A study of the temporal flow of passenger and cargo transport in a Brazilian region: use of different statistical models for better forecasts and discovery of important factors that affect the efficiency of the Brazilian air transport system

Ms. Luiz Rodrigo Bonette [1], Dr. Walther Azzolini Junior *[1,2], Dr. Jorge Alberto Achcar [1,3]

[1]University Araraquara (UNIARA) *[1,2]Universidade de São Paulo (USP), *[1,3] University Araraquara

Abstract: *Better efficiency of the air transport system of a country at the national level, especially in terms of its capacity to generate value for passenger flow and cargo transport, effectively depends on the identification of the demand generation potential of each hub for this type of service. This requires the mapping of the passenger flow and volume of cargo transport of each region served by the system and the number of connections. The main goal of this study was to identify important factors that account for the great variability (demand) of regional hubsof the airport modal system in operation in the State of São Paulo, the most populated and industrialized in the Southeast region in Brazil. For this purpose, datasets for each airport related to passengers or cargo flow were obtained from time series data in the period ranging from January 01, 2008 to December 31, 2014. Different data analysis approaches could imply in better mapping of the flow of the air modal system from the evaluation of some factors related to operations/volume. Therefore, different statistical models - such as multiple linear regression with normal errors and new stochastic volatility (SV) models - are introduced in this study, to provide a better view of the operation system in the four main regional hubs, within a large group of 32 airports reported in the dataset.*

Keywords*: hub-and-spoke, multiple linear regression modeling, Bayesian analysis, volatility stochastic models.*

I. Introduction

The initial studies in hub-and-spoke transport networks always related themes addressed to the transport of passengers and hub location problems, the so-called p-hub. In the 2000 and 2010 decades, new studies associating an extension of the main related issues, such as transport of air cargo, were introduced in the literature as a strategy to identify the regional hub profile, consequently leading to new research themes.

In this direction, Alumur and Kara (2008) introduced the modeling of a network that operates with the hub concept, that is, a new class of models denoted as the network hub location models (NHLM).

In another paper, Farahani et al. (2009) introduced a good review of methods and applications for a special class of models denoted as Hub Localization Problems (HLP). Later, due to the importance of these models, Campbell and O'Kelly (2012) provided an overview of the first motivations of the academy regarding specific studies with emphasis on the analysis of Hub Localization Problems (HLP) as the highlight for future research.

In this context, the issue highlighted for research in this paper focuses on how to determine strategic hubs linking the flow of passengers to the volume of air cargo transported in a given region. In this case, the mapping of the key factors of regional hubs through the statistical analysis of airport operations and economic covariates compose the general goal of this paper.

The specific goals associated to air cargo in Brazil and the world are the application of different statistical models to analyze the variables and covariates assuming a dataset linked to a group of 32 airports in the State of São Paulo obtained from the statistical handling reports of the São Paulo Airway Department (DAESP-http://www.daesp.sp.gov.br/), in Brazil. DAESP is an office linked to the air transport in the State of São Paulo. In this study, we consider the comparison between the development and the seasonality of passenger and air cargo and the determination of strategic factors that form regional hubs from the comparative findings between passenger and air cargo in terms of airport economic operations. From the 32 airports reported in the dataset, our study considers the four most important ones, which provides a good overview of the sector.

The shortage of hub-and-spoke studies associated with air cargo transport justifies this research, and its main contribution is to analyze the composition of the formation of regional hubs through airport operations and economic factors. The use of statistical models such as stochastic volatility time series models and linear regression models brings us the relevant findings using the dataset from DAESP.

1.1 A brief review of the literature

The customization of p-hub designs for air cargo (Morrel and Pilon, 1999) and location distribution of air cargo systems due to transient hubs are described by Kara and Tansel (2001). Lin et al. (2003) study the indirect connections to large postal cargo hub networks in general and the multiple use of aircraft capabilities to consolidate center-to-center hubs and the development of flow models to generate improvements in the strategic mapping of hub networks for air cargo.

Gardiner, Humphreys and Ison (2005), Gardiner and Ison (2008), Gardiner, Humphreys and Ison (2016) quantify the importance of an individual airport through the concept of "airport hierarchy" (primary hubs, secondary hubs, tertiary hubs). Alumur et al. (2007, 2009) emphasize the need to determine potential flows at airports through the criterion of the analysis of volumes of cargo operations in a multi-modal and hierarchical network.

Several studies emphasize issues related to the capacity of an airport (see, for example, Smilowitz and Daganzo, 2007; Tan and Kara, 2007; Scholz and Cossel, 2011; Petersen, 2007; Bartodziej et al., 2009; Leung et al., 2009; Li et al., 2009; Wang and Kao, 2008; Amaruchkul and Lorchirachoonhkul, 2011).

Assaf and Gillen (2012) highlight, according to the literature on measuring and understanding factors affecting airport efficiency, that new studies have been introduced considerably over the last few years using different methodologies. The authors emphasize the use of complex analytical methods, such as Data Envelopment Analysis (DEA) and stochastic frontier (SF) distance functions.

Wu, et al. (2011) reported the growth of hub-and-spoke networking, allowing for major airports to limit the size of passenger demand in the air capture to become the main hubs in their respective regions for cargo. Gardiner and Ison (2008) reported the existence of three classes of important decisions for the air cargo operator to choose the airport that will operate and the geographical location of the airport, financial return and airport security operations. Scholz and Cossel (2011) point out three important points that serve as an argument and reflection for this research:

- ✓ The growth of air cargo tonnages changing the relationship between passengers and cargo has become a significant source of revenue for airlines and airports;
- ✓ The importance of individual airports on a network usually assessed by measures based on passengers, cargo and operating numbers;
- ✓ The combined passenger and cargo services operated in the major airlines.

Heinicke (2006, 2007), Scholz and Cossel (2011) and Onghena (2011) point out that the literature is scarce on this subject and this type of research has been mostly focused on the integration between airlines and not on the integration of hub-and-spoke networks. Some important studies on the subject are also introduced by other authors (Grosso and Shepherd, 2011; Alumur et al., 2012; Bowen, 2012; Oktal and Özger, 2013; Lakew and Tok, 2015). Feng et al. (2015) describe that cargo transport is more complex than carrying passengers because the former involves more actors, more sophisticated processes combining volume, varying priority services, integration strategies and consolidation and various itineraries in a transmission system (see also Huang and Lu, 2015).

Alamo and Brinati (2006), Lin et al. (2003), Tan and Kara (2007), Sholz and Cossel (2011), Fraga (2011), Costa et al. (2011), Oliveira and Correia (2011), Grosso and Sheperd (2011), Alumur et al. (2012), Torquato and Junior (2012, 2014), Oktal and Özger (2013), Lakew and Tok (2015), Feng et al. (2015), Huang and Lu (2015) used one or all of the three decisive points that influenced the method of this research:

- ✓ The use of different statistical models in the analysis of the data;
- ✓ The use of response variables and covariates;
- ✓ The use of secondary data obtained from official offices in the airline industry and economy variables to their studies.

The paper is organized as follows: in section 2, we present the goals of the research and the dataset; in section 3, we present the proposed methodology; in section 4, we present the results considering the different models; in section 5, we present a detailed discussion of the obtained results for each airport considered in the study; in section 6, we present the fit of the proposed stochastic volatility (SV) model; finally, in section 7, we present some concluding remarks.

II. Goals of the study and the times series data

A study to determine strategic hubs associated with the flow of passengers and the volume of air cargo transported in a region is a starting point for research, which is still scarce in many regions of the world, and this approach could generate important strategic contributions to air transport networks and impacts on the local and regional economy of these airports.

The statistical method used in this work seeks to analyze the temporal time series data obtained from statistical reports of DAESP for the period ranging from 2008 to 2014 extracted from a set of 10,572 observations formed by 32 airports, two response variables (volume of passengers and cargo) associated with two economic covariates, the dollar exchange rate and the unemployment rate obtained from the economic Brazilian office of the Commercial Association of São Paulo (ACSP-http://www.acsp.com.br/).

This way, we consider two statistical approaches for the data analysis: the use of multiple linear regression with normal errors for the logarithm transformed dataset and new stochastic volatility (SV) models considering the four main regional hubs, within the large group of 32 airports reported in the dataset. The four airports belong to the following locations: Ribeirão Preto, São José do Rio Preto, Bauru/Arealva and Presidente Prudente.

III. Methodological framework

In this section, we introduce the different modeling approaches to be used in the data analysis: multiple linear regression model with normal errors and stochastic volatility (SV) models.

3.1 Use of a multiple linear regression models

The use of regression models is important to report a response variable in conjunction with several factors that may be related to this response. The construction of general regression models is made empirically, and each model must be checked for fit from the residual analysis of the model. The use of regression models allows to statistically identify which of these factors (use of hypothesis testing) significantly affect the response. In addition, a regression model is also used to forecast future values of the response dataset of covariates.

For the setting of the models, let $N \geq 1$ be a fixed integer number that records the amount of observed data (in our case, it will represent the counting of passengers or cargo for the airports of the State of São Paulo). Also, let $Y_j(t)$, $t = 1, 2, \ldots, N$, $j = 1, 2, \ldots, K$, indicating the times series in the logarithm scale recording for the counting of passengers or cargo in the t^{th} month, $t = 1, 2, \ldots, N$ and j^{th} airport, $j = 1,2,3,4$. Here $N = 84$ months and $K=4$ ($j = 1$ for Ribeirão Preto; $j = 2$ for São José do Rio Preto; $j = 3$ for Bauru/Arealva; $j = 4$ for Presidente Prudente).

We consider the multiple linear regression model in each of the four main airports $j = 1,2,3,4$:

$$Y_j(t) = \beta_{jo} + \beta_{j1}\text{dollar exchange.rate} + \beta_{j2}\text{unemployment.rate} + \beta_{j3}\text{months} + \beta_{j4}\text{years} + \epsilon_j(t), \quad (1)$$

where the error term ϵ_j is assumed as a random variable with a normal distribution with mean equal to zero and constant variance σ^2.

Linear regression models are often fitted using the least squares approach. When using more than one explanatory variable to predict the behavior of a variable response, the model is denoted in the literature as a multiple regression model (Draper and Smith, 1981). In multiple linear regression analysis, the overall effect of covariates on the Y response is verified (see, for example, Draper and Smith, 1981; Seber and Lee, 2003; Montgomery and Runger, 2011).

3.2 Use of a Stochastic volatility (SV) models under a Bayesian approach

Stochastic volatility (SV) models have been extensively used to analyze financial time series (see Danielsson,1994; Yu, 2002) as a powerful alternative for existing auto-regressive models such as ARCH (autoregressive conditional heteroscedastic) introduced by Engle (1982) and the generalized autoregressive conditional heteroscedastic (GARCH) models introduced by Bollerslev (1986) but rarely used in transport or other engineering applications (see also Ghysels, 1996; Kim & Shephard, 1998; or Meyer & Yu, 2000).

In the presence of heteroscedasticity, that is, variances depending on the t time, assume that the time series $Y_j(t)$, $t = 1, 2, \ldots, N$; $j = 1,2,3,4$ can be written as

$$Y_j(t) = \beta_{jo} + \beta_{j1}\text{dollar exchange.rate} + \beta_{j2}\text{unemployment.rate} + \beta_{j3}\text{months} + \beta_{j4}\text{years} + \sigma_j(t)\epsilon_j(t) \quad (2)$$

where $\epsilon_j(t)$ is a noise considered to be independent and identically distributed with a normal distribution $N(0, \sigma_\epsilon^2)$ and $\sigma_j(t)$ is the square root of the variance of (2) (for simplicity, it is assumed that $\sigma_\epsilon^2 = 1$, since in our case the obtained inference results do not have significative changes). The variance of $Y_j(t)$ is assumed to be given by the model $\sigma_\epsilon^2 e^{hj(t)}$ where $h_j(t)$ depends on a latent variable or unobserved variable.

It is interesting to observe that usually a stochastic volatility process $Y_j(t)$ in finance applications is given by a special case of the equation (2), that is, given by the model $Y_j(t) = \sigma_j(t)\epsilon_j(t)$ where $Y_j(t)$ is the logarithm of returns and

$\sigma_j(t)$ is a strictly stationary sequence of positive random variables which is independent of the independent identically distributed noise sequence $\epsilon_j(t)$.

The independence of the noise $\epsilon_j(t)$ and the volatility $\sigma_j(t)$ allows for a much simpler probabilistic structure than that of a GARCH (Generalized Autoregressive Conditional Heteroscedasticity) process, which includes explicit feedback of the current volatility with previous volatilities and observations.

This is one of the advantages of stochastic volatility (SV) models. In this case, the mutual independence of the sequences $Y_j(t)$ and $\sigma_j(t)$ and their strict stationarity immediately imply that $Y_j(t)$ is strictly stationary. Conditions for the existence of a stationary GARCH process are much more difficult to establish (see, for example, Nelson, 1990 and Bougerol & Picard, 1992).

To analyze the dataset, we assumed a latent variable (non-observed variable) defined by an autoregressive model AR(2) given, for t =1,2,3,…,N; j=1,2,…,K, assuming the following SV model:

$$h_j(1) = \mu_j + \zeta_j(1),\ t = 1,$$
$$h_j(2) = \mu_j + \phi_{1j}[h_j(1) - \mu_j] + \zeta_j(2) \quad (3)$$
$$h_j(t) = \mu_j + \phi_{1j}[h_j(t-1) - \mu_j] + \phi_{2j}[h_j(t-1) - \mu_j] + \zeta_j(t),\quad t = 3,4,\ldots,N,$$

where $\zeta_j(t)$ is a noise with a Normal distribution N(0, $\sigma^2_{j\zeta}$), which is associated to the latent variable $h_j(t)$. The quantities $\sigma^2_{j\zeta}$, μ_j, ϕ_{1j} and ϕ_{2j}, j =1,2,3,4 are unknown parameters that should be estimated ; also $|\phi_{1j}| < 1$ and $|\phi_{2j}| < 1$.

This way, $Y_j(t)$ has a normal distribution given by:

$$Y_j(t) \sim N(g_j, \sigma^2_\epsilon e^{hj(t)}), \quad (4)$$

where $g_j = \beta_{jo} + \beta_{j1}$dollar exchange.rate+ β_{j2}unemployment.rate+ β_{j3}months + β_{j4}years;

$$h_j(1) \sim N(\mu_j, \sigma^2_{j\zeta});$$
$$h_j(2)|\ h_j(1) \sim N(\mu_j + \phi_{1j}[h_j(1) - \mu_j],\ \sigma^2_{j\zeta}); \quad (5)$$

$$h_j(t)|\ h_j(t-1) \sim N(\mu_j + \phi_{1j}[h_j(t-1) - \mu_j] + \phi_{2j}[h_j(t-1) - \mu_j],\ \sigma^2_{j\zeta}),\ t = 3,4,\ldots, N$$

The likelihood function of the SV defined by (2) given $h_j(t)$ which depends on a latent variable or unobserved variable is given for each j = 1,2,3,4, by,

$$L = \prod_{t=1}^{N} f[y_j(t) / h_j(t)] \quad (6)$$

where, f $[y_j(t) / h_j(t)]$ is the density function of a normal distribution N(g_j, $\sigma^2_\epsilon e^{hj(t)}$).

Bayesian inference based on Markov Chain Monte Carlo (MCMC) methods (see, for example, Gelfand & Smith, 1990, or Smith & Roberts, 1993) has been considered to analyze stochastic volatility (SV) models. The main reason for the use of Bayesian methods is that we may have great difficulties when using a standard classical inference approach.

Those difficulties may appear in the form of high dimensionality and likelihood function (note that, under a classical approach, we should eliminate the latent variables in $h_j(t)$ by integration) with no closed form, among other factors.

For a Bayesian analysis of the models, it is assumed that the prior distributions for the parameters μ_j, ϕ_{vj} and $\sigma^2_{j\zeta}$, v =1,2; j =1,2,3,4 are, respectively, a Normal N(0,a_j^2) distribution, a Beta(b_j,c_j) distribution and a Gamma(d_j,e_j) distribution, where Beta(b,c) denotes a Beta distribution with mean b/(b+c) and variance bc/[$(b+c)^2$(b+c+1)] and Gamma(d,e) denotes a Gamma distribution with mean d/e and variance d/e^2. The hyperparameters a_j, b_j, c_j, d_j and e_j are considered to be known and are specified latter. These prior distributions were chosen by observing the ranging of values in each parameter.

IV. Results assuming the different models

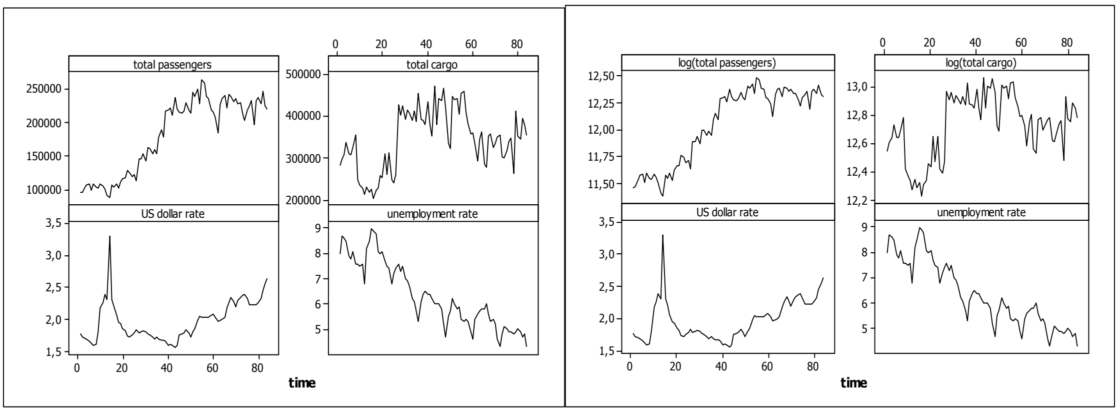
4.1 Use of multiple linear regression models

The study data refers to the monthly movement of passengers and cargo in the period from January 01, 2008 to December 31, 2014 at 32 airports in the State of São Paulo (Andradina, Araçatuba, Araraquara, Assis, Avaré/Arandu, Barretos, Bauru, Bauru/Arealva, Botucatu Bragança Paulista, Campinas (Amaral), Dracena, Franca, Itanhaém, Jundiaí, Lins, Marília, Ourinhos, Penápolis, Piracicaba, Presidente Epitácio, Presidente Prudente, Registro, Ribeirão Preto, São Carlos, São José do Rio Preto, São Manuel, Sorocaba, Tupã, Ubatuba, Urubupungá, Votuporanga) based on statistical year reports reported by DAESP (2015).

In addition to the counts (passenger/cargo), we also consider some covariates that may be correlated with the flow (passenger/cargo), such as the dollar exchange rate and the unemployment rate obtained from the economic Brazilian office of ACSP. For two airports (Urubupungá and Registro), we have a lot of missing data, therefore the data from these two airports aren't considered in the statistical analysis.

In Figure 1, we present the graphs of the time series for the four variables (passenger/cargo counts, dollar exchange rate and unemployment rate reported in the period ranging from January 01, 2008 to December 31, 2014 considering the total flow for all airports.

Figure 1: Time series passenger/cargo counts, dollar exchange rate and unemployment rate on original and logarithmic scale.



From the graphs of Figure 1, we observe that:

- ✓ Apparently, the number of passengers increases over time; then decreases a little;
- ✓ Apparently, the amount of cargo increases over time; then decreases a little;
- ✓ Apparently, the monthly number of passengers decreases with the increased dollar exchange rate.

To study the relationship between the variables and to find the most important factors affecting the variability of passenger/cargo count in the period from January 1, 2008 to December 31, 2014, we considered a first statistical analysis assuming multiple regression models relating the covariates dollar exchange rate, unemployment rate, years and months with the responses given by passengers and cargo counts.

4.1.1 Use of a multiple regression model for analysis of data considering the total of all airports

To satisfy some necessary assumptions of the regression model (normality of errors and constant variance) a multiple linear regression model with the response given in the logarithmic scale (1) was considered in this study.

Assuming the responses (passengers/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

$$\log(\text{passengers}) = 12.5 - 0.341\ \text{dollarexchangerate} - 0.0659\ \text{unemploymentrate} + 0.0084\ \text{month} + 0.138\ \text{year} \quad (7)$$

In Table 1, we present the least squares estimates (LSE) for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero.

Table I: LSE estimators for the regression parameters (all airports)

Passengers (s=0.11549; R^2=89.5%)

predictor	LSE	SE of coef	T	p
constant	12.5313	0.2646	47.36	< 0.001
dollar exchange rate	-0.34118	0.05114	-6.67	< 0.001
unemployment rate	-0.06589	0.03064	-2.15	0.035
months	0.008469	0.005254	1.61	0.111
years	0.13831	0.02003	6.90	< 0.001

Cargo (s=0.15216; R^2=54.8%)

predictor	LSE	SE of coef	T	p
constant	13.9794	0.3486	40.10	< 0.001
dollar exchange rate	-0.40585	0.06738	-6.02	< 0.001
unemployment rate	-0.08800	0.04037	-2.18	0.032

months	0.003634	0.006922	0.52	0.601
years	0.02048	0.02639	0.78	0.440

The necessary assumptions for the model (normal residuals and constant variance) were verified from residual plots. This was also done for all multiple linear regression models considered in this study. From the results in Table 1, we have:

- ✓ Years, dollar exchange rate and unemployment rate affect the number of passengers (p-value < 0.05), that is, regression coefficients are statistically different from zero at a significance level of 5%.
- ✓ We observe a positive value for the regression parameter related to years (0.13831), which implies that there is a significant increase in the number of passengers over the years (2008-2014).
- ✓ We observe a negative value for the regression parameter related to dollar exchange rate (-0.34118), which implies that there is a significant decrease in the number of passengers with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ We observe a negative value for the regression parameter related to unemployment rate (-0.06589), which implies that there is a significant decrease in the number of passengers with increased unemployment (January 01, 2008 to December 31, 2014).
- ✓ Approximately 89.5% of the data variability is explained by the model, that is, an excellent fit.
- ✓ An estimator for the standard deviation of error is given by 0.1155.

Assuming the responses (cargo/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

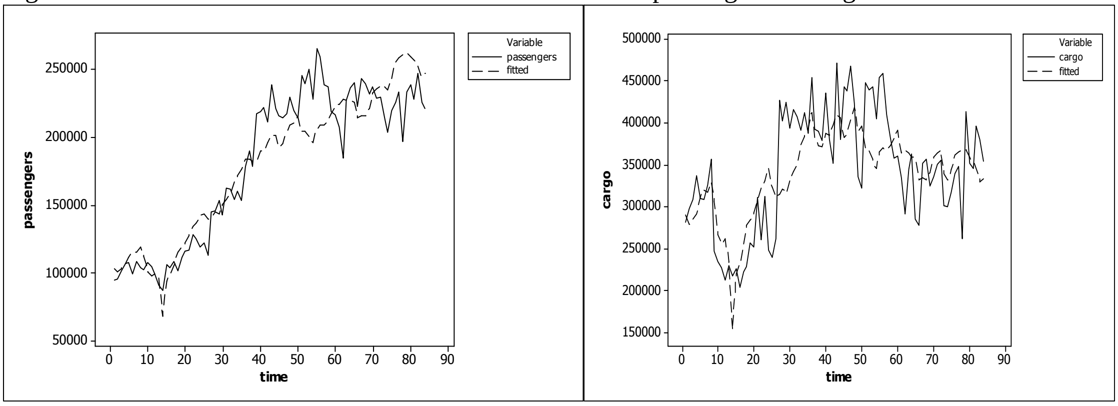
$$\begin{aligned}\log(\text{cargo}) = 14.0 &- 0.406\ \text{dollarexchangerate} \\ &- 0.0880\ \text{unemploymentrate} \\ &+ 0.00363\ \text{month} + 0.0205\text{year}\end{aligned}$$
(8)

In Table 1, we also have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators for the cargo flow, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero. From the results of Table 1, we have,

- ✓ The unemployment rate and dollar exchange rate affect the transport of cargo (p-value < 0.05), that is, the regression coefficients are statistically different from zero. The months and years do not affect the transport of cargo (p-value > 0.05).
- ✓ We observe a negative value for the regression parameter related to dollar exchange rate (-0.40585), that is, there is a significant decrease in cargo transport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ We observe a negative value for the regression parameter related to unemployment rate (-0.08800), that is, a significant decrease in cargo volume with increased unemployment (January 01, 2008 to December 31, 2014).
- ✓ Approximately 55% of the data variability is explained by the model.
- ✓ An estimate for the standard deviation of the error is given by 0.1521.

Using the fitted models (7) and (8), we can make predictions for future observations. For example, if we consider month = 1 (January), year = 8 (2015), dollar exchange rate = 3.30 and unemployment rate = 9.50, the expected value is given in the original scale by 146,444 for passengers and 158,198 for cargo. In Figure 2, we have the graphs of the time series for passenger and cargo counts reported monthly in the period (January 01, 2008 to December 31, 2014) considering the total for all airports and the fitted values obtained by the regression models (7) and (8). A good fit of the model to the data is observed.

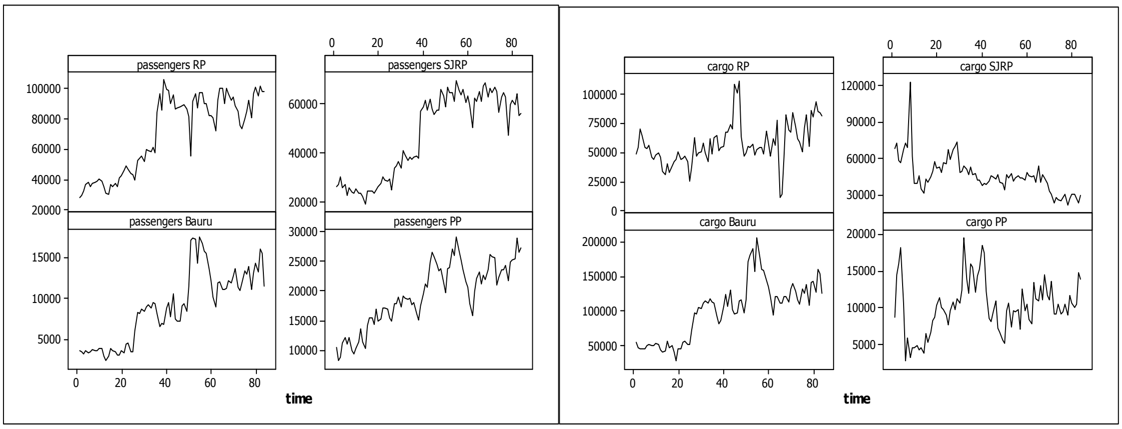
Figure 2: Time series for observed and fitted values for total passenger and cargo.



4.1.2. Use of a multiple regression model for the data analysis considering the four airports

In Figure 3, we have the graphs of the time series for passenger and cargo counts in the airports of Ribeirão Preto, São José do Rio Preto, Bauru/Arealva and Presidente Prudente, the most important airports of the region, reported monthly in the period (January 01, 2008 to December 31, 2014).

Figure 3:Time series for passenger/cargo counts for the airports of Ribeirão Preto, São José do Rio Preto, Bauru/Arealva and Presidente Prudente.



From Figure 3, we have:

- ✓ Apparently, the number of passengers increases over time for all airports; then decreases a little.
- ✓ Apparently, the amount of cargo increases over time to the airports of Ribeirão Preto, Bauru, and São José do Rio Preto, then decreases. The behavior for Presidente Prudente airport is different from the other airports.

4.1.2.1 Response log (passengers/month) and log (cargo/month) - Ribeirão Preto airport

Assuming the responses (passengers/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

$$\log(\text{passRP}) = 11.7 - 0.427 \text{ dollarexchangerate} - 0.0812 \text{ unemploymentrate} + 0.0140 \text{ month} + 0.158 \text{ year} \quad (9)$$

In Table II, we have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero.

Table II: LSE for the regression parameters (Ribeirão Preto)

Passengers (s=0.15594; R^2=86.5%)

predictor	LSE	SE of coef	T	p
constant	11.7171	0.3573	32.80	< 0.001
dollar exchange rate	-0.42681	0.06905	-6.18	< 0.001
unemployment rate	-0.08115	0.04137	-1.96	0.053
months	0.014048	0.007094	1.98	0.051
years	0.15831	0.02705	5.85	< 0.001

Cargo (s=0.32759; R^2=22.6%)

predictor	LSE	SE of coef	T	p
constant	11.8536	0.7505	15.79	< 0.001
dollar exchange rate	-0.1714	0.1451	-1.18	0.241
unemployment rate	-0.11469	0.08691	-1.32	0.191
months	0.01812	0.01490	1.22	0.228
years	0.00031	0.05682	0.01	0.996

Assuming the response (cargo/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year (obtained with the Minitab® software), we have:

$$\log(\text{cargoRP}) = 11.9 - 0.171\ \text{dollarexchangerate} - 0.115\ \text{unemploymentrate} + 0.0181\text{months} + 0.0003\ \text{years} \quad (10)$$

In Table 2, we have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter (cargo) is equal to zero.

4.1.2.2 Response log (passengers/month) and log (cargo/month) – São José do Rio Preto airport

Assuming the responses (passengers/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

$$\log(\text{passSJRP}) = 11.4 - 0.363\ \text{dollarexchangerate} - 0.100\ \text{unemploymentrate} + 0.00297\text{month} + 0.147\ \text{year} \quad (11)$$

In Table III, we have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero.

Table III: LSE estimates for the regression parameters (São José do Rio Preto)

Passengers (s=0.14981; R^2=87.4%)

predictor	LSE	SE of coef	T	p
constant	11.4273	0.3432	33.29	< 0.001
dollar exchange rate	-0.36318	0.06634	-5.47	< 0.001
unemployment rate	-0.10045	0.03975	-2.53	0.013
months	0.00297	0.00682	0.44	0.664
years	0.14711	0.02599	5.66	< 0.001

Cargo (s=0.18647; R^2=65.3%)

predictor	LSE	SE of coef	T	p
constant	11.5128	0.4272	26.95	< 0.001
dollar exchange rate	-0.43256	0.08257	-5.24	< 0.001
unemployment rate	0.04099	0.04947	0.83	0.410
months	0.000013	0.008483	0.00	0.999
years	-0.05743	0.03234	-1.78	0.080

Assuming the responses (cargo/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

$$\log(\text{cargoSJRP}) = 11.5 - 0.433\ \text{dollarexchangerate} + 0.0410\ \text{unemploymentrate} + 0.00001\text{months} - 0.574\ \text{years} \quad (12)$$

4.1.2.3 Response log (passengers/month) and log (cargo/month) – Bauru/Arealva airport

Assuming the responses (passengers/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

$$\log(\text{passBauru}) = 9.45 - 0.397\text{dollarexchangerate} - 0.105\ \text{unemploymentrate} + 0.0128\ \text{months} + 0.217\ \text{years} \quad (13)$$

In Table VI, we have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero.

Table IV: LSE estimators for the regression parameters (Bauru/Arealva)

predictor	LSE	SE of coef	T	p
Passengers ($s=0.25129$; $R^2=82.1\%$)				
constant	9.4487	0.5757	16.41	< 0.001
dollar exchange rate	-0.3965	0.1113	-3.56	0.001
unemployment rate	-0.10493	0.06667	-1.57	0.120
months	0.01276	0.01143	1.12	0.268
years	0.21712	0.04359	4.98	< 0.001
Cargo ($s=0.23620$; $R^2=76.7\%$)				
predictor	LSE	SE of coef	T	p
constant	12.6354	0.5411	23.35	< 0.001
dollar exchange rate	-0.3451	0.1046	-3.30	0.001
unemployment rate	-0.16404	0.06267	-2.62	0.011
months	0.00339	0.01075	0.32	0.753
years	0.12271	0.04097	3.00	0.004

Assuming the responses (cargo/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

$$\log(\text{cargoBauru}) = 12.6 - 0.345\ \text{dollarexhangerate} - 0.164\ \text{unemploymentrate} + 0.0034\ \text{months} + 0.123\ \text{years} \quad (14)$$

4.1.2.4 Response log(passengers) and log(cargo) – Presidente Prudente airport

Assuming the responses (passengers/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

$$\log(\text{passPP}) = 9.08 - 0.279\ \text{dollarexchangerate} + 0.056\ \text{unemploymentrate} + 0.0339\ \text{months} + 0.180\ \text{years} \quad (15)$$

In Table V, we have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero.

Table V. LSE estimates for the regression parameters (Presidente Prudente)

predictor	LSE	SE of coef	T	p
Passengers ($s=0.14254$; $R^2=80.7\%$)				
constant	9.0773	0.3266	27.80	< 0.001
dollar exchange rate	-0.27866	0.06312	-4.41	< 0.001
unemployment rate	0.05634	0.03782	1.49	0.140
months	0.033933	0.006485	5.23	< 0.001
years	0.18048	0.02472	7.30	< 0.001
Cargo ($s=0.34393$; $R^2=29.8\%$)				
predictor	LSE	SE of coef	T	p
constant	8.0213	0.7880	10.18	< 0.001
dollar exchange rate	-0.6919	0.1523	-4.54	< 0.001
unemployment rate	0.21114	0.09125	2.31	0.023
months	0.03557	0.01565	2.27	0.026

years	0.23222	0.05966	3.89	< 0.001

Assuming the response (cargo/month) transformed to a logarithmic scale and the covariates dollar exchange rate, unemployment rate, month and year in a multiple regression model fitted by the least squares method (obtained with the Minitab® software), we have:

$$\log(\text{cargoPP}) = 8.02 - 0.692\ \text{dollarrate} + 0.211\ \text{unemploymentrate} + 0.0356\ \text{months} + 0.232\ \text{years} \quad (16)$$

4.2. Statistical analysis of the data of the airports in the State of São Paulo using stochastic volatility (SV) models under a Bayesian approach

Let us assume Beta (1,1) prior distributions for ϕ_{1j}, uniform U(0,0.1) prior distributions for ϕ_{2j}, Gamma(1.1) prior distributions for $\sigma^2_{j\zeta}$, normal N(0,1) prior distributions for μ_j ,normal N(10,1) prior distributions for β_{jo} and normal N(0,1) prior distributions for β_{jl}, j =1,2,3,4; l =1,2,3,4.

A burn-in sample with 21,000 samples is considered to eliminate the effect of the initial values in the iterative method; after that, another 90,000 samples are generated taking samples from 10 to 10 totaling a final sample size of 9,000 to get the posterior summaries of interest (see Table 6).

In the simulation of samples of the joint posterior distribution of interest the OpenBugs software is used (Spiegelhalter et al, 2003). The convergence of the Gibbs sampling algorithm was verified from standard traceplots of the generated Gibbs samples.

Considering the cargo volume in the four airports, the same prior distributions assumed for the passenger case are assumed, that is, Beta (1.1) prior distributions for ϕ_{1j}, uniform U (0.1) prior distributions for ϕ_{2j}, Gamma (1.1) prior distributions for $\sigma^2_{j\zeta}$,normal N (0.1) prior distributions for μ_j , normal N (10.1) prior distributions for β_{jo} and N (0.1) prior distributions for β_{jl}, j = 1,2,3,4; l = 1,2,3,4.

Also using the OpenBugs software, a burn-in sample size of 21,000 is considered; then, we generated another 90,000 samples taking samples from 10 to 10 totaling a final sample size of 9,000 to get the posterior summaries of interest (see Table VII).

Table VI: Posterior summaries – passengers

	mean	sd	LL2.5%	UL97.5%
β_{10}	11.07	0.3404	10.46	11.81
β_{11}	-0.3385	0.1178	-0.621	-0.1739
β_{12}	-0.03404	0.03757	-0.1123	0.04636
β_{13}	0.02243	0.007268	0.007812	0.03657
β_{14}	0.1752	0.02824	0.1241	0.2374
β_{20}	11.06	0.4434	10.23	11.91
β_{21}	-0.2935	0.09021	-0.5058	-0.1609
β_{22}	-0.08124	0.03879	-0.1575	-0.007583
β_{23}	0.007795	0.007734	-0.007614	0.02227
β_{24}	0.1667	0.02578	0.1146	0.2149
β_{30}	8.675	0.3949	7.922	9.469
β_{31}	-0.3248	0.1028	-0.5348	-0.1395
β_{32}	-0.04081	0.04516	-0.1344	0.04643
β_{33}	0.02598	0.007805	0.009535	0.04048
β_{34}	0.2378	0.02762	0.1825	0.2922
β_{40}	9.234	0.2451	8.707	9.697
β_{41}	-0.22	0.04789	-0.3116	-0.124
β_{42}	0.04343	0.0281	-0.006466	0.1049
β_{43}	0.02887	0.003767	0.022	0.0369
β_{44}	0.1445	0.01909	0.1093	0.1855
μ_1	-2.629	0.9792	-4.1	-0.4999
μ_2	-2.749	1.018	-4.044	-0.5457
μ_3	-2.416	0.7627	-3.637	-0.6911
μ_4	-0.9206	0.8072	-2.464	0.6866
ϕ_{11}	0.8498	0.1268	0.5029	0.9881
ϕ_{12}	0.7309	0.2486	0.1116	0.979
ϕ_{13}	0.8444	0.08187	0.66	0.9713
ϕ_{14}	0.9536	0.02412	0.9035	0.9945
ϕ_{21}	0.05135	0.02852	0.002768	0.0977
ϕ_{22}	0.0503	0.02851	0.002886	0.09743

ϕ_{23}	0.04909	0.02867	0.002549	0.097
ϕ_{24}	0.0658	0.0244	0.01118	0.09863
ζ_1	2.268	1.005	0.8512	4.77
ζ_2	2.924	1.331	1.094	6.248
ζ_3	1.909	0.843	0.7078	3.997
ζ_4	2.78	1.148	1.111	5.537

(sd: standard deviation; LL2.5%: lower limit; UL97.5%: upper limit)

Table VII: Posterior summaries – cargo

	mean	sd	LL2.5%	UL97.5%
β_{10}	10.76	0.5436	9.8	11.81
β_{11}	-0.1753	0.06556	-0.3101	-0.04929
β_{12}	-0.01155	0.05656	-0.1158	0.09262
β_{13}	0.03134	0.01098	0.008729	0.05112
β_{14}	0.09188	0.03801	0.01714	0.1584
β_{20}	11.54	0.3735	10.83	12.33
β_{21}	-0.378	0.1028	-0.5565	-0.1454
β_{22}	0.02489	0.04428	-0.0643	0.1138
β_{23}	-0.001191	0.008008	-0.01742	0.01418
β_{24}	-0.06743	0.03149	-0.1328	-0.005542
β_{30}	11.86	0.3814	11.06	12.58
β_{31}	-0.3563	0.09295	-0.5425	-0.1762
β_{32}	-0.08597	0.04153	-0.1671	9,80E-01
β_{33}	0.02059	0.006896	0.0064	0.03353
β_{34}	0.155	0.02563	0.106	0.2081
β_{40}	8.312	0.6196	7.038	9.519
β_{41}	-0.5819	0.1427	-0.8501	-0.2911
β_{42}	0.1567	0.06345	0.03276	0.2824
β_{43}	0.0363	0.0102	0.01685	0.05694
β_{44}	0.1883	0.04512	0.1027	0.2724
μ_1	-1.871	1.084	-3.403	0.4202
μ_2	-3.147	0.6257	-3.828	-1.247
μ_3	-2.339	0.8223	-3.67	-0.4718
μ_4	-0.6796	0.8618	-2.327	1.011
ϕ_{11}	0.7816	0.1544	0.4134	0.9768
ϕ_{12}	0.5057	0.2594	0.0409	0.9401
ϕ_{13}	0.847	0.08634	0.6412	0.9786
ϕ_{14}	0.8933	0.05804	0.7487	0.9788
ϕ_{21}	0.04734	0.0285	0.002384	0.09696
ϕ_{22}	0.05025	0.02852	0.002646	0.09717
ϕ_{23}	0.05043	0.02857	0.002579	0.09753
ϕ_{24}	0.06185	0.02479	0.009566	0.0983
ζ_1	1.141	0.5293	0.4524	2.451
ζ_2	2.306	1.117	0.8192	5.143
ζ_3	1.801	0.7774	0.7008	3.687
ζ_4	2.811	1.261	1.054	5.9

(sd: standard deviation; LL2.5%: lower limit; UL97.5%: upper limit)

5. Discussion of the obtained results for each airport using the two proposed models

5.1 Ribeirão Preto airport

5.1.1 Use of a multiple linear regression model

From the results of Table 2, we observe that:

- ✓ Years, months, dollar exchange rate and unemployment rate affect the number of passengers in Ribeirão Preto airport (p-value < 0.05 or very close to 0.05), that is, the regression coefficients are statistically different from zero at a significance level of 5%.
- ✓ There is a positive value for the regression parameter related to years (0.15831), which implies that there is

a significant increase in the number of passengers in Ribeirão Preto airport over the years (2008-2014).

- ✓ There is a negative value for the regression parameter related to dollar exchange rate (-0.42691), which implies that there is a significant decrease in the number of passengers in Ribeirão Preto airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ There is a negative value for the regression parameter related to unemployment rate (-0.08115), which implies that there is a significant decrease in the number of passengers in Ribeirão Preto airport with increased unemployment (January 01, 2008 to December 31, 2014).
- ✓ There is a positive value for the regression parameter related to months (0.014048), which implies that there is a significant increase in the number of passengers over the months. The end of each annual period leads to a significant increase in the number of passengers.
- ✓ Approximately 86.5% of the variability of the data (number of passengers) is explained by the model. That is, we have an excellent fit.
- ✓ An estimator for the standard deviation of error is given by 0.1559.

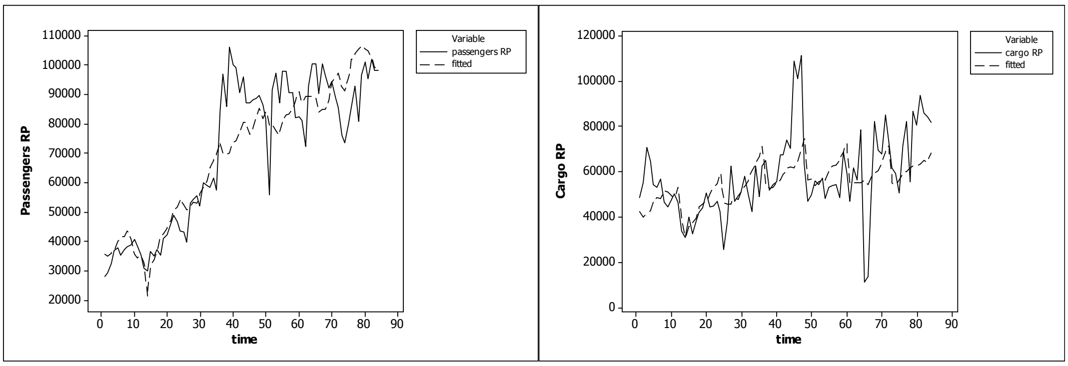
From the results of Table 2, we have:

- ✓ No covariate (month, year, unemployment rate and dollar exchange rate) affects cargo transport in Ribeirão Preto airport (p-value > 0.05), that is, the regression coefficients are not statistically different from zero. There is stability in cargo transport in Ribeirão Preto airport regardless of the variability of the covariates (month, year, unemployment rate and dollar exchange rate). This fact can also be seen in Figure 3.
- ✓ Approximately 22.6% of the data variability is explained by the model.
- ✓ An estimator for the standard deviation of the error is given by 0.3276.

Using the fitted models (9) and (10) and considering month = 1 (January), year = 8 (2015), dollar exchange rate = 3.30 and unemployment rate = 9.50, the forecast value is given in the original scale by 49,901 for passengers and 27,424 for cargo volume in Ribeirão Preto airport.

In Figure 4, we present the graphs of the time series for passenger and cargo counts reported in the period (January 01, 2008 to December 31, 2014) for Ribeirão Preto airport and the fitted values given by the regression models (9) and (10). A good fit of the models to the data is observed.

Figure 4: Time series for observed values and fitted values for passengers and cargo-Ribeirão Preto.



5.1.2 Use of a stochastic volatility (SV) model

From the results in Table 6, similar results as obtained using a multiple linear regression model are obtained under a Bayesian approach for the passengers case, assuming a stochastic volatility (SV) model:

- ✓ Years, months and dollar exchange rate affect the number of passengers in Ribeirão Preto airport (the 95% credible interval does not contain zero), that is, the regression coefficients are statistically different from zero. Observe that a 95% credible interval corresponds to a 95% confidence interval under a classical inference approach, that is, a 5% significance level to test if each regression parameter is equal to zero against an alternative to be different from zero. Under a Bayesian approach, we get the inferences using the credible intervals.
- ✓ A positive value is observed for the estimator of the regression parameter related to years (0.1752), which implies that there is a significant increase in the logarithm of the number of passengers in Ribeirão Preto airport over the years (2008/2014).
- ✓ A negative value is observed for the estimator of the regression parameter related to the dollar exchange rate (-0.3385), which implies that there is a significant decrease in the

logarithm of the number of passengers in Ribeirão Preto airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).

- ✓ A positive value is observed for the estimator of the regression parameter related to months (0.02243), which implies that there is a significant increase in the logarithm of the number of passengers over the months. End of year leads to a significant increase in the number of passengers.

From the results in Table 7, some important interpretations are obtained for the cargo case:

- ✓ Years, months and dollar exchange rate affect the cargo transport in Ribeirão Preto airport (the 95% credible interval does not contain zero), that is, the regression coefficients are statistically different from zero.
- ✓ A negative value is observed for the estimator of the regression parameter related to the dollar exchange rate (-0.1753), which implies that there is a significant decrease in the logarithm of cargo transport in Ribeirão Preto airport with increased dollar exchange rate.
- ✓ A positive value is observed for the estimator of the regression parameter related to months (0.03134), which implies that there is a significant increase in the logarithm of cargo transport in Ribeirão Preto airport (end of year leads to increased cargo volume).
- ✓ A positive value is observed for the estimator of the regression parameter related to years (0.09188), which implies that there is a significant increase in the logarithm of cargo transport in Ribeirão Preto airport over the years.

5.2 São José do Rio Preto airport

5.2.1 Use of a multiple linear regression model

From the results of Table 3, we have:

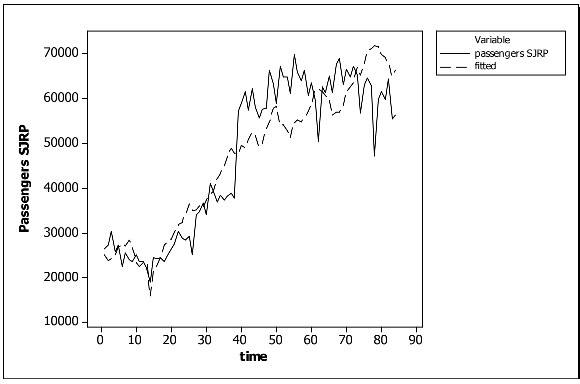
- ✓ Years, dollar exchange rate and unemployment rate affect the number of passengers in São José do Rio Preto airport (p-value < 0.05), that is, the regression coefficients are statistically different from zero at a significance level of 5%.
- ✓ We observe a positive estimate for the regression parameter related to years (0.14711), which implies that there is a significant increase in the number of passengers in São José do Rio Preto over the years (2008/2014).
- ✓ We observe a negative value for the regression parameter related to dollar exchange rate (-0.36318), which implies that there is a significant decrease in the number of passengers in São José do Rio Preto airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ We observe a negative value for the regression parameter related to unemployment rate (-0.10045), which implies that there is a significant decrease in the number of passengers in São José do Rio Preto airport with increased unemployment (January 01, 2008 to December 31, 2014).
- ✓ Approximately 87.4% of the data variability is explained by the model, which is an excellent fit.
- ✓ An estimator for the standard deviation of the error is given by 0.1498.

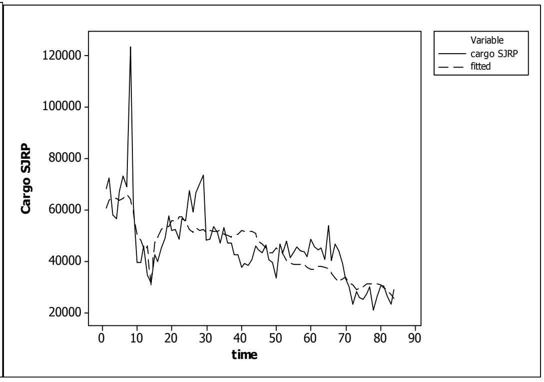
In Table 3, we have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero. From the results of Table 3, we have:

- ✓ The dollar exchange rate affects the transport of cargo (p-value < 0.05), that is, the regression coefficients are statistically different from zero at a significance level of 5%. For years we have a significance effect considering a significance level of 10% (p-value <0.10).
- ✓ We observe a negative estimate for the regression parameter related to years (-0.05743), which implies that there is a significant decrease in cargo transport in São José do Rio Preto airport over the years (2008-2014).
- ✓ We observe a negative value for the regression parameter related to dollar exchange rate (-0.43256), which implies that there is a significant decrease in cargo transport in São José do Rio Preto airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ Approximately 65.3% of the data variability is explained by the model.
- ✓ An estimator for the standard deviation of the error is given by 0.1865.

Using the fitted models (11) and (12) and considering month = 1 (January), year = 8 (2015), dollar exchange rate = 3.30 and unemployment rate = 9.50, the forecast value is given in the original scale by 34.697 for passengers and 22.366 for the cargo volume for São José do Rio Preto airport. In Figure 5, we have the graphs of the time series for monthly passenger and cargo counts reported in the period (January 01, 2008 to December 31, 2014) at São Jose do Rio Preto airport and the fitted values obtained by the regression models (11) and (12). We observe a reasonable fit of the models for the data.

Figure 5:Time series for the observed and fitted values - passenger and cargo, São José do Rio Preto.





5.2.2 Use of a stochastic volatility (SV) model

From the results in Table 6, some interpretations under a Bayesian approach for the passengers case, assuming a stochastic volatility model are given as follows:

- ✓ Years, dollar exchange rate and unemployment rate affect the number of passengers in São José do Rio Preto airport (the 95% credible interval does not contain zero), that is, the regression coefficients are statistically different from zero.
- ✓ A positive value is observed for the estimator of the regression parameter related to years (0.1667), which implies that there is a significant increase in the logarithm of the number of passengers in São José do Rio Preto airport over the years (2008-2014).
- ✓ A negative value is observed for the estimator of the regression parameter related to the dollar exchange rate (-0.2935), which implies that there is a significant decrease in the logarithm of the number of passengers in São José do Rio Preto airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ A negative value is observed for the estimator of the regression parameter related to the unemployment rate (-0.08124), which implies that there is a significant decrease in the logarithm of the number of passengers in São José do Rio Preto airport with increased unemployment (January 01, 2008 to December 31, 2014).

From the results in Table 7, some important interpretations are obtained for the cargo case:

- ✓ Years and dollar exchange rate affect cargo transport in São José do Rio Preto airport (the 95% credible interval does not contain zero), that is, the regression coefficients are statistically different from zero. The other factors are not significant, possibly due to the presence of some outlier that could affect the obtained inference.
- ✓ A negative value is observed for the estimator of the regression parameter related to the dollar exchange rate (-0.3780), which implies that there is a significant decrease in the logarithm of cargo transport in São José do Rio Preto airport with increased dollar exchange rate.
- ✓ A negative value is observed for the estimator of the regression parameter related to years (-0.06743), which implies that there is a significant decrease in the logarithm of cargo transport in São José do Rio Preto airport over the years.

5.3 Bauru/Arealva airport

5.3.1 Use of a multiple linear regression model

From the results of Table 4, we have:

- ✓ Years and dollar exchange rate affect the number of passengers (p-value < 0.05), that is, the regression coefficients are statistically different from zero at a significance level of 5%.
- ✓ We observe a positive estimate for the regression parameter related to years (0.21712), which implies that there is a significant increase in the number of passengers in Bauru airport over the years (2008-2014).
- ✓ We observe a negative estimate for the regression parameter related to dollar exchange rate (-0.3965), which implies that there is a significant decrease in the number of passengers in Bauru airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ Approximately 82.1% of the data variability is explained by the model. That is, we have an excellent fit.
- ✓ An estimator for the standard deviation of error is given by 0.2513.

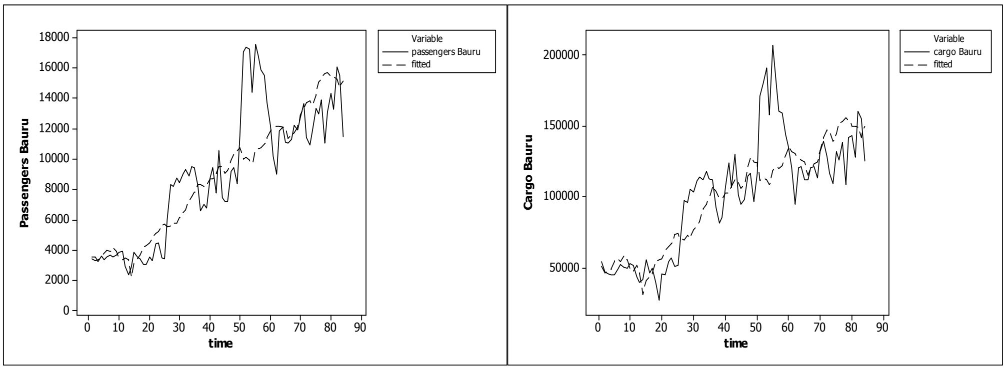
In Table 4, we have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero. From the results of Table 4, we have:

- ✓ Dollar exchange rate and unemployment rate affect cargo transport (p-value < 0.05), that is, the regression coefficients are statistically different from zero at a significance level of 5%.
- ✓ We have a positive estimate for the regression parameter relatad to years (0.12271), which implies that there is a significant increase in cargo transport in Bauru airport over the years (2008/2014).
- ✓ We have a negative estimate for the regression parameter related to dollar exchange rate (-0.3451), which implies that there is a significant decrease in cargo transport in Bauru airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ We have a negative value for the regression parameter related to unemployment rate (-0.16404), which implies that there is a significant decrease in cargo transport in Bauru airport with increased unemployment (January 01, 2008 to December 31, 2014).
- ✓ Approximately 76.7% of the data variability is explained by the model.
- ✓ An estimator for the standard deviation of the error is given by 0.2362.

Using the fitted models (13) and (14) and considering month = 1 (January), year = 8 (2015), dollar exchange rate = 3.30 and unemployment rate = 9.50, the forecast value is given in the original scale by 7.282 for passengers and 55.453 for cargo volume in the Bauru/Arealva airport.

In Figure 6, we have the graphs of the time series for passenger and cargo counts reported in the period (January 01, 2008 to December 31, 2014) for the Bauru/Arealva airport and the fitted values obtained from the regression models (13) and (14). A good fit of the models for the data is observed.

Figure 6: Time series for observed values and fitted values-passengers and cargo, Bauru/Arealva.



5.3.2 Use of a stochastic volatility (SV) model

From the results in Table 6, some interpretations under a Bayesian approach for the passengers case, assuming a stochastic volatility model are given as follows:

- ✓ Years, months and dollar exchange rate affect the number of passengers in Bauru airport (the 95% credible interval does not contain zero), that is, the regression coefficients are statistically different from zero.
- ✓ A positive value is observed for the estimator of the regression parameter related to years (0.2378), which implies that there is a significant increase in the logarithm of the number of passengers in Bauru airport over the years (2008-2014).
- ✓ A negative value is observed for the estimator of the regression parameter related to the dollar exchange rate (-0.3248), which implies that there is a significant decrease in the logarithm of the number of passengers in Bauru airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ A positive value is observed for the estimator of the regression parameter related to months (0.02598), which implies that there is a significant increase in the logarithm of the number of passengers over the months. End of year leads to a significant increase in the number of passengers.

From the results in Table 7, it is observed for the cargo case:

- ✓ Years, months and dollar exchange rate affect cargo transport in Bauru airport (the 95% credible interval does not contain zero), that is, the regression coefficients are statistically different from zero.

- ✓ A negative value is observed for the estimator of the regression parameter related to the dollar exchange rate (-0.3563), which implies that there is a significant decrease in the logarithm of cargo transport in Bauru airport with increased dollar exchange rate.
- ✓ A positive value is observed for the estimator of the regression parameter related to months (0.02059), which implies that there is a significant increase in the logarithm of cargo transport in Bauru airport over the months (end of year leads to increased cargo volume).
- ✓ A positive value is observed for the estimator of the regression parameter related to years (0.1550), which implies that there is a significant increase in the logarithm of cargo transport in Bauru airport over the years.

5.4 Presidente Prudente airport

5.4.1 Use of a multiple linear regression model

From the results of Table 5, we have:

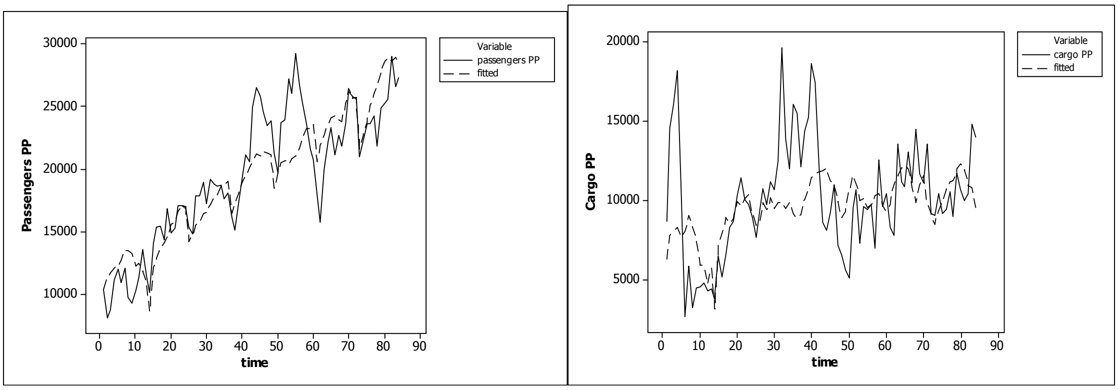
- ✓ Years, months and dollar exchange rate affect the number of passengers (p-value < 0.05), that is, the regression coefficients are statistically different from zero at a significance level of 5%. In this airport, a significant effect of unemployment rate in the number of passengers is not observed (p-value > 0.05).
- ✓ We observe a positive estimate for the regression parameter related to years (0.18048), which implies that there is a significant increase in the number of passengers in Presidente Prudente airport over the years (2008/2014).
- ✓ We observe a negative estimate for the regression parameter related to dollar exchange rate (-0.27866), which implies that there is a significant decrease in the number of passengers in Presidente Prudente airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ We observe a positive estimate for the regression parameter related to months (0.03393), which implies that there is a significant increase in the number of passengers over the months (end of year leads to increased number of passengers).
- ✓ Approximately 80.7% of the data variability is explained by the model, which is an excellent fit.
- ✓ An estimator for the standard deviation of error is given by 0.1425.

In Table 5, we have the least squares estimates for the regression coefficients, the standard errors (SE) of the estimators, the T-Student statistics values and the p-values to test if each regression parameter is equal to zero. From the results of Table 5, we have:

- ✓ Years, months, dollar exchange rate and unemployment rate affect the air transport of cargo (p-value <0.05), that is, the regression coefficients are statistically different from zero at a significance level of 5%.
- ✓ A positive estimate is observed for the regression parameter related to years (0.2322), which implies that there is a significant increase in cargo transport in Presidente Prudente airport over the years (2008 to 2014).
- ✓ There is a positive estimate for the regression parameter related to months (0.03557), which implies that there is a significant increase in cargo transport in Presidente Prudente airport over the months.
- ✓ We have a negative estimate for the regression parameter related to dollar exchange rate (-0.6919), which implies that there is a significant decrease in cargo transport in Presidente Prudente airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ We have a positive estimate for the regression parameter related to unemployment rate (0.21214), which implies that there is a significant increase in cargo transport in Presidente Prudente airport with increased unemployment (January 01, 2008 to December 31, 2014). Note that the unemployment rate is related to the entire State of São Paulo.
- ✓ Approximately 29.8% of the data variability is explained by the model. It is important to note that other factors not included in the model can also have significant effects on the variability of cargo transport in Presidente Prudente airport.
- ✓ An estimator for the standard deviation of the error is given by 0.3439. Using the fitted models (9) and (10) and considering month = 01 (January), year = 8 (2015), dollar exchange rate = 3.30 and unemployment rate = 9.50, the forecast value is given in the original scale by 26.129 passengers and 15.323 in the cargo volume in Presidente Prudente airport.

In Figure 7, we have the graphs of the time series for monthly passenger and cargo counts reported in the period (January 01, 2008 to December 31, 2014) in Presidente Prudente airport and the fitted values by the regression models (15) and (16). We observe a reasonable fit of the models to the data.

Figure 7: Time series for observed values and fitted values - passenger and cargo, Presidente Prudente.



5.4.2 Use of a stochastic volatility (SV) model

From the results in Table 6, some important observations are reported in the following subsections.

- ✓ Years, months and dollar exchange rate affect the number of passengers in Presidente Prudente airport (the 95% credible interval does not contain zero), that is, the regression coefficients are statistically different from zero.
- ✓ A positive value is observed for the estimator of the regression parameter related to years (0.1445), which implies that there is a significant increase in the logarithm of the number of passengers in Presidente Prudente airport over the years (2008-2014).
- ✓ A negative value is observed for the estimator of the regression parameter related to the dollar exchange rate (-0.2200), which implies that there is a significant decrease in the logarithm of the number of passengers in Presidente Prudente airport with increased dollar exchange rate (January 01, 2008 to December 31, 2014).
- ✓ A positive value is observed for the estimator of the regression parameter related to months (0.1445), which implies that there is a significant increase in the logarithm of the number of passengers in Presidente Prudente airport over the months (end of year leads to increased number of passengers).

From the analyses of the results in Table 7, some important observations are reported in the following subsections.

- ✓ Years, months, unemployment rate and dollar exchange rate affect cargo transport in Presidente Prudente airport (the 95% credible interval does not contain zero), that is, the regression coefficients are statistically different from zero.
- ✓ A negative value is observed for the estimator of the regression parameter related to the dollar exchange rate (-0.5819), which implies that there is a significant decrease in the logarithm of cargo transport in Presidente Prudente airport with increased dollar exchange rate.
- ✓ A positive value is observed for the estimator of the regression parameter related to months (0.0363), which implies that there is a significant increase in the logarithm of cargo transport in Presidente Prudente airport over the months (end of year leads to increased cargo volume).
- ✓ A positive value is observed for the estimator of the regression parameter related to years (0.1883), which implies that there is a significant increase in the logarithm of cargo transport in Presidente Prudente airport over the years.
- ✓ A positive value is observed for the estimator of the regression parameter related to unemployment rate (0.1567), which implies that there is a significant increase in the logarithm of cargo transport in Presidente Prudente airport with increased unemployment.

6. Stochastic volatility model fit and the estimated volatilities

From the fitted model for the dataset related to the four airports (Ribeirão Preto, São José do Rio Preto, Bauru/Arealva and Presidente Prudente), Figures 8, 9, 10 and 11 present the fitted means and the observed values for passenger and cargo transport in the period ranging from January 01, 2008 to December 31, 2014.

Figure 8: Time series for observed values and fitted values - passengers and cargo, Ribeirão Preto.

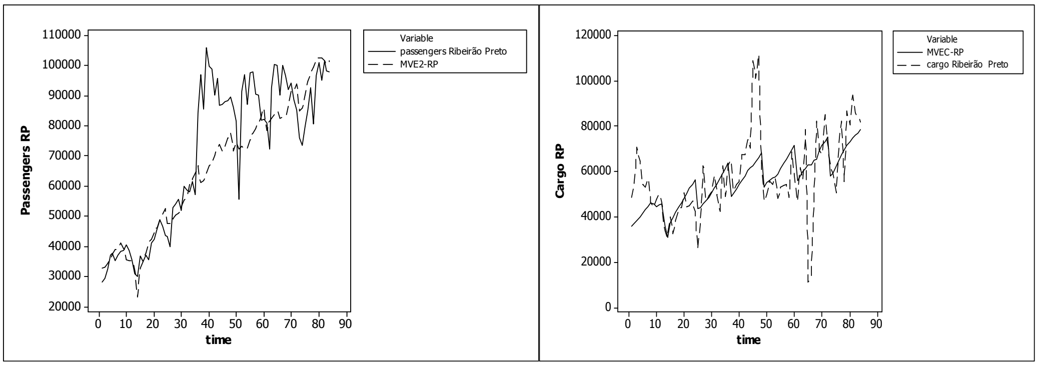


Figure 9: Time series for observed values and fitted values - passengers and cargo, São José do Rio Preto.

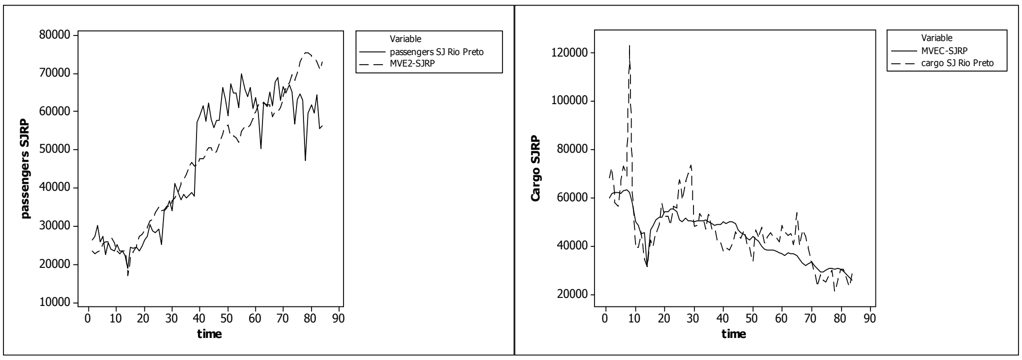


Figure 10: Time series for observed values and fitted values - passengers and cargo, Bauru/Arealva.

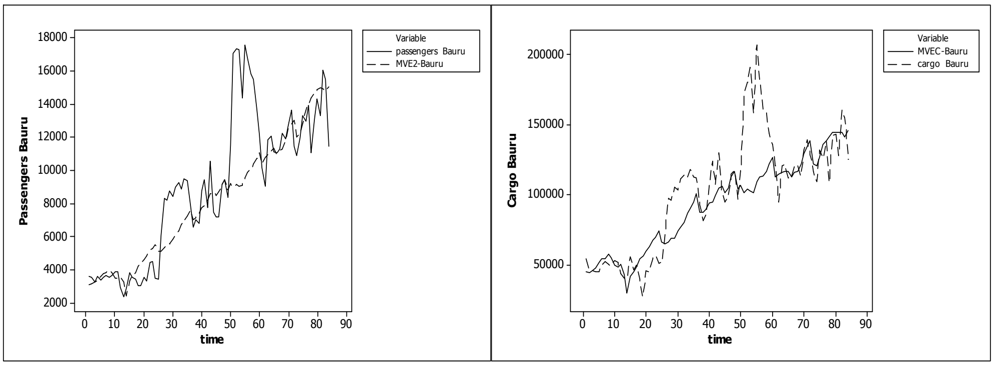
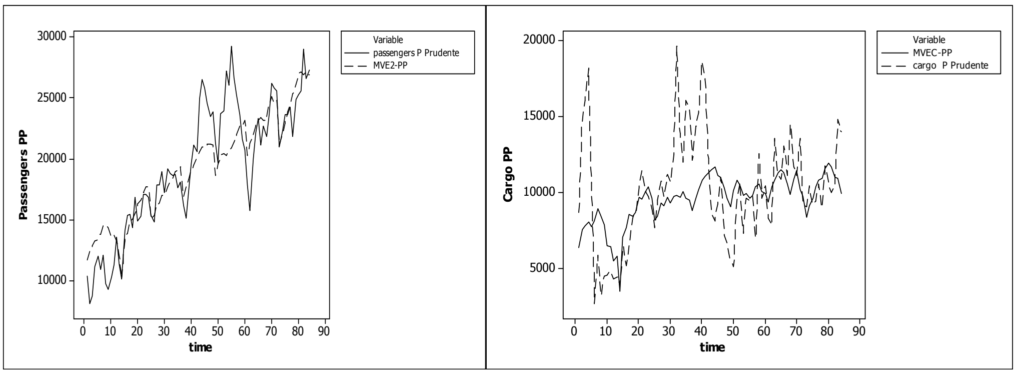
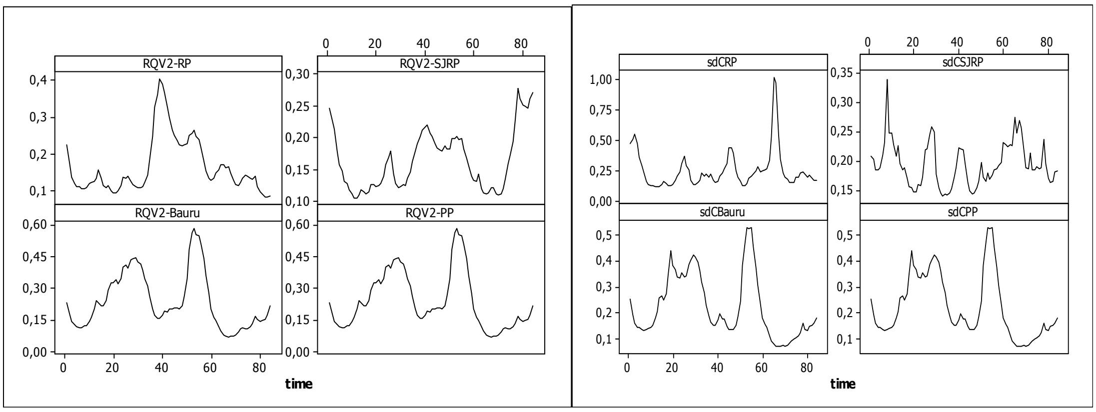


Figure 11:Time series for observed values and fitted values - passengers and cargo, Presidente Prudente.



In Figure 12, we present the graphs of the square roots for the volatilities considering the dataset of the four airports (monthly volume of passengers and cargo).

Figure 12: Square roots for volatilities (passengers/cargo) of the airports.



From the plots of Figure 12, we get some important interpretations:

Passengers:

- ✓ There is great volatility regarding the number of passengers in Ribeirão Preto airport between month 40 (April, 2011) and month 50 (February, 2012). Outside this period, there is little volatility.
- ✓ Similar behavior is observed for the number of passengers in São José do Rio Preto airport; the volatility increases at the end of the observed period, that is, close to the end of 2014 (year in which serious economic problems start in Brazil).
- ✓ Bauru airport has great volatility regarding the number of passengers close to month 25 (January, 2010) and month 55 (end of 2012).
- ✓ Presidente Prudente airport has similar behavior regarding volatility of the number of passengers, as observed for the Bauru airport.

Cargo:

- ✓ Ribeirão Preto airport has seasonal behavior regarding volatilities in cargo transport (in some periods of the year there is greater volatility); more volatility is observed close to month 65, which corresponds to 2013.
- ✓ Similar behavior is observed regarding cargo transport in São José do Rio Preto airport.
- ✓ Bauru and Presidente Prudente airports have great volatilities regarding cargo transport, especially between months 10 and 35 (end of 2008 and end of 2010) and close to month 55 (middle of 2012).

The volatility of the data shows the dependence of the effective use of air transport on the variability of passenger demand (tourism and corporate business) and cargo (potential of local production that requires the use of air transport for the flow of products), directing efforts to the economic and financial viability of the national air transport, with emphasis on the best practices inherent to the concept of a hub.

7. Concluding remarks

From the results obtained in this study, it was possible to observe,with both assumed: models multiple linear regression model with normal errors or the SV model, that some periods of the year imply great volatilities of the air transport of passengers and cargo for the four studied airports in the State of São Paulo during the study period. It was also possible to identify important factors affecting the volume of passengers and cargo for these airports using the multiple linear regression model introduced in Section 3 and the SV models introduced in Section 4.

Although the obtained interpretations in terms of inference results are similar with both proposed models, multiple linear regression model with normal errors or the SV model. The SV model has an important advantage over the standard multiple linear regression model in identification of important factors affecting the passengers and cargo flow in airports in São Paulo state, and in the modeling fit: the estimation of the volatilities of flow in each airport. This could be an important result for airport managers.

Another important result was that the fitted models used in this study can be of great importance in the prediction of the volume of passengers and cargo related to the economic factors considered. These results are of great interest to airport managers planning to build new large hubs for passengers or cargo as an alternative for the large airports located in São Paulo and Campinas. Other structures of SV models could also be considered to analyze the dataset considering autoregressive model AR(L) structures larger than 2, to get better fit of the model for the data. In addition, it is important to point out that the use of a Bayesian approach with Markov Chain Monte Carlo (MCMC) methods is facilitated using free software available, such as OpenBugs, which gives a great simplification in the computational work to get the posterior summaries of interest.

The use of the new stochastic volatility (SV) model proposed in this manuscript was very important in the discovery of the most important factors affecting the means of the time series in each one of the four airports located in the State of São Paulo and also in the modeling of the volatilities associated to the flow of passengers and the volume of cargo during the time period assumed in this study. As observed in the results obtained, some of the airports are more dependent on unemployment rates than others. It is important to point out that some of the airports are located in regions with more economic stability.For future studies, it would be possible to consider other volatility structures and to compare the results obtained with results from other statistical approaches, for example using generalized linear models to count data in the original scale and not in the logarithm scale, as considered in this study.

It is also important to point out that, in future studies, the authors could apply the proposed statistical model for new datasets from the Brazilian office of the National Civil Aviation Agency (ANAC) linked to the major passenger and cargo airports in Brazil in order to identify the dependency level of the demand of these airports related to the volatility of each airport, as a criterion to plan the logistics project of air transport in Brazil.

References

[1.] Alamo, J.A.T., Brinati, M. A., 2006. Modelagem para localização de hubs no transporte de encomendas expressas. Produção. 16(3), 470-480.

[2.] Alumur, S.A., Kara, B. Y., 2007. A new model for the hazardous waste location-routing problem. Computers & Operations Research, 34(5), 1406-1423.

[3.] Alumur, S.A., Kara, B. Y., 2008. Network hub location problems: The state of the art. Eur. J. Oper. Res. 190, 1-21.

[4.] Alumur, S.A., Kara, B.Y, Karasan, O.E., 2009. The design of incomplete single allocation hub networks. Transportation Res. Part B, 43, 936-951.

[5.] Alumur, S.A.; Yaman H.; Kara B.Y., 2012. Hierarchical multimodal hub location problemwith time-definite deliveries, Transportation Research. v. Part E 48, 1107-1120.

[6.] Amaruchkul,K.;Lorchirachoonkul,V., 2011. Air-cargocapacityallocationformultiplefreightforwarders.Transp.Res.PartE47(1), 30-40.

[7.] Assaf, A.G., Gillen, D., 2012. Measuring the joint impact of governance form and economic regulation on airport efficiency. European Journal of Operational Research 220, 187-198.

[8.] Bartodziej, P., Derigs, U., Malcherek, D., Vogel, U., 2009. Models and algorithms for solving combined vehicle and crew scheduling problems with restconstraints:anapplicationtoroadfeederserviceplanninginaircargo transport.ORSpectrum,31(2),405-429.

[9.] Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 31, 307-327.

[10.] Bougerol, P., Picard, N., 1992. Stationarity of GARCH processes and of some nonnegative time series. Journal of Econometrics, 52,115-127.

[11.] Bowen, J. T. Jr., 2012. A spatial of Fedex and UPS: hubs, spokes, and network structure. Journal of Transport Geography, 24, 419-431.

[12.] Campbell, J.F., O'Kelly, M.E., 2012. Twenty-five years of hub location research. Transportation Science, 46, 153-169.

[13.] Costa, T. F. G., Lohmann, G., Oliveira, A. V. M., 2011. Mensuração de Concentração e Identificação de Hubs no Transporte Aéreo. Journal of Transport Literature, 5(2),106-133. DAESP. (Departamento de Aviação do Estado de São Paulo).

[14.] Danielsson, J.,1994. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. Journal of Econometrics, 61, 375-400.

[15.] Draper, N.R., Smith, H., 1981. Applied regression analysis. Wiley series in probability and mathematical statistics.

[16.] Engle, R.F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987-1007.

[17.] Farahani, R.Z., Asgari, N., Davarzani, H., 2009. Supply Chain and Logistics in National, International and Governmental Environment-Concepts and Models, Springer.

[18.] Feng, B., Li, Y., Shen, Z.J, 2015. Air cargo operations: Literature review and comparison with pratices. Transportation Research Part C, 56, 263-280.

[19.] Fraga, R., 2011. Mercado Doméstico de Carga Aérea: Análise Descritiva e Econométrica do Setor. Journal of Transport Literature, 5(3), 256-280.

[20.] Gardiner, J., Humphrey, I., Ison, St., 2005. Freighter operators' choice of airport: A three-stage process. TransportReviews, 25(1), 85-102.

[21.] Gardiner, J., Ison, St., 2008. The geography of non-integrated cargo airlines: An international study. Journal of Transport Geography, 16(1), 55-62.

[22.] Gardiner,J., Ison,St, Humphreys,I., 2016. Factorsinfluencingcargoairlines'choice of airport: An international survey. Journal of Air Transport Management, 11(6), 393-399.

[23.] Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, 85(410), 398-409.

[24.] Ghysels, E., 1996. Stochastic volatility. In: Statistical methods on finance, North-Holland.

[25.] Grosso, M. G., Shepherd, B., 2011. Air cargo transport in APEC: Regulation and effects on merchandise trade. Journal of Asian Economics, 203-212.

[26.] Heinicke, K. World Air Cargo Forecast. (2006/2007) <freight.transportation.org/ doc/DCA_WACF_2006_Review.ppt>

[27.] Huang, K.L.U., 2015. A linear programming-based method for the network revenue management problem of air cargo. Transportation Research Procedia 7, 459-473.

[28.] Kara, B.Y, Tansel, B., 2001. The latest arrival hub location problem, Manage Sci 47, 1408-1420.

[29.] Kim, S., Shephard, N., 1998. Stochastic volatility: likelihood inference and comparison with arch models. Review of Economic Studies, 65, 361-393.

[30.] Lakew, P.A., Tok, Y.C.A., 2015. Determinants of air cargo traffic California. Transportation Research Part A, 80, 134-150.

[31.] Leung, L.C., Van Hui, Y., Wang, Y., Chen, G., 2009. A 0–1 LP model for the integration and consolidation of air cargo shipments. Oper. Res., 57(2), 402-412.

[32.] Li, Y., Tao, Y., Wang, F., 2009. A compromised large-scale neighborhood search heuristic for capacitated air cargo loading planning. Eur. J. Oper. Res. 199(2),553-560.

[33.] Lin, C.C., Lin, Y.J., Lin, D.Y., 2003. The economic effects of center-to-center directs on hub-and-spoke networks air express common carriers. Journal of Air Transport Management, 9, 255-265.

[34.] Meyer, R., Yu, T., 2000. Bugs for a Bayesian analysis of stochastic volatility models. Econometrics Journal, 3, 198-215.

[35.] Morrel, P. S., Pilon, R.V., 1999. KLM and Northwest: a survey of the impact of a passenger aliance on cargo servisse characteristics. Journal of Air Transport Management, 153-160.

[36.] Montgomery, D. C., Runger, G. C., 2011. Applied statistics and probability for engineers. Fifty Ed. New York: Wiley.

[37.] Nelson, D., 1990. Stationarity and persistence in the GARCH(1,1) model, Econometric Theory, 6, 318–334.

[38.] Oktal, H., Ozger, A., 2013. Hub location air cargo transportation: A case study. Journal of Air Transport Management, 27, 1-4.

[39.] Oliveira, D.S., Correia, A.R., 2011. Estudo do desempenho operacional dos aeroportos brasileiros relativo ao movimento de cargas. Journal of Transport Literature, 5(3), 141-162.

[40.] Onghena,E., 2011. Integratorsinachangingworld.Criticalis suesin air transport economics and business. London:Routledge, 112-132.

[41.] Petersen,J., 2007.Airfreightindustry–whitepaper.ResearchReport,GeorgiaIns tituteofTechnology.

[42.] Seber, G. A. F., Lee, A. J., 2003. Linear regression analysis. Second edition. Wiley series in probability and mathematical statistics.

[43.] Scholz, A.B., Cossel, J., 2011. Assessing the importance of hub airports for cargo carriers and its implications for a sustainable airportmanagement. Research in Transportation Business & Management, 1, 62-70.

[44.] Smilowitz, K. R., Daganzo, C. F., 2007. Continuum approximation techniques for the design of integrated package distribution systems. Networks, 50, 183-196.

[45.] Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. Journal of the Royal Statistical Society. Series B. Methodological, 55(1), 3–23.

[46.] Spiegelhalter, D.J., Thomas, A., Best, N.G., Lund, D., 2003. Winbugs user manual. Cambridge. United Kingdon: MRC Biostatistics Unit.

[47.] Tan,P.Z.,Kara,B.Y., 2007. Ahub-covering modelforcargodeliverysystems.Networks, 49, 28-39.

[48.] Torquato, T. L. L. e A. A. Raia Jr., 2012. Modelo de geração de viagens para condomínios residenciais horizontais, In: XXVI Congresso de Pesquisa e Ensino em Transportes, Joinville.

[49.] Torquato, T.L.L., Junior, A.A.R., 2014. Modelos de geração de viagens para condomínios residenciais horizontais, TRANSPORTES, v. 22, n. 1, p. 56-64.

[50.] Yu, J., 2002. Forecasting volatility in the New Zeland stock market. Applied Financial Economics, 12,193-202.

[51.] Wu, C.J.H., Hayashi, Y., 2011. Airport Attractiveness Analysis through a Gravity Model: A Case Study of ChubuInternational AirportinJapan, Proceedings of the Eastern Asia Society for Transportation Studies, Vol.8.

[52.] Wang, Y.J., Kao, C.S., 2008. An application of a fuzzy knowledge system for air cargo overbooking under uncertain capacity. Comput. Math. Appl., 56(10),2666-2675