Structural Equation Modeling Applied to Internet Consumption Forecast in Brazil

MATHEUS PEREIRA LIBÓRIO[1], **PETR YA EKEL**[1], **RENATA DE MELLO LYRIO**[2],
PATRÍCIA BERNARDES[1], **GUSTAVO LUÍS SOARES**[1], **AND**
THIAGO MELO MACHADO-COELHO[3]

[1]Pontifical Catholic University of Minas Gerais, Belo Horizonte 30535-012, Brazil
[2]Federation of Industries of Minas Gerais, Belo Horizonte 35501-001, Brazil
[3]Graduate Program in Electrical Engineering, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil

Corresponding author: Thiago Melo Machado-Coelho (thmmcoelho@ufmg.br)

ABSTRACT The growth of Internet services over the past few years has increased the importance of Internet consumption forecasting models in telecommunications network planning activity. Most of these models use historical data of variables that, when reduced in size, form context variables: economic, innovation, technology, and Internet supply. Although they include multiple context variables, the techniques employed in using these models are limited to the analysis of a single dependency relationship. Thus, these models fail to understand the interrelationships between context variables and their effects (direct and indirect) on Internet consumption. This research uses Partial-Least-Square Structural Equations Modeling to understand Internet consumption in Brazil between 2002 and 2017. The results show, for example, that the economic context produces direct effects in the context of innovations and indirect effects in the context of technologies and Internet consumption. These results contribute to increase the knowledge of the relationships between the context variables that influence Internet consumption, providing inputs for the development of more accurate forecasting models, thus contributing to the activity of telecommunications network planning.

INDEX TERMS Structural equation modeling, internet consumption, telecommunications networks planning.

I. INTRODUCTION

Telecommunications network planning is more than the capacity of allocation and Internet/traffic routing [1] Studies show that most telecommunications network planning problems are associated with forecasting future Internet consumption [2]. Researchers suggest that such forecasts are important in ensuring that the planned network can meet rising Internet consumption [3] In this context, several studies have been dedicated to forecasting Internet consumption. Among these studies, the most commonly found models are those that assume that past consumption is a good indication of future consumption [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Shunfeng Cheng.

Most Internet consumption forecast models by historical standards use a small number of independent variables (time series) [5]. This is because the presence of multicollinearity in the time series regressions [6] interferes with the results of the number of coefficients, errors, parameter reliability, and forecast accuracy [7]. Besides, time series models usually distort long-term forecasts since such models: I) do not capture emerging consumption behaviors [8]; II) ignore that future hitherto unknown technologies and innovations may change Internet consumption behavior and create new demands [9]; III) consider that Internet consumption is an independent variable of supply, ignoring that Internet supply enables the development of technologies and innovations that stimulate Internet consumption, that is, they disregard that supply creates its demand [10], [11]; and IV) are sensitive

to the quality of historical Internet consumption data, which in turn is questionable [12].

In turn, Internet consumption forecast models not based on historical data enable forecasts in short-term changes in the amount of Internet traffic over five months [13] or to understand the concentration and distribution of consumption [14], but fail to replicate and/or generalize results to longer scenarios and/or to other geographic dimensions, such as cities and countries.

Considering the limitations discussed above, researchers claim that traditional models are unreliable to forecast phenomena embedded in complex contexts and that new models should represent such contexts in forecasting Internet consumption [1] by including variables such as, for example, the profile and number of Internet consumers [15]. Also, it must be considered that these variables, which represent the context of demand for connectivity and accessibility, correlate with economic factors that vary over time and space [16]. Loskot *et al.* [4] add that forecast models should incorporate variables of different types and natures, not just variables of an economic nature. The literature brings studies that correlate Internet consumption with demographic [8], technological and innovation [9], [17] variables and the supply of technologies based on Internet [11]. However, according to the OECD [12], the inclusion of variables in the model is to be based on theoretical bases capable of representing an Internet consumption profile consistent with reality. Therefore, how to represent the various contexts that permeate Internet consumption, their interrelationships, and effects on future Internet consumption?

This research addresses the problem of understanding future Internet consumption by approaching the multiple dependency relationships among variables of various types and natures associated with Internet consumption. Therefore, a simultaneous equation model is employed that combines factor analysis (dimensional reduction) and Partial-Least-Squares (PLS) regression [18] the so called Partial-Least-Squares Structural Equations Modeling (PLS-SEM).

In addition to this Introduction, the paper presents the theoretical framework of the Internet consumption forecast model in Section II. Section III describes the operational structure of PLS-SEM. Section IV details the 32 variables that represent the contexts: economics, technology, innovation, and Internet supply as well as the methods which process these variables: factor analysis, PLS regression, and exponential smoothing. The obtained results are presented in Section V and the conclusions of this study are given in section VI.

II. THEORETICAL FRAMEWORK OF THE INTERNET CONSUMPTION FORECAST

In PLS-SEM, the simultaneous estimation of dependency relationships requires the presence of a pre-established theoretical model that indicates the combination and order of dependency relationships as coherent justifications derived from the literature [26]. In this sense, the proposed model is formed by four exogenous or latent variables that represent

the innovation, technology, economy, and Internet supply environments, as well as the endogenous variable of Internet consumption. The model assumes that Internet consumption is positively correlated with: I) Technology and innovation factors [9], [17]; II) Internet supply [10], [11]; III) The economic context, including its demographics [8], [16]. The model further assumes that the economic context is a relevant predictor in understanding Internet supply and technology and innovation factors over time [4]. Finally, the model considers that technologies are dependent on innovations [9]. Fig. 1 illustrates the structural model of Internet consumption in Brazil based on these assumptions.
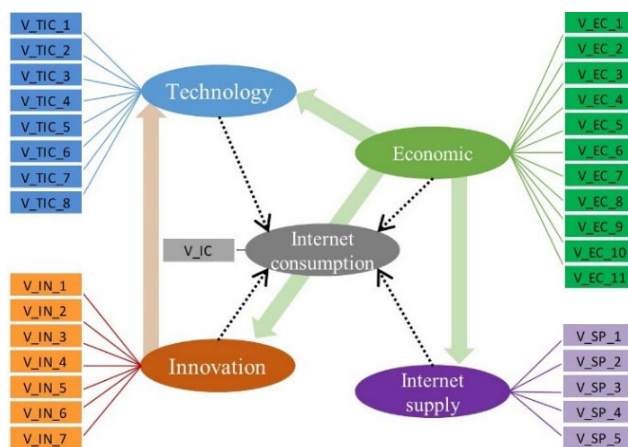

FIGURE 1. Structural model PLS-SEM of Internet consumption.

This structural model includes thirty-two observed variables (detailed in Table 1/Section IV), five constructs or latent variables, four dependency relationships between latent variables (continuous arrows), and four dependency relationships between latent variables and independent variables (dotted arrows).

The proposed structural model is associated with a method of maximizing the explained variance of exogenous and endogenous constructs [18]. PLS-SEM retains a larger number of variables for each construct, allowing to simultaneously increase the forecast accuracy of endogenous variables [36]. Also, the sample size may be smaller in PLS-SEM than in CB-SEM, and there is no need for normal data distribution [26]. Thus, PLS-SEM is an appropriate method for conducting exploratory analyzes with smaller samples, more complex models, rich data, and weak theory [37]–[39]. In this sense, the proposed PLS-SEM structural model explores the interrelationships between the latent variables mentioned in the literature, and their respective direct and indirect effects on Internet consumption. Thus, it seeks to strengthen the theory of how the various contexts interrelate with Internet consumption over time.

III. OPERATIONAL FRAMEWORK OF THE INTERNET CONSUMPTION FORECAST

The basics of SEM are presented in the studies of Galton [19] and Pearson and Lee [20]. However, it is in

Spearman's research [21] that the factor model (exploratory and confirmatory factor analysis) was presented. From this, Jöreskog [22] develops a general model containing two approaches in which factors (latent variables) correlate to a specific subset of observed variables [23]. In the measurement approach, the observed variables are related to the latent variables through confirmatory factor analysis [24]. Based on the covariance technique, this approach is called Covariance-based SEM (CB-SEM). By allowing the representation of constructs by factors, CB-SEM is more used in studies that seek confirmation of theories [25]. In the structural approach, latent variables, that is, variables constructed from observed variables, are related to each other through systems of simultaneous equations [18], [26]. Based on the partial least squares technique, this approach is called PLS-SEM. By enabling the representation of significance and strength (correlation coefficients) of factor relationships, PLS-SEM is most commonly used in studies that explore and analyze causal relationships [24], [27], [28].

SEM is an approach focused on the analysis of complex phenomena [29], and brings several advances concerning traditional methods (factor analysis, principal components, discriminant analysis, and multiple regression) [26], [27]. SEM enables: first, the inclusion of observed and/or latent variables [29]. Second, the analysis of more than one dependency relationship [30]. Third, the incorporation of estimation/measurement errors into the model [31]. Fourth, the confirmation of theoretical propositions, including them automatically to the model [32]. Fifth, through the interaction between theory and data, statistically testing theoretical assumptions by modeling relationships between predictor and multiple criterion variables [33]. In summary, unlike traditional methods, SEM allows the analysis of theories developed in previous research from an unlimited number of dependent variables and/or relationships [31], [32]. Therefore, SEM-based temporal data forecasts [34], [35] capture the dynamic interrelationships between the constructs examined, and offer a broader understanding of the evolution of observed and latent variables over time.

IV. MATERIALS AND METHODS

The process of building a forecast of Internet consumption includes three stages: data collection, use of the PLS-SEM, and utilization of a forecast model.

The first stage: eleven data sources [40]–[50] were consulted to reproduce the conceptual model presented in Section III. In total, 32 variables were taken into account, which are related to four dimensions or contexts: Economic, Innovation, Internet supply, and Technologic. The data of the 32 variables tested in the model cover the period from 2002 to 2017 and are presented in Table 1.

The second stage: the collected data were processed within the PLS-SEM, whose use combines the methods of factor analysis and partial least squares regression [38]. Factor analysis is a statistical method that allows representing a set of variables observed from a small number of unobserved

TABLE 1. Observed variables tested in the Internet consumption model.

Code	Description	Ref.
Economic context variables		
V_EC_1	Domestic credit to private sector[1]	
V_EC_2	GDP[2]	
V_EC_3	GDP per capita[2]	
V_EC_4	Gross capital formation[2]	
V_EC_5	Gross domestic income[3]	
V_EC_6	Gross domestic savings[2]	[40]
V_EC_7	Gross national expenditure[2]	
V_EC_8	Gross savings[2]	
V_EC_9	Net investment in nonfinancial assets[2]	
V_EC_10	Population, total	
V_EC_11	Unemployment rate	
Innovation context variables		
V_IN_1	R&D employees	[41]
V_IN_2	R&D Firms	[41]
V_IN_3	Firms that spend on R&D[4]	[42]
V_IN_4	R&D expenditure[1]	[43]
V_IN_5	Researchers in R&D[5]	[43]
V_IN_6	Scientific and technical journal articles	[44]
V_IN_7	Technicians in R&D[5]	[43]
Internet supply context variables		
V_SP_1	Households with computers	[45]
V_SP_2	Households connected to the Internet	[45]
V_SP_3	Fixed broadband subscriptions	[46]
V_SP_4	Fixed telephone subscriptions	[46]
V_SP_5	Individuals using the Internet[6]	[46]
Technologic context variables		
V_TIC_1	Telecommunications employees	[41]
V_TIC_2	Telecommunications companies	[41]
V_TIC_3	Telecommunications service index	[47]
V_TIC_4	ICT employees	[41]
V_TIC_5	ICT companies	[41]
V_TIC_6	Investment in ICT with private participation[7]	[48]
V_TIC_7	ICT activity index	[49]
V_TIC_8	High technology exports[7]	[50]
V_IC	Internet consumption[8] (dependent variable)	[51]

variables, called factors [26]–[28], [52]. The factor analysis model is given as follows:

$$X = \mu + LF + e \qquad (1)$$

where X is the measurement vector, $p \times 1$; μ is the vector of means, $p \times 1$; L is the matrix of factorial loads, $p \times m$; F is the vector of common factors, $m \times 1$; e is the vector of residues, $p \times 1$; p is the number of measurements; m is the number of common factors.

Applying the factor analysis, the 32 observed variables presented in Table 1, were reduced in their dimensions to four context variables: economy, technology, innovation, and Internet supply. Such a decrease in dimensionality makes it possible to reduce the probability of collinearity between independent variables in the forecast model [6], [14], [17].

The partial least squares regression is a well-known technique that minimizes the sum of the squares of the residuals of the regression ($Y = a+bx$) [26]–[28], [52] within the following model:

$$Y_i = a + bx_i + e_i \qquad (2)$$

where $i = 1, \ldots, n$ is the observation number and e_i is the ith residual.

The minimization of $\sum_{i=1}^{n} e_i^2$ permits one to adjust the regression model to the observed data.

In the use of the partial least squares regression, the significance and strength of the correlations between the context variables were calculated and tested [36], [39]. The significance and correlation of the independent variables indicate which context variables are to be forecasted to calculate Internet consumption.

The factor analysis combined with the partial least squares regression was performed by the PLS-SEM. In turn, the PLS-SEM can be operationalized in different types of software such as SmartPLS, EQS, MPLUS, PROC CALIS (in SAS), HLM, SIMPLIS, and GLAMM [23] as well as the popular SPSS AMOS [38] or by the pioneer LISREL [22]. The Internet consumption model developed in this research was operationalized by the SmartPLS Software [39], following a usage tutorial presented in [36], and using license available at www.smartpls.de.

The exploratory analysis procedure of the relationships between multiple observations (independent variables) and multiple constructs (dependent variables) was realized in two steps [36]. The first step in applying the PLS-SEM Algorithm concerns to the factor analysis. The observed variables that influences the latent variable measured through the factor loading remain in the model. The factor loading must be greater than 0.7 [26]–[28]. The latent variables that present acceptable convergent validity measured through the Average Variance Extracted (AVE) remain in the model. The AVE threshold must be greater than 0.5 to explain, on average, more than half of the variance of the observed variables [26].

Once the observed and latent variables are selected, the reliability of the model is analyzed by the cross-loadings, Cronbach's Alpha (CA), and the Dillon-Goldstein Rho (Rho) criteria [52], [53] The factor loadings of the observed variables must be higher in their respective latent variables than in others [33]. Differences close to 2.5% are acceptable, but it is possible to increase the rigor of discriminant validity by excluding observed variables with high correlations in two latent variables [36]. The internal consistency is measured through CA. Its test threshold varies by type of survey. For instance, for exploratory studies, CA values must be greater than 0.60 [27], [28]. Rho is considered the best reliability measure test [52]. Its values must be greater than 0.70 [53].

The second step in applying the PLS-SEM Algorithm concerns to the internal validity of model in which the PLS regression is performed. The observed variables not statistically significant (t-test < 1.96) to the latent variables are excluded from the model [18]. The correlations between latent variables that are not significant with a p-value greater than 0.05 do not remain in the model. The correlations between latent variables not consistent with the theory, for example, a negative correlation between Internet consumption and the context of technology and innovation do not remain in the model as well [36]. The significance and strength of the regressions are evaluated by Student's t-test

and the coefficient of determination R^2 [26]. At the end of the PLS-SEM, path coefficients are generated, which are the standardized weights of the regression variables [36], [37].

The third stage: after the running the two steps of applying the PLS-SEM Algorithm, the latent variables with significant correlations with the dependent variables of Internet consumption are projected by the use of exponential smoothing autoregressive model.

Exponential smoothing is a traditional linear statistical model that allows forecasting future values based on the past observations [54], [55] as follows:

$$s_t = \alpha x_t + (1 - \alpha) s_{t-1} = s_{t-1} + \alpha (x_t - s_{t-1}) \qquad (3)$$

where α $(0 < \alpha < 1)$ is the smoothing factor.

The smoothed statistic s_t is a simple weighted average of the observation x_t and the previous smoothed statistic s_{t-1}. The value of α defines the smoothing level so that the smoothing level will be zero when $\alpha = 1$. The mean squared error (MSE) and mean absolute deviation (MAD) are applied to measure the forecasting accuracy. The ideal model has MSE and MAD values close to zero [54].

Internet consumption was forecasted considering the future values of the latent variables generated through exponential smoothing and the coefficients of these latent variables generated through the PLS-SEM. An overview of the three stages of the Internet consumption forecast model is illustrated in Fig. 2.
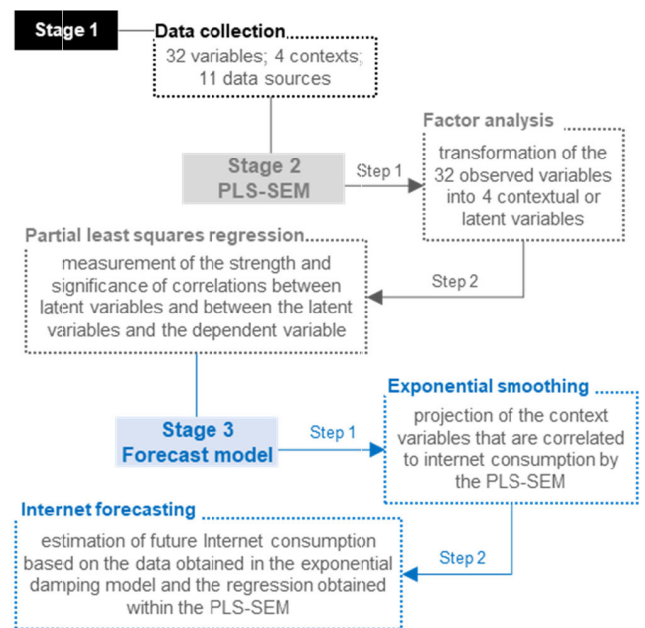
FIGURE 2. Overview of the three stages of the Internet consumption forecast model.

Finally, the results of Internet consumption based on the multiple relationships between the variables were compared with the results obtained with the use of the traditional internet forecast model that considers only direct effects.
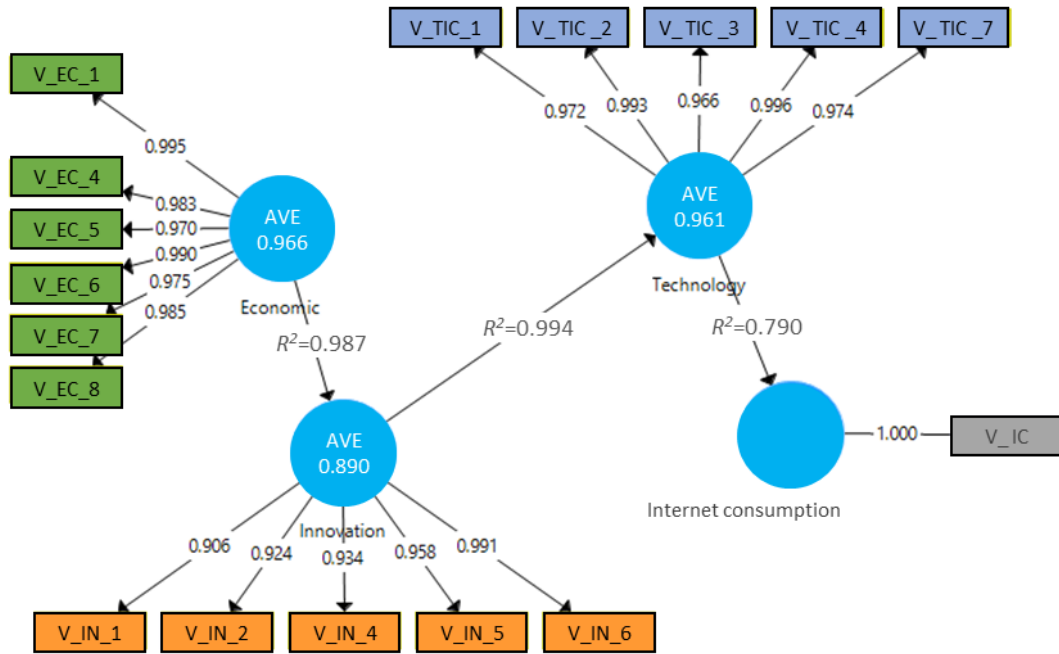
FIGURE 3. Average variance extracted (AVE) and coefficient of determination of the Internet consumption model.

V. RESULTS

After running the first step of the PLS-Algorithm, the latent variable of Internet Supply did not meet the AVE 0.50 threshold and did not remain in the model. The others latent variables: economic, innovation, and technology meet the AVE thresholds and remained in the model. Six observed variables of economic context, two observed variables of innovation context, and three observed variables of the technology context did not meet the factor loading threshold. The observed variables of each context, that is, of each latent variable that remained in the model, are given in Table 2 and reflect the reliability of the model by the cross-loadings criterion.

TABLE 2. Model reliability - cross loading criterion.

Variables	Economic	Innovation	Technology	Internet Consumption
V_EC_1	1.00	0.98	0.94	0.69
V_EC_4	0.98	0.96	0.93	0.67
V_EC_5	0.97	0.95	0.89	0.61
V_EC_6	0.99	0.99	0.97	0.77
V_EC_7	0.98	0.99	1.00	0.88
V_EC_8	0.99	0.99	0.96	0.76
V_IN_1	0.92	0.91	0.88	0.70
V_IN_2	0.94	0.92	0.88	0.67
V_IN_4	0.91	0.93	0.96	0.82
V_IN_5	0.93	0.96	1.00	0.92
V_IN_6	0.99	0.99	0.98	0.80
V_TIC_1	0.99	0.99	0.97	0.78
V_TIC_2	0.95	0.98	0.99	0.90
V_TIC_3	0.86	0.93	0.97	0.97
V_TIC_4	0.97	1.00	0.99	0.87
V_TIC_7	0.96	0.98	0.97	0.84
V_IC	0.74	0.83	0.89	1.00

Caption: the red text indicates which variables do not have higher cross-loadings in their respective constructions.

The factor loadings of the observed variables must be higher in their respective latent variables than in others [33]. Only two of the seventeen observed variables do not meet the criterion of discriminant validity. The differences in cross-loadings of variables V_IN_1 (0.91 and 0.92) and V_TIC_4 (0.99 and 1.00) are less than 2.5% which is an acceptable difference [33]. Table 3 shows the reliability of the model measured by the CA, Rho, and AVE criteria for each of the three latent variables maintained in the model.

TABLE 3. Reliability of the model.

Description	CA	Rho	AVE
Economic	0.994	0.994	0.966
Innovation	0.976	0.977	0.890
Technology	0.992	0.992	0.961

Considering the results of Tables 2 and 3, the V_IN_1 and V_TIC_4 variables were maintained in their original constructs, establishing the structural model of Internet consumption. Fig. 3 provides a synthesis of the PLS-SEM-based structural model of Internet consumption, showing the variance extracted from the observed and latent variables, as well as the strength of these relationships (coefficient of determination).

By analyzing the internal validity of the structural model measured by the coefficients of determination (R^2) [26], which presented values $R^2 = 0.987$ (Economic $\rightarrow$ Innovation), $R^2 = 0.994$ (Innovation $\rightarrow$ Technology), and $R^2 = 0.790$ (Technology $\rightarrow$ Internet consumption), it can be said that the structural model is strong, since $R^2 > 0.67$ indicate a strong model; $R^2 > 0.33$ and $R^2 < 0.67$ indicate a moderate model; and $R^2 < 0.19$ indicate a weak
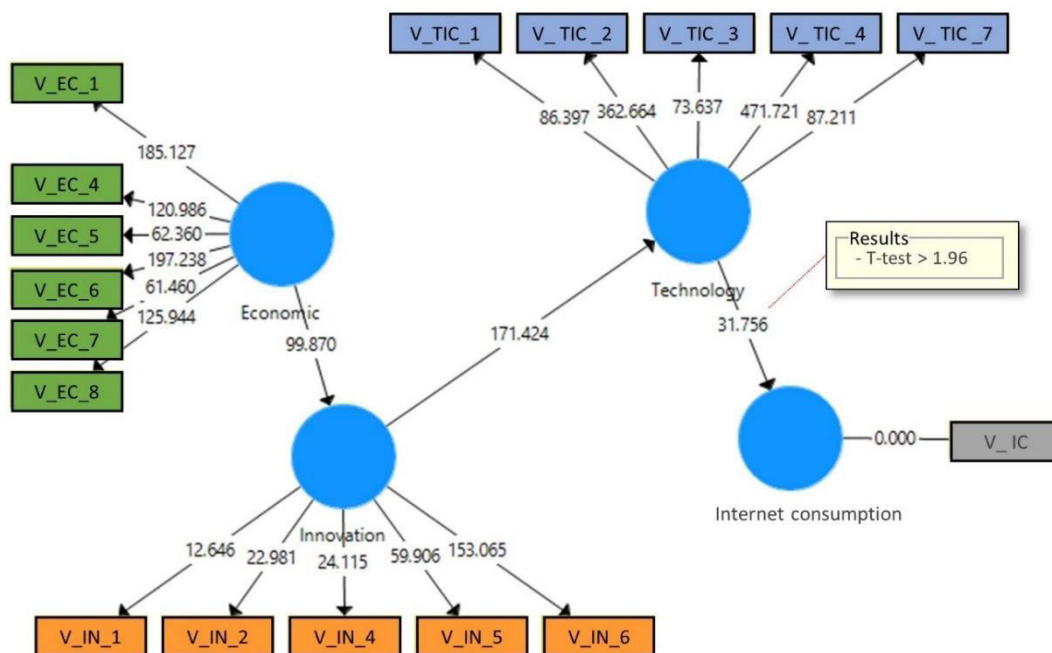
FIGURE 4. Significance of model correlations and regressions (Student's t-test).

TABLE 4. Latent variables correlations.

Latent variables	Economic	Innovation	Technology	Internet consumption
Economic	1.000	0.994	0.965	0.740
Innovation	0.994	1.000	0.997	0.830
Technology	0.965	0.997	1.000	0.889
Internet consumption	0.740	0.830	0.889	1.000

model [33]. Table 4 summarizes the correlations between the latent variables uses in the model.

The significance of the correlations and regressions were evaluated by Student's *t*-test [26]. Fig. 4 shows that all relationships between variables showed *t*-test > 1.96, indicating that correlations and regressions are acceptable.

Once the PLS-SEM parameters were estimated, the relationships between the latent variables indicated that the economic variables are strongly correlated with innovation (R = 0.994), technology (R = 0.965), and Internet consumption ($R = 0.740$), supporting the understanding that variables of an economic nature are relevant predictors for understanding other variables over time [4]. From the path coefficients obtained in the model, the effects that the latent variables have on each other are estimated. This effect is estimated from the variation of the Standard Deviation (SD). In the present model, the variation of one SD in the economic variable produces effects of 0.994 SD in the innovation variable, 0.990 SD in the technology variable, and 0.880 SD in the variable of Internet consumption.

The relationships between the latent variables also indicates that the innovation variable are strongly correlated with technology (R = 0.997) and Internet consumption

(R = 0.830) variables, supporting the understanding that innovations drive Internet consumption [17] as well as the creation of technologies [9]. Considering the path coefficients extracted from the model, it is estimated that the effect of the variation of one SD in the innovation variable produces an effect of: 0.997 SD in the technology variable, and 0.886 SD in the Internet consumption variable.

Finally, the relationships between the latent variables indicate that the technology variables are strongly correlated with Internet consumption (R = 0.889), supporting the understanding that Internet-based technologies also drive Internet consumption [11], [17]. The path coefficients indicate that the effect of the variation of one SD on the technology variable produces an effect of 0.889 SD in the internet consumption variable.

It is noteworthy that the model also indicates that the internet supply variable may not stimulate Internet consumption as some research suggests [10], [11]. Taking into account this set of results, this research provides new evidence of the applicability of factor analysis to reduce the number of variables observed in a few latent variables, and to avoid multicollinearity between such variables in time series [6], [14], [17].

Besides, the results of the PLS-SEM also indicate that the effect between latent variables and Internet consumption may not always be direct. In fact, the results show the presence of indirect effects between: I) Innovation → Technology → Internet consumption (R = 0.886); II) Economic → Innovation → Technology → Internet consumption ($R = 0.880$); and I) Economic → Innovation → Technology ($R = 0.990$).

Table 5 shows the path coefficients, indicating what are the effects (direct and indirect) between the latent variables.

TABLE 5. Path coefficients (direct and indirect effects).

Variable	Innovation	Technology	Internet consumption
Economic	0.994	0.990	0.880
Innovation		0.997	0.886
Technology			0.889

The indirect effects alter the curve slope of the independent variables, and, therefore, the forecast of Internet consumption and the statistical parameters of the model. Fig. 5 shows how indirect effects influence the forecast of Internet consumption, while Table 6 shows the influence of indirect effects on the model's statistical parameters.
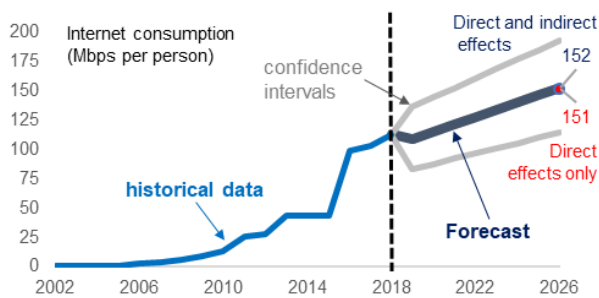

FIGURE 5. Internet forecasting considering direct and indirect effects.

TABLE 6. Statistics of the models.

Model / parameter	Direct effects	Direct and indirect effects
R^2	0.76	0.79
R^2-adjusted	0.74	0.78
MAD	0.09	0.06
MSE	0.02	0.01
Internet consumption[1]	150.9	152.1

Caption: [1]Internet consumption per person in Mbps.

Table 6 shows that considering indirect effects on the model can improve to some extent the accuracy of the forecasting, generating higher R^2 and R^2-adjusted coefficients and reduce the model forecast errors, generating smaller MAD and MSE. The obtained results suggest that understanding the interrelationships between latent variables, their respective factor loadings, correlations, coefficients, and direct and indirect effects may improve the forecast of Internet consumption.

Taking all results and their analysis into account, this research contributes to broadening understanding of the direct and indirect effects of exogenous factors on Internet consumption over time. By providing a broader understanding of the causal relationships of the phenomenon under study, this research provides insights that enable future Internet consumption to be forecasted with a greater theoretical foundation and accuracy.

VI. CONCLUSION

This research showed that, at least in Brazil, independent variables (economy, innovation, technology) do not always have a direct effect on future Internet consumption. This study also demonstrates that the latent variables that represent economy and innovation, by influencing the latent technology variable, have an indirect effect on Internet consumption. This shows that in the PLS-SEM, the independent variable (technology) can assume the role of the dependent variable of the other independent variables (economy and innovation). Thus, the multiple dependency ratios, correlation coefficients, and regression significance are summarized in a structural model. This contributes to broadening the understanding of Internet consumption over time. Getting together multiple contexts, the structural model brings the forecast of Internet consumption closer to reality, opening the possibility for better planning of telecommunications networks.

Given these possibilities, it is suggested that future works will be directed at comparing the results of different Internet consumption forecasting models. In addition, to try to generalize the paper results to other countries, it is suggested to evaluate the impacts of the different results of these forecasts on the planning of telecommunications networks.

REFERENCES

[1] B. Todd and J. Doucette, "A novel long term telecommunication network planning framework," *J. Netw. Syst. Manage.*, vol. 25, no. 1, pp. 47–82, Jan. 2017.
[2] E. K. Doolittle, H. L. M. Kerivin, and M. M. Wiecek, "Robust multiobjective optimization with application to Internet routing," *Ann. Oper. Res.*, vol. 271, no. 2, pp. 487–525, Dec. 2018.
[3] P. Papanikolaou, K. Christodoulopoulos, and E. Varvarigos, "Optimization techniques for incremental planning of multilayer elastic optical networks," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 10, no. 3, pp. 183–194, Mar. 2018.
[4] P. Loskot, M. A. M. Hassanien, F. Farjady, M. Ruffini, and D. Payne, "Long-term drivers of broadband traffic in next-generation networks," *Ann. Telecommun. Annales des Télécommunications*, vol. 70, nos. 1–2, pp. 1–10, Feb. 2015.
[5] I. C. Lima, S. Moro, and F. G. J. Junior, "Ciclos e previsão cíclica: Um modelo de indicadores antecedentes para a economia brasileira," in *Proc. Anais 34th Encontro Nacional de Economia 34th Brazilian Econ. Meeting*, 2006, pp. 1–21.
[6] H. H. Kim and N. R. Swanson, "Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence," *J. Econometrics*, vol. 178, pp. 352–367, Jan. 2014.
[7] A. Afifi, S. May, and V. A. Clark, *Practical Multivariate Analysis*. Boca Raton, FL, USA: CRC Press, 2011.
[8] M. P. Libório, S. Laudares, I. S. Silva, C. R. Cunha, P. Bernardes, and P. S. Saviott, "Dinâmica espaço tempo da demanda por Internet: As cidades do futuro no cenário brasileiro de telecomunicações," *Raega-O Espaço Geográfico em Análise*, vol. 44, pp. 8–20, May 2018.
[9] H. S. Na, D. Lee, J. Hwang, and C. Lee, "Research on the mutual relations between ISP and ASP efficiency changes for the sustainable growth of the Internet industry," *Appl. Econ.*, vol. 50, no. 11, pp. 1238–1253, Mar. 2018.
[10] R. Kenny and T. Broughton, *Domestic Bandwidth Requirements in Australia: A Forecast for the Period 2013–2023*. Communications Chambers, for The CIE. Accessed: Jan. 2, 2020. [Online]. Available: https://www.communications.gov.au/sites/g/files/net301/f/Forecasting-Australian-Per-Household-Bandwidth-Demand-Commun.pdf
[11] X. Han, V. Martinez, and A. Neely, "Service in the platform context: A review of the state of the art and future research," in *Collaborative Value Co-Creation in the Platform Economy*. Singapore: Springer, 2018, pp. 1–27.

[12] *Broadband Policies for Latin America and the Caribbean: A Digital Economy Toolkit*. OECD Publishing, Paris, France, 2017. Accessed: Jan. 2, 2020. [Online]. Available: http://www.oecd.org/Internet/broadband/lac-digital-toolkit/

[13] A. Mozo, B. Ordozgoiti, and S. Gómez-Canaval, "Forecasting short-term data center network traffic load with convolutional neural networks," *PLoS ONE*, vol. 13, no. 2, Feb. 2018, Art. no. e0191939.

[14] E. Jenelius and H. N. Koutsopoulos, "Urban network travel time prediction based on a probabilistic principal component analysis model of probe data," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 436–445, Feb. 2018.

[15] T. Karagiannis, M. Molle, and M. Faloutsos, "Long-range dependence ten years of Internet traffic modeling," *IEEE Internet Comput.*, vol. 8, no. 5, pp. 57–64, Sep. 2004.

[16] E. A. Mack and R. Maciejewski, "A profile of visual analytical toolkits for understanding the spatio-temporal evolution of broadband provision," *Telecommun. Policy*, vol. 39, nos. 3–4, pp. 320–332, May 2015.

[17] M. P. Libório, T. M. Machado-Coelho, P. Bernardes, A. M. C. Machado, P. Y. Ekel, and G. L. Soares, "Forecasting Internet Demand Using Public Data: A Case Study in Brazil," *IEEE Access*, vol. 6, pp. 65974–65980, 2018.

[18] C. B. Astrachan, V. K. Patel, and G. Wanzenried, "A comparative Studyof CB-SEM and PLS-SEM for theory development in family firm research," *J. Family Bus. Strategy*, vol. 5, no. 1, pp. 116–128, 2014.

[19] F. Galton, *Hereditary Genius: An Inquiry into its Laws and Consequences*, Vol. 27. New York, NY, USA: Macmillan, 1869.

[20] K. Pearson and A. Lee, "On the laws of inheritance in man: I. Inheritance of physical characters," *Biometrika*, vol. 2, no. 4, pp. 357–462, Nov. 1903.

[21] C. Spearman, "'General Intelligence' objectively determined and measured," *Amer. J. Psychol.*, vol. 15, no. 2, pp. 201–292, 1904.

[22] K. G. Jöreskog, "A general method for estimating a linear structural equation system," *ETS Res. Bull. Ser.*, vol. 1970, no. 2, pp. 1–41, Dec. 1970.

[23] P. Tarka, "An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences," *Qual. Quantity*, vol. 52, no. 1, pp. 313–354, Jan. 2018.

[24] P. B. Lowry and J. Gaskin, "Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it," *IEEE Trans. Prof. Commun.*, vol. 57, no. 2, pp. 123–146, Jun. 2014.

[25] D. Kaplan, "Structural equation modeling. Foundations and extensions," in *Advanced Quantitative Techniques in the Social Sciences Series*, vol. 10. Newbury Park, CA, USA: Sage, 2000.

[26] J. F. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks, CA, USA: Sage, 2014.

[27] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a silver bullet," *J. Marketing Theory Pract.*, vol. 19, no. 2, pp. 139–152, Apr. 2011.

[28] J. F. Hair, C. M. Ringle, and M. Sarstedt, "Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance," *Long Range Planning*, vol. 46, nos. 1–2, pp. 1–12, Feb. 2013.

[29] A. L. M. Codes, "Modelagem de equações estruturais: Um método para a análise de fenômenos complexos," *Caderno CRH*, vol. 18, no. 45, pp. 471–484, 2005.

[30] J. B. Ullman and P. M. Bentler, "Structural equation modeling," in *Handbook of Psychology*, vol. 2, 2nd ed. New York, NY, USA: Wiley, 2012.

[31] L. R. Fabrigar, R. D. Porter, and M. E. Norris, "Some things you should know about structural equation modeling but never thought to ask," *J. Consum. Psychol.*, vol. 20, no. 2, pp. 221–225, 2010.

[32] R. P. Bagozzi and Y. Yi, "Specification, evaluation, and interpretation of structural equation models," *J. Acad. Marketing Sci.*, vol. 40, no. 1, pp. 8–34, Jan. 2012.

[33] W. W. Chin, "Commentary: Issues and opinion on structural equation modeling," *Manage. Inf. Syst. Quart.*, vol. 22, no. 1, pp. 7–16, 1998.

[34] E. Ferrer and J. J. McArdle, "Alternative structural models for multivariate longitudinal data analysis," *Struct. Equation Model.*, vol. 10, no. 4, pp. 493–524, 2003.

[35] T. Asparouhov, E. L. Hamaker, and B. Muthén, "Dynamic structural equation models," *Struct. Equation Modeling, A Multidisciplinary J.*, vol. 25, no. 3, pp. 359–388, May 2018.

[36] C. Ringle, S. D. Da, and D. Bido, "Structural equation modeling with the SmartPLS," *Brazilian J. Marketing*, vol. 12, no. 2, pp. 56–73, 2014.

[37] J. F. Hair, M. Sarstedt, C. M. Ringle, and J. A. Mena, "An assessment of the use of partial least squares structural equation modeling in marketing research," *J. Acad. Marketing Sci.*, vol. 40, no. 3, pp. 414–433, May 2012.

[38] B. M. Byrne, *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming*. Abingdon, U.K.: Routledge, 2016.

[39] M. Sarstedt and J.-H. Cheah, "Partial least squares structural equation modeling using SmartPLS: A software review," *J. Marketing Anal.*, vol. 7, no. 3, pp. 196–202, Sep. 2019.

[40] (2019). *World Bank, World Development Indicators*. Accessed: Jan. 2, 2020. [Online]. Available: https://databank.worldbank.org/source/world-development-indicators

[41] IBGE. (2019). *Unidades Locais e Pessoal Ocupado Por Atividade*. Accessed: Jan. 2, 2020. [Online]. Available: https://sidra.ibge.gov.br/tabela/988

[42] (2019). *World Bank, Enterprise Surveys*. Accessed: Jan. 2, 2020. [Online]. Available: http://www.enterprisesurveys.org

[43] UNESCO. (2019). *Institute for Statistics, Science, Technology & Innovation*. Accessed: Jan. 2, 2020. [Online]. Available: http://uis.unesco.org

[44] NSF. (2019). *Science and Engineering Indicators*. Accessed: Jan. 2, 2020. [Online]. Available: https://www.nsf.gov/statistics/

[45] IBGE. (2019). *Domicílios Particulares Permanentes (Existência de Microcomputador, Acesso à Internet e Tipo de Telefone)*. Accessed: Jan. 2, 2020. [Online]. Available: https://sidra.ibge.gov.br/tabela/2387

[46] ITU. (2019). *World Telecommunication/ICT Development Report and Database*. Accessed: Jan. 2, 2020. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx

[47] ANATEL. (2019). *Valores do Ìndice de Serviços de Telecomunicações (IST)*. Accessed: Jan. 2, 2020. [Online]. Available: https://anatel.gov.br/setorregulado/tarifas-e-precos/valores-do-indice-de-servicos-de-telecomunicacoes-ist

[48] World Bank. (2019). *Private Participation in Infrastructure Project Database*. Accessed: Jan. 2, 2020. [Online]. Available: http://ppi.worldbank.org

[49] IBGE. (2019). *Série Encadeada do Índice de Volume Trimestral da Atividade de Informação e Comunicação*. Accessed: Jan. 2, 2020. [Online]. Available: https://sidra.ibge.gov.br/tabela/1620

[50] United Nations. (2019). *Comtrade Database Through the WITS Platform*. Accessed: Jan. 2, 2020. [Online]. Available: https://comtrade.un.org

[51] ITU. (2018). *Measuring the Information Society Report 2017*. ITU, Geneva, Switzerland. Accessed: Jan. 2, 2020. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2018/MISR-2018-Vol-1-E.pdf

[52] V. E. Vinzi, W. W. Chin, J. Henseler, and H. Wang, *Handbook of Partial Least Squares: Concepts, Methods and Applications*. Berlin, Germany: Springer, 2010.

[53] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *J. Marketing Res.*, vol. 18, no. 1, pp. 39–50, Feb. 1981.

[54] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing.*, vol. 50, pp. 159–175, Jan. 2003.

[55] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting With Exponential Smoothing: The State Space Approach*. Berlin, Germany: Springer-Verlag, 2008.

MATHEUS PEREIRA LIBÓRIO received the B.S. degree in economic sciences and the B.S. degree in management from the College of Management Studies (FEAD), in 2011 and 2016, respectively, the P.B. degree in geoprocessing from the Pontifical Catholic University of Minas Gerais, Brazil, in 2013, the P.B. degree in strategic management and competitive intelligence from the University of Araquara, in 2014, and the M.Sc. degree in geography from the Pontifical Catholic University of Minas Gerais, in 2016, where he is currently pursuing the Ph.D. degree in management. His research interests include decision-making models and methods in economics, business management, and geography.

PETR YA EKEL received the Ph.D. degree from the Kiev Polytechnic Institute, National Technical University of Ukraine, and the D.Sc. (Habilitation) degree from the Institute of Electrodynamics, Ukrainian Academy of Sciences. Since 2004, he has been the President and a Principal Consultant with Advanced System Optimization Technologies. He is currently a Full Professor with the Pontifical Catholic University of Minas Gerais, Belo Horizonte, Brazil. He is also with the Federal University of Minas Gerais. He has published numerous articles and two research monographs related to these topics. His main research interests include operational research, decision making, computational intelligence, involve problems of modeling and optimization under different levels of uncertainty, and multicriteria decision making, including decision making in a fuzzy environment, fuzzy identification and control, processing of heterogeneous information, and risk evaluation and management. He was a member of numerous program and an advisory committees of international conferences in system modeling, optimization, and control and decision making. He is a member of the Ukrainian Academy of Engineering Sciences. He serves as an Editorial Board Member for numerous international journals, including *Information Sciences*, *Information Fusion*, and *Group Decision and Negotiation*.

PATRÍCIA BERNARDES received the M.Sc. degree in economics, in 1993, and the Ph.D. degree in management from the Graduate Center for Research in Management, Faculty of Economic Sciences, Federal University of Minas Gerais. She is currently an Adjunct Professor with the Graduate Program in Management, Pontifical Catholic University of Minas Gerais. She is an Ad Hoc Servant with the INEP/National Institute of Educational Studies and Research. She is also an Economist. She is a member of management councils, academics, and curators. She is also a member of BASIS-Bank of Evaluators of the National Higher Education Assessment System. She was a recipient of the Minas de Economia Prize, being honored with the medals of Inconfidência and Santos Dumont.

GUSTAVO LUÍS SOARES received the degree in electrical engineering and the M.Sc. degree from the Federal University of Minas Gerais (UFMG), in 1995 and 1997, respectively, and the Ph.D. degree from UFMG and the University of Brest, Brest, France, in 2008. He is currently an Adjunct Professor with the Pontifical Catholic University of Minas Gerais. His activities include consulting and project design in the field of electrical engineering for buildings and small industries. His main research interests include neural networks and hybrid algorithms using interval analysis, robust optimization, and multicriteria evolutionary computation.

RENATA DE MELLO LYRIO received the B.S. degree in management from the University Center, UNA, in 2015. She is currently pursuing the M.Sc. degree with the Knowledge Organization and Management Graduate Program, Federal University of Minas Gerais (UFMG). She is also a Competitive Intelligence Analyst with Federation of Industries of the State of Minas Gerais (FIEMG).

THIAGO MELO MACHADO-COELHO received the B.Sc. degree in computer science and the M.Sc. degree in electrical engineering from the Pontifical Catholic University of Minas Gerais, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with the Federal University of Minas Gerais. His research interests include optimization, evolutionary computation, computational intelligence, machine learning, and interval analysis.

• • •