

FERNANDO FERNANDES NETO

**ANÁLISE E APLICAÇÃO DE MODELOS AUTO-REGRESSIVOS PARA PREVISÃO DE  
DEMANDA EM UM SERVIÇO PÚBLICO CONCESSIONADO**

São Paulo

2013

FERNANDO FERNANDES NETO

**ANÁLISE E APLICAÇÃO DE MODELOS AUTO-REGRESSIVOS PARA PREVISÃO DE  
DEMANDA EM UM SERVIÇO PÚBLICO CONCESSIONADO**

v.1

São Paulo

2013


FERNANDO FERNANDES NETO

**ANÁLISE E APLICAÇÃO DE MODELOS AUTO-REGRESSIVOS PARA PREVISÃO DE  
DEMANDA EM UM SERVIÇO PÚBLICO CONCESSIONADO**

Monografia apresentada à Escola Politécnica da  
Universidade de São Paulo, para obtenção do  
título de MBA em Engenharia Financeira.

Área de Concentração:

Engenharia Financeira

Orientador: Prof. Dr. Claudio  GARCIA

São Paulo

2013



**UNIVERSIDADE DE SÃO PAULO**

**Reitoria:** Prof. Dr. João Grandino Rodas

**Vice-Reitor:**



**ESCOLA POLITÉCNICA**

**Diretor:** Prof. Dr. José Roberto Cardoso

**Vice-Diretor:** Prof. Dr. José Roberto Castilho Piqueira

**COMISSÃO DE PÓS-GRADUAÇÃO**

**Presidente:** Prof<sup>a</sup>. Dr. Fernando José Barbin Laurindo

**DIVISÃO DE BIBLIOTECA**

**Diretora Técnica:** Maria Cristina Olaio Villela

**Editoração:**

Fernando Fernandes Neto

## **DEDICATÓRIA**

Dedico este trabalho à minha família pelo carinho e apoio de sempre, pois eu nada seria sem ela.

## **AGRADECIMENTOS**

Ao professor Dr. Claudio Garcia, pelas orientações, pela minúcia, pelo cuidado e esmero, fatores estes que possibilitaram a realização de um trabalho de qualidade.

## RESUMO

Este trabalho tem como objetivo a discussão, comparação, verificação e aplicação de diferentes tipos de modelos auto-regressivos para a previsão da demanda de serviços públicos concessionados, tendo grande importância tanto na engenharia financeira de novos projetos, quanto na gestão do fluxo de caixa dos mesmos.

É feita uma ampla revisão bibliográfica sobre as técnicas de métodos de previsão de séries temporais, e todas as soluções são feitas sobre programas de computador cujo código fonte é aberto, tornando possível a replicação dos mesmos por qualquer um sem maiores recursos.

São discutidos e calibrados modelos univariados (abordagem escalar) e multivariados (abordagem vetorial), onde os modelos que tiveram melhor desempenho no *backtesting* foram os segundos.

Foram obtidos resultados satisfatórios em termos de aderência e erros, onde o erro médio absoluto dos melhores modelos foi de aproximadamente três por cento, para um intervalo de seis meses fora da amostra utilizada na inferência.

Por fim, os códigos fontes dos modelos estatísticos são fornecidos nos apêndices.

**Palavras-chave:** VAR; ARIMA; SARIMA; séries temporais; estabilidade; *backtesting*; demanda e rodovias.

## **ABSTRACT**

This study aims the discussion, comparison, checking and application of different autoregressive models for public services demand forecasting, keeping in mind the great importance of it in terms of the financial engineering of new projects, and of the cash flow management.

It is made in this work a wide and extensive bibliographic revision over the time series forecasting methods, and all solutions here provided are made using open and free software, making available the replication of them by anyone without any major resources.

Here are discussed and calibrated univariate (scalar approach) and multivariate (vectorial approach) models, where the best were the models of the second type.

The results obtained were satisfactory in terms of fitting and out-of-sample errors, achieving an absolute mean deviation of three percent for an interval of six months out-of-sample.

Finally, all source codes of the statistic models are made available in the appendices.

**Keywords:** VAR; ARIMA; SARIMA; time series; stability; backtesting; demand e toll-roads.



## Índice de Figuras

Figura 1 - Uma Típica Estrutura de Capital.....	10
Figura 2 – Consumo de Energia em kWh no Brasil.....	18
Figura 3 – Modelo de Séries Temporais para Explicação das Ocorrências e Previsão do Consumo no Próximo Período.....	19
Figura 4 – Função de Densidade de Probabilidade Normal e sua respectiva probabilidade acumulada (área escura abaixo da curva).....	21
Figura 5 – Comparação entre um processo estacionário versus um processo não estacionário.....	25
Figura 6 – Representação Gráfica do Sistema Anchieta-Imigrantes.....	37
Figura 7 – Ajuste do modelo SARIMA1.....	69
Figura 8 – Ajuste do modelo SARIMA2.....	70
Figura 9 – Ajuste do modelo VAR(1).....	71
Figura 10 – Ajuste do modelo VEC(1).....	72
Figura 11 – Autovalores do modelo VAR(1).....	74
Figura 12 – Autovalores do modelo VEC(1).....	75
Figura 13 – Descrição da Sobre Parametrização.....	84

## Glossário Técnico

SQR – Soma Quadrática dos Resíduos

AIC – *Akaike Information Criterion*, Critério de Informação de Akaike

SIC – *Schwartz Information Criterion*, Critério de Informação de Schwarz

BIC – *Bayesian Information Criterion*, Critério de Informação Bayesiano (este é outro nome para o critério de informação de Schwartz)

Matriz de Desenho – Matriz que quando multiplicada por um vetor, tal que esta multiplicação seja igualada a outro vetor, forma um sistema linear de equações

WACC – *Weighted Average Cost of Capital*, Custo Médio Ponderado de Capital, que é a média ponderada entre capital de terceiros e capital próprio (dos acionistas)

VAR – *Vector Auto Regression*, Vetor Auto Regressivo, modelo multivariado vetorial auto-regressivo

VEC – *Vector Error Correction*, Correção de Erros Vetoriais, modelo multivariado vetorial auto-regressivo com termo de correção de erros decorrente de relações de co-integração.

## **Sumário**

<b>1. Introdução .....</b>	<b>1</b>
<b>2. Referencial Teórico .....</b>	<b>4</b>
2.1. Revisão Bibliográfica .....	4
2.2. Histórico e motivações para previsões .....	6
2.3. A atividade empresarial, o fluxo de caixa e a geração de valor ao proprietário .....	7
2.4. A natureza das receitas operacionais.....	11
2.5. Estimando a receita bruta operacional .....	15
2.5.1. Modelos Econométricos Vs Modelos de Séries Temporais .....	15
2.5.2. Séries temporais.....	16
2.5.3. Modelos Estocásticos.....	17
2.5.4. Conceitos Fundamentais em Análise de Séries Temporais.....	20
2.5.4.1. Função de Distribuição de Densidade de Probabilidades .....	20
2.5.4.2. Resíduo.....	21
2.5.4.3. Regressão Linear .....	21
2.5.4.4. Medidas de Variação, Dispersão e Dependência.....	22
2.5.4.5. Estacionariedade.....	24
2.5.4.6. Critério de Informação de Schwarz .....	26
2.5.4.7. Os principais tipos de modelos de série temporal auto- regressivos .....	27
2.5.4.8. Modelos ARMA .....	28
2.5.4.9. Modelos Vetoriais Auto Regressivos (VAR).....	29
2.5.4.10. Modelos Vetoriais de Correção de Erros (VEC) .....	30
2.5.4.11. Análise de Estabilidade e Representações como VAR(1) .....	32
2.6. Conclusão .....	35
<b>3. Metodologia.....</b>	<b>36</b>
3.1. Apresentação da Empresa .....	36
3.2. Obtenção dos Dados .....	39
3.3. Transformações Matemáticas nos Dados.....	39
3.4. Softwares Empregados na Modelagem.....	40

3.5. Critérios estatísticos para escolha da estrutura ótima dos modelos.....	41
3.6. Critérios estatísticos para escolha do melhor modelo .....	44
3.7. Conclusão .....	45
4. Resultados.....	46
4.1. Modelos ARIMA Sazonais .....	47
4.2. Estratégia de Identificação de modelos ARIMA Sazonais.....	48
4.3. Identificação de Estrutura de Retardos de Modelos VAR.....	52
4.4. Comparativo de Desempenho Empírico .....	65
4.5. Análise de Estabilidade.....	71
4.6. Conclusão .....	75
5. Conclusão do Trabalho .....	76
Referências .....	79
Apêndice A – Derivação do Método dos Mínimos Quadrados .....	82
Apêndice B – Código Fonte (R) – Modelos Univariados .....	86
Apêndice C – Código Fonte (GRET) – Modelo VAR(1) .....	88
Apêndice D – Código Fonte (GRET) – Modelo VEC(1).....	89

## 1. Introdução

Um dos problemas mais recorrentes na gestão de serviços concessionados e na regulação dos mesmos é o alcance do retorno prospectado em *business plan*, em conformidade com a proposta de remunerar o proprietário/acionistas.

Considerando as informações disponíveis a partir dos sistemas contábeis e financeiros, uma das maneiras efetivas para o alcance do retorno prospectado é o controle eficiente de custos variáveis em função da receita, tal que o lucro basicamente consiste na diferença entre ambos, descontando outras despesas operacionais, juros e impostos.

