Economic and climatic models for estimating coffee supply

Adriana Ferreira de Moraes-Oliveira(1), Lucas Eduardo de Oliveira Aparecido(2) and Sérgio Rangel Fernandes Figueira(1)

(1)Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Departamento de Economia, Administração e Educação, Via de Acesso Prof. Paulo Donato Castellane, s/n, CEP 14884-900 Jaboticabal, SP, Brazil. E-mail: adriana_fmorae@hotmail.com, figueira@fcav.unesp.br (2)Unesp, Faculdade de Ciências Agrárias e Veterinárias, Departamento de Ciências Exatas, Via de Acesso Prof. Paulo Donato Castellane, s/n, CEP 14884-900 Jaboticabal, SP, Brazil. E-mail: lucas-aparecido@outlook.com

Abstract – The objective of this work was to estimate the coffee supply by calibrating statistical models with economic and climatic variables for the main producing regions of the state of São Paulo, Brazil. The regions were Batatais, Caconde, Cássia dos Coqueiros, Cristais Paulista, Espírito Santo do Pinhal, Marília, Mococa, and Osvaldo Cruz. Data on coffee supply, economic variables (rural credit, rural agricultural credit, and production value), and climatic variables (air temperature, rainfall, potential evapotranspiration, water deficit, and water surplus) for each region, during the period from 2000–2014, were used. The models were calibrated using multiple linear regression, and all possible combinations were tested for selecting the variables. Coffee supply was the dependent variable, and the other ones were considered independent. The accuracy and precision of the models were assessed by the mean absolute percentage error and the adjusted coefficient of determination, respectively. The variables that most affect coffee supply are production value and air temperature. Coffee supply can be estimated with multiple linear regressions using economic and climatic variables. The most accurate models are those calibrated to estimate coffee supply for the regions of Cássia dos Coqueiros and Osvaldo Cruz.

Index terms: Coffea arabica, climate, econometrics, modelling, rural credit.

Modelos econômicos e climáticos para estimativa da oferta cafeeira

Resumo – O objetivo deste trabalho foi estimar a oferta cafeeira por meio da calibração de modelos estatísticos, com variáveis econômicas e climáticas, das principais regiões produtoras do Estado de São Paulo. As regiões estudadas foram Batatais, Caconde, Cássia dos Coqueiros, Cristais Paulista, Espírito Santo do Pinhal, Marília, Mococa e Osvaldo Cruz. Foram utilizados dados de oferta cafeeira, variáveis econômicas (crédito rural, crédito rural na agricultura e valor da produção) e variáveis climáticas (temperatura do ar, precipitação pluvial, evapotranspiração potencial, deficiência e excedente hídrico) de cada região, para o período de 2000–2014. Os modelos foram calibrados com uso de técnicas de regressão linear múltipla, e todas as combinações possíveis foram testadas para a seleção das variáveis. A oferta cafeeira foi a variável dependente, e as demais, as independentes. A acurácia e a precisão dos modelos foram analisadas pelo erro percentual médio e pelo coeficiente de determinação ajustado, respectivamente. As variáveis que mais influenciam a oferta cafeeira são o valor de produção e a temperatura do ar. É possível estimar a oferta cafeeira com regressões lineares múltiplas por meio de variáveis econômicas e elementos climáticos. Os modelos mais acurados são os calibrados para estimar a oferta cafeeira das regiões de Cássia dos Coqueiros e Osvaldo Cruz.

Termos para indexação: Coffea arabica, clima, econometria, modelagem, crédito rural.

Introduction

Coffee (Coffea arabica L.) is a commodity that generates employment and income in several Brazilian regions (Aparecido et al., 2015) and is one of the country’s main agricultural exports (Resende et al., 2009; Barbosa et al., 2012). Worldwide, it is the most consumed beverage after water (Zelber-Sagi et al., 2015). Brazilian production is 51.94 million 60-kg sacks (Acompanhamento..., 2016). A method for estimating supply would undoubtedly be a useful strategic marketing tool, considering the importance of this commodity.

Supply is the quantity of a product or service available for purchase over a period of time (Saylor, 1974; Alves & Bacchi, 2004). Identifying the variables that determine the variation in coffee supply is fundamental for understanding the sector’s response...
in the future markets (Aparecido et al., 2017). These determinants of supply may be economic, such as product price, production value, and public credit (Goedert et al., 1980), and climatic, including air temperature, rainfall, and water deficit (Martins et al., 2015).

Simulation models can be used to efficiently evaluate the economic and climatic effects on crop supply (Vianna & Sentelhas, 2014; Shao et al., 2015). The models integrate information from several areas (Aparecido et al., 2014), from economic variables to crop agrometeorological conditions (Jame & Cutforth, 1996; Gouranga & Kumar, 2014). The models help producers in their strategic planning (Lucio et al., 2016) and decision-making in the production chain (Chipanshi et al., 2015). Simulation models were used to generate and analyse temporal projections (Pinto et al., 2015) for Brazilian coffee cultivation (Aparecido et al., 2017).

Previous studies have used econometric analyses. Shikida et al. (2007), for example, evaluated the variation of sugar and alcohol supplies in the state of Paraná from 1980 to 2015 and found that the supply of sugar varied inversely with the price of alcohol; the quantity of sugar on the market decreased by 2% with a 1% increase in the average price of alcohol. Satolo & Bacchi (2009) evaluated the role of sudden changes in supply and demand in the recent evolution of sugarcane production for the state of São Paulo and concluded that the variation of the price of sugarcane had an impact of >40% on the productivity of the supply of sugarcane. A few studies, however, have boldly estimated coffee supply based on economic and climatic conditions.

The objective of this work was to estimate coffee supply by calibrating statistical models with economic and climatic variables for the main producing regions of the state of São Paulo, Brazil.

Materials and Methods

Historical series of economic, climatic, and phytotechnical data of arabica coffee planted in the state of São Paulo, Brazil, were used. Representative regions of the state were selected for estimating coffee production: Batatais (BTT), Caconde (CCD), Cássia dos Coqueiros (CDC), Cristais Paulista (CTP), Espírito Santo do Pinhal (ESP), Marília (MAR), Mococa (MCC), and Osvaldo Cruz (OVC).

Multiple linear regression models (RLM) were used for the modeling of coffee production supply. In the present study, the equation used was:

\[ Y = CL + a \times X_1 + b \times X_2 + c \times X_3 + d \times X_4 + \varepsilon, \]

in which \( Y \) is the coffee supply of the municipalities (bags of 60 kg); \( a, b, c, \) and \( d \) are the parameters of the model (weight); \( X_1, X_2, X_3, \) and \( X_4 \) are the economic and climatic variables selected; \( CL \) is the linear coefficient (constant term); and \( \varepsilon \) is the random error.

Economic and climatic variables were used as independent variables in the construction of the models. The economic variables were the total rural credit of the municipality (real), the rural credit of the municipality’s agriculture (real), and the value of coffee production, which consisted of the unit price multiplied by the quantity produced (current thousand real). The climatic variables were air temperature (ºC), rainfall (mm per year), potential evapotranspiration (mm per year), water deficit (mm per year), and water surplus (mm per year). The coffee-supply and economic data were for 2000-2014 and were obtained from the Instituto de Economia Agrícola (IEA, 2017) and Fundação Sistema Estadual de Análise de Dados (São Paulo, 2015). The series of climatic data was obtained from the National Institute of Meteorology.

All possible combinations of up to five independent variables were tested to avoid problems of stabilization of local errors and to obtain highly consistent analyses (Walpole et al., 2012). The method of estimation was the ordinary least squares (OLS) regression with generalised reduced gradient optimisation. The model assessed was fit by analysing the possible effect of collinearity between the exploratory variables (multicollinearity) and testing the assumptions of the normality of errors and the homoscedasticity of the variables (Gujarati & Porter, 2011). Pearson’s correlation analysis was used to identify any multicollinearity of the explanatory variables. The exploratory variables were extracted with the Pearson’s correlation coefficient (\( r, \geq|0.7| \)). Collinearity between exploratory variables can be a problem in models when analysing the weights of the coefficients (elasticity/sensitivity) (Gujarati & Porter, 2011). Normality was verified by the Kolmogorov-Smirnov test (K-S), and the assumption of the homoscedasticity of the variables was verified by the White test (1980), also known as a covariance-matrix test.
Correlations (r) were estimated between coffee supply and the economic and climatic variables, in order to identify those with the largest effect on the coffee supplies of the municipalities. The climatic and economic variables with the highest Pearson's correlation coefficients (r=1) were also subjected to a simple linear regression analysis to explore and infer the relationship of these variables as a function of coffee supply.

A sensitivity analysis, also known as “elasticity”, was applied with the calibrated models (Gujarati & Porter, 2011). In this analysis, the angular coefficients (weight) of the independent variables are compared, and the higher the weight, the greater the effect of that variable on coffee supply.

The models were calibrated using visual basic for applications. The models were selected by an evaluation of the accuracy indicated by MAPE (%) and of the precision indicated by the adjusted coefficient of determination (adjusted $R^2$), according to the equations:

$$MAPE = \frac{\sum_{i=1}^{n} \left( \frac{\text{Yest}_i - \text{Yobs}_i}{\text{Yobs}_i} \times 100 \right)}{n},$$

in which $\text{Yest}_i$ is an estimated variable; $\text{Yobs}_i$ is an observed variable; and $n$ is the number of datapoints (years).

Adjusted $R^2 = 1 - \left( \frac{1 - R^2 \times (n - 1)}{n - k - 1} \right)$,

in which $n$ is the number of datapoints (years), and $k$ is the number of independent variables in the regression. Only the significant regressions by the F-test, at 5% probability, were selected.

**Results and Discussion**

The economic variables evaluated had a direct relationship with coffee supply in the state of São Paulo (Table 1). Coffee supply increased with increases in total rural credit (RC), rural credit from agriculture (RCA), and the value of coffee production (VP), because the increases in the credits and the price of the product allowed coffee growers to invest in their crops, which led to increased production (Tosi et al., 2007).

The climatic variables were variously correlated with coffee supply (Table 1). Coffee supply decreased with increases in mean air temperature (T), potential evapotranspiration (ETo), and water deficit (WD). Other studies, such as that of Martins et al. (2015), have also reported that high T and intense WD negatively affected coffee production.

The economic variables were also variously correlated with the climatic conditions. P and WD were directly and indirectly correlated, respectively, with RC, RCA, and VP, showing that the financing

### Table 1. Pearson's correlation coefficients (r) between the economic and climatic variables for coffee (Coffea arabica) supply of the state of São Paulo, Brazil.

<table>
<thead>
<tr>
<th>Supply</th>
<th>RC</th>
<th>RCA</th>
<th>VP</th>
<th>Temperature</th>
<th>Rainfall</th>
<th>ETo</th>
<th>WD</th>
<th>SUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>0.286*</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCA</td>
<td>0.340*</td>
<td>0.988*</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP</td>
<td>0.808*</td>
<td>0.261*</td>
<td>0.319*</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>-0.552*</td>
<td>-0.387*</td>
<td>-0.418*</td>
<td>-0.437*</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall</td>
<td>0.101</td>
<td>0.078</td>
<td>0.085</td>
<td>0.016</td>
<td>-0.432</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETo</td>
<td>-0.249*</td>
<td>-0.080</td>
<td>-0.083</td>
<td>-0.179</td>
<td>0.437*</td>
<td>-0.045</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>WD</td>
<td>-0.208*</td>
<td>-0.217*</td>
<td>-0.218*</td>
<td>-0.067</td>
<td>0.570*</td>
<td>-0.471*</td>
<td>0.731*</td>
<td>1.0</td>
</tr>
<tr>
<td>SUR</td>
<td>0.162</td>
<td>0.064</td>
<td>0.069</td>
<td>0.115</td>
<td>-0.528*</td>
<td>0.811*</td>
<td>-0.488*</td>
<td>-0.564*</td>
</tr>
</tbody>
</table>

*RC, total rural credit by municipality (real); RCA, rural credit of the municipality’s agriculture (real); VP, value of coffee production (real); T, air temperature (°C); R, rainfall (mm per year); ETo, potential evapotranspiration (mm per year); WD, water deficit (mm per year); and SUR, water surplus (mm per year).*Significant by the t-test, at 5% probability.
of crops by public agencies was higher in years with high water rates and product prices. These effects of climatic conditions on the credits have already been highlighted in other studies, such those of as Rossetti (2001) and Guanziroli & Guanziroli (2015), who also found that the financing sources varied their capital investments as a function of meteorological conditions.

Coffee supply was more highly correlated with the economic than the climatic variables (Figure 1 A). It was most correlated with VP, followed by T, with high coefficients of 0.808 and -0.552, respectively. Identifying the variables that were highly correlated with coffee supply allowed to evaluate these relationships more thoroughly. VP had a direct relationship, which explained 83% of the variation in coffee supply. This value of VP was high, because production increased by 2,820 sacks of 60-kg for every 1,000 real increase in VP (Figure 1 B).

Testing all possible combinations with up to five variables produced models with a total of 1,744 combinations of independent variables (Figure 2 A). This method of selecting variables by testing all combinations was efficient; p decreased as the precision increased (adjusted $R^2 \approx 1.00$) and the mean absolute percentage error (MAPE) gradually decreased, ultimately reaching zero (Figure 2 B).

This classification was essential for selecting the most accurate models (more accurate and with less tendency), because only those with the lowest MAPEs and the highest precision (adjusted $R^2$) with $p<0.05$ were selected for estimating coffee supply.

The calibrated models were efficient; the accuracy and precision for all regions averaged 5.04% (MAPE) and 0.90 (adjusted $R^2$) (Table 2). These values are considered adequate by crop modellers (Santos & Camargo, 2006; Savin et al., 2007; Rosa et al., 2010). Chipanshi et al. (2015) used accurate estimation models to show that producers could improve their decision-making and strategic planning. The calibrated model for estimating the coffee supply of the CDC region (model 3) had a low MAPE and a high adjusted $R^2$ of 2.30 and 0.97%, respectively. An estimate of the supply of a region with an error of only 2.30% is considered very low (Aparecido et al., 2017); an average supply of 25,000 sacks of coffee in a municipality would vary by only 575 sacks. RC, VP, T, and annual rainfall had the greatest effect on the quantification of the coffee supply in CDC (Table 2). Of these, T showed the greatest effect and was inversely proportional to the supply.

The largest difficulty in estimating coffee supply was the high bienniality of coffee between
harvests, which is due to the physiology of the crop (Rodrigues et al., 2013; Melke & Fetene, 2014). The calibrated models for the coffee regions followed the variability of coffee supply. Notably, the models calibrated with economic and climatic variables, in addition to accompanying the bienniality, also followed the tendencies of increases and decreases in coffee supply, as in BTT and OVC, respectively (Figure 3).

The highest performances were in the models calibrated for BTT, CDC, and OVC, with adjusted R² of 0.97, 0.93, and 0.91, respectively (Figure 4), which are considered high in the calibration of estimation models (Savin et al., 2007). Moreto & Rolim (2015) also reported precisions >0.95 (adjusted R²) in calibrated models for estimating orange [Citrus sinensis (L.) Osbeck] production for the state of São Paulo. The model developed for ESP performed the worst, with 0.75.

Table 2. Calibrated models to estimate the coffee (Coffea arabica) supply of the state of São Paulo, Brazil, as affected by economic and climatic parameters.

<table>
<thead>
<tr>
<th>Region</th>
<th>Models</th>
<th>p-value</th>
<th>Calibration MAPE (%)</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) BTT</td>
<td>Y = 0.00165CR − 0.00166CRA + 3.419VP + 5686.20T − 4.063SUR − 121479.74</td>
<td>0.001</td>
<td>6.5</td>
<td>0.89</td>
</tr>
<tr>
<td>(2) CCD</td>
<td>Y = −0.00119CRA + 3.86VP − 51.20R − 113.23ETo − 169.09WD + 41615.47</td>
<td>0.002</td>
<td>4.6</td>
<td>0.96</td>
</tr>
<tr>
<td>(3) CDC</td>
<td>Y = 0.000438CRA + 3.17VP − 1660.79T − 1.260R + 46934.7</td>
<td>0.004</td>
<td>2.3</td>
<td>0.97</td>
</tr>
<tr>
<td>(4) CTP</td>
<td>Y = 0.0661CR − 0.0655CRA − 0.840VP − 45223T − 151.34SUR − 1154451</td>
<td>0.002</td>
<td>3.7</td>
<td>0.93</td>
</tr>
<tr>
<td>(5) ESP</td>
<td>Y = 0.00039CR + 1.245VP − 221.16ETo − 59.95WD + 456296.40</td>
<td>0.004</td>
<td>10.7</td>
<td>0.75</td>
</tr>
<tr>
<td>(6) MAR</td>
<td>Y = 0.0021CR + 0.00301CRA + 1.376VP + 1.900ETo − 17.762WD + 12526.20</td>
<td>0.002</td>
<td>4.2</td>
<td>0.90</td>
</tr>
<tr>
<td>(7) MCC</td>
<td>Y = 0.00097CR + 10.37VP − 55267.4T − 219.9R + 206.26SUR + 1453247</td>
<td>0.002</td>
<td>3.3</td>
<td>0.92</td>
</tr>
<tr>
<td>(8) OVC</td>
<td>Y = −0.00145CR + 3.899VP − 1282.95T − 2.155SUR + 7.213WD + 58424.29</td>
<td>0.001</td>
<td>2.9</td>
<td>0.94</td>
</tr>
</tbody>
</table>

BTT, Batatais; CCD, Caconde; CDC, Càssia dos Coqueiros; CTP, Cristais Paulista; ESP, Espírito Santo do Pinhal; MAR, Marilia; MCC, Mococa; and OVC, Osvaldo Cruz. Y, coffee supply (sacks); RC, total rural credit by municipality (real); RCA, rural credit of the municipality’s agriculture (real); VP, value of coffee production (real); T, air temperature (°C); R, rainfall (mm per year); ETo, potential evapotranspiration (mm per year); WD, water deficit (mm per year); and SUR, water surplus (mm per year). MAPE, mean absolute percentage error.
Figure 3. Coffee (*Coffea arabica*) supply estimated by the calibrated models for the state of São Paulo, Brazil, in: A, Batatais; B, Caconde; C, Cássia dos Coqueiros; D, Cristais Paulista; E, Espírito Santo do Pinhal; F, Marília; G, Mococa; and H, Osvaldo Cruz.
Figure 4. Performance of the coffee (*Coffea arabica*) supply estimation models for the state of São Paulo, Brazil, in: A, Batatais; B, Caconde; C, Cássia dos Coqueiros; D, Cristais Paulista; E, Espírito Santo do Pinhal; F, Marília; G, Mococa; and H, Osvaldo Cruz. Constant line is the 1:1 line.
an accuracy of 0.30 (Figure 4 E). ESP had the largest deviations in the estimates of the coffee supply, due to the effect of climatic factors not used in the calibration of the models, e.g., low minimum air temperature, cold winds, and frost.

Conclusions

1. Coffee (Coffea arabica) supply is controlled by economic and climatic variables.

2. The value of coffee production and air temperature have the greatest effect on the variation of coffee supply in the state of São Paulo, Brazil.

3. Coffee supply increases with an increase in production value.

4. Coffee supply can be estimated with multiple linear regressions of economic and climatic variables.

5. The models calibrated with economic and climatic variables are accurate and precise, and the models calibrated for estimating the coffee of the regions of Cássia dos Coqueiros and Osvaldo Cruz are the most accurate.

References


