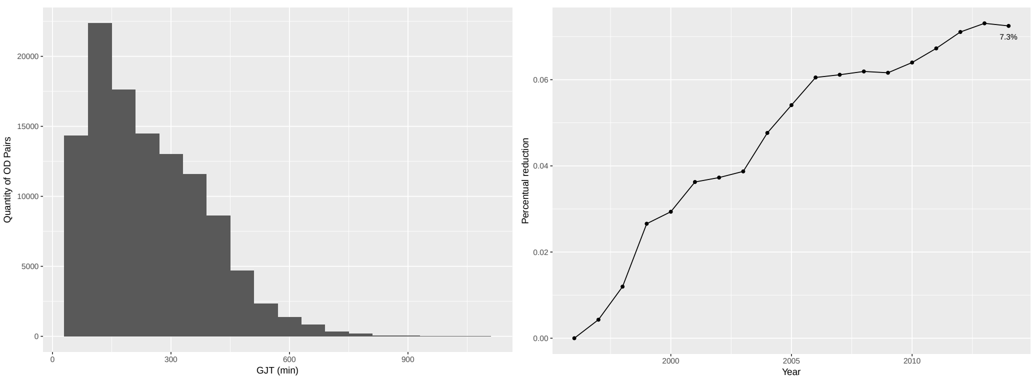
Generalised Journey Time - GJT

In this work, the GJT measure represents the quality-related drivers of demand. It comprises the journey time plus the frequency and interchange penalties, as defined in PDFH, and it is measured in minutes.

To understand the dynamics of this variable Figure 3.5 brings two dimensions of it: an overview of the GJT of the OD pairs considered in the study and the accumulated average reduction of GJT over the years, taking the year 1996 as the base.

Figure 3.5a shows the overall distribution of routes by GJT. As a general characterization, it is possible to observe that the mass of routes peaks around 150 minutes and decreases as the GJT increases, achieving extreme values over 1,000 minutes.

Additionally, Figure 3.5b shows how, in average, routes are reducing their GJT over the years. The values plotted in the graph are accumulated reductions with respect to the year 1996. It is observed an average reduction of 7.3% in the GJT in 19 years (1996-2014).



(a) Distribution of GJT across the routes (b) Accumulated Reduction of GJT

Figure 3.5: Exploratory Analysis on *GJT*

Source: Own work

3.1.2 Subsetting markets in the dataset

The availability or not of a different set of tickets across the routes creates different markets environments. When there are more types of fares the competition increases and the market share of tickets is affected simply because there are more options at one's disposal.

Therefore, the existence of different sets of fares in a route affects the fare elasticities. For example, the fare elasticity of reduced tickets with respect to the demand for full tickets may differ whether there is or not the first class ticket.

The dataset was subsetted into four markets to properly estimate the fare elasticities for each circumstance. The two first will regard only the standard class, and in the other two, first class will be added.

With the introduction of first class tickets in the Markets 3 and 4, their complexity has increased significantly. To overcome that, the coverage of the dataset was shortened to a regional level to reduce the unobserved heterogeneity and provide a more well-behaved data. The regions [1] chosen were Scotland and Yorkshire and Humber, so only routes which both origin and destination in these regions were considered. There was no strong justification for choosing these regions beyond the fact that any interference with London flows was avoided.

An alternative to keep the broad coverage and reduce the heterogeneity that may exist across regions could be the adoption of dummy variables to capture these unobserved characteristics. However, this option was not adopted because in the Bayesian framework it would represent an increase of 10 marginal posterior distributions (from the 11 regions), in each regression model. It was considered, thus, that such complexity was needless since this is an introductory study of Bayesian econometrics in the field. It is important to be aware, however, that a cut in the dataset is a stronger isolation of unobserved characteristics than regressing with dummy variables.

Table 3.2 presents summary information on the resultant dataset for each market.

Table 3.2: Market subsetting

Market	Fares	OD Pairs	Time Series	Obs	Coverage
1	2F, 2R	2,074	1996 - 2014	15,481	Great Britain
2	2F, 2R, 2A	1,550	2000 - 2014	5,909	Great Britain
3	1N, 2F, 2R, 2A	516	1996 - 2014	5,783	Scotland - Yorkshire and Humber
4	1F, 1R, 1A, 2F, 2R, 2A	251	1996 - 2014	1,738	Scotland - Yorkshire and Humber

Source: Own work

[1]NUTS 1 level regions.

It should not be expected, however, that the quantity of OD pairs shown in Table 3.2 to be constant over the years because the availability of tickets for a given route is not fixed in time. It may happen that the quantity of ODs pairs in each market varies from year to year. For example, if in a given route the first class ticket started being commercialised in the year 2000 onwards, this OD pair was considered in market 1 or 2 before 2000 and in market 3 or 4 after that. The important is that the pool of each market is coherent with their actual competition conditions. For completeness, Figure 3.6 illustrates the variation of OD pairs in each market.

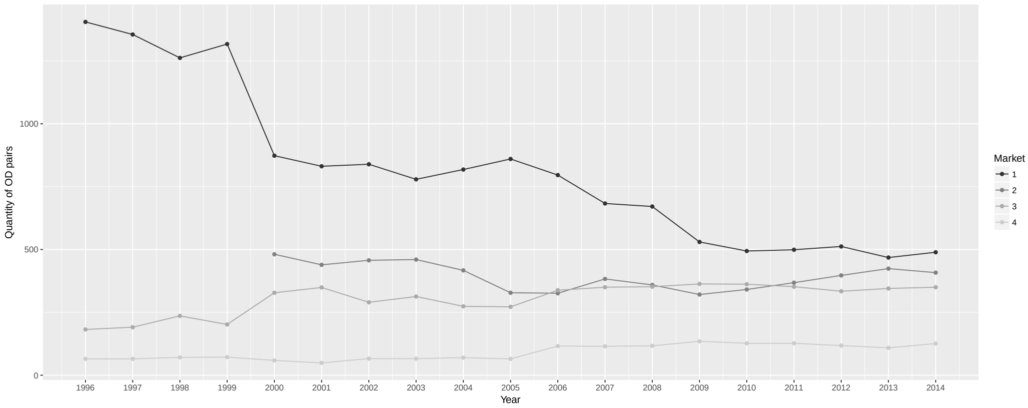


Figure 3.6: Quantity of OD pairs by market along the years

Source: Own work

Another aspect that may be relevant to analyse the results is the market share of fares. Table 3.3 shows these numbers.

It is noticeable that in all markets the dominance of the *standard reduced* fare, followed by the *standard full*. It is also noticeable that the *advance* ticket has been gaining space, growing from 1% to 8% in 15 years, in Market 2, and from 4% to 12% in Market 3. The *first class* tickets presented a small market share, in decline in both Markets 3 and 4, from 7% to 3% and 5-6% to 2-0%, respectively. It is observable that the *First Class* tickets when disaggregated by fare type show very small market shares, even less than 1%.

Table 3.3: Market share of fares by market segmentation

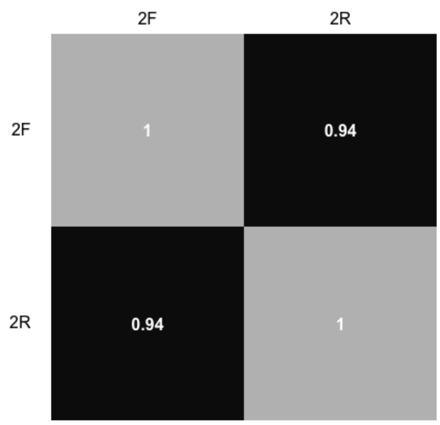
	Market 1		Market 2			Market 3				Market 4					
Year	2F	2R	2F	2R	2A	1N	2F	2R	2A	1F	1R	1A	2F	2R	2A
1996	41%	59%	-	-	-	7%	28%	61%	4%	5%	5%	6%	7%	66%	11%
1997	39%	61%	-	-	-	7%	32%	55%	6%	4%	9%	6%	7%	60%	15%
1998	38%	62%	-	-	-	7%	31%	57%	5%	4%	8%	5%	12%	58%	13%
1999	37%	63%	-	-	-	9%	26%	58%	7%	5%	8%	6%	7%	59%	15%
2000	39%	61%	26%	73%	1%	2%	34%	58%	6%	4%	0%	0%	26%	58%	11%
2001	38%	62%	31%	68%	1%	2%	34%	59%	5%	4%	0%	0%	27%	60%	8%
2002	39%	61%	25%	73%	2%	3%	32%	59%	6%	4%	0%	0%	26%	60%	10%
2003	38%	62%	33%	66%	1%	2%	28%	64%	6%	3%	0%	0%	21%	65%	10%
2004	39%	61%	32%	67%	1%	3%	26%	64%	8%	3%	0%	1%	17%	68%	11%
2005	38%	62%	32%	66%	2%	2%	25%	67%	6%	3%	0%	1%	17%	68%	11%
2006	38%	62%	35%	62%	2%	2%	27%	64%	6%	3%	0%	1%	21%	65%	11%
2007	38%	62%	38%	60%	2%	2%	31%	62%	5%	3%	0%	1%	23%	61%	11%
2008	39%	61%	36%	61%	3%	3%	31%	62%	4%	3%	0%	1%	26%	62%	8%
2009	43%	57%	38%	59%	3%	3%	29%	59%	8%	3%	2%	1%	25%	57%	12%
2010	43%	57%	38%	58%	4%	3%	26%	63%	8%	2%	1%	1%	20%	64%	12%
2011	45%	55%	40%	56%	4%	3%	26%	62%	9%	1%	1%	1%	20%	63%	13%
2012	47%	53%	38%	56%	6%	3%	23%	65%	9%	1%	1%	1%	16%	68%	12%
2013	46%	54%	35%	58%	6%	3%	31%	55%	12%	2%	0%	2%	28%	49%	19%
2014	43%	57%	37%	55%	8%	3%	33%	52%	12%	2%	0%	2%	28%	48%	20%

Source: Own work

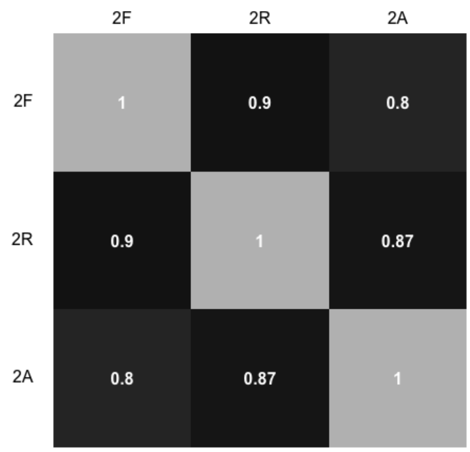
3.1.3 Data correlation

As mentioned in Chapter 2, the correlation has been a problem in the previous studies in fares elasticities estimation.

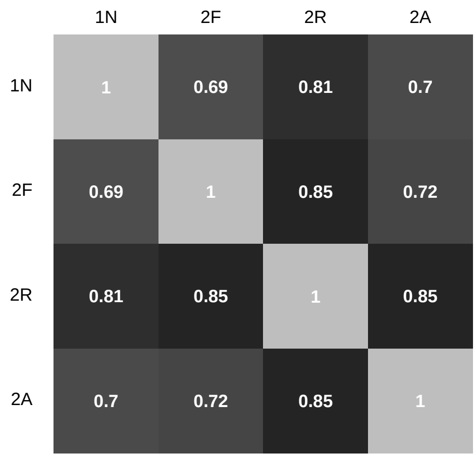
Checking the data used in this study, it is noticed that correlation is also present, as should be expected. Figure 3.7 presents correlation matrixes of the fares variables for each market after the segmentation of the dataset.



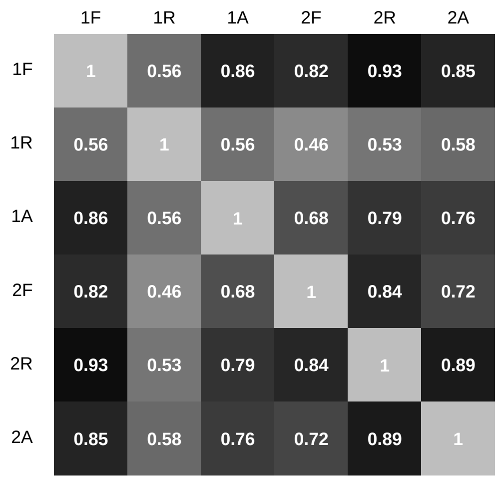
(a) Market 1



(b) Market 2



(c) Market 3



(d) Market 4

Figure 3.7: Correlation of fares by market

Source: Own work

In Market 1, there is a high correlation between fares, computed as 0.94.

Also for the Market 2 the correlations are still very high, above of 0.80. For Market 3, correlation assumes lower values, but still for the *reduced* and *advance*, and *reduced* and *full* tickets of the standard class they are very high, deemed as 0.85 for both. For Market 4, correlations above 0.80 for two-fifths of the coefficients.

Despite the presence of high correlation among the fares, it should not be an issue in Bayesian regression when strong informative priors are provided, as discussed in Chapter 2.

3.2 Model

Bayesian models are usually wrote formally as shown in Equation 3.1. This format is useful to identify the elements that must conceptualised when building a model: the linear model, the likelihood, and the priors. Next subsections will work on these elements.

$$\begin{aligned} \text{outcome}_i \sim& Normal(\mu_i, \sigma) && \text{[likelihood]} \\ \mu_i =& \beta \times \text{predictor}_i && \text{[linear model]} \\ \beta \sim& Normal(\mu_\beta, \sigma_\beta) && \text{[prior]} \\ \sigma \sim& uniform(a, b) && \text{[prior]} \end{aligned} \tag{3.1}$$

3.2.1 Linear model

The demand model adopted in this study is the one presented Equation 2.1, Chapter 2.

A stochastic version of this model is given simply by adding the error term into Equation 2.1, as shown by Equation 3.2.

$$V_i = a\ GVA_i^g\ P_i^f\ GJT_i^\gamma\ e^{u_i} \tag{3.2}$$

A linear alternative to express exponential relationships is the double-log form, as shown by Equation 3.3. The linear alternative is convenient because it allows estimation by linear regression models and the estimated coefficients can be interpreted as elasticities.

$$lnV_i = ln\ a + \ g\ lnGVA_i + f\ lnP_i + \gamma\ GJT_i + u_i \tag{3.3}$$

Because the interest is to estimate price elasticities of each specific fares, the price component of the model and its elasticity will vary according to the range of available fares in the considered market. Therefore, for Market 1, the price component will be expanded to $f_{2F}lnP_{2F}$ and $f_{2R}lnP_{2R}$, for fares $2F$ and $2R$ respectively. Also, different systems of equations will be estimated for different markets according to the fares available. For instance, in Market 1, there will be two equations: one to estimate the demand of fare 2F and one to estimate the demand of fare 2R. Therefore, the generic form of the equations to be estimated will be, as shown by Equation 3.4.

$$lnV_{ki} = ln\ a + \ g\ lnGVA_i + \sum f_k\ lnP_{ki} + \gamma\ GJT_i + u_i \tag{3.4}$$

where

k is the available type of fares in the market.

For completeness, the relation of outcomes and predictor variables in each market is shown in the Appendix A.