Historicamente, na gestão de negócios, os gestores e gerentes junto à operação não dispõem de ferramentas adequadas para a projeção da receita, seja em razão do seu leque de conhecimentos, seja em razão dos sistemas de informação a eles disponíveis.

Sendo assim, propõe-se uma ferramenta simples de previsibilidade da receita em função dos dados históricos, facilmente coletados a partir das informações contábeis da empresa, que podem ser processados a partir de ferramentas amplamente disponíveis, como o Microsoft Excel, gerar cenário e, por conseguinte, permitir aos gestores o controle efetivo dos custos variáveis e remunerar o proprietário em consonância com as prospecções do investimento.

Assim, basicamente o trabalho visa responder a seguinte pergunta: “Como gerar valor junto ao proprietário de maneira a reduzir os riscos, aumentar o controle e verificar a rentabilidade do negócio dentro do setor de concessão de rodovias?”.

Esta pergunta se torna bastante pertinente nos dias de hoje, pois com a crescente abertura de capital das grandes empresas deste setor, o

acionista/gestor de fundos cada vez se torna mais exigente em relação ao atingimento/cumprimento de metas; tem uma aversão cada vez maior a riscos, os quais ele não consegue compreender – reflexo de uma profissionalização crescente da indústria de fundos; assim como capaz de relacionar e avaliar melhor seus portfólios, dada a crescente disponibilização de informações e de demais variáveis, apesar das limitações cognitivas e da assimetria de informações, às quais eles também estão sujeitos.

Ao longo deste trabalho é narrada a experiência obtida durante o processo de elaboração de métodos de previsão em séries temporais, aplicados em novos investimentos feitos dentro da ARTESP (Agência Reguladora de Transportes do Estado de São Paulo), junto às empresas controladoras dos serviços concessionados de transportes no Estado de São Paulo e também, dos pressupostos teóricos e técnicos que asseguram a assertividade da previsão em conformidade com observações reais, conhecimentos estes obtidos ao longo do curso de Engenharia Financeira, principalmente nas disciplinas de Estatística, Análise de Séries Temporais e Sistemas Dinâmicos, tal que objetiva-se demonstrar os princípios matemáticos, econômicos e financeiros aplicados, assim como os pressupostos para suas respectivas aplicações.

De maneira a apresentar o trabalho de forma sistemática, em suma tem-se:

No **Capítulo 2**, tem-se a introdução ao referencial teórico e como as diferentes áreas de Séries Temporais, Sistemas Dinâmicos e Estatística dialogam entre si.

Já no **Capítulo 3**, tem-se:

Apresentação de um caso prático, onde é feita uma breve apresentação da empresa envolvida, dos processos empregados nas

estimações, explicitando as diversas conexões existentes entre a teoria absorvida durante o curso de pós-graduação.

Ainda no Capítulo 3 está a Metodologia, em que são abordadas as técnicas utilizadas para a obtenção dos dados, assim como seu respectivo tratamento para a produção de informação, tanto para a construção da solução proposta, assim como para o estudo da identificação do problema e viabilidade da mesma.

No **Capítulo 4** são abordados os resultados obtidos após a aplicação dos métodos e a comparação entre eles.

Por conseguinte, na **Conclusão** são abordados os possíveis impactos da proposta de trabalho, assim como comentários em relação à informação gerada pelos modelos propostos em possíveis trabalhos futuros.

Ao longo dos **Apêndices**, busca-se demonstrar alguns dados e seus respectivos tratamentos, assim como alguns pressupostos especiais, os quais são enunciados ao longo do Capítulo 3 e que não são comumente ensinados ao longo do curso de graduação em Administração de Empresas, mas que, porém fornecem informações e conhecimentos substanciais a respeito de sua natureza, sendo desenvolvidos objetivando a transmissão do conhecimento com maior profundidade, mas que não são necessários para o entendimento deste trabalho, assim como sua respectiva aplicação.

## **2. Referencial Teórico**

As áreas do conhecimento abrangidas neste trabalho são:

- I. Matemática e Estatística (cálculo, sistemas dinâmicos, inferência estatística, estatística descritiva, análise de séries temporais);
- II. Finanças (índices econômicos financeiros e criação de valor, matemática financeira);
- III. Economia (Teoria Econômica dos Ciclos de Negócio, efeito acelerador e multiplicador); e
- IV. Gestão de custos e de suprimentos.

### **2.1. Revisão Bibliográfica**

Os autores escolhidos são contemporâneos, de diversas áreas do conhecimento, que têm visões suplementares a respeito da Engenharia Financeira, visto que ela é composta por um amplo conjunto de conhecimentos de diversas áreas.

De maneira a mostrar como os textos selecionados estão relacionados, e consequentemente como foram estruturados e sistematizados os conhecimentos aplicados na proposta aqui discorrida, são apresentados os textos de finanças e engenharia econômica.

Partindo das leituras de Brigham & Ehrhardt (2005), Matarazzo (2006) e Pilão & Hummel (2003), consegue-se claramente visualizar a importância dos cálculos do valor econômico do projeto, de maneira a remunerar o proprietário de acordo com o custo de oportunidade do capital empregado no empreendimento, tal que conforme será discorrido posteriormente, para efeito de simplicidade tanto de análise quanto de



gestão financeira, foram estimadas (levando em conta tais conceitos enunciados) as margens de retorno do empreendimento.

Também, destacando aspectos adjacentes à gestão financeira de empresas, Brigham & Ehrhardt (2005) destacam a importância de ferramentas de gestão de custo e a importância de políticas de gestão destes custos, decorrentes da gestão operacional e conseqüentemente, dos insumos envolvidos no processo, cuja empresa detém o conhecimento e a partir do qual gera valor, políticas estas esmiuçadas por Chopra & Meindl (2003), que não só discorrem acerca destas políticas de gestão, mas expõem um importante ferramental quantitativo, que objetiva a tomada de decisão.

Todos esses conhecimentos podem ser, em graus diferentes, aplicados ou observados durante o processo de formação das propostas vencedoras dos certames, disponíveis no sítio da ARTESP, anexadas nos contratos celebrados entre o Governo do Estado de São Paulo e as concessionárias. No entanto, a fim de estender a aplicação deles (especialmente dos modelos quantitativos expostos por Brigham & Ehrhardt (2005)) sob a forma da proposta de melhoria, foram buscados autores contemporâneos de outras áreas do conhecimento, como a Análise de Séries Temporais e a Estatística, tal que as premissas dos primeiros se apoiam sobre estas.

Neste contexto, a escolha de modelos Auto-Regressivos, dentro do campo da Análise de Séries Temporais, sofre grande influência da teoria econômica, principalmente a partir da análise de textos de métodos de econometria/estatística empíricos, quando contrapostos à luz dos pressupostos matemáticos-econômicos dos principais modelos teóricos expostos por Blanchard (2011) e Gujarati *et. al*(2011).

A partir deste ponto, Chopra & Meindl (2003) “dialogam” com Morettin & Toloí (2006), tal que os últimos explicam conceitualmente as premissas

de obtenção das ideias gerais de modelagem dos primeiros (tal que como será visto mais adiante, a razão para a adoção do modelo proposto pelos primeiros se dá pela facilidade de implementação e não pura e simplesmente pela natureza dos dados históricos).

Por fim, para pressupor a validade do modelo explicado por Morettin & Tolo (2006), partiu-se de conceitos estatísticos fundamentais, estes percorridos por Box & Jenkins (1976), Hamilton (1994) e Greene (2002), tal que se obtém um modelo de séries temporais, que não tem um distanciamento muito grande dos pressupostos econométricos.

## **2.2. Histórico e motivações para previsões**

Uma das grandes vontades do homem ao longo da história foi o desejo de prever o futuro, de maneira a antecipar os acontecimentos e, por conseguinte, mudar o curso dos acontecimentos a seu favor, com o intuito de ser bem sucedido no desempenho e exercício de suas atividades e relações ao longo de sua vida, eliminando os riscos e incertezas às quais está submetido.

Sendo assim, conforme exposto por Heer (1968), diversas ciências foram desenvolvidas com o intuito de estudar os fenômenos e antever os acontecimentos, desde pressupostos que diversas antigas civilizações (e por incrível que pareça, ainda um substancial número de pessoas nos dias de hoje) acreditavam reger a vida das pessoas, como a astrologia, até complexas ciências que demonstram enorme eficácia em descrever fenômenos naturais, como os ramos da Física, que nos possibilitaram, em grande parte, entender os acontecimentos e fazer previsões com certa segurança. Foi neste sentido que se desenvolveram diversos conhecimentos, como a Matemática e seus ramos, que possibilitaram não somente prever com acurácia diversos fenômenos naturais, que posteriormente nos permitiram a transformação do mundo ao nosso

redor, mas que também forneceram uma considerável coleção de conhecimentos, que habilita prever acontecimentos de natureza social, de forma a permitir o aumento de chances de sucesso no desempenho de nossas atividades e relações em uma sociedade complexa.

Assim se desenvolveram diversos ramos do conhecimento como Finanças, Sociologia, Economia, entre outros, que não somente se apoiam em pressupostos inerentes às suas naturezas, mas também encontram significativo apoio na matemática, para o delineamento e quantificação das grandezas envolvidas na predição de cenários futuros.