3.2.2 Likelihood

The first thing to consider to define the likelihood is the scale type of the outcome variable - whether metric, ordinal, nominal or count. Because things can be measured in different scales, different probability measure may apply and "the likelihood function must specify a probability distribution on the appropriate scale" (Kruschke 2014, p. 423).

For this study, the outcome variable - *journeys* - can be considered as metric scale, since its value actually provides a quantity of what is being measured - even though it is transformed to logarithmic form.

For this kind of scale, the most usual probability distribution is the normal. McElreath (2016) discuss two main reasons that can justify it.

The first one regards the fact the normal distribution has the property to represent phenomenons that are sum of fluctuations of other phenomenons, irrespective of their original distribution. "Repeatedly adding finite fluctuations results in a distribution of sums that shed all information about the underlying process, aside from mean and spread" (McElreath 2016, p. 75). In this sense, it might be plausible to interpret the number of rail journeys as a resultant phenomenon of fluctuations of the subprocess. People often make travel decision based on a general range of factors, and the way they usually vary in time and region may represent the fluctuations mentioned.

The second reason regards the fact that the normal distribution "represents a particular state of ignorance" (McElreath 2016, p. 75). This represents the most convenient form to express the lack of knowledge about a variable because its shape can comprise several different assumptions.

This also seems to be applicable to our variable. Putting in another way, one may not have evidence that this should not assume a normal distribution and this ignorance may suggest normal it is appropriate.

Therefore, the likelihood which will be applied for all models will be such as $lnV_i \sim Normal(\mu_i, \sigma)$.

3.2.3 Priors

When defining a prior distribution, one must be aware of two aspects of it: first it regards the definition of an interval for credible values that the parameter can assume; second, regards the shape of the probability density in this interval. To consider these elements for the current case, one may first recover the parameters presented in Equation 2.1, summarised by Table 3.4.

Table 3.4: Model's parameters that demand prior distributions

Parameter	Measure	Expected Sign
f	price elasticity of demand. It can regard the own or cross elasticity.	own: negative cross: positive
g	GVA elasticity of demand.	positive
γ	GJT elasticity of demand.	negative
σ	standard deviation of the outcome variable.	N.A.

Source: Own work

Starting for the price elasticity, it will be convenient divide it in two discussions: the own elasticity prior and the cross elasticity prior. This is relevant because of the assumption adopted from the PDFH that the differently available fares are competitors, thus the cross elasticity should expect to be positive, and the own elasticity should expect to be negative.

Recovering that the price elasticity measures how much the demand responds to a change in price, in percentage terms, one may consider plausible that credible values for the own price elasticities of demand in this study should be somewhere between zero and -1. This interval covers circumstances which go from a complete inelastic demand, which does not vary irrespective of a change in price, to an elastic demand, in which case a change in price causes a proportional change in the demand in the opposite direction. It eventually could be lower than -1, for a very price-sensitive demand, but it would not be reasonable considering that there is no lower boundary, even though setting one may be arbitrary.

For the cross elasticity, the rationale is analogous but the values are symmetric, so a credible interval would be between zero and 1. Again, it would eventually be greater than 1, even though it may be arbitrary setting an upper boundary.

In what regards the shape of the probability density across the credible values for price elasticities, economic theory provides no evidence abaout it, which may lead to the adoption of a uniform probability. However, the guidances from the statistical package used in this work - Stan (Stan 2017*c*), suggests as a general principle that uniform priors should not be applied "unless the boundaries represent true constraints" (Stan 2017*b*). Despite one of the boundaries actually represent a constraint - the upper boundary is zero for own elasticity and the lower boundary is also to zero for the cross elasticity - the others boundaries would be defined arbitrarily. Therefore, adopting uniform distribution would be not a goodd practice.

The suggested solution is to stablish a normal distribution centred in the credible interval. Indeed, there are specific guidances for elasticities estimates in double-log regressions that recommends that normal distribution with mean 0.5 and standard deviation of 0.5 is a good default prior.

To ensure that the theoretical constraints will be respected, a one-sided restriction will be applied, as pertinent. As introduced in Chapter 2, the possibility of applying constraints is one of the advantages of Bayesian inference.

For own price elasticities, the prior will be defined as $f \sim Normal(-0.5, 0.5)$, truncated at zero at the upper boundary ($T[\quad, 0]$); and for cross price elasticities, $f_x \sim Normal(0.5, 0.5)$, trucated at zero at the lower bound ($T[0, \quad]$).

For the GVA elasticity of demand, the prior knowledge suggests a positive polarity between the rail demand and the GVA, as already discussed in the exploratory analysis of the data in this Chapter. Even though the GVA elasticity is expected to be positive, this single piece of information is not enough to define the interval for credible values and the shape of probability density function. It will be convenient, thus, follow the guidances from the statistical package - Stan. A suggested generic prior is recommended to be Normal(0,1). The guidance highlights that this prior may limit extreme values because the normal distribution has shorter tails than other bell-shaped distributions.

The GJT elasticity of demand is analogous to the GVA but in the opposite direction. As discussed in the exploratory analysis, an increase in GJT tends to cause decrease in the demand, so a negative coefficient should be expected. For the same reasons presented for the GVA elasticity, it will also be adopted the generic prior Normal(0,1) to the GJT parameter.

Lastly, as done for the price elasticity priors, to ensure right algebraic signs, the prior distribution for GVA and GJT elasticities will be constrained by a lower and upper boundary at zero, respectively.

Therefore, for the GVA and the GJT's elasticities the prior applied will be $g \sim Normal(0, 1)$, truncated at zero at the lower boundary ($T[0, \quad]$) and $\gamma \sim Normal(0, 1)$, truncated at zero at the upper boundary ($T[\quad , 0]$), respectively.

For the σ prior there is no evident definition for a range of credible values, despite the fact that it is by nature a positive value. It was adopted the default distribution from Stan - uniform prior on $(-\infty, \infty)$ (Stan 2017*d*) - constrained by the positive domain. Therefore, the resulting σ's prior was $\sigma \sim uniform(0, \infty)$.

3.3 Calibration and target measures

As mentioned, the Bayesian models were estimated with the aid of the *Stan*, a software package. In this package the posterior distributions are approximated using a Markov Chain of Monte Carlo called *Hamiltonia Monte Carlo*, a variation of the Metropolis algorithm.

Some elements of the estimation process that are important to interpret the results regard the calibration of the process in terms of number of iterations, warm-up period, thinness, number of chains, initial values, effective sample and autocorrelation measures, and convergence. A brief definition and the default adopted in this work are presented in the following.

The quantity of iterations regards the length of the Markov Chain that will be run to build the posterior distribution. The longer it is, the more defined the posterior gets. However, it does not mean that one needs to exhaust computation resources to get a reasonable draw of posterior distributions. For this work, the standard number of iterations adopted was $2,000$. It is supposed to be enough to achieve convergence without autocorrelated samples, for non-complex models. Eventually, more iterations may be demanded.

Still regarding the number of iterations, is the warm-up period, which is "the practice of discarding early iterations in Markov Chain (...) to diminish the influence of starting values" (Gelman et al. 2014). The discarded ratio may vary according to the circumstances, but for a conservative approach, it will be adopted to discard the first half of the iterations. Therefore, running $2,000$ iterations, the first $1,000$ will be discarded.

Another aspect of the Markov Chain regards thinness of steps, which is the rule of which steps are kept as a draw for the posterior. It is usually defined as 1, so each iteration is kept. Nevertheless, it may be increased to reduce autocorrelation (Kruschke 2011) - instead of keeping each iteration, one keeps only every n^{th} step. An implication of increasing the thinness of an estimation is that it reduces the iterations kept, which might demand as a counterpart the increase of the number of iterations proportionally. In this study it will be adopted thin = 1. If it eventually demands to be increased to n, the interactions will be $2,000 \cdot n$.

Despite defining the number of iterations and its related attributes, it is also necessary to define the number of chains. After the warm-up period the variance within the chain is supposed to be stable, so one may consider that it has achieved convergence. However, for a reliable inference, it is also important to check whether different independent sequences will converge to the same distribution. In this study, it will be adopted the default of 3 chains. When discussing the results a measure for convergence will be the $\hat{R}$, the potential scale reduction factor. When convergence is achieved, $\hat{R} = 1$.

Another element that needs to be calibrated in the estimation procedure is the initial values from which the Markov Chains will start. For this study the default will not be explicitly defined. Instead, they will be randomly generated by the software (Stan 2017*d*).

Lastly, there is the the effective sample size (n_{eff}), which is not an element to be calibrated in the process but an output used to judge the quality of the resulting posterior distribution in combination with the $\hat{R}$. As long it will be used in Chapter 4, it worths introducing its concept.

When samples are autocorrelated in the chain the estimation might become inefficient because it does not fully explore the probability space of the parameters, instead it may get stuck. The n_{eff} is a measure that discount autocorrelated samples, providing a sense of the number of independent samples drawn to estimate the posterior distribution. As a default rule, Gelman et al. (2014) suggests running the simulation until n_{eff} is at least ten times the number of chains. Therefore, since the estimation will run with 3 chains, the n_{eff} must be higher than 30.

4 Results

4.1 General guidances on interpretation of the results

To interpret the results one must keep in mind that the estimated coefficients are elasticities. The practical implication of this is that the coefficient values represent the percentual change in the demand for a one percent change in the variable (untransformed to log). Therefore, a coefficient of value 1 will mean a one percent change in the demand for a one percent change in the predictor variable - fares, GVA or GJT.

Beyond the point estimates, the Bayesian regression provides a probability density function (pdf) of credible values that the coefficients may assume. In fact, the point estimate is the mean of the posterior distribution.

An advantage of a pdf output over a point estimate is that it is possible to directly interpret the probabilities of values in a given interval. For instance, one can take two values and compute the probability of the coefficient being in the interval. This is fundamentally different from the frequentist confidence interval, which only regards the probability of a sampling procedure resulting in an interval which contains the true value of the estimated parameter.

The posterior distributions are convenient to discuss the uncertainty of estimates. The usual procedure is taking the high-density interval (HDI) to consider the boundaries of credible values for the coefficients. The HDI is the smallest interval to achieve a given probability, which may differs from the symmetrical density interval, as illustrated by Figure 4.1, when the distribution is not symmetrical.

In this work, a HDI of 95% will be adopted in the analysis. Full information on the estimated HDI is presented on Appendix C .

The analysis of the results will be presented with the aid of marginal posterior distribution's plots. This format is useful since it is a bi-dimensional figure, but it worths recovering that the actual posterior distribution, as discussed in Chapter 2 is a multi-dimensional space, where the dimensions depend on the number of predictor variables.

An additional consideration when interpreting the posterior distribution plots is that the y axis would eventually be greater than one. This is apparently counter-intuitive since probabilities are defined in the [0,1] domain. However, because it is a continuous function, the actual probability of the

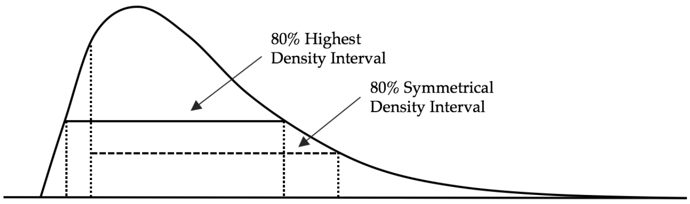


Figure 4.1: High Density Interval *versus* Symmetrical Density Interval

Source: Meredith & Kruschke (n.d.) [adapted]

estimate assuming a point value is zero. So the y axis should be not be used to directly read the probabilities from values in the x axis. Instead, a probability can only be computed integrating a given an interval. Integrating the full density function it would be approximately 1 (Kruschke 2014).

4.2 Comparative Estimates: SURE/OLS Models

For a comparison purpose, frequentist estimates will be provided along the Bayesian estimates. The most plausible approach would be the seemingly unrelated regression estimation (SURE), nevertheless, despite it seems natural to think about the demand estimation for each market as a system of demand equations (one for each type of fare), as taught by Kennedy (2003), SURE estimation becomes identical to OLS when all the predictive variables are the same, which happens to be the case in this study.

The models were estimated based on the linear models presented in Table A.1. As it should be expected, the estimation of SURE/OLS models reinforces previous studies showing that "freely estimating models will not yield robust elasticities" (ITS & Systra 2016, p. 37). Despite statistically significant - with some exceptions - several estimated elasticities presented unexpected signs, according to the assumptions of the PDFH. Table 4.1 summarises the occurrences.

Table 4.1: Summary of SURE/OLS outputs with respect to algebraic sign of coefficients

Market	Dependent Variable	Elasticities Estimates			
		Expected Sign		Unexpected Sign	
1	lnV_{2F}	4	f_{2F}, f_{2R}, g, γ	-	-
	lnV_{2R}	3	f_{2F}, f_{2R}, γ	1	g
2	lnV_{2F}	4	f_{2F}, f_{2A}, g, γ	1	f_{2R}
	lnV_{2R}	4	f_{2F}, f_{2R}, g, γ	1	f_{2A}
	lnV_{2A}	3	f_{2R}, f_{2A}, g	2	f_{2F}, γ
3	lnV_{1N}	6	f_{1N}, f_{2F}, f_{2R}, f_{2A}, g, γ	-	-
	lnV_{2F}	4	f_{2F}, f_{2A}, g, γ	2	f_{1N}, f_{2R}
	lnV_{2R}	4	f_{2F}, f_{2A}, g, γ	2	f_{1N}, f_{2R}
	lnV_{2A}	5	f_{2F}, f_{2R}, f_{2A}, g, γ	1	f_{1N}
4	lnV_{1F}	8	f_{1F}, f_{1R}, f_{1A}, f_{2F}, f_{2R}, f_{2A}, g, γ	-	-
	lnV_{1R}	6	f_{1R}, f_{2F}, f_{2R}, f_{2A}, g, γ	2	f_{1F}, f_{1A}
	lnV_{1A}	6	f_{1F}, f_{2F}, f_{2R}, f_{2A}, g, γ	2	f_{1R}, f_{1A}
	lnV_{2F}	3	f_{2A}, g, γ	5	f_{1F}, f_{1R}, f_{1A}, f_{2F}, f_{2R}
	lnV_{2R}	8	f_{1F}, f_{1R}, f_{1A}, f_{2F}, f_{2R}, f_{2A}, g, γ	-	-
	lnV_{2A}	7	f_{1F}, f_{1A}, f_{2F}, f_{2R}, f_{2A}, g, γ	1	f_{1R}

Source: Own work

4.3 Bayesian models with weakly informative priors

Because it is safer to estimate Bayesian models gradually building in complexity, Bayesian models with weakly informative priors were estimated before the constrained models. The priors established for these models have followed the same guidances from Chapter 3, but without the truncation.

The models were run without any issues. However, this estimation was of little help in generating estimates with correct algebraic signs resulting in estimates very similar to the OLS. This may be explained due to the high volume of data that makes the likelihood prevails over the prior distribution (Kruschke 2014). Therefore, the valuable prior information added to the models was the restriction applied in the domain of the elasticities' probability densities, as discussed in the next section.

For completeness, the Appendix B presents the results of estimated models with weakly informative priors.

4.4 Bayesian constrained models

4.4.1 Market 1

i. Autocorrelation and convergence

The first market, with only two fares competing, is the simpler one. The model converged with low autocorrelation in the posterior samples, as shown by the statistics in Table 4.2.

Table 4.2: Autocorrelation and convergence measures - Market 1

	Dependent variable:			
	$ln\ V_{2F}$		$ln\ V_{2R}$	
	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}
f_{2F}	1.0	1650	1.0	402
f_{2R}	1.0	1692	1.0	821
g	1.0	1561	1.0	719
γ	1.0	1778	1.0	432
Constant	1.0	1477	1.0	702
Iter	2,000		2,000	
Thin	1		1	

Source: Own work

It is relevant, however, to consider that a warning of divergent transitions after the warm-up was reported for both regressions. The issue around it is that these warnings signalise that the estimates might be biased (Stan 2017*a*), even though they have converged to stable and equivalent variances across the chains.

The divergent transitions are likely to be related to the imposition of constraints in the domain of the probability density. This might have prevented the MCMC to explores the parameter space with plausible values according to the likelihood, but constrained by the prior. Therefore, it appears that the divergent transitions are part of the adopted solution of applying restrictions to prevent unfeasible estimates.

It will be assumed, therefore, that biased estimates are the best information at hand and should not be discarded. As taught by Gelman (2017), in the context of Bayesian data analysis, one should recognise that "unbiasedness" is a very idealistic concept, and because it may hardly be achieved, one should not refrain oneself to make usage of the available information.

ii. OLS comparison

A comparison between the Bayes and OLS estimates is presented in Table 4.3. As expected, the all Bayesian coefficients have coherent signs. It has clearly improved the estimation of g with respect to the V_{2R} demand, for which the OLS estimation has failed.

Table 4.3: Comparison of bayesian and SURE/OLS estimates - Martket 1

	Bayes		*SURE/OLS*	
	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2F}$	$ln\ V_{2R}$
f_{2F}	-1.32	0.70	-1.28	0.81
	(0.05)	(0.05)	(0.05)	(0.05)
f_{2R}	0.17	-0.89	0.17	-0.89
	(0.05)	(0.05)	(0.05)	(0.05)
g	0.77	0.06	0.49	-0.52
	(0.05)	(0.04)	(0.7)	(0.7)
γ	-0.89	-0.88	-0.98	-1.06
	(0.04)	(0.05)	(0.04)	(0.05)
Constant	6.10	11.27	9.14	17.6
	(0.57)	(0.47)	(0.71)	(0.73)

Source: Own work

iii. Uncertainty

Figures 4.2 and 4.3 present the marginal posterior distribution of the estimated elasticities coefficients. The continuous lines represent the Bayesian mean, the red line represents the OLS mean and the dotted lines represent the boundaries of the 95% HDI.

As one may notice, the small standard deviations are reflected in a posterior distribution with a decimal scale for all estimates, except the constant - which usually will not be covered in the analysis. These small ranges of credible values for the coefficients are useful since these are elasticities measures. It is important that they do not allow the distribution to spread much, becoming low-quality information.

This is also reflected in the HDI boundaries, which are very close to the mean value demonstrating that there is a small range of credible values. As a general rule, it will be considered that an HDI up 0.30 is satisfactorily small.

Above this value, the uncertainty would be unfavourable to any practical application since this decimal variation in the coefficient, in fact, is a percent impact on the demand. Detailed information of the HDI is presented in the Appendix C.

With respect to the V_{2f} demand, Figure 4.2, all elasticities distributions are smooth bell-shaped curves, demonstrating that they have not faced any strong constraint. Indeed, the OLS estimate already had the correct signs, so there was no conflict in the estimation.

Even though the OLS estimates had correct signs, some of them were not considered credible values according to the Bayesian estimation: the g elasticity lied out of the HDI and the γ was at the edge.

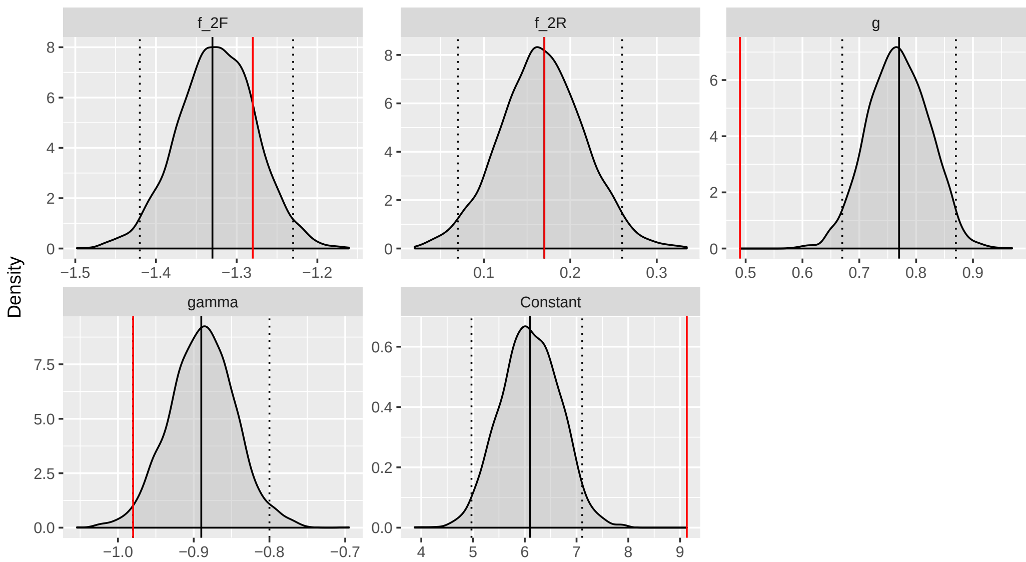


Figure 4.2: Posterior density function of elasticities w.r.t V_{2F} - Market 1

Source: Own work

With respect to the V_{2R} demand, Figure 4.3, the curves are bell-shaped, except for the g which is skewed towards the left. It is noticeable that this is the sign-reverted elasticity. It appears that the constraint at zero, which blocked the distribution to assume negative values, was an actual barrier. Indeed this may explain the warnings of divergent transitions.

All HDI were satisfactorily small, from 0.14 to 0.20.

The OLS estimates of f_{2F}, g and γ are out of the HDI, demonstrating that these are not credible values according to the Bayesian estimation. Opposedly, the OLS estimate for f_{2R} was coincident.

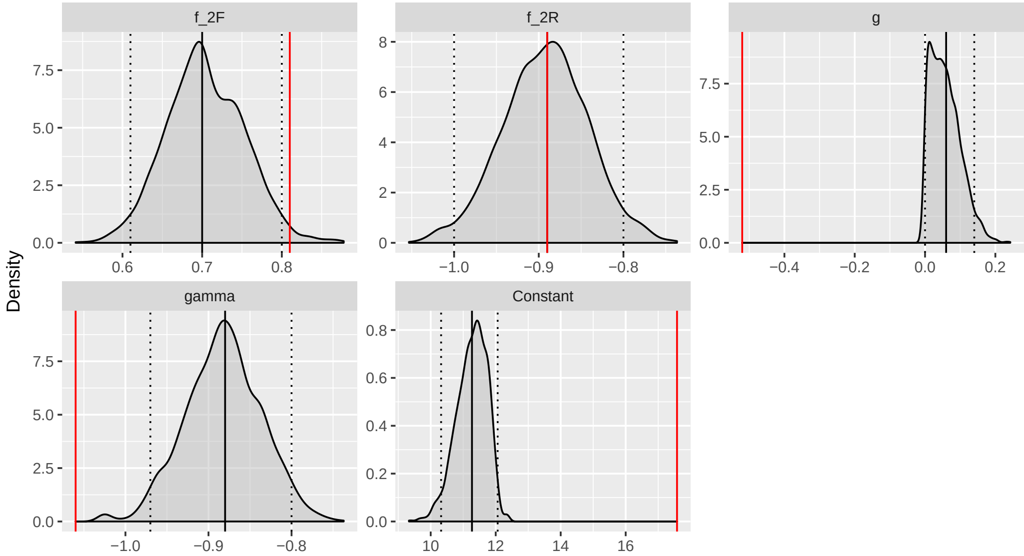


Figure 4.3: Posterior density function of elasticities w.r.t V_{2R} - Market 1

Source: Own work

iv. Magnitude of estimates

Analysing the magnitude and distributions of coefficients one can notice that the *Full* fare is more sensitive to its own price, with mean -1.3, then the *Reduced* fare, with mean -0.9. This difference may be reasonable if one considers the *Reduced* fare as a basic service, accessible for everyone without time restrictions, whilst the *Full* fare would offer a plus of convenient schedules in the peak hours. The price differentiation of *Reduced* and *Full* and its different sensitiveness to price would, therefore, be in accordance with economic theory which says that superfluous goods are more elastic than essential goods.

The magnitude of cross elasticities can be also interpreted by this rationale of basic and more convenient services. Therefore, the demand of *Reduced* fare is more affected by variations in the price of the *Full* fare than the contrary, which means that everything constant, when the *Full* fare increases more passenger change for the *Reduced* fare than vice-versa.

In which regards the GVA elasticity it is unexpectedly below 1 for the demand of *Full*, with mean 0.7, and even more unexpected for the demand of *Reduced*, being very close to zero. According to the PDFH, the GDP elasticities - for which GVA is assumed as an equivalent measure of income/wealth (PDFC 2011, p. 4, Chapter B1), are expected to be around 1.1 (PDFC 2011, p. 9, Chapter B1).

Irrespective of their magnitude, the fact that $g_{2F} > g_{2R}$ should be expected because there is a higher proportion of business trips in the *Full* fare than in the *Reduced*, for non-London long distance journeys (PDFC 2011).

Lastly, the quality elasticity was very similar in both *Full* and *Reduced* demands, with mean -0.9 and similar standard deviation. These values are within the expected interval of -0.7 to -1.1, according to the PDFH (PDFC 2011). However, business travellers are more sensitive to quality rather than price because "it is undertaken at the time and at the expense of the employer" (PDFC 2011, p. 1, Chapter A1). Because of that, $\gamma_{2F} > \gamma_{2R}$ should be expected, which was not observed.

4.4.2 Market 2

i. Autocorrelation and convergence

In Market 2, the estimation was more complex. Both the V_{2R} and V_{2A} regressions demanded longer iterations (10,000 instead of 2,000) to converge with an acceptable amount of effective samples. Nevertheless, the estimation was successful, as demonstrated by the statistics in Table 4.4.

As in Market 1, there also were divergent transitions reported in all three regressions of Market 2. The previous interpretation applies.

Table 4.4: Autocorrelation and Convergence Measures - Market 2

	Dependent variable:					
	$ln\ V_{2F}$		$ln\ V_{2R}$		$ln\ V_{2A}$	
	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}
f_{2F}	1.0	405	1.0	464	1.0	529
f_{2R}	1.0	251	1.0	414	1.0	493
f_{2A}	1.0	614	1.0	702	1.0	634
g	1.0	315	1.0	561	1.0	555
γ	1.0	400	1.0	525	1.0	912
Constant	1.0	310	1.0	548	1.0	500
Iter	2,000		10,000		10,000	
Thin	1		1		1	

Source: Own work

ii. OLS comparison

A comparison between the Bayes and OLS estimates is presented in Table 4.5. As expected, the all Bayesian coefficients have coherent signs. It has clearly improved the estimates since the OLS present four wrong sign elasticities.

Table 4.5: Comparison of bayesian and SURE/OLS estimates - Martket 2

	Bayes			*SURE/OLS*		
	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$
f_{2F}	-1.29	0.04	0.05	-1.24	0.03†	-0.18
	(0.05)	(0.03)	(0.04)	(0.06)	(0.06)	0.10)
f_{2R}	0.05	-0.80	1.01	-0.06	-0.60	1.03
	(0.04)	(0.05)	(0.08)	(0.08)	(0.08)	(0.13)
f_{2A}	0.14	0.01	-0.46	0.15	-0.10	-0.47
	(0.04)	(0.01)	(0.06)	(0.04)	(0.04)	(0.06)
g	1.68	0.96	0.29	1.75	0.39	0.66
	(0.07)	(0.07)	(0.08)	(0.11)	(0.10)	(0.17)
γ	-1.28	-1.07	-0.03	-1.23	-1.21	0.33
	0.06	(0.06)	(0.03)	(0.07)	(0.06)	(0.10)
Constant	-0.28	5.46	-0.61	-1.09†	11.72	-5.53
	(0.76)	(0.71)	(0.82)	(1.15)	(1.09)	(1.77)

Note: † $p > 0.1$

Source: Own work

iii. Uncertainty

Figures 4.4, 4.5 and 4.6 illustrate the probability densities of the estimated elasticities coefficients. The elements in the graphs remain as explained for Market 1.

Again, the small standard deviations were very important in constraining the range of credible values for the estimates in the decimal, even centesimal, scale. The HDI was generally satisfactorily small with the exception of two elasticities, with respect the demand V_{2A}. Nevertheless, they were very close to 0.30.

With respect to the V_{2F} demand, Figure 4.4, all HDI were in generally satisfactorily small, from 0.14 to 0.28, demonstrating low uncertainty since the credible values are concentrated in small intervals. It is observed that the fare elasticities had the smaller HDI - 0.17, 0.14 and 0.15 for *Full*, *Reduced* and *Advance*, respectively, in opposition to the g and γ which were above 0.20. Detailed information of the HDI is presented in the Appendix C.

The OLS estimates lied inside the HDI interval, which indicates that the OLS estimates can be deemed credible values according to the Bayesian estimation. An exception is made for the f_{2R}, for which it should already be expected since its OLS estimate was on an unfeasible domain.

All elasticities distributions are bell-shaped curves, although with some deformity, except for the f_{2R}, skewed towards the left with an abrupt cut at zero. It is noticeable that this is the sign-reverted elasticity, which may indicate that the constraint to a positive domain was an actual barrier. What apparently happens is that, because the likelihood considers feasible values on the negative side but the prior does not allow this domain, they tend to agree until zero, so samples are extracted until the limit of the constraint. Beyond that point, they clash, and since the prior assumes zero probability for those values, it is reflected in the posterior distribution. Indeed this may explain the warnings of divergent transitions.

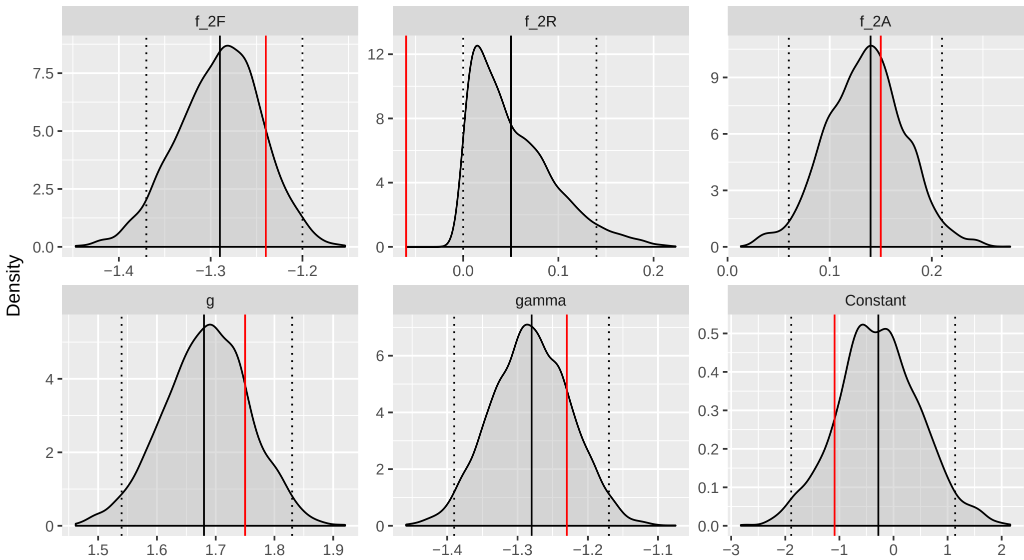


Figure 4.4: Posterior density function of elasticities w.r.t V_{2F} - Market 2

Source: Own work

With respect to the demand V_{2R}, Figure 4.5, there were very short HDI intervals, as for f_{2F} and f_{2A}, 0.09 and 0.03 respectively, indicating a very precise estimation. The other fare elasticity can also be considered with low HDI, ranging 20%. Again the g and γ had higher intervals. Detailed information of the HDI is presented in the Appendix C.

The OLS estimation lied inside the HDI interval only for the f_{2F} elasticity, which indicates big discrepancies between the Bayesian and frequentist's outputs.

Again, the sign-reversed coefficient, f_{2A}, had its distribution finishing abruptly at the zero bound. The cut at zero also appears for the f_{2F} elasticity. Here it is also applied the interpretation of the sharp edge made for the f_{2R}, with respect to the demand V_{2F}.

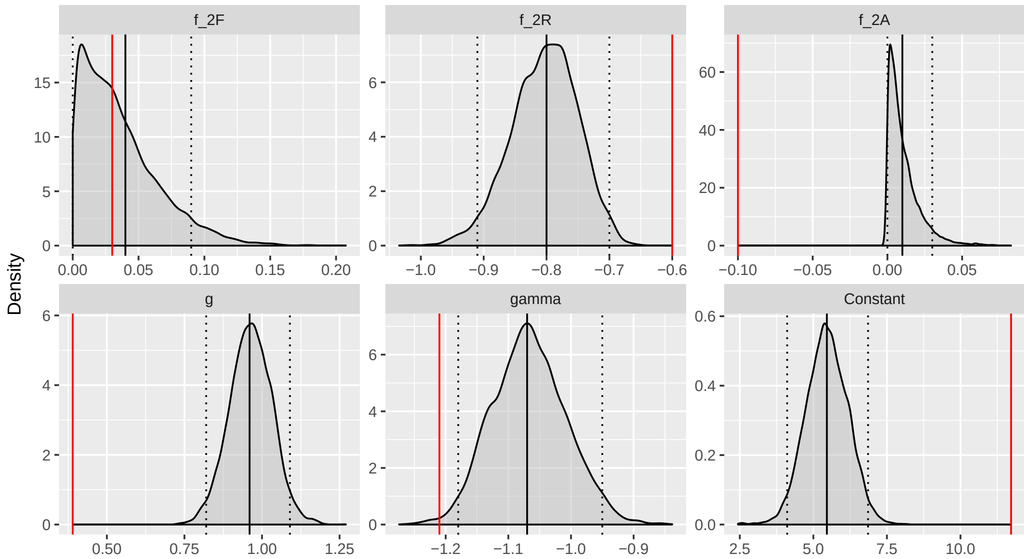


Figure 4.5: Posterior density function of elasticities w.r.t V_{2R} - Market 2

Source: Own work

With respect to the demand V_{2A}, Figure 4.6, the HDI's range was satisfactorily small for f_{2F}, f_{2A} and γ, with a surprisingly precise estimate ranging only 0.08. For the other elasticities, however, they were above 0.30, even though they are close to it. It must be highlighted that one can always trade-off certainty for precision, which means that reducing the target probability of HDI will also provide more precise intervals. Detailed information of the HDI is presented in the Appendix C.

The analysis of the position of OLS estimates relatively to the HDI shows divergences between the two estimation methods. Only for f_{2R} and f_{2A} the OLS mean lies inside the HDI.

Again, the sign-reversed coefficients, f_{2F} and g presented a sharp edge at the zero value. The previous interpretation applies.

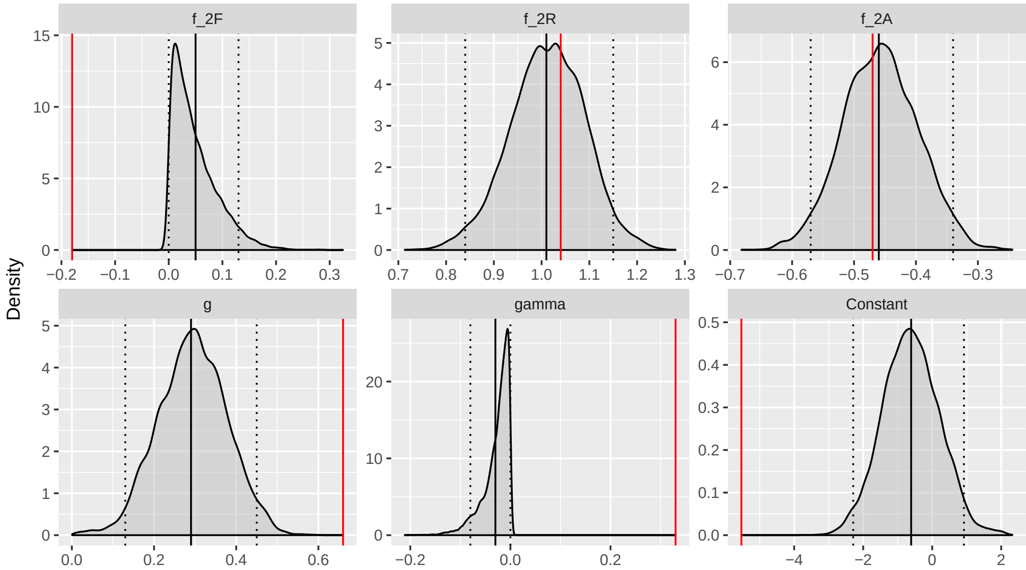


Figure 4.6: Posterior density function of elasticities w.r.t V_{2A} - Market 2

Source: Own work

iv. Magnitude of estimates

Analysing of the magnitude of coefficients, the most sensitive fare with regard its own price is *Full*, with mean -1.29, followed by the *Reduced*, with mean -0.80. These values are close to the estimates in Market 1, which is interesting since both do not have *First Class* tickets. The *Advance* fare presented a lower own elasticity, -0.46.

Regarding the cross elasticities, for the *Full* fare they were very small - 0.04 and 0.05, indicating a very low impact in the other ticket demands.

Conversely, for the *Reduced*, a notably high cross elasticity of 1.01 - higher than the own elasticity, was computed with respect to the *Advance* demand. The other cross elasticity was a very small effect - -0.05.

Lastly, for the *Advance* fare, both cross elasticities were considerably small effects - 0.014 and 0.01.

The GVA elasticity in this market are significantly higher than in the Market 1, which is difficult to interpret. However, a sign of consistency is noticed since there still is the pattern of the GVA affecting more the *Full* demand, whose mean is 1.68, then the *Reduced*, whose mean is 0.96. For the *Advance* demand the elasticity was of 0.29, but because the advance fares are a mix of promotional fares from both *Full* and *Reduced* it is difficult to interpret it as well.

Lastly, the GJT elasticities were also higher than in Market 1, which is difficult to reasonably interpret. Additionally, they did not follow the same pattern from the previous market, in which they coefficients had very similar means. Instead, for the *Full* demand the mean was deemed as -1.28, and for the *Reduced* it was -1.07.

It is interesting, however, that the GJT elasticity of the *Advance* ticket has a very tight distribution close to zero, being far from the expected interval of -0.7 and -1.1 (PDFC 2011) without an explicit reason.

4.4.3 Market 3

i. Autocorrelation and convergence

In the third market, the estimation was more complex than in previous markets. The estimation of V_{2F}, V_{2R} and V_{2A} demanded longer iterations (10,000 instead of 2,000) to converge with an acceptable amount of effective samples. Additionally, to reduce autocorrelation, the thinness was increased to 5. The resultant estimation was successful, with low autocorrelation, as shown in Table 4.6.

As in Market 1 and 2, there also were divergent transitions reported in all four regressions of Market 3. The previous interpretation applies.

Table 4.6: Autocorrelation and Convergence Measures - Market 3

	Dependent variable:							
	$ln\ V_{1N}$		$ln\ V_{2F}$		$ln\ V_{2R}$		$ln\ V_{2A}$	
	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}
f_{1N}	1.0	2712	1.0	492	1.0	826	1.0	939
f_{2F}	1.0	2350	1.0	278	1.0	637	1.0	1031
f_{2R}	1.0	2203	1.0	205	1.0	561	1.0	948
f_{2A}	1.0	3000	1.0	294	1.0	678	1.0	1187
g	1.0	2128	1.0	147	1.0	376	1.0	668
γ	1.0	2708	1.0	367	1.0	559	1.0	945
Constant	1.0	2130	1.0	140	1.0	409	1.0	756
Iter	2,000		10,000		10,000		10,000	
Thin	1		5		1		5	

Source: Own work

ii. OLS comparison

A comparison between the Bayes and OLS estimates is presented in Table 4.7. As expected, the all Bayesian coefficients have coherent signs. It has clearly improved the estimates since the OLS present four wrong sign elasticities.

Table 4.7: Comparison of bayesian and SURE/OLS estimates - Martket 3

	Bayes				*SURE/OLS*			
	$ln\ V_{1N}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$	$ln\ V_{1N}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$
f_{1N}	-0.82	0.00	0.01	0.01	-0.84	-0.43	-0.44	-0.50
	(0.06)	(0.00)	(0.01)	(0.01)	(0.06)	(0.05)	(0.05)	(0.07)
f_{2F}	1.04	-1.40	0.16	1.04	1.03	-1.12	0.23	1.07
	(0.09)	(0.05)	(0.05)	(0.10)	(0.09)	(0.07)	(0.07)	(0.10)
f_{2R}	1.73	0.02	-0.06	1.97	1.80	-0.04	0.34	2.28
	(0.11)	(0.02)	(0.05)	(0.11)	(0.11)	(0.09)	(0.09)	(0.13)
f_{2A}	0.44	0.18	0.10	-0.54	0.43	0.33	0.15	-0.39
	(0.06)	(0.05)	(0.04)	(0.07)	(0.06)	(0.05)	(0.05)	(0.07)
g	1.53	1.72	1.24	1.03	1.50	1.66	0.80	1.54
	(0.07)	(0.07)	(0.07)	(0.08)	(0.11)	(0.09)	(0.09)	(0.13)
γ	-3.25	-1.55	-2.05	-1.94	-3.30	-1.60	-2.31	-2.01
	(0.08)	(0.06)	(0.06)	(0.09)	(0.082)	(0.07)	(0.07)	(0.10)
Constant	0.20	1.56	5.80	-2.37	$0.62^{\dagger}$	2.85	11.66	-6.72
	(0.76)	(0.72)	(0.68)	(0.85)	(1.23)	(1.00)	(0.95)	(1.45)

Note: † $p > 0.1$

Source: Own work

iii. Uncertainty

Figures 4.7, 4.8, 4.9 and 4.10 illustrate the probability densities of the estimated elasticities coefficients. The elements in the graphs remain as previously explained.

In this estimation, the standard deviations were not satisfactorily small as in the previous markets. Neither was the amplitude of HDI intervals: for 7 out of 24 elasticities it was above 0.30. As discussed before, is always possible to trade-off certainty for precision. Nevertheless, in general, the estimation clearly lost in quality in this market.

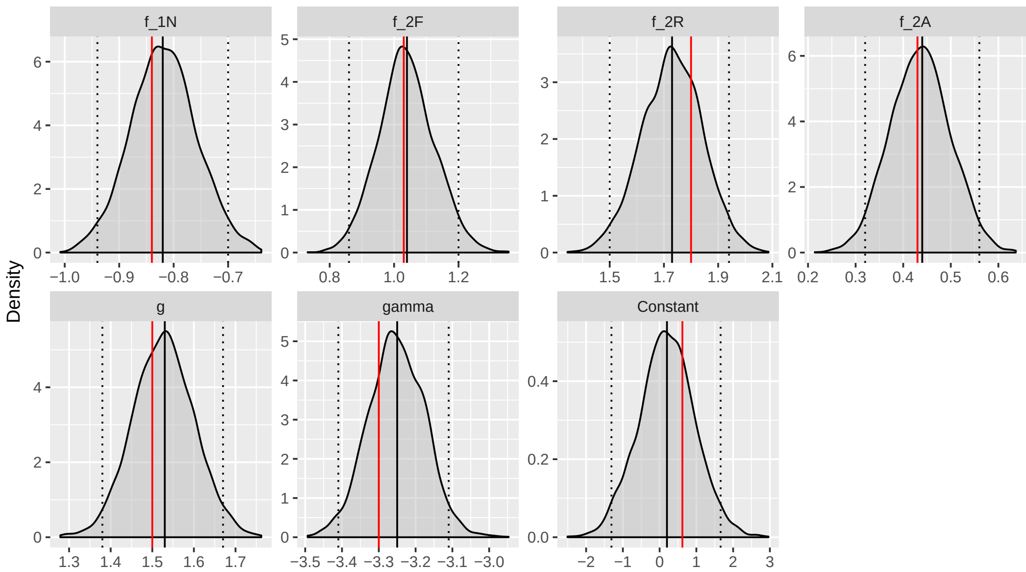


Figure 4.7: Posterior density function of elasticities w.r.t V_{1N} - Market 3

Source: Own work

With respect to the V_{1N} demand, Figure 4.7, half of the coefficients had HDI's amplitude below 0.30 - f_{1N}, g and γ, and half were above this value - f_{2F}, f_{2R} and γ, achieving up to 0.43. Detailed information of the HDI is presented in the Appendix C.

As one may notice, for this estimation, there were no discrepancies between the OLS mean estimates and the Bayesian ones - all were very close. This may have occurred because the OLS estimates have already had correct algebraic signs. Therefore, the constraints applied to the prior did not represent actually barriers to help to shape the posterior distributions. Indeed, there were no sharp edges in any distribution.

Since the constraint was not effective, the priors become simply non-informative priors, and because of the large number of observations, the likelihood prevails, causing the estimation to be similar to the OLS. Indeed, all OLS estimates are lying inside the HDI interval.

The inefficiency of the prior may also explain the large standard deviations. As one may recover from Chapter 2, in the presence of high correlation of variables, strong priors are important because defining the boundaries of one covariant automatically shapes the correlated one. Thus, the mixed and blurred joint effect of two correlated variables takes a cut-off point.

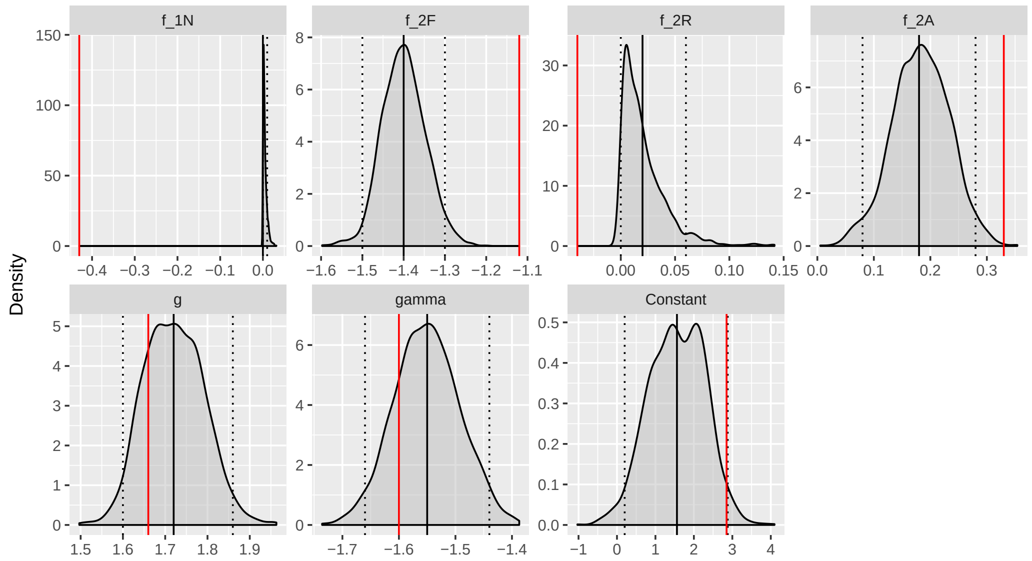


Figure 4.8: Posterior density function of elasticities w.r.t V_{2F} - Market 3

Source: Own work

With respect to the V_{2F} demand, Figure 4.8, the elasticities' HDI were satisfactorily small. The sign-reversed coefficients f_{1N} and f_{2N}notably presented HDI with very small amplitude, 0.01 and 0.06, respectively. Detailed information of the HDI is presented in the Appendix C.

With these short HDI intervals, the OLS estimates lied out of the range of credible values for all estimates, but g. This shows that the precision of bayesian estimates was enough to distinguish even OLS estimates that already have correct algebraic signs from being credible values.

It is noticeable that, once again, for the sign-reversed coefficients, the shape of the posterior distribution is skewed towards the zero value, showing that the imposition of constraint was fundamental to avoid the wrong sign estimates indicated by the likelihood.

With respect to the V_{2R} demand, Figure 4.9, the elasticities' HDI were again satisfactorily small. It is noticeable that, as in the previous model, the f_{1N} elasticity was very close to zero HDI's range of 0.02. Detailed information of the HDI is presented in the Appendix C.

The OLS estimates lied out of the credible values of for g and γ, besides the sign-reversed coefficients, for which it should already be expected.

Also, similarly to the last model, the sign-reversed coefficients had skewed distributions towards the zero constraints. A similar interpretation is applicable in this regard.

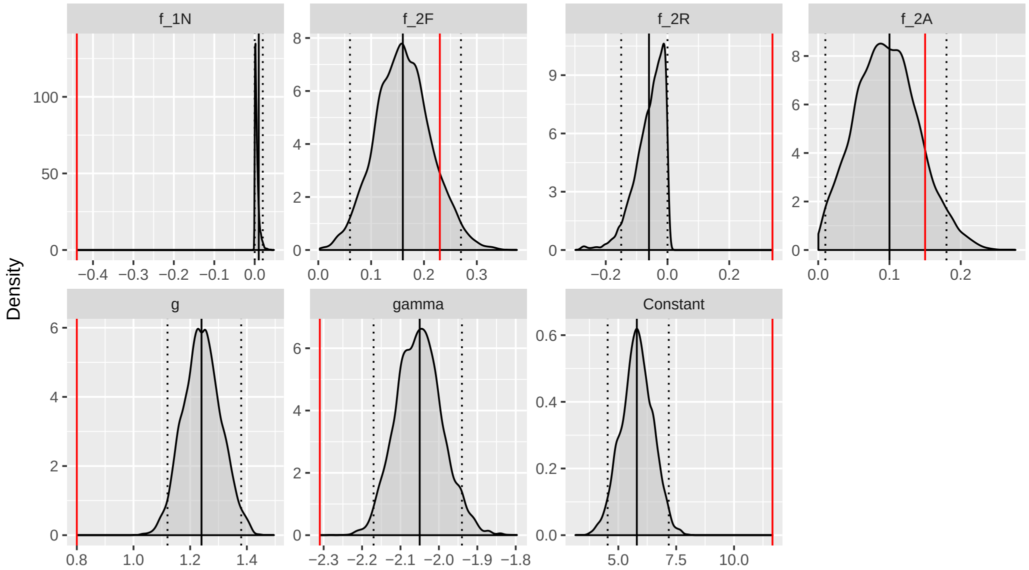


Figure 4.9: Posterior density function of elasticities w.r.t V_{2R} - Market 3

Source: Own work

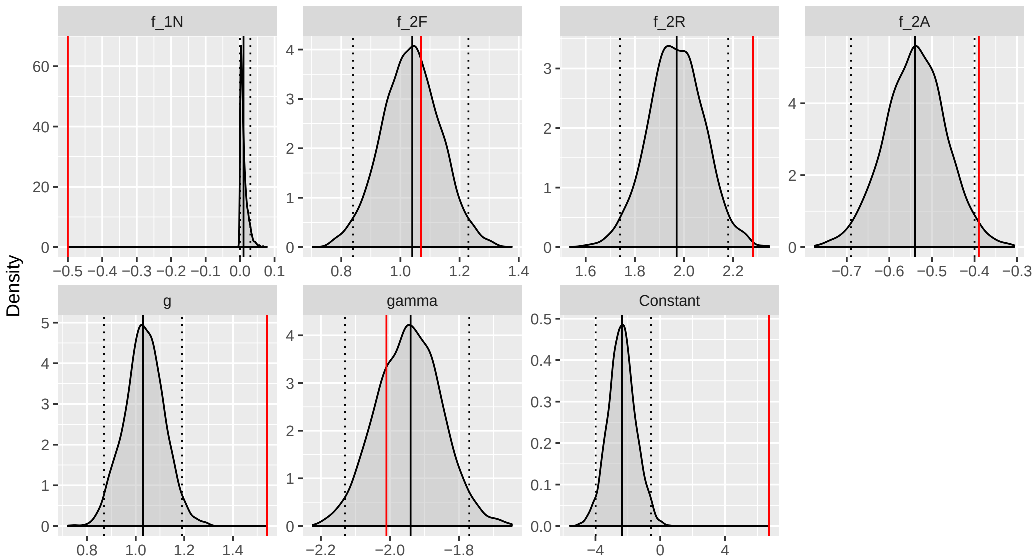


Figure 4.10: Posterior density function of elasticities w.r.t V_{2A} - Market 3

Source: Own work

With respect to the V_{2A} demand, Figure 4.10, the HDI are again presenting ranges above 0.30 up to 0.44 (f_{2R}). excepted for f_{1N} and f_{2A}. The f_{2A} was at the edge, with 029. Conversely, it is noticeable that the f_{1N}, as in the last model, it presented a very short HDI of 0.03. Detailed information of the HDI is presented in the Appendix C.

Even though the large HDI intervals, the OLS estimates lied out of the credible values for the f_{1N}, which was already expected since the OLS estimates were in the wrong domain, f_{2R}, f_{2A} and g.

Once again, the sign-reversed coefficient had skewed distributions towards the zero constraints. A similar interpretation is applicable in this regard.

iv. Magnitude of estimates

Analysing the magnitude of coefficients, the most sensitive fare with regard its own price is the *Full*, with mean -1.4, followed by the *First Class*, with mean -0.82, *Advance*, with mean -0.46 and *Reduced*, with a very low mean of -0.06. Notably, the *Full* fare remains very elastic in this market.

Regarding the cross elasticities, there were very extreme values - small effects very close to zero and high cross elasticities, bigger than 1.

Some notable observations deserves to be highlighed: the virtually null cross elasticities for the *First Class* for all demands; the high cross elasticities of the *Full* fare with respect to *First Class* and *Advance*, 1.04 for both and the also high effects of the *Reduced* fare with respect to the *First Class* and *Advanced* demands - 1.73 and 1.97, respectively. That may deserve further investigation. Eventually other constraints may be applied to test the propability of these elasticities being smaller.

The GVA elasticity was above unit for all type of demands, which may be deemed reasonable since PDFH states a general expected value of 1.1 (PDFC 2011, p. 9, Chapter B1) and also considers that "the elasticities to GDP (...) can be expected to vary by ticket type" (PDFC 2005, p. 7, Chapter B2). It worths noticing that $g_{2F} > g_{2R}$, which is consistent with the fact that the percentage of business journeys is higher in *Full* tickets. However, it is intriguing that $g_{1N} < g_{2F}$ because *First class* is usually associated with business journeys. Nevertheless, it should be highlighted that the 1N fare aggregates *Full*, *Reduced* and *Advance* tickets from the first class. In other words, 1N mixes the types of fares for which PDFH have stated the shares of trip purposes - there is only differentiation of trip purpose between *Full* and *Reduced*, making it unclear to judge the elasticity magnitude by this terms.

For the GJT effects, the elasticities were considerably high: from -1.55, for the *Full* demand, to -3.25, for the *First Class*. Even though there is the same GVA's reservation that states that elasticities can be expected to vary by ticket type (PDFC 2005), it is markedly above general interval expected by PDFH - -0.7 to -1.1 (PDFC 2011).

It is remarkable that the most quality sensitive fare is the *First Class*, which is reasonable since the first class is a quality differentiation, but the relative magnitude of the others are difficult to interpret.

4.4.4 Market 4

i. Autocorrelation and convergence

The estimation was complex again, as it was in the third. Except for the V_{2R}, the models demanded longer iterations (10,000 instead of 2,000) to converge with an acceptable amount of effective samples. Additionally, to reduce autocorrelation, the thinness was increased to 5 for V_{1A} and V_{2F}. Nevertheless, the estimation was successful, as demonstrated by the statistics in Table 4.8.

As in all previous markets, there also were divergent transitions reported in all regressions of Market 4. The previous interpretation applies.

Table 4.8: Autocorrelation and Convergence Measures - Market 4

	Dependent variable:											
	$ln\ V_{1F}$		$ln\ V_{1R}$		$ln\ V_{1A}$		$ln\ V_{2F}$		$ln\ V_{2R}$		$ln\ V_{2A}$	
	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}	$\hat{R}$	n_{eff}
f_{1F}	1.0	707	1.0	373	1.0	622	1.0	173	1.0	98	1.0	408
f_{1R}	1.0	534	1.0	273	1.0	548	1.0	186	1.0	339	1.0	473
f_{1A}	1.0	663	1.0	409	1.0	632	1.0	139	1.0	225	1.0	435
f_{2F}	1.0	826	1.0	260	1.0	739	1.0	147	1.0	278	1.0	515
f_{2R}	1.0	668	1.0	308	1.0	527	1.0	133	1.0	206	1.0	284
f_{2A}	1.0	251	1.0	354	1.0	799	1.0	121	1.0	213	1.0	410
g	1.0	442	1.0	311	1.0	507	1.0	95	1.0	211	1.0	221
γ	1.0	596	1.0	292	1.0	533	1.0	149	1.0	268	1.0	320
Const.	1.0	485	1.0	280	1.0	469	1.0	103	1.0	197	1.0	404
Iter	10,000		10,000		10,000		10,000		2,000		10,000	
Thin	1		1		5		5		1		1	

Source: Own work

ii. OLS comparison

A comparison between the Bayes and OLS estimates is presented in Table 4.9. As expected, all Bayesian coefficients have coherent signs. It clearly improved the estimates, reverting the ten wrong sign elasticities from the OLS estimates.

Table 4.9: Comparison of bayesian and SURE/OLS estimates - Martket 4

	Bayes						*SURE/OLS*					
	$ln\ V_{1F}$	$ln\ V_{1R}$	$ln\ V_{1A}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$	$ln\ V_{1F}$	$ln\ V_{1R}$	$ln\ V_{1A}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$
f_{1F}	-1.02	0.04	1.10	0.02	0.16	0.46	-0.94	-0.09†	1.20	-0.50	0.01†	0.54
	(0.19)	(0.04)	(0.19)	(0.02)	(0.12)	(0.20)	(0.22)	(0.27)	(0.27)	(0.20)	(0.19)	(0.26)
f_{1R}	0.05	-0.49	0.02	0.01	0.21	0.03	0.06†	-0.43	-0.25	-0.14	0.22	-0.11
	(0.04)	(0.07)	(0.02)	(0.01)	(0.05)	(0.03)	(0.06)	(0.08)	(0.08)	(0.06)	(0.05)	(0.07)
f_{1A}	0.24	0.02	-0.09	0.04	0.39	0.80	0.28	-0.99	0.09†	0.33	0.44	0.88
	(0.11)	(0.02)	(0.07)	(0.03)	(0.09)	(0.13)	(0.125)	(0.151)	(0.154)	(0.114)	(0.108)	(0.147)
f_{2F}	0.91	0.46	0.72	-1.04	0.45	0.83	0.85	0.71	0.74	-0.59	0.51	0.82
	(0.13)	(0.13)	(0.16)	(0.09)	(0.12)	(0.16)	(0.15)	(0.18)	(0.18)	(0.14)	(0.13)	(0.17)
f_{2R}	0.87	0.13	0.43	0.05	-1.32	0.64	0.67	0.00†	0.10†	-0.32†	-1.32	0.46†
	(0.22)	(0.10)	(0.23)	(0.04)	(0.19)	(0.26)	(0.27)	(0.33)	(0.34)	(0.25)	(0.23)	(0.32)
f_{2A}	0.16	1.28	0.82	0.04	0.31	-1.08	0.18†	2.37	1.04	0.03†	0.35	-0.99
	(0.12)	(0.17)	(0.19)	(0.03)	(0.14)	(0.20)	(0.18)	(0.22)	(0.22)	(0.16)	(0.16)	(0.21)
g	1.91	1.21	0.69	2.37	1.87	1.30	2.74	1.17	1.12	2.86	1.91	1.86
	(0.09)	(0.17)	(0.10)	(0.08)	(0.08)	(0.09)	(0.18)	(0.22)	(0.22)	(0.17)	(0.16)	(0.22)
γ	-2.66	-1.48	-2.18	-2.14	-2.07	-2.05	-2.39	-1.79	-2.03	-1.71	-2.07	-1.90
	(0.13)	(0.15)	(0.16)	(0.10)	(0.12)	(0.16)	(0.15)	(0.18)	(0.18)	(0.14)	(0.13)	(0.18)
Constant	-2.40	-0.71	-0.84	-2.27	-0.01	-1.17	-11.98	-3.34†	-5.89	-8.38	-0.38†	-7.47
	(0.89)	(0.92)	(0.94)	(0.79)	0.81	(0.04)	(2.09)	(2.53)	(2.57)	(1.91)	(1.80)	(2.46)

Note: † $p > 0.1$

Source: Own work

iii. Uncertainty

Figures 4.11, 4.12, 4.13, 4.14, 4.15 and 4.16 illustrate the probability densities of the estimated elasticities coefficients. The elements in the graphs remain as previously explained.

As for Market 3, the estimation in Market 4 has not reported small standard deviations. Neither was the amplitude of HDI intervals: for 35 out of 48 elasticities coefficients it was above 0.30. Once again, it worths highlight that there is always a trade-off between the precision of the HDI interval and the probability mass covered, which may be useful for practical applications.

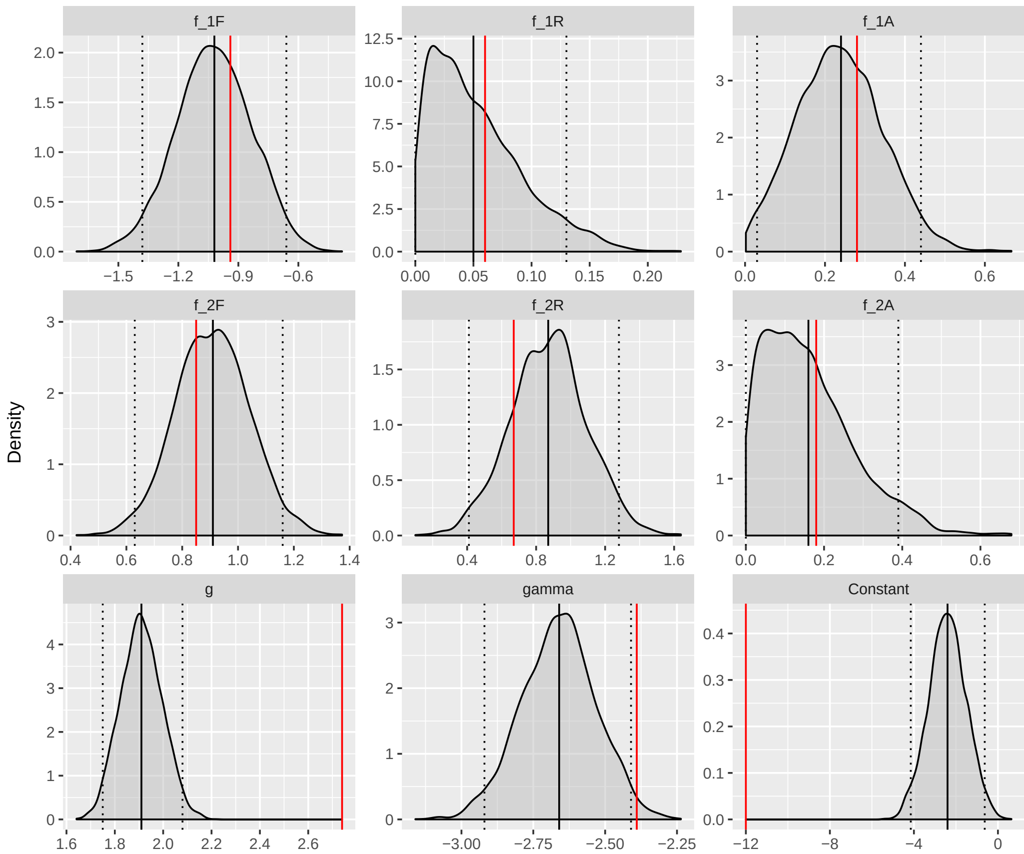


Figure 4.11: Posterior density function of elasticities w.r.t V_{1F} - Market 4

Source: Own work

With respect to V_{1F} demand, Figure 4.11, the only elasticity with short HDI was the f_{1R}, which amplitude is 0.13. For all others coefficients, the HDI range was significantly high. Detailed information of the HDI is presented in the Appendix C.

What happened in this estimation was similar to the V_{1N} demand, in Market 3. The prior constraints had no effect on the likelihood and became non-informative priors since the OLS estimates already had correct signs.

Nevertheless, two exceptions can be commented on that. It is noticeable that both f_{1R} and f_{2A} have an abrupt cut-off at zero, which may show that even though the mean value of OLS estimates already have correct signs, they were spreading to unfeasible values, which was cut by the prior constraint. Therefore, even though, the prior has not directly affected the mean value, it had a soft pressure in the distribution as a whole.

That, however, appears to be a weak pressure given the large standard deviations resulting and estimates to similar to the OLS, as explained in Market 3. Except by the g and γ, all OLS estimates are lying inside the HDI interval.

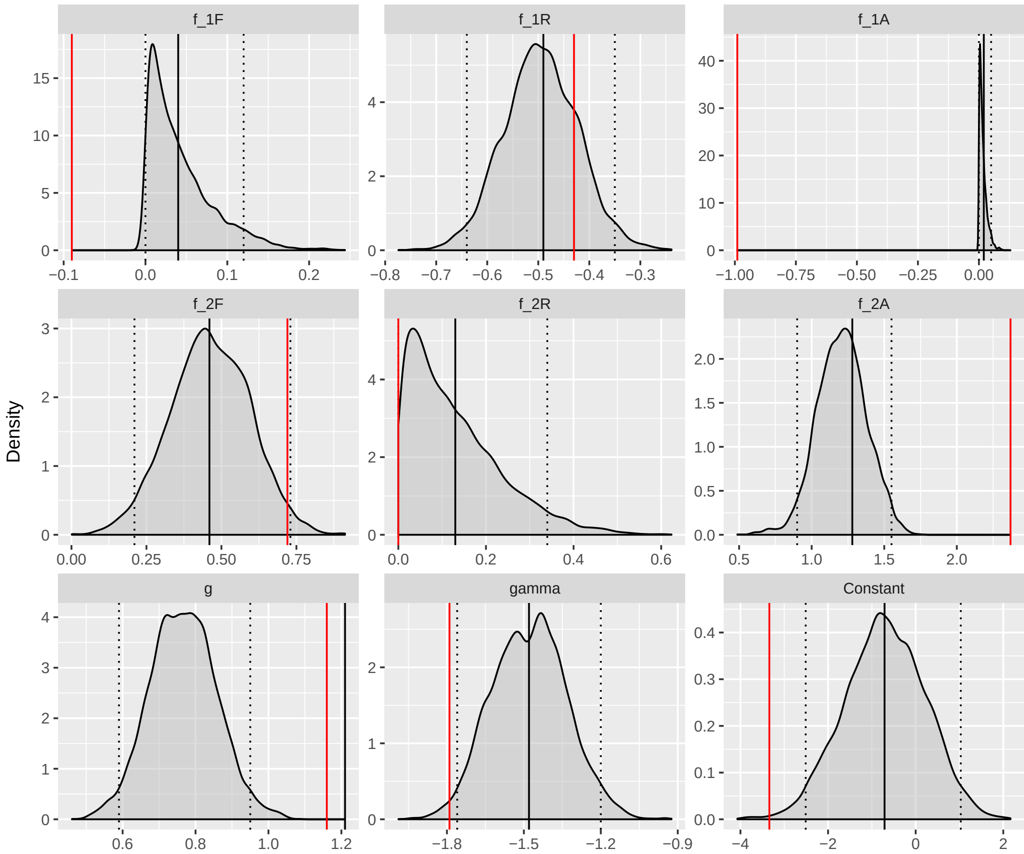


Figure 4.12: Posterior density function of elasticities w.r.t V_{1R} - Market 4

Source: Own work

With respect to V_{1R} demand, Figure 4.12, three elasticities presented satisfactorily small HDI: f_{1F}, f_{1R} and f_{1A}, with 0.12, 0.29 and 0.05, respectively. It's noticeable that the shortest interval was the one from the sign-reversed elaticities, f_{1F} and f_{1A}. For all others, it was above 0.30, up to 0.65 for the f_{2R} elasticity. Detailed information of the HDI is presented in the Appendix C.

Even though the large HDIs, it was possible to distinguish the bayesian from OLS estimates: in addition to the sign-reversed coefficient, which should already be expected, the OLS lied out of the HDI for f_2A, g and γ.

Similarly to previous estimations, some distributions are skewed towards the zero constraints with sharp edges. Analogous interpretation is applicable in this regard.

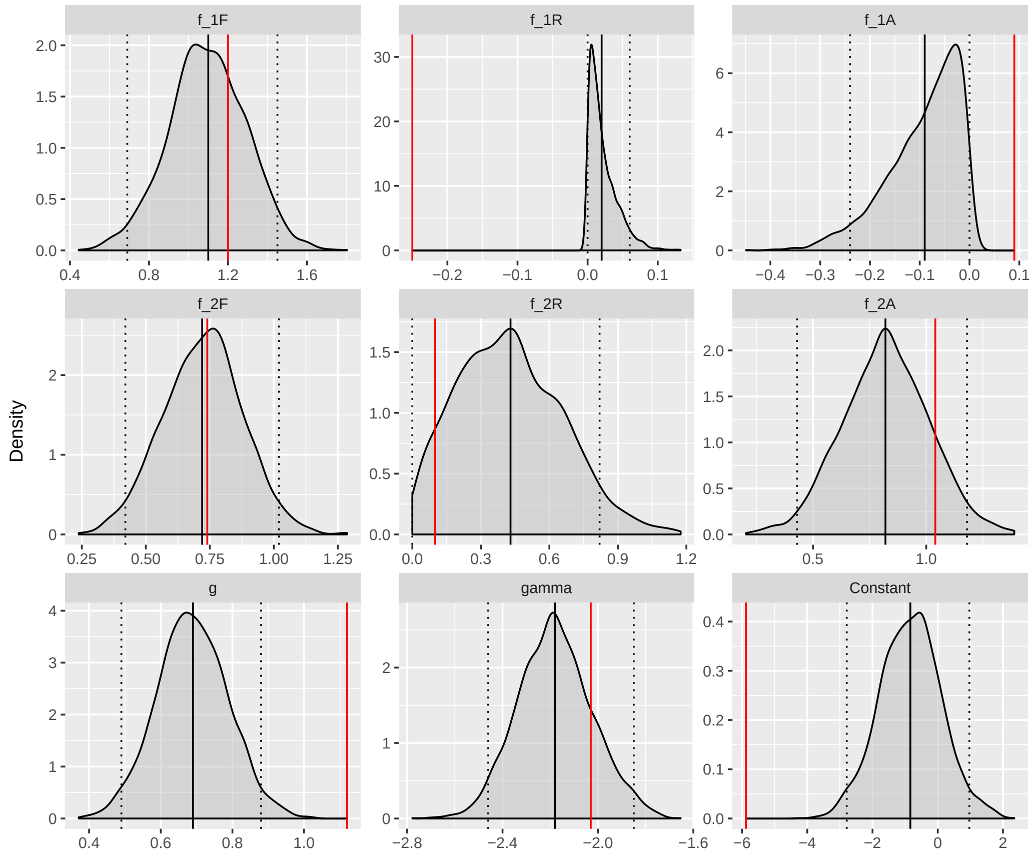


Figure 4.13: Posterior density function of elasticities w.r.t V_{1A} - Market 4

Source: Own work

With respect to V_{1A} demand, Figure 4.13, only two elasticities presented short HDI's amplitude - 0.06 for f_{1R} and 0.24 for f_{1A}, which turn to be the constrained sign-reversed elasticities. Once again, skewness is observed and the sharp edge at zero has a similar interpretation from previous models. For all others coefficients, it was above 0.30, up to 0.82 for the f_{2F}. Detailed information of the HDI is presented in the Appendix C.

As it should be expected the OLS estimate lied out of the HDI for the sign-reversed elasticities. For all other elasticities, but g, the HDI was very large and the OLS estimates lied inside the range of 95% probable values.

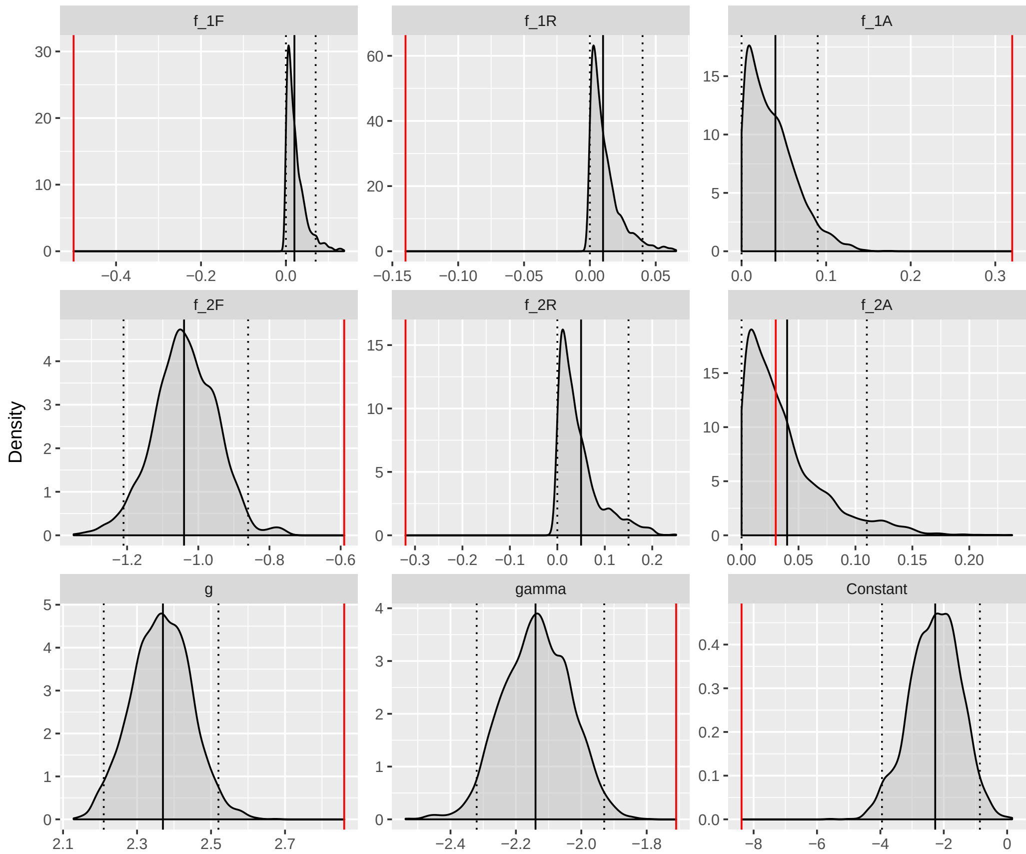


Figure 4.14: Posterior density function of elasticities w.r.t V_{2F} - Market 4

Source: Own work

The V_{2F} model, Figure 4.14, was the estimation with the highest number of actual constraints. It is noticeable that the zero boundary was an actual barrier to the f_{1F}, f_{1R}, f_{1A}, f_{2R} and f_{2A} elasticities and these were the five OLS estimates wrong sign coefficients that were reverted in the Bayesian estimation. With strong restrictions on the priors, the HDI was significantly shorter than in other estimations in this market, with five HDI ranging below 0.30. Even the ones higher 0.30 were around it, up to 0.38. Detailed information of the HDI is presented in the Appendix C.

Only the OLS elasticity of f_{2A} lied inside the HDI interval, which demonstrates that squeezing the credible interval brought relevant information to the analysis.

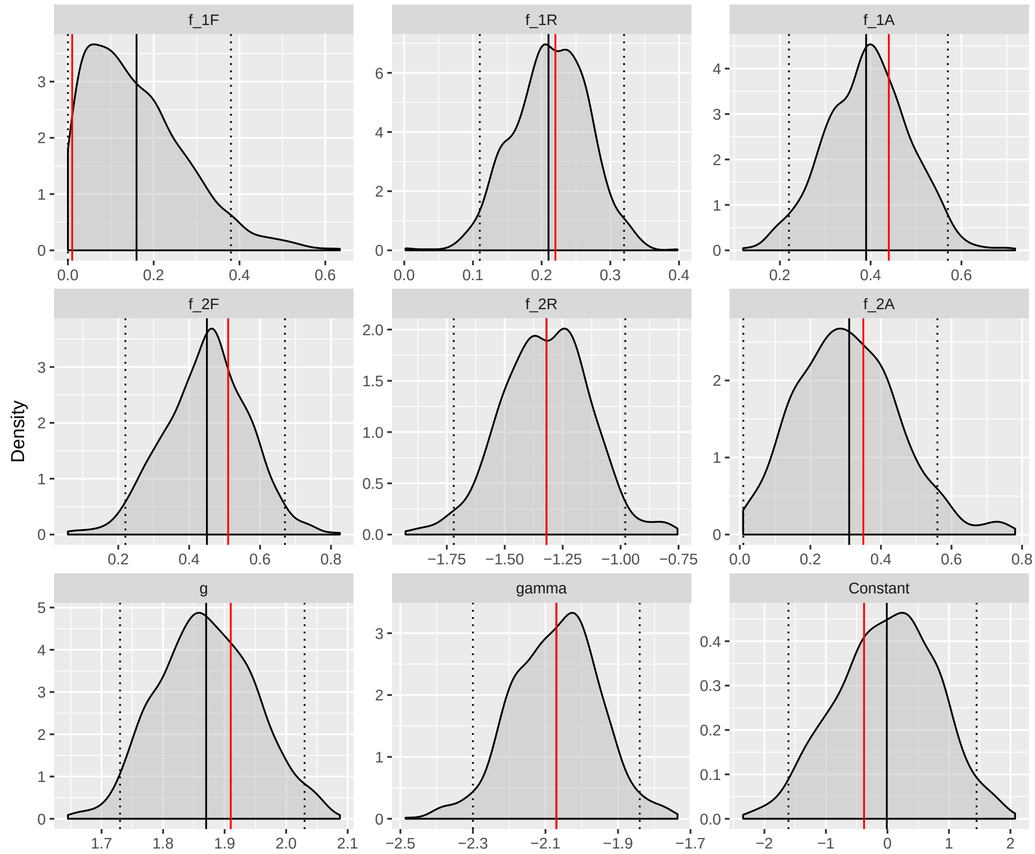


Figure 4.15: Posterior density function of elasticities w.r.t V_{2R} - Market 4

Source: Own work

The estimation of V_2R demand, Figure 4.15, was very similar to the V_{1F} since any OLS estimates have wrong-sign. As explained before, without strong constraints, the likelihood prevailed and so the difficulty to address the correlation problems. Analogous interpretation is applied in this regard.

The result was, once more, very large HDI, which means high uncertainty in the estimated elasticities. Only the f_{1R} coefficient had HDI lower than 0.30.

Given that, the resultant estimates were very similar to the OLS estimates - all of them were inside the HDI. Even the f_{2R} had a coincident mean.

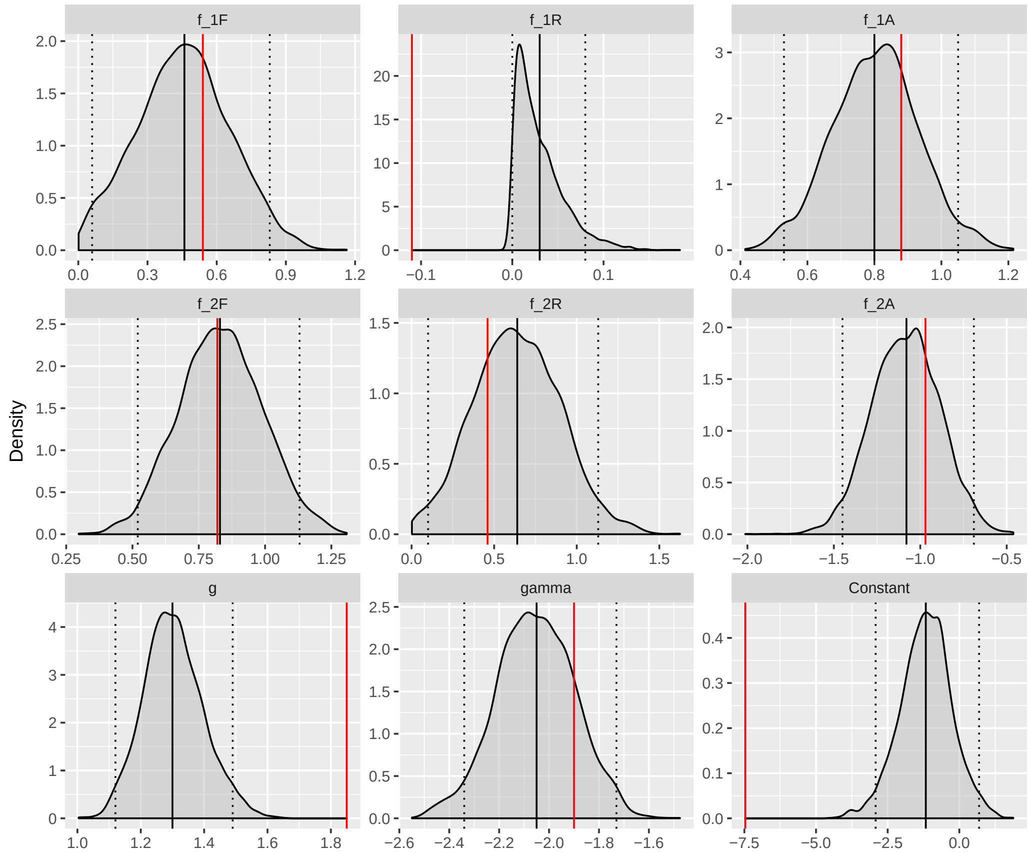


Figure 4.16: Posterior density function of elasticities w.r.t V_{2A} - Market 4

Source: Own work

With respect to V_2A demand, Figure 4.16, only the f_{1R} elasticity suffered an actual constraint by the prior. As one may notice, it was, as usual, a wrong-sign OLS estimate reversed in the Bayesian estimation, resulting in a skewed distribution towards its boundaries.

It appears, however, that because it is a very complex model, this was not enough to shape the other variables since all of HDI remained very large, up to 1.02. Once again, OLS lied inside the HDI, which makes one consider them as credible values, except for the f_{1R}, which should already be expected since it is the reversed coefficient, and the g elasticities.

iv. Magnitude of estimates

Analysing the magnitude of coefficients, the most sensitive fare with regard its own price is the *Standard Reduced*, with mean -1.32, followed by the *Standard Advance*, *Standard Full* and *First Class Full*, with almost proportional effects - -1.08, -1.04 and -1.02, respectively. Smaller own elasticities were observed for the other *First Class*, -0.51 and -0.09, which suggests less elastic demands. These values are considerably different from Market 3.

Regarding to the cross elasticities, the *First Class Full* fare has very small - almost null - price effects in the demand of the *First Class Reduced* and *Standard Full*, 0.02 and 0.05, respectively. Higher impacts, but still less than proportional occur in the demand of *Standard Reduced* and *Standard Advance*, 0.16 and 0.46, respectively. The *First Class Advance* is surprisingly high, with 1.10, suggesting that, for first class passengers, the *Advance* demand is very sensitive to changes in the *Full* fare. This may sign there is a cost for passengers to plan their trips in advance.

For the *First Class Reduced*, the most significant cross elasticities regarded the impact in the V_{2R} demand, 0.21, all the other were very close to zero.

For the *First Class Advance*, the impact in the V_{1R} and V_{2F} demand was also very close to zero. Higher effects were observed impacting the *First Class Full* and *Standard Reduced*, being of 0.24 and 0.39, respectively. The highest impact regards the *Standard Advance* demand, deemed as 0.80.

Conversely, the *Standard Full* have not presented such small cross elasticities. The smallest one regarding the impact in the *Standard Reduced* demand, comparable to the effect in the *First Class Reduced*, deemed as 0.46. The highest impact was 0.91 with regard to the *First Class Full* demand.

For the *Standard Reduced*, the cross elasticities with respect to the *First Class Reduced* and *Standard Full* demand were the smallest, 0.13 and 0.05, respectively. The biggest effect is the impact in the *First Class Full*, deemes as 0.87.

The *Standard Advance* had a small cross effect with respect to the *First Class Full* and *Standar Full*, 0.16 and 0.04, respectively. A very high impact was observed for the *First Class Reduced* 1.28 - even higher than the own elasticity.

Generally, cross elasticities have ranged for very small values, close to zero, to very high ones, above 1. It was not possible to understand the interaction pattern neither even why some cross elasticities are higher than the own-elasticity.

The GVA elasticity it worths noticing that the $g_F > g_R$ for both *First* and *Standard Class*, which is consistent with the fact that the percentage of business journeys is higher in *Full* tickets. Additionally, the *Standard Class* elasticities are higher relative to the *First Class*, which may suggest that increase in productivity impacts more the mass of medium level professionals relatively to the white collar professional, that demand first class services.

For the GJT effects, the elasticities were considerably high: from -1.42, for the *First Class Reduced* demand, to -2.66, for the *First Class*. Even though there is the same GVA's reservation that states that elasticities can be expected to vary by ticket type (PDFC 2005), it is markedly above general interval expected by PDFH - -0.7 to -1.1 (PDFC 2011).

5 Discussion

5.1 Conclusion

Overall, the application of Bayesian econometrics in the estimation of rail fare elasticities have presented itself as a potential alternative to the current methods. As a proof of concept, this work has successfully worked Bayesian regressions estimating fare elasticities with coherent signs differentiating up to six fares in a market.

The primary advantage is its efficiency to estimate correct algebraic sign elasticities. All elasticities estimates were in accordance to the what is expected from the PDFH. Even for the complex Market 4, in which the *First Class* tickets were broken into *Full*, *Reduced* and *Advance*, the Bayes estimation was able to revert ten wrong signs from the SURE/OLS estimates.

The secondary advantage was the correlation being reverted to a useful feature that has helped to address the effects among correlated predictors due to the anti-correlation property of the estimates. When strong priors were defined the result was, generally, smaller standard deviations, which in turn means shorter ranges of credible values and more precise estimates.

Another advantage of Bayesian statistics regards its inherent way of interpreting probabilities, since the interpretation of posterior distribution as actual probabilities interval is much more intuitive than interpreting confidence intervals, from the sampling theory approach.

Even though the benefits, there were some issues that must be mentioned. Firstly, the occurrence of diverted transitions, which is an indicator of biased estimates. This kind of warning is supposed to raise a flag on the quality of the resulting posterior distributions. However, their occurrence may be directly related to the constraints applied in the domain of the prior distributions, which have prevented the MCMC to fully explore some regions of the parameter space. In this sense, this issue may be regarded a necessary evil to the easy application of constrained Bayes estimation in the data set.

Additionally, it was clear that the estimation has lost in quality as the models increased in complexity. This can be noticed from the difference of precision between the Markets 1 and 2, with short HDI, against Markets 3 and 4, with wider ones. Nevertheless, the later estimates can still be considered useful if one is willing to trade-off uncertainty for precision. In this study the HDI adopted regarding a range of values that represented 95% of the probability. If the degree of certainty is reduced, for instance for an HDI of 80%, so it will be reduced the range of credible values.

Also, in which regard the adequacy of the magnitude os estimates, even though it is recognised that it demands further development given the complexity of interactions affecting the fare elasticities, it must be pointed that some occurrences were notably out of expectations. Extremely high cross elasticities and cross elasticities bigger than the own elasticities, for instance, may deserve attention. These issues must be further investigated.

Nevertheless, acknowledging the weaknesses, the overall result was positive and with a potential practical application.

5.2 Next frontiers

As an introductory work on Bayesian econometrics applied to rail fare elasticities estimation there is plenty of further considerations that can improve the method.

In general aspects, issues that have already considered in previous studies as dynamic elasticities and the quota control for *Advance* tickets were out of scope. These elements are, however, of undeniable relevance. The first one regards the differentiation between short and long-run elasticities which helps understanding how the price effects across time, which part of the price sensitiveness is immediately and which is reflected in the long-run. Some studies have reported also a cumulative effect (immediate + future) - for instance, the impact of promotions (Kopalle et al (1999) in Liu et al. (2009)). The second one regards the fact that *Advance* tickets are not available for all passenger, since they may sell out quickly. This supply restriction must be taken into account.

Specifically, regarding Bayesian econometrics, another possibility of should to be explored is the adoption of hierarchical models. As discussed in Chapter 2, hierarchical models are being successfully applied in the retail market to estimate cross elasticities of competitors products. The rail fares have an inherent meaningful hierarchical nature that can be explored. Analogously to the retail market, in which national elasticities are decomposed to store-to-store elasticities, being able to keep regional market features, so can the broad fare elasticities be decomposed, even to route-specific elasticities. This could be a promising method to increase the ability of train operating companies in managing fares.

Bibliography

Butcher, L. (2017), Rail fares and ticketing, Technical report, University of Zurich, Department of Informatics.
URL: *http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN01904*

DfT (2016), 'Rail demand forecast estimation'.
URL: *https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/610059/phase2-rail-demand-forecasting-estimation-study.pdf*

Gelman, A. (2017), 'Theres no such thing as unbiased estimation. and its a good thing, too.'.
URL: *http://andrewgelman.com/2015/05/11/theres-no-such-thing-as-unbiased-estimation-and-its-a-good-thing-too/*

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014), *Bayesian data analysis*, Vol. 2, CRC press Boca Raton, FL.

Geweke, J. (1986), 'Exact inference in the inequality constrained normal linear regression model', *Journal of Applied Econometrics* **1**(2), 127–141.

Griffiths, W. E. et al. (1988), 'Bayesian econometrics and how to get rid of those wrong signs', *Review of Marketing and Agricultural Economics* **56**(1), 36–56.

Gujarati, D. N. & Porter, D. C. (2009), *Basic Econometrics*, 5th edn, New York: McGraw-Hill.

ITS & Systra (2015), 'Conditional elasticity of demand to fares - Proposal'.

ITS & Systra (2016), 'Conditional elasticity of demand to fares - Final Report'.

Kennedy, P. (2003), *A guide to econometrics*, MIT press.

Koop, G. (2003), *Bayesian econometrics*, Wiley-Interscience.

Kruschke, J. (2011), 'Doing bayesian data analysis - blog'.
URL: *http://doingbayesiandataanalysis.blogspot.co.uk/2011/11/thinning-to-reduce-autocorrelation.html*

Kruschke, J. (2014), *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*, Academic Press.

Kruschke, J. K., Aguinis, H. & Joo, H. (2012), 'The time has come: Bayesian methods for data analysis in the organizational sciences', *Organizational Research Methods* **15**(4), 722–752.

Liu, Q., Otter, T. & Allenby, G. M. (2009), 'Measurement of own-and cross-price effects', *Handbook of Pricing Research in Marketing* p. 61.

McElreath, R. (2016), *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, 1th edn, CRC Press.

Meredith, M. & Kruschke, J. (n.d.), 'Package HDInterval'.
URL: *https://cran.r-project.org/web/packages/HDInterval/HDInterval.pdf*

Montgomery, A. L. & Rossi, P. E. (1999), 'Estimating price elasticities with theory-based priors', *Journal of Marketing Research* pp. 413–423.

PDFC (2005), Passenger demand forecast handbook, Technical report, Passenger Demand Forecasting Council.

PDFC (2011), Passenger demand forecast handbook, Technical report, Passenger Demand Forecasting Council.

Rail Delivery Group (2015), 'Passenger demand forecasting council'.
URL: *https://www.raildeliverygroup.com/pdfc.html*

SRA (2003), Fares review conclusions - 2003, Technical report, United Kingdom - Strategic Rail Authority.

Stan (2017*a*), 'Brief guidances on stan's warnings'.
URL: *http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup*

Stan (2017*b*), 'Prior choice recommendation'.
URL: *https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations*

Stan (2017*c*), 'Stan'.
URL: *http://mc-stan.org/users/documentation/*

Stan (2017*d*), Stan Modelling Language - user's guide and reference manual, Technical report.
URL: *http://mc-stan.org*

Studenmund, A. H. (2011), *Using econometrics: a practical guide*, 6th edn, Addison-Wesley, Boston.

Wardman, M. & Toner, J. (2003), 'Econometric modelling of competition between train ticket types'.

A Linear Models

Table A.1: Linear models by market

Market	Equations
1	$lnV_{2Fi} = ln\ a + g\ lnGVA_i + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + \gamma\ lnGJT_i$
	$lnV_{2Ri} = ln\ a + g\ lnGVA_i + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + \gamma\ lnGJT_i$
2	$lnV_{2Fi} = ln\ a + g\ lnGVA_i + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{2Ri} = ln\ a + g\ lnGVA_i + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{2Ai} = ln\ a + g\ lnGVA_i + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
3	$lnV_{1Ni} = ln\ a + g\ lnGVA_i + f_{1N}\ lnP_{1Ni} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{2Fi} = ln\ a + g\ lnGVA_i + f_{1N}\ lnP_{1Ni} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{2Ri} = ln\ a + g\ lnGVA_i + f_{1N}\ lnP_{1Ni} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{2Ai} = ln\ a + g\ lnGVA_i + f_{1N}\ lnP_{1Ni} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
4	$lnV_{1Fi} = ln\ a + g\ lnGVA_i + f_{1F}\ lnP_{1Fi} + f_{1R}\ lnP_{1Ri} + f_{1A}\ lnP_{1Ai} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{1Ri} = ln\ a + g\ lnGVA_i + f_{1F}\ lnP_{1Fi} + f_{1R}\ lnP_{1Ri} + f_{1A}\ lnP_{1Ai} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{1Ai} = ln\ a + g\ lnGVA_i + f_{1F}\ lnP_{1Fi} + f_{1R}\ lnP_{1Ri} + f_{1A}\ lnP_{1Ai} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{2Fi} = ln\ a + g\ lnGVA_i + f_{1F}\ lnP_{1Fi} + f_{1R}\ lnP_{1Ri} + f_{1A}\ lnP_{1Ai} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{2Ri} = ln\ a + g\ lnGVA_i + f_{1F}\ lnP_{1Fi} + f_{1R}\ lnP_{1Ri} + f_{1A}\ lnP_{1Ai} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$
	$lnV_{2Ai} = ln\ a + g\ lnGVA_i + f_{1F}\ lnP_{1Fi} + f_{1R}\ lnP_{1Ri} + f_{1A}\ lnP_{1Ai} + f_{2F}\ lnP_{2Fi} + f_{2R}\ lnP_{2Ri} + f_{2A}\ lnP_{2Ai} + \gamma\ lnGJT_i$

Source: Own work

B Bayes estimation with weakly informative priors

Table B.1: Comparison of elasticities estimates - Martket 1

	Bayes		*SURE/OLS*	
	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2F}$	$ln\ V_{2R}$
f_{2F}	-1.33	0.70	-1.28	0.81
	(0.05)	(0.05)	(0.05)	(0.05)
f_{2R}	0.17	-0.88	0.17	-0.89
	(0.05)	(0.05)	(0.05)	(0.05)
g	0.77	0.04	0.49	-0.52
	(0.06)	(0.05)	(0.7)	(0.7)
γ	-0.88	-0.88	-0.98	-1.06
	(0.04)	(0.04)	(0.04)	(0.05)
Constant	6.05	11.43	9.14	17.6
	(0.58)	(0.57)	(0.71)	(0.73)

Source: Own work

Table B.2: Comparison of elasticities estimates - Martket 2

	Bayes			*SURE/OLS*		
	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$
f_{2F}	-1.24	-0.02	-0.11	-1.24	0.03†	-0.18
	(0.06)	(0.06)	(0.09)	(0.06)	(0.06)	0.10)
f_{2R}	-0.04	-0.66	1.02	-0.06	-0.60	1.03
	(0.08)	(0.08)	(0.12)	(0.08)	(0.08)	(0.13)
f_{2A}	0.15	-0.10	-0.47	0.15	-0.10	-0.47
	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.06)
g	1.68	0.97	0.28	1.75	0.39	0.66
	(0.07)	(0.07)	(0.09)	(0.11)	(0.10)	(0.17)
γ	-1.25	-1.05	0.23	-1.23	-1.21	0.33
	(0.06)	(0.06)	(0.09)	(0.07)	(0.06)	(0.10)
Constant	-0.37	5.39	-1.34	-1.09†	11.72	-5.53
	(0.74)	(0.75)	(0.86)	(1.15)	(1.09)	(1.77)

Note: † $p > 0.1$

Source: Own work

Table B.3: Comparison of elasticities estimates - Martket 3

	Bayes				*SURE/OLS*			
	$ln\ V_{1N}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$	$ln\ V_{1N}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$
f_{1N}	-0.82	-0.42	-0.41	-0.47	-0.84	-0.43	-0.44	-0.50
	(0.06)	(0.05)	(0.05)	(0.07)	(0.06)	(0.05)	(0.05)	(0.07)
f_{2F}	1.03	-1.13	0.20	1.13	1.03	-1.12	0.23	1.07
	(0.08)	(0.07)	(0.07)	(0.10)	(0.09)	(0.07)	(0.07)	(0.10)
f_{2R}	1.73	-0.06	0.23	2.23	1.80	-0.04	0.34	2.28
	(0.10)	(0.09)	(0.08)	(0.13)	(0.11)	(0.09)	(0.09)	(0.13)
f_{2A}	0.44	0.33	0.19	-0.42	0.43	0.33	0.15	-0.39
	(0.06)	(0.05)	(0.05)	(0.07)	(0.06)	(0.05)	(0.05)	(0.07)
g	1.52	1.78	1.29	1.11	1.50	1.66	0.80	1.54
	(0.08)	(0.07)	(0.07)	(0.08)	(0.11)	(0.09)	(0.09)	(0.13)
γ	-3.25	-1.57	-2.17	-2.06	-3.30	-1.60	-2.31	-2.01
	(0.08)	(0.06)	(0.06)	(0.09)	(0.082)	(0.07)	(0.07)	(0.10)
Constant	0.23	1.51	6.15	-2.21	$0.62^{\dagger}$	2.85	11.66	-6.72
	(0.78)	(0.73)	(0.70)	(0.82)	(1.23)	(1.00)	(0.95)	(1.45)

Note: † $p > 0.1$

Source: Own work

Table B.4: Comparison of elasticities estimates - Martket 4

	Bayes						*SURE/OLS*					
	$ln\ V_{1F}$	$ln\ V_{1R}$	$ln\ V_{1A}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$	$ln\ V_{1F}$	$ln\ V_{1R}$	$ln\ V_{1A}$	$ln\ V_{2F}$	$ln\ V_{2R}$	$ln\ V_{2A}$
f_{1F}	-1.02	0.05	1.10	0.02	0.16	0.46	-0.94	-0.09†	1.20	-0.50	0.01†	0.54
	(0.19)	(0.05)	(0.19)	(0.02)	(0.12)	(0.20)	(0.22)	(0.27)	(0.27)	(0.20)	(0.19)	(0.26)
f_{1R}	0.05	-0.51	0.02	0.01	0.21	0.03	0.06†	-0.43	-0.25	-0.14	0.22	-0.11
	(0.04)	(0.08)	(0.02)	(0.01)	(0.05)	(0.03)	(0.06)	(0.08)	(0.08)	(0.06)	(0.05)	(0.07)
f_{1A}	0.24	0.02	-0.09	0.04	0.39	0.80	0.28	-0.99	0.09†	0.33	0.44	0.88
	(0.11)	(0.02)	(0.07)	(0.03)	(0.09)	(0.13)	(0.125)	(0.151)	(0.154)	(0.114)	(0.108)	(0.147)
f_{2F}	0.91	0.59	0.72	-1.04	0.45	0.83	0.85	0.71	0.74	-0.59	0.51	0.82
	(0.13)	(0.16)	(0.16)	(0.09)	(0.12)	(0.16)	(0.15)	(0.18)	(0.18)	(0.14)	(0.13)	(0.17)
f_{2R}	0.87	-0.13	0.43	0.05	-1.32	0.64	0.67	0.00†	0.10†	-0.32†	-1.32	0.46†
	(0.22)	(0.23)	(0.23)	(0.04)	(0.19)	(0.26)	(0.27)	(0.33)	(0.34)	(0.25)	(0.23)	(0.32)
f_{2A}	0.16	1.28	0.82	0.04	0.31	-1.08	0.18†	2.37	1.04	0.03†	0.35	-0.99
	(0.12)	(0.17)	(0.19)	(0.03)	(0.14)	(0.20)	(0.18)	(0.22)	(0.22)	(0.16)	(0.16)	(0.21)
g	1.91	0.74	0.69	2.37	1.87	1.30	2.74	1.17	1.12	2.86	1.91	1.86
	(0.09)	(0.10)	(0.10)	(0.08)	(0.08)	(0.09)	(0.18)	(0.22)	(0.22)	(0.17)	(0.16)	(0.22)
γ	-2.66	-1.42	-2.18	-2.14	-2.07	-2.05	-2.39	-1.79	-2.03	-1.71	-2.07	-1.90
	(0.13)	(0.15)	(0.16)	(0.10)	(0.12)	(0.16)	(0.15)	(0.18)	(0.18)	(0.14)	(0.13)	(0.18)
Constant	-2.40	-0.64	-0.84	-2.27	-0.01	-1.17	-11.98	-3.34†	-5.89	-8.38	-0.38†	-7.47
	(0.89)	(0.93)	(0.94)	(0.79)	0.81	(0.04)	(2.09)	(2.53)	(2.57)	(1.91)	(1.80)	(2.46)

Note: † $p > 0.1$

Source: Own work

C Bayes estimation with constrained priors - HDI

Table C.1: Higher Density Intervals - Market 1

	V_{2F}			V_{2R}		
	L	U	A	L	U	A
f_{2F}	-1.42	-1.23	0.19	0.61	0.80	0.19
f_{2R}	0.07	0.26	0.19	-1.00	-0.80	0.20
g	0.67	0.87	0.20	0.00	0.14	0.14
γ	-0.98	-0.80	0.17	-0.97	-0.80	0.18
Constant	4.97	7.11	2.14	10.32	12.06	1.74

L = lower bound, U = upper bound, A = amplitude

Source: Own work

Table C.2: Higher Density Intervals - Market 2

	V_{2F}			V_{2R}			V_{2A}		
	L	U	A	L	U	A	L	U	A
f_{2F}	-1.37	-1.20	0.17	0.00	0.09	0.09	0.00	0.13	0.13
f_{2R}	0.00	0.14	0.14	-0.91	-0.70	0.20	0.84	1.15	0.31
f_{2A}	0.06	0.21	0.15	0.00	0.03	0.03	-0.57	-0.34	0.24
g	1.54	1.83	0.28	0.82	1.09	0.27	0.13	0.45	0.32
γ	-1.39	-1.17	0.22	-1.18	-0.95	0.23	-0.08	-0.00	0.08
Constant	-1.89	1.14	3.03	4.11	6.86	2.75	-2.29	0.92	3.21

L = lower bound, U = upper bound, A = amplitude

Source: Own work

Table C.3: Higher Density Intervals - Market 3

	V_{1N}			V_{2F}			V_{2R}			V_{2A}		
	L	U	A	L	U	A	L	U	A	L	U	A
f_{1N}	-0.94	-0.70	0.24	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.03	0.03
f_{2F}	0.86	1.20	0.34	-1.50	-1.30	0.20	0.06	0.27	0.21	0.84	1.23	0.39
f_{2R}	1.50	1.94	0.43	0.00	0.06	0.06	-0.15	-0.00	0.15	1.74	2.18	0.44
f_{2A}	0.32	0.56	0.24	0.08	0.28	0.20	0.01	0.18	0.17	-0.69	-0.40	0.29
g	1.38	1.67	0.29	1.60	1.86	0.26	1.12	1.38	0.26	0.87	1.19	0.32
γ	-3.41	-3.11	0.30	-1.66	-1.44	0.22	-2.17	-1.94	0.23	-2.13	-1.77	0.36
Constant	-1.31	1.66	2.98	0.20	2.88	2.68	4.54	7.18	2.65	-3.99	-0.58	3.41

L = lower bound, U = upper bound, A = amplitude

Source: Own work

Table C.4: Higher Density Intervals - Market 4 [Part 1/2]

	V_{1F}			V_{1R}			V_{1A}		
	L	U	A	L	U	A	L	U	A
f_{1F}	-1.38	-0.66	0.72	0.00	0.12	0.12	0.69	1.45	0.76
f_{1R}	0.00	0.13	0.13	-0.64	-0.35	0.29	0.00	0.06	0.06
f_{1A}	0.03	0.44	0.41	0.00	0.05	0.05	-0.24	-0.00	0.24
f_{2F}	0.63	1.16	0.52	0.21	0.73	0.52	0.42	1.02	0.61
f_{2R}	0.41	1.28	0.87	0.00	0.34	0.34	0.00	0.82	0.82
f_{2A}	0.00	0.39	0.39	0.90	1.55	0.65	0.43	1.18	0.74
g	1.75	2.08	0.33	0.59	0.95	0.36	0.49	0.88	0.39
γ	-2.92	-2.41	0.51	-1.76	-1.20	0.56	-2.46	-1.85	0.61
Constant	-4.15	-0.63	3.52	-2.51	1.03	3.54	-2.79	0.97	3.75

L = lower bound, U = upper bound, A = amplitude

Source: Own work

Table C.5: Higher Density Intervals - Market 4 [Part 2/2]

	V_{2F}			V_{2R}			V_{2A}		
	L	U	A	L	U	A	L	U	A
f_{1F}	0.00	0.07	0.07	0.00	0.38	0.38	0.06	0.83	0.77
f_{1R}	0.00	0.04	0.04	0.11	0.32	0.21	0.00	0.08	0.08
f_{1A}	0.00	0.09	0.09	0.22	0.57	0.35	0.53	1.05	0.52
f_{2F}	-1.21	-0.86	0.34	0.22	0.67	0.45	0.52	1.13	0.61
f_{2R}	0.00	0.15	0.15	-1.72	-0.98	0.74	0.10	1.13	1.02
f_{2A}	0.00	0.11	0.11	0.01	0.56	0.54	-1.45	-0.69	0.76
g	2.21	2.52	0.31	1.73	2.03	0.30	1.12	1.49	0.37
γ	-2.32	-1.93	0.38	-2.30	-1.84	0.46	-2.34	-1.73	0.61
Constant	-3.95	-0.86	3.09	-1.61	1.45	3.07	-2.92	0.69	3.61

L = lower bound, U = upper bound, A = amplitude

Source: Own work