No caso deste trabalho em especial, é apresentada uma concessão de rodovias (podendo este trabalho ser estendido a outras empresas que estejam em conformidade com os pressupostos aqui estabelecidos), que está submetida a diversos riscos, sejam eles de natureza operacional ou financeiro, e se discorrer acerca de um conjunto de conhecimentos e fundamentações teóricas para o desenvolvimento de sistemas de predição, de maneira a auxiliar os gestores de negócios a antever possíveis cenários futuros e criar alternativas mitigatórias, avaliando o risco e as consequências que as variações nas receitas possam gerar, objetivando o aumento de valor da gestão e o alcance das metas pré-estabelecidas na prospecção do negócio, conforme proposto por Stulz (2003) e Brigham & Ehrhardt (2005).

### **2.3. A atividade empresarial, o fluxo de caixa e a geração de valor ao proprietário**

O que motiva a atividade empresarial é a adição de valor e a consequente remuneração ao proprietário/acionista, decorrente da propriedade do meio de produção (e/ou do capital empregado na produção).

Não obstante, a geração de fluxo de caixa e a rentabilidade esperada junto ao proprietário têm suma importância, tanto na lei de contratações públicas quanto na lei de concessões, respectivamente lei 8.666/93 e lei 8.797/95, que balizam a relação entre privado e Estado.

Dada a complexidade das atividades e relações entre organizações no mundo moderno, há uma considerável diversidade na natureza dos custos, despesas e receitas inerentes às atividades empresariais e, conseqüentemente, há um aumento considerável na necessidade de processamento das informações e de tomadas de decisão inerentes a estas entradas e saídas de caixa.

Assim, segundo Brigham & Ehrhardt (2005), dada a capacidade cognitiva limitada dos gestores e suas mais diversas formações profissionais, ferramentas que os auxiliam no processamento de informações e na tomada de decisão aumentam a eficiência da gestão da riqueza do proprietário, de maneira a considerar os riscos externos aos quais a empresa está submetida e o volume de dados para geração de informação gerencial.

Neste sentido, as decisões tomadas não somente impactam na geração de valor presente, mas também no posicionamento futuro da empresa (visto que muitas decisões de investimento geram uma “inércia estratégica”, como colocado por Fredrickson e Laquinto (1989)).

Também vale destacar que segundo Froot, Scharfstein & Stein (1994), a maneira mais barata de financiamento das operações de uma empresa é a própria geração de fluxo de caixa, tal que dada uma estrutura de capital mais flexível, há uma significativa diminuição de alguns riscos aos quais a empresa está submetida, como riscos financeiros e de crédito, em concordância com a diminuição do que Fredrickson e Laquinto (1989) colocam como “inércia estratégica”.

No entanto, esta visão é contraposta em parte por Brigham & Ehrhardt (2005), pois os autores defendem que as empresas devem ter um nível de financiamento ótimo em sua estrutura de capital, objetivando a redução do custo de capital, de maneira a aumentar o valor da empresa.

Este ponto pode melhor ser elucidado através de métodos de engenharia econômica, como aqueles demonstrados por Pilão & Hummel (2003), que é possível claramente visualizar a importância dos cálculos do valor econômico dos projetos, de maneira a remunerar o proprietário de acordo com o custo de oportunidade do capital empregado no empreendimento, este refletido pelo custo médio ponderado do capital.

Para melhor ilustrar as proposições acima, é possível considerar a seguinte formulação para o custo médio ponderado do capital (WACC – “Weighted Average Cost of Capital”) (Brigham & Ehrhardt (2005)).

$$WACC = (1 - T) \cdot W_d \cdot D + W_e \cdot E$$

Em que:

$E$  – “Equity”, ou seja, o custo do capital próprio, determinado pelo acionista, que pode ser determinado pelo modelo CAPM (Capital Asset Pricing Model).

$W_e$  - Proporção do “Equity” (Patrimônio Líquido) no passivo da empresa

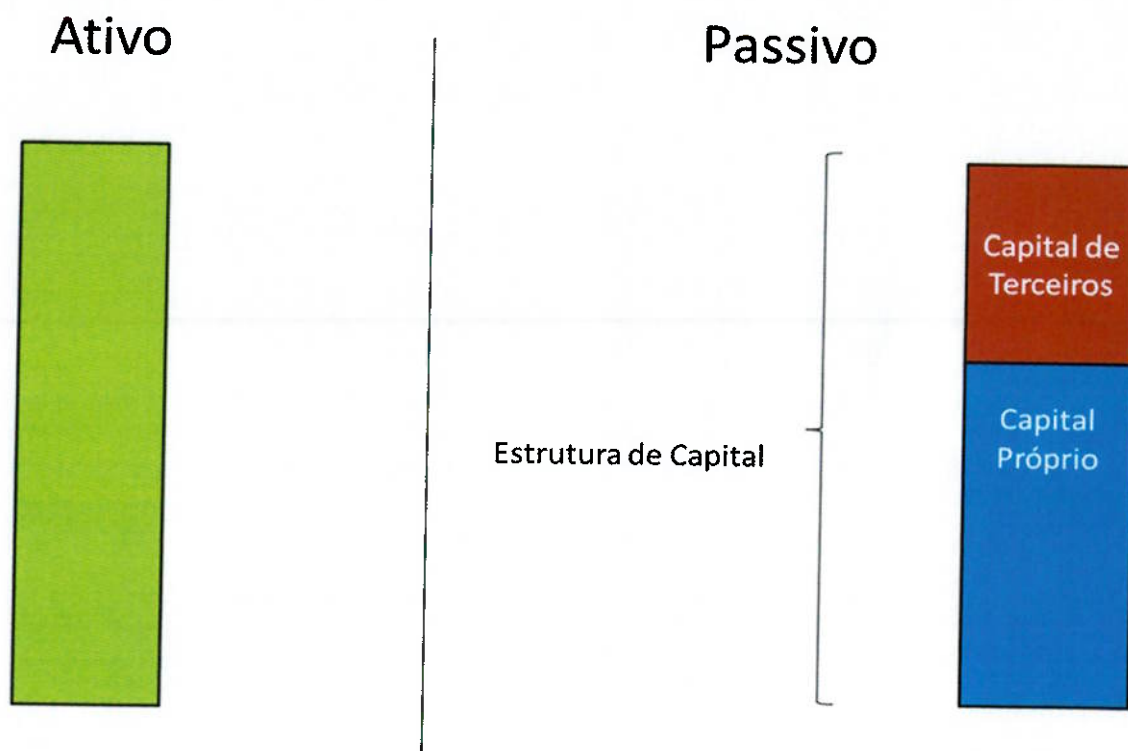
$W_d$  - Proporção do “Debt” (Capital de Terceiros) no passivo da empresa

$D$  – “Debt”, ou seja, o custo do capital de terceiros, que para a empresa é o custo financeiro médio para dispor deste capital

$T$  – “Taxes”, ou seja, a alíquota de imposto de renda a qual a empresa está submetida, uma vez que as despesas financeiras geram

impacto na apuração do lucro e consequentes desencaixes de natureza de imposto.

Assim, a estrutura de capital (passivo) pode ser apresentada conforme da Figura 1.



**Figura 1 - Uma Típica Estrutura de Capital.**

Fonte: Elaborado pelo autor com base em Brigham & Ehrhardt (2005)

Pode-se ver que a partir da Figura 1 e da expressão do “WACC”, há um nível de endividamento ótimo, a fim de melhor gerar valor ao proprietário, que é o que Brigham & Ehrhardt (2005) expõem.

Porém, também é possível ver que, dada a natureza da atividade empresarial, a existência dos sistemas de controle de pagamentos de crédito, e em consonância com o que é estabelecido no código civil, as execuções das obrigações com terceiros, têm prioridade maior do que aquelas com os proprietários, tal que isso gera uma inflexibilidade na adaptabilidade da empresa frente a mudanças no contexto sócio-

econômico no qual está inserida, compondo parte do que Fredrickson e Laquinto (1989) colocam como “inércia estratégica”.

Mesmo assim, depois de tais considerações, pode-se destacar que é de senso comum o fato de que os proprietários de empresas, dentro do ambiente econômico do Brasil em especial, possuem grande aversão ao uso de capitais de terceiros em sua estrutura de capital, em decorrência de cenários de instabilidade política e econômica, que acabam por elevar os riscos financeiros e de crédito.

Assim, em todas as direções temporais possíveis, a gestão de fluxo de caixa operacional é uma das principais variáveis para a gestão de valor da atividade empresarial e, conforme mencionado acima, tem sua importância aumentada quando se trata de micro e pequenas empresas no ambiente brasileiro.

Por conseguinte, saber adequar a empresa frente às demandas e desafios que surgem no decorrer de suas operações é de fundamental importância. E como destacam Chopra & Meindl (2003), um bom planejamento começa pela gestão dos custos a partir do ingresso de recursos operacionais (receitas).

## **2.4. A natureza das receitas operacionais**

Segundo Matarazzo:

A Receita Operacional Bruta da empresa é constituída pelo valor bruto faturado. O faturamento representa o ingresso bruto de recursos externos provenientes das operações normais de venda a prazo ou à vista, no mercado nacional e exterior, de produtos, mercadorias ou serviços (MATARAZZO, 2006, p. 67).

Por conseguinte, tendo em vista que naturalmente as receitas operacionais são um dos elementos principais da gestão do fluxo de

caixa, tal que é a partir delas que são abatidos os custos, despesas e obrigações inerentes à atividade empresarial, estabelece-se neste trabalho estas (as receitas operacionais brutas) como variável-objeto para a construção de um modelo de predição, que permite a geração de informação gerencial, que auxilie na gestão do fluxo de caixa da empresa, como exposto nos tópicos anteriores.

Conforme o excerto destacado anteriormente, é possível fazer uma ponte entre o enunciado ingresso de recursos externos à empresa e o que leva os agentes econômicos a efetuarem as transações junto à companhia (no caso o uso de um serviço público), assim como a natureza das escolhas e preferências de consumo (uso).

No caso de concessões de rodovias, o desempenho macroeconômico em nível nacional e global tende a influenciar as demandas pela utilização do serviço público (transporte sobre a rodovia pública), uma vez que se há um aumento no Produto Interno Bruto (PIB), pela ótica da demanda, haverá tanto uma maior demanda por materiais a fim de serem transformados em um produto acabado, assim como de equipamentos necessários aos respectivos investimentos, que capacitam o atendimento de novas demandas.

Portanto, é possível vislumbrar um simples e hipotético aumento, tanto no fluxo de bens de capital (da indústria para o campo) quando no de matérias primas (do campo para a indústria).

Por outro lado, como também é sabido, há uma natural defasagem temporal entre a apuração do PIB e o que vem sendo consumido e investido no mesmo instante de tempo. Tal defasagem implica no grande desafio da criação de indicadores precedentes do desempenho do produto, como o IBC-Br, criado pelo Banco Central do Brasil, ou alguns outros como em Notini (2009).



Sabendo que as rodovias compõem aproximadamente 58% do transporte de cargas no Brasil, conforme Lopes *et. Al* (2008), que em larga medida se traduzem em mercadorias (consumo) e equipamentos (investimento), é natural (e será base deste trabalho) a investigação de modelos de séries temporais vetoriais, onde uma série é uma candidata natural a explicar o comportamento das demais séries que compõem o vetor.

Portanto, tendo em vista os argumentos supramencionados, se torna interessante a tentativa de modelagem da predição da receita bruta através de métodos de série temporal (como visto mais adiante neste capítulo).

Para tal, espera-se que as escolhas dos agentes não sejam feitas de maneira aleatória, mas sim dentro de um contexto macroeconômico, que tenda a seguir uma evolução de acordo com os estados da economia imediatamente anteriores (a série não é um “ruído branco” e possivelmente deve apresentar auto-correlação); e tampouco tendam a sofrer mudanças grandes ao longo do tempo (estabilidade dos parâmetros).

Sobre a primeira característica do parágrafo anterior, a auto-correlação, ela deve ser esperada basicamente em função da dinâmica evolutiva econômica. Se hipoteticamente, para efeito de exemplificação, considera-se o modelo acelerador-multiplicador para o PIB de Samuelson (1939), que consiste na seguinte relação de identidade:

$$Y_t = C_t + G_t + I_t$$

(Produto Interno Bruto é igual à soma de tudo o que foi gasto, investido e consumido no país, de maneira simplificada), em que  $Y_t$  denota o PIB,  $C_t$  o consumo,  $G_t$  os gastos do governo e  $I_t$  os investimentos, todos em um mesmo instante de tempo. Assumem-se os seguintes pressupostos:

$$G_t = G_0$$

(Gastos do governo constantes)

$$C_t = C_0 + c \cdot Y_{t-1}$$

(Consumo atual é função de um nível base de consumo e da renda [produto sob a ótica da renda gerada] obtida no período anterior)

$$I_t = I_0 + i \cdot (C_t - C_{t-1}) \rightarrow I_0 + i \cdot (c \cdot Y_{t-1} - c \cdot Y_{t-2})$$

(Investimento atual é função de um nível base de investimentos e da propensão do empresariado em expandir seus investimentos em função de um aumento no consumo)

Obtém-se a seguinte expressão com base nos pressupostos e na relação de identidade macroeconômica:

$$Y_t = G_0 + C_0 + c \cdot Y_{t-1} + i \cdot (c \cdot Y_{t-1} - c \cdot Y_{t-2})$$

Por conseguinte, com um simples argumento de natureza econômica, é natural se esperar que o comportamento do produto apresente autocorrelação, visto que a relação acima pode ser vista como um processo Auto Regressivo de ordem dois (AR(2)), conforme Box & Jenkins (1976).

Dependendo do embasamento teórico subjacente, é possível verificar que outros processos Auto-Regressivos de ordem superior podem ser facilmente obtidos, e que não obstante, por conta desta natureza Auto-Regressiva encontrada nas mais diversas fundamentações teóricas, os modelos Auto-Regressivos (AR) e modelos Auto-Regressivos Vetoriais (VAR) são tidos como as principais ferramentas empíricas na economia, conforme Greene (2002), Gujarati *et al* (2011) e Morettin (2011).

Outros fatores como inflação também podem ser facilmente contornados, seja através da deflação da série do PIB pelo deflator oficial (IPCA), seja através da modelagem das receitas através da previsão da

quantidade de veículos (ou itens vendidos) e posteriormente multiplicados por um preço, cuja base monetária se encontre no mesmo período de tempo que a base do deflator do PIB.

## **2.5. Estimando a receita bruta operacional**

Neste trabalho, dado o conhecimento desenvolvido ao longo do curso, tem-se como foco principal a estimação da receita bruta operacional, através do conjunto de ferramentas conhecido como “Análise de Séries Temporais”.

Para melhor entender o que é uma série temporal, são expostos os seguintes conceitos.

### **2.5.1. Modelos Econométricos Vs Modelos de Séries Temporais**

De acordo com Morettin & Tolói:

No primeiro [econométrico], o analista se baseia fortemente na teoria econômica para construir um modelo, incluindo muitas variáveis, enquanto que no segundo [séries temporais], não há esta limitação, dado que o estatístico deixa “os dados falarem por si” para construir seu modelo, estando preparado para usar um modelo que não se harmonize com a teoria econômica, desde que produza melhores previsões (MORETTIN & TOLÓI, 2006, p. 7).

Assim, embora tenham sido discutidas nesse capítulo algumas premissas de natureza econômica (em especial os pressupostos acerca da natureza da escolha dos agentes econômicos, e o que isso acarreta na escolha de um modelo de séries temporais para a estimação da receita bruta), o modelo será predominantemente baseado em séries

temporais, embora se busque a manutenção da harmonia com a teoria, e não um modelo econométrico, embora possa ser considerado um modelo “híbrido”.

De maneira a reforçar as preocupações anteriores com os pressupostos econômicos assumidos e a natureza das séries temporais, vale ressaltar os trabalhos de Ashley e Granger (1979), que segundo Morettin & Tolói (2006, p. 7): “sugerem adotar um enfoque híbrido”, que basicamente reflete nossa preocupação, visto que, ainda segundo estes autores:

a conclusão de vários autores é que os modelos econométricos se ajustam melhor aos dados, enquanto que certos modelos, como aqueles da classe ARIMA fornecem melhores previsões. Bhattacharyya (1980) *apud* (MORETTIN & TOLÓI, 2006, p. 7).

## **2.5.2. Séries temporais**

Segundo Morettin & Tolói (2006, p. 1), “Uma série temporal é qualquer conjunto de observações ordenadas no tempo.” Ainda segundo os autores (p. 3), os objetivos de uma série temporal são:

- “(a) investigar o mecanismo gerador da série temporal ...”
- “(b) fazer previsões de valores futuros da série ...”
- “(c) descrever apenas o comportamento da série ...”
- “(d) procurar periodicidades relevantes nos dados ...”

No caso, as receitas brutas ao longo do tempo constituem as respectivas observações dispostas ao longo de uma linha temporal, compondo a série temporal propriamente dita, de maneira que o estudo em si da mesma visa propor uma solução ao problema identificado,

resultante da modelagem e da consequente previsão de valores futuros (extrapolação da série em direção ao futuro).

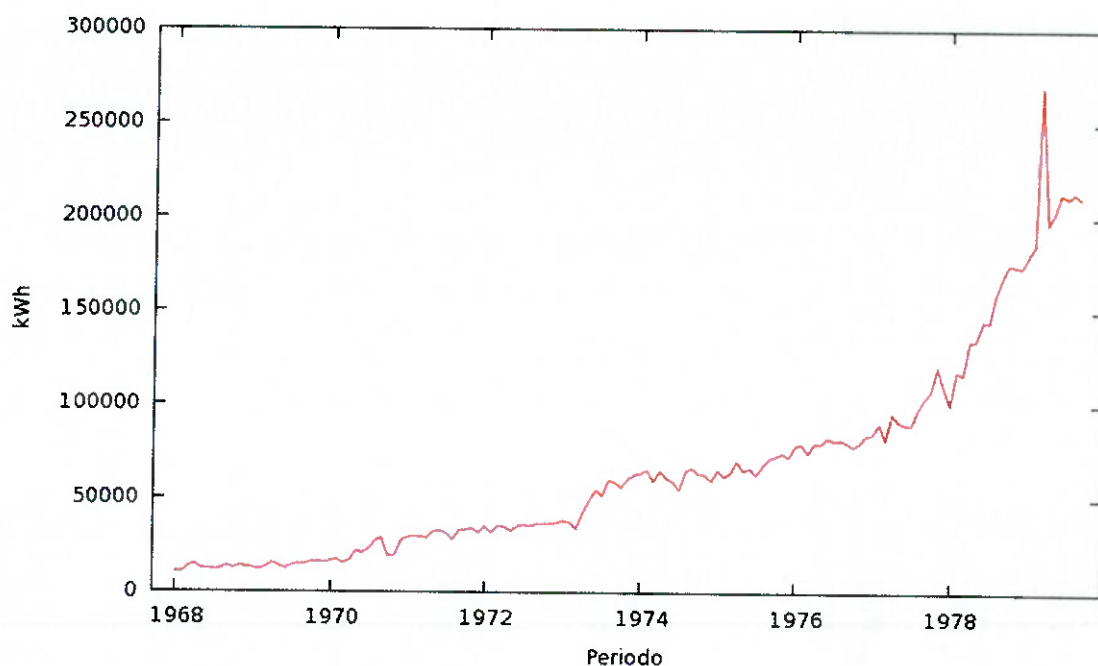
### **2.5.3. Modelos Estocásticos**

Morettin & Tolói (2006, p. 19) afirmaram que: “Os modelos utilizados para descrever séries temporais são estocásticos, isto é, processos controlados por leis probabilísticas”.

Logo, o modelo gerado que tenta explicar e servir como base para a previsão dos acontecimentos futuros leva em conta a distribuição dos erros, assim como a sua minimização (leis probabilísticas), independentemente do método.

No caso, as abordagens acerca das leis probabilísticas diferem basicamente em como melhor explicar as ocorrências e extrapolar o futuro, sejam estes processos auto-regressivos (que partem do pressuposto de que há um distanciamento do passado ótimo e uma respectiva extrapolação deste passado para melhor prever o próximo acontecimento), e de modelos de regressão normais, que partem de pressupostos acerca da homocedasticidade dos dados (no caso os erros de uma previsão são normalmente distribuídos e dispersos ao longo de funções de regressão) e da minimização da variância.

A fim de melhor ilustrar modelos de séries temporais e o que é uma série temporal, vide Figuras 2 e 3.

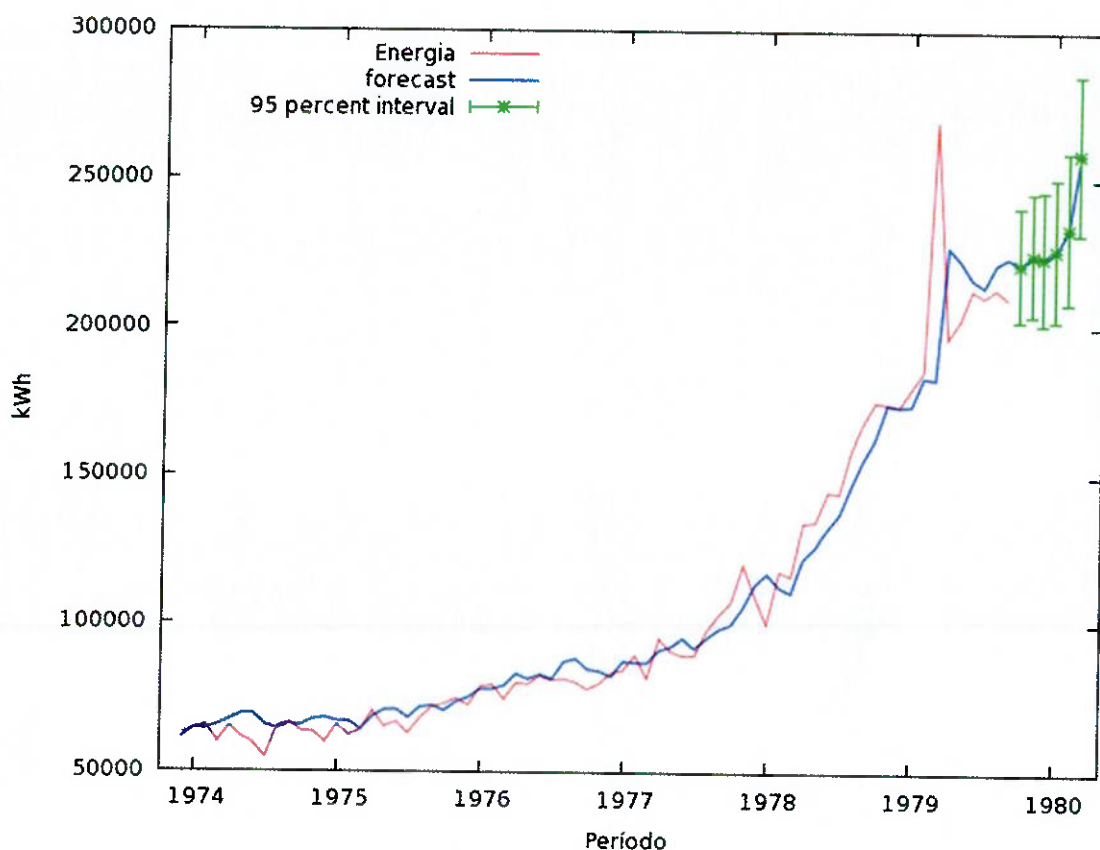


**Figura 2 - Consumo de Energia em kWh no Brasil**

Fonte: (<http://www.ime.usp.br/~pam/>) – conforme Morettin & Tolói (2006, p. 15)

Na Figura 2 pode-se ver a sequência de dados ao longo do tempo, compondo a série, conforme a definição. Os dados ilustram o consumo de energia ao longo de um intervalo de tempo. Assim, tenta-se elucidar que diversos tipos de dados, de diversas naturezas podem ser estudados conforme sua evolução ao longo do tempo, a fim de tentar prever as ocorrências futuras.

Esta série em especial foi escolhida como uma tentativa didática de explicitar situações comuns, do dia-a-dia, que estão dispostas sob a forma de uma série temporal. Vale ressaltar que a escolha desta série temporal foi feita de maneira arbitrária, pois ela apresenta uma menor complexidade de análise, se avaliada do ponto de vista da estacionariedade, tendência e dependência das observações anteriores. Tendo em vista a mencionada série temporal, a fim de se ilustrar como exemplo de um modelo de séries temporais gerado a partir da série original, foi utilizado um modelo ARIMA(1,1,1) (a ser explicado mais adiante), de acordo com a abordagem Box & Jenkins (1976) para ilustrar a previsão das ocorrências futuras.



**Figura 3 - Modelo de Séries Temporais para Explicação das Ocorrências e Previsão do Consumo no Próximo Período**

Fonte: Autor, gerado no programa R (Modelo ARIMA(1,1,1))

Note que a linha em vermelho explica com razoável acurácia os acontecimentos das ocorrências de consumo de energia (em azul), tal que é possível, a partir desta, a previsão do consumo de energia para o próximo período com razoável acurácia, o que permite, no caso para os agentes tomadores de decisão, como o governo, por exemplo, a elaboração de medidas de planejamento, assim como outras possíveis medidas mitigatórias em relação à administração dos recursos empregados na geração de energia.

## **2.5.4. Conceitos Fundamentais em Análise de Séries Temporais**

Nesta subseção são apresentados alguns conceitos estatísticos fundamentais acerca da Teoria de Análise de Séries Temporais, que são imprescindíveis para a compreensão e contextualização das principais características dos modelos construídos sob tal abordagem.

### **2.5.4.1. Função de Distribuição de Densidade de Probabilidades**

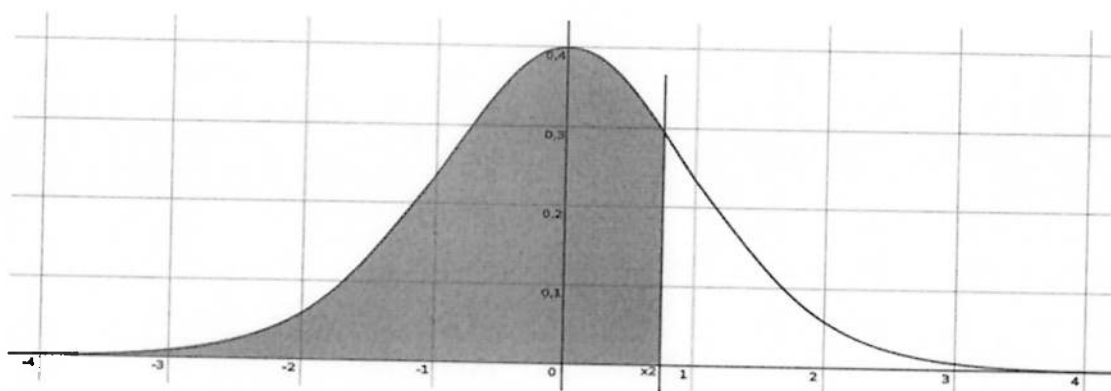
Em uma função de distribuição de probabilidades, ao se introduzir o valor de uma variável aleatória, ela retorna a sua respectiva probabilidade de ocorrência / frequência.

Exemplo: suponha que a altura dos indivíduos do sexo masculino seja normalmente distribuída, com média 1,70 metros e desvio padrão de 0,1 metros. Ao se colocar como entrada na função de probabilidade cumulativa<sup>1</sup> uma determinada altura arbitrariamente definida, após normalização (que é a transformação desta altura em uma variável normal com média 0 e desvio padrão um), tem-se como saída a frequência / probabilidade de se encontrar uma pessoa com altura menor ou igual àquela colocada como entrada.

---

<sup>1</sup> Que é a função da área abaixo da função de densidade de probabilidade, conforme Figura 4.





**Figura 4 – Função de Densidade de Probabilidade Normal e sua respectiva probabilidade acumulada (área escura abaixo da curva).**

Fonte: Google Imagens

#### 2.5.4.2. Resíduo

O resíduo é a diferença entre o quanto um modelo projetou para uma determinada observação e a própria observação.

#### 2.5.4.3. Regressão Linear

Uma regressão linear é a obtenção de uma equação que visa explicar uma variável em função de outras variáveis, através de relações lineares simples, do tipo:

$$y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \dots + \beta_n \cdot X_n$$

em que  $y$  é a variável que se deseja explicar;  $\beta_i$  são os pesos que o modelo atribue a cada variável explicativa e  $X_i$  são as próprias variáveis explicativas.

A obtenção dos coeficientes  $\beta_i$  deste modelo pode ser vista em Gujarati et al (2011), podendo ser feita através de diversos processos,

dos quais o mais tradicional e usual é o Método dos Mínimos Quadrados, conforme apêndice A.

#### 2.5.4.4. Medidas de Variação, Dispersão e Dependência

A **covariância** é uma medida estatística referente ao quanto duas variáveis mudam conjuntamente. Para tal, é empregada a seguinte fórmula:

$$Covar(x, y) = \frac{1}{N - 1} \cdot \sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

em que  $x_i$  e  $y_i$  são as  $i$ -ésimas observações das variáveis  $x$  e  $y$ , respectivamente,  $\bar{x}$  e  $\bar{y}$  são as médias amostrais das variáveis  $x$  e  $y$ , respectivamente e  $N$  é o número de observações utilizado.

Assim, em outras palavras, o que se calcula nesta medida estatística é:

- Se as maiores observações de  $x$  em relação à sua própria média ocorrerem ao mesmo tempo que as maiores observações de  $y$  em relação à sua respectiva média, tal medida será um número positivo.
- Se as menores observações de  $x$  em relação à sua própria média ocorrerem ao mesmo tempo que as maiores observações de  $y$  em relação à sua respectiva média – ou vice-versa – tal medida será um número negativo.

Caso se substitua  $y$  por  $x$ , tem-se a medida de **variância**, que é uma medida da média da distância quadrática de cada observação em relação à sua própria média amostral. E conseqüentemente, ao se

calcular a raiz quadrada da variância, tem-se uma medida estatística denominada “**desvio padrão**”, que visa colocar tal distanciamento proposto pela medida de variância na mesma unidade de grandeza das variáveis observadas.

Como a medida de variância é uma quantidade elevada ao quadrado, há uma alteração nas grandezas envolvidas. Por conseguinte, ao se calcular a raiz quadrada, tal distorção é corrigida, e se tem uma medida de dispersão na mesma unidade de grandeza da variável original.

Com isso, conhecendo-se os desvios padrão de cada uma das variáveis e conhecendo-se a covariância entre duas variáveis, é possível normalizar a covariância e obter uma medida chamada **correlação**, conforme a fórmula:

$$\rho(x, y) = \frac{\text{covar}(x, y)}{\text{desv. pad.}(x) \cdot \text{desv. pad.}(y)}$$

Assim obrigatoriamente se obterá uma medida estatística que varia entre -1 e 1, possibilitando a análise da força da dependência entre duas variáveis, em que:

-1 é uma relação de correlação negativa forte. Entende-se por correlação negativa o fato de que quando uma varia positivamente, a outra tende a variar negativamente;

0 é interpretado como independência entre as variáveis – ou seja, quando uma varia, nada se pode dizer a respeito da outra variável;

1 é uma relação de correlação positiva forte. Entende-se por correlação positiva o fato de que quando uma varia positivamente, a outra também variará positivamente.

Logo, é possível definir o conceito de **autocorrelação**, que é a medida de correlação aplicada a observações atrasadas da mesma série, de maneira a medir o quanto uma observação depende das observações imediatamente anteriores.

Exemplo: a autocorrelação de sétimo atraso de uma série de veículos equivalentes de uma rodovia de São Paulo indica o grau de dependência explicativo entre quantos carros passaram no mesmo dia da semana passada e hoje. Já a autocorrelação de primeiro atraso da mesma série indica o grau de dependência explicativo entre quantos carros passaram hoje e ontem.

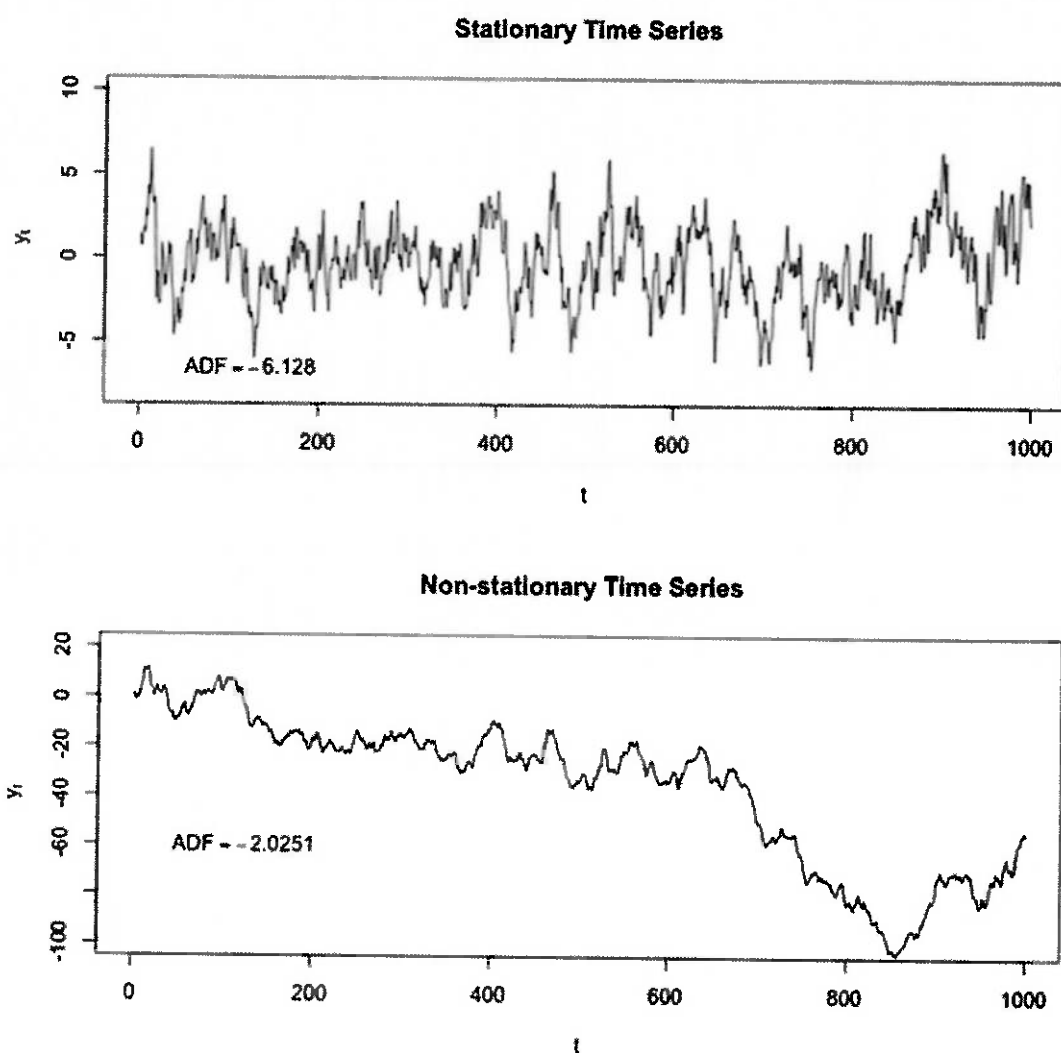
Por fim, é definido o conceito de **autocorrelação parcial**, que é a medida de dependência entre duas observações distantes temporalmente na série – como a autocorrelação – com a devida exclusão da interdependência entre as observações existentes entre elas.

Como pode ser visto em Morettin & Tolói (2004), Gujarati et. al (2011), Hamilton (1994) e Box & Jenkins (1976), tal medida pode ser calculada simplesmente por meio de uma regressão linear entre a variável e suas respectivas observações retardadas, em que o valor do coeficiente da observação mais atrasada representa o respectivo valor da autocorrelação parcial.

#### **2.5.4.5. Estacionariedade**

Um processo / série é dito estacionário quando a distribuição de probabilidades de ocorrência de diferentes valores é a mesma em qualquer ponto no tempo. Logo, se uma série, por exemplo, tem uma

tendência ou tem comportamento instável, a probabilidade de ocorrência dos valores muda ao longo do tempo e logo ela não é estacionária.



**Figura 5 –Comparação entre um processo estacionário versus um processo não estacionário**

Fonte: Google Imagens

Na Figura 5, a primeira série temporal é estacionária, pois a probabilidade de ocorrência de diferentes valores tende ser a mesma em diferentes pontos dela, já que a média e a dispersão dos valores aparentemente se mantêm constantes; enquanto que na segunda, o valor médio da série muda ao longo dela.

#### 2.5.4.6. Critério de Informação de Schwarz

O Critério de Informação de Schwarz tem por característica a imposição de penalidades para a imposição de coeficientes a serem estimados – visto que o aumento gradual de coeficientes diminui a confiabilidade dos próprios parâmetros estimados e contribui para o fenômeno de “Overfitting”<sup>2</sup> – ponderados pela Verossimilhança do modelo, que por sua vez é uma medida de aderência do modelo através da análise da soma dos resíduos quadráticos.

A fórmula a seguir, que “condensa” em um único número o total da distância entre ocorrido e previsto, evitando compensações entre erros positivos e negativos, representa a função de soma dos resíduos quadráticos:

$$SQR = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$$

onde  $N$  é o número de observações,  $Y_i$  é o observado e  $\hat{Y}_i$  é o previsto pelo modelo.

Já a função de Verossimilhança é dada por:

$$V = -\frac{N}{2} (1 + \log(2\pi) + \log(SQR))$$

<sup>2</sup>Falta de capacidade do modelo em prever pontos fora da amostra utilizada para sua inferência. No caso de um sistema dinâmico deste tipo, caracteriza a incapacidade do modelo em prever o futuro, somente em descrever o ocorrido.

Assim, é possível escrever o critério de informação de Schwarz (1978) como:

$$SIC = \frac{(k \log(N) - 2V)}{N}$$

em que  $k$  é o número de parâmetros a serem estimados.

Consequentemente, quanto menor o valor do SIC, melhor o modelo, e assim, baseado neste critério, é possível escolher a melhor estrutura de atrasos para um modelo autorregressivo, por exemplo.

#### **2.5.4.7. Os principais tipos de modelos de série temporal auto-regressivos**

Segundo Morettin & Tolói (2006, p. 6), fazendo uma referência direta a Priestley (1979) “não há algo chamado de método de previsão”. No entanto, há diversas abordagens comuns para a criação de tais modelos, dos quais se destaca a publicação de Box & Jenkins (1976), que busca desenvolver uma abordagem generalista para o desenvolvimento de modelos de séries temporais, conhecido como família de modelos ARIMA (*Auto-Regressive Integrated Moving Average*)  $(p,d,q)$ , onde denota-se  $p$  pela dependência das observações anteriores,  $q$  pela dependência dos erros das observações *versus* previsto para cada uma delas e  $d$  pela diferenciação entre observações de maneira a tornar o processo estacionário (que se desenvolve ao longo de uma média “bem comportada”), assim como modelos vetoriais que os generalizam, modelos VARIMA, de maneira a estruturar também relações entre as variáveis dos vetores e não somente de uma série apenas.

Neste trabalho são apresentados e testados os modelos ARIMA, VAR (generalização vetorial do modelo AR) e VEC – este último uma extensão dos modelos VAR, com os coeficientes de correção de erros decorrentes de relações de co-integração entre as variáveis; assim como algumas definições necessárias ao entendimento do emprego destas metodologias.

#### 2.5.4.8. Modelos ARMA

Desconsiderando as diferenças implícitas nas séries que o modelador pode escolher de maneira a obter uma série estacionária, os modelos auto-regressivos de médias móveis têm a seguinte formulação:

$$X_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + k$$

em que  $X_{t-i}$  denota as observações anteriores,  $\varepsilon_{t-i}$  os erros de predição anteriores (inovações no processo) e  $k$  uma constante qualquer.

Portanto, tal modelo é capaz de relacionar as observações atuais às observações anteriores (evolução determinística) e à compensação de erros de previsões anteriores (que podem ser interpretados como choques ou inovações no processo, tendo natureza estocástica).

Quando o modelo só possui as componentes auto-regressivas (função das observações passadas), ordem  $p$  diferente de zero e  $q$  igual a zero, este pode ser estimado pelo tradicional método dos mínimos quadrados, conforme Hamilton (1994), que é deduzido e se encontra no apêndice deste trabalho.



Por outro lado, se o modelo possuir as componentes de média móvel, este deverá ser estimado por algum procedimento de máxima verossimilhança, que deve considerar a distribuição teórica das inovações, conforme Box & Jenkins (1976). Usualmente se adota um procedimento numérico do tipo Levenberg-Marquardt, cuja função objetivo a ser minimizada é a soma das inovações quadráticas, pressupondo-se uma distribuição normal destas inovações, podendo inclusive ser facilmente implementado em planilha Excel através da funcionalidade "Solver".

#### 2.5.4.9. Modelos Vetoriais Auto Regressivos (VAR)

Já os modelos Vetoriais Auto Regressivos, como já dito anteriormente, são extensões vetoriais dos modelos auto-regressivos escalares.

Os modelos VAR, assim como os modelos auto-regressivos escalares [modelos ARIMA de ordem  $(p,d,0)$ ], visam modelar a parte determinística de um sistema dinâmico.

É possível expressá-los na forma matricial por:

$$Y_t = BY_{t-1} + CX_t + \beta_0 + \beta_1 \cdot t + U$$

em que:

- $Y_t$  é a matriz de observações cujas linhas são as observações em cada amostra no tempo e as colunas denotam cada variável a ser explicada;
- $B$  é a matriz dos coeficientes do sistema dinâmico para as variáveis endógenas (retardos das variáveis dependentes);

- $C$  é a matriz dos coeficientes do sistema dinâmico para as variáveis exógenas (possíveis outras variáveis/intervenções necessárias no modelo);
- $Y_{t-1}$  é a matriz dos dados de entrada (observações com atraso) das variáveis endógenas;
- $X$  é a matriz dos dados de entrada (observações) das variáveis exógenas;
- $\beta_0$  é a matriz de constantes determinísticas a serem adicionadas no sistema;
- $\beta_1$  é a matriz de constantes que modelam a tendência linear determinística do sistema (se ela existir); e
- $U$  é a matriz de choques externos.

#### **2.5.4.10. Modelos Vetoriais de Correção de Erros (VEC)**

Os modelos Vetoriais de Correção de Erros (do inglês *Vector Error Correction Models*) – VEC, são extensões naturais do modelo VAR discutido anteriormente [de acordo com Hendry & Juselius (2000, 2001) e Lutkepohl (1991)], com o acréscimo de um termo de correção de erros, que visa corrigir eventuais erros e inconsistências produzidos por possíveis relações de co-integração, conforme os trabalhos de Johansen (1988) e Engle & Granger (1987).

O conceito de co-integração, ainda de acordo com estes autores, basicamente consiste de duas ou mais variáveis não-estacionárias, que possuem uma relação de equilíbrio de longo prazo – ou seja, são descorrelacionadas, não estacionárias, mas com alguma relação entre si.

Para verificar possíveis relações de co-integração, de acordo com os trabalhos de Engle & Granger (1987), se testa a não estacionariedade das séries usando o teste de Dickey-Fuller (1979) Extendido e

posteriormente se as séries não forem estacionárias, testa-se a estacionariedade dos resíduos de uma regressão de uma das variáveis contra as demais. Caso esses resíduos sejam estacionários, as séries são co-integradas.

A fim de ilustrar como a estrutura de um modelo VEC é obtida, é possível partir de um sistema dinâmico de duas variáveis, onde ambas são co-integradas por hipótese, seguindo os passos expostos em Hendry & Juselius (2000,2001), Lutkepohl (1991) e Morettin (2011).

Assim, sejam  $Y_{1,t}$  e  $Y_{2,t}$  duas séries não estacionárias cointegradas, e assumamos que haja uma relação de equilíbrio entre as duas dada por:

$$Y_{1,t} - \beta \cdot Y_{2,t} = \mu_t \sim N(0, \sigma)$$

Logo, se considerado que as variações em  $Y_{1,t}$  e  $Y_{2,t}$  dependem dos desvios deste equilíbrio no instante anterior, segue que:

$$\Delta Y_{1,t} = \alpha_1 (Y_{1,t-1} - \beta \cdot Y_{2,t-1}) + a_{1,t} : a_{1,t} \sim N(0, \sigma_1)$$

$$\Delta Y_{2,t} = \alpha_2 (Y_{1,t-1} - \beta \cdot Y_{2,t-1}) + a_{2,t} : a_{2,t} \sim N(0, \sigma_1)$$

Tendo em mente a expressão anterior, é possível generalizar este modelo de correção de erros em uma forma mais geral, onde a evolução das variáveis não só depende do desvio em relação ao equilíbrio mas também de possíveis autocorrelações, como:

$$\Delta Y_{1,t} = \alpha_1 (Y_{1,t-1} - \beta \cdot Y_{2,t-1}) + \phi_{1,1} \cdot \Delta Y_{1,t-1} + \phi_{1,2} \cdot \Delta Y_{2,t-1} + a_{1,t} \\ : a_{1,t} \sim N(0, \sigma_1)$$

$$\Delta Y_{2,t} = \alpha_2 (Y_{1,t-1} - \beta \cdot Y_{2,t-1}) + \phi_{2,1} \cdot \Delta Y_{1,t-1} + \phi_{2,2} \cdot \Delta Y_{2,t-1} + a_{2,t} \\ : a_{2,t} \sim N(0, \sigma_1)$$

Assim, se obtém um modelo VEC(1) – Vetor de Correção de Erros de ordem 1. De maneira análoga aos processos auto regressivos apresentados anteriormente, também se obtém modelos VEC de ordem  $n$ .

#### **2.5.4.11. Análise de Estabilidade e Representações como VAR(1)**

Conforme pode ser visto em Hamilton (1994), a estabilidade dos processos da família ARIMA depende exclusivamente das raízes características da parte auto-regressiva do polinômio.

O cálculo das raízes características consiste basicamente em resolver a equação da parte auto-regressiva como se fosse um polinômio algébrico qualquer. Quando o polinômio possui ordem elevada, este cálculo pode ser facilitado, obtendo-se uma matriz cujos autovalores correspondam às raízes da equação algébrica do polinômio.

Assim, por exemplo, ao considerar o seguinte polinômio:

$$p(x) = x^4 - 10x^2 + 9$$

Sua correspondente matriz para cálculo de raízes é dada por:

$$M = \begin{bmatrix} 0 & 0 & 0 & -9 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 10 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

As raízes do polinômio são obtidas ao calcular os autovalores da matriz  $M$ , através da expressão:

$$\det(M - \lambda I) = 0$$

Por conseguinte, o processo é estável se e somente se todas as raízes (autovalores) tiverem módulo menor do que um. A estabilidade é um critério importante para a escolha do modelo, pois, conforme Morettin (2011), garante que os choques que eventualmente impactem o sistema dinâmico em questão sejam dissipados ao longo da evolução do sistema.

Matematicamente, pensando em um modelo AR(2), por exemplo, significa que ao absorver um choque  $a_t$ :

$$Y_t = \phi_1 \cdot Y_{t-1} + \phi_2 \cdot Y_{t-2} + a_t$$

e que nos demais períodos, por hipótese, não haverá outros choques, em  $Y_{t \rightarrow \infty}$ , para ser estável, a contribuição deste choque (perturbação)  $a_t$  deverá ser zero.

Portanto, como pode ser visto Monteiro (2011), a estabilidade de um sistema discreto é garantida quando todos seus autovalores têm módulo menor que um.

De maneira a estender este método aos sistemas multivariados, todo e qualquer sistema dinâmico linear com coeficientes constantes pode ser reduzido a um sistema de ordem 1, de maneira a obter uma única matriz, cujas sub-matrizes remetam às matrizes de cada uma das equações.

Por exemplo, um VAR( $p$ ) pode ser escrito sob a forma de um VAR(1) através da transformação a seguir. Seja o VAR( $p$ ) dado pela expressão:

$$Y_t = C + A_1 \cdot Y_{t-1} + A_2 \cdot Y_{t-2} + \dots + A_n \cdot Y_{t-n} + \varepsilon_t$$

Este poderá ser escrito sob a forma de um VAR(1), através da seguinte expressão:

$$\begin{aligned} Z_t &= [Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_{t-n}]^T \\ Z_t &= [C, 0]^t + \begin{bmatrix} A_1 & \dots & A_n \\ I & & 0 \end{bmatrix} \cdot Z_{t-1} + \varepsilon_t \\ \therefore Z_t &= K_t + B \cdot Z_{t-1} + \varepsilon_t \end{aligned}$$

Logo, a mesma análise feita antes pode ser realizada a partir dos autovalores da matriz  $B$ .

Já para se avaliar a estabilidade dos modelos  $VEC(p)$ , é necessário recorrer ao teorema da representação de Engle-Granger (1987), que basicamente consiste na representação destes sob a forma VAR.

A fim de ilustrar sinteticamente o teorema, é possível reconsiderar o  $VEC(1)$  exposto anteriormente, o qual pode ser reescrito sob a forma de um VAR(1), sob a forma:

$$\Delta Y_t = \alpha \beta' + A \cdot Y_{t-1} + a_t$$

onde:

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \beta' = [1 \quad \beta], A = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix}$$

E, portanto, também podendo ter sua estabilidade analisada de acordo com os mesmos critérios expostos anteriormente.

## 2.6. Conclusão

Espera-se que, conforme discutido ao longo deste capítulo, seja possível a obtenção de um modelo híbrido (séries temporais e econométrico) a ser percorrido nos demais capítulos, adequado ao ferramentário e aos recursos disponíveis junto aos gestores das empresas do ramo, assim como de outras empresas que tenham problemas com estas características (cabendo algumas mudanças na implementação), de maneira a gerar valor ao proprietário, atingir as metas prospectadas em *business plan*, tal que torne possível as adequações das estimativas de custos, proporcionais às receitas brutas operacionais, garantindo a margem de lucro bruto.

Sob o ponto de vista tanto dos pressupostos econômicos (acerca da relação com o PIB) estabelecidos, quanto dos pressupostos estatísticos, espera-se a presença de uma estrutura auto-regressiva que motive o uso de modelos auto-regressivos univariados e multivariados.

Também foram apresentados critérios estatísticos para a escolha dos melhores modelos em cada uma das classes de modelos apresentadas (VAR, VEC, ARIMA), assim como técnicas matemáticas de análise de estabilidade destes modelos. Estes referenciais teóricos são indispensáveis para a fundamentação da metodologia a ser discutida.

No próximo capítulo são discutidas as metodologias e softwares empregados na identificação de modelos ARIMA, VAR e VEC, assim como a obtenção dos dados utilizados e possíveis transformações de ordem matemática/estatística a serem feitas nas séries.

### 3. Metodologia

A metodologia deste trabalho consiste na elaboração de um estudo de caso real, a partir dos dados obtidos junto à ARTESP (Agência Reguladora de Transportes do Estado de São Paulo) a respeito da concessionária Ecovias, que administra o sistema Anchieta-Imigrantes no Estado de São Paulo.

Assim, para este caso real, primeiramente é apresentada a concessionária Ecovias, para uma melhor contextualização da situação-problema e posteriormente, são percorridas a obtenção dos dados, técnicas de determinação de ordem de modelos das classes VAR, ARIMA e VEC com base nos critérios estatísticos discutidos no capítulo anterior; apresentados os *softwares* utilizados para a obtenção de cada modelo – cabendo ressaltar que todos os programas de computador aqui utilizados têm seus respectivos código-fontes abertos; e a utilização de critérios estatísticos para a escolha dos melhores modelos.

#### 3.1. Apresentação da Empresa

A concessionária Ecovias é uma empresa brasileira, constituída no Estado de São Paulo, que, conforme pode ser visto no sítio da empresa:

Em operação desde 1998, a Ecovias dos Imigrantes é a concessionária que opera o Sistema Anchieta-Imigrantes, ligação entre a região metropolitana de São Paulo e o Porto de Santos - o maior da América Latina - , o Polo Petroquímico de Cubatão, as indústrias do ABCD e a Baixada Santista. Com 176,8 km de extensão e movimento anual superior a 30 milhões de veículos pedagiados, é de vital importância para a economia brasileira, pois atende ao mais importante corredor de exportação da América Latina, além de área turística na Baixada Santista (sítio da Ecovias).

Ainda, segundo o sítio da companhia, o Sistema Anchieta-Imigrantes, é composto pelas seguintes rodovias:



- **SP 150 - Rodovia Anchieta**, do km 9, mais 700 metros, até o km 65, mais 600 metros. Um total de 55,90 quilômetros;
- **SP 160 - Rodovia dos Imigrantes**, do km 11, mais 460 metros, até o km 70. Um total de 58,54 quilômetros;
- **SP 040/150 – Interligação Planalto**. Tem 8 km de extensão, ligando as rodovias Anchieta e Imigrantes no alto da Serra, altura do km 40;
- **SP 059/150 - Interligação Baixada**. Tem 1,8 quilômetros de extensão, ligando as rodovias Anchieta, altura do km 59, com a Imigrantes, altura do km 62;
- **SP 248/55 – Rodovia Cônego Domênico Rangoni**, mais conhecida como Piaçaguera-Guarujá, com 30,6 quilômetros – do km 270 ao km 248 de Cubatão e do km 1 ao km 8, no Guarujá;
- **SP 055 – Rodovia Padre Manoel da Nóbrega**, com 21,60 quilômetros do km 270 de Cubatão ao km 292 da Praia Grande.

A Figura 6 mostra o mapa do sistema Anchieta Imigrantes administrado pela Ecovias.



**Figura 6 –Representação Gráfica do Sistema Anchieta-Imigrantes.**

Fonte: Sítio do Guia Quatro Rodas – Editora Abril

Esta empresa é controlada pelo grupo EcoRodovias S/A, que possui capital aberto, com ações negociadas na Bolsa de Valores de São Paulo sob o código “ECOR3.SA”, que também controla outras rodovias dentro e fora do Estado de São Paulo, como por exemplo a concessionária EcoPistas (Rodovia Carvalho Pinto / Ayrton Senna).

Também vale mencionar que ela é regulada pela ARTESP (Agência Reguladora de Transportes do Estado de São Paulo), conforme a lei de número 914, de 14 de Janeiro de 2002, que conforme o sítio da instituição:

[A ARTESP foi] instituída como autarquia de regime especial, dotada de autonomia orçamentária, financeira, técnica, funcional, administrativa e poder de polícia, com a finalidade de regulamentar e fiscalizar todas as modalidades de serviços públicos de transporte autorizados, permitidos ou concedidos a entidades de direito privado, no âmbito da Secretaria de Estado de Logística e Transportes, pelo decreto 46.708, de 22 de abril de 2002. (sítio da ARTESP)

### **3.2. Obtenção dos Dados**

A obtenção dos dados referentes ao tráfego foi feita a partir da observação direta das contagens fornecidas pela concessionária Ecovias junto à ARTESP, que pela lei federal de acesso à informação de número 12.527 de 18 de Novembro de 2011, regulamentada pelo decreto estadual de número 58.052 de 16 de Maio de 2012, foram obtidas junto à ARTESP.

Estes dados, na concessionária, são obtidos através de contagem direta feita por dispositivos eletromecânicos instalados nas praças de pedágio e enviados à Agência Reguladora, para fins de auditoria e verificação dos repasses (função da receita bruta) estabelecidos nos contratos de concessão, que a mesma regula.

Já os dados referentes às séries mensais do Produto Interno Bruto (PIB), foram extraídos a partir do sítio do Instituto de Pesquisas Econômicas Aplicadas (IPEA), que divulga mensalmente as estimativas do Banco Central do Brasil (BACEN).

### **3.3. Transformações Matemáticas nos Dados**

As seguintes transformações matemáticas foram feitas:

- Transformação das séries de tráfego pedagiado em séries de tráfego pedagiado em veículos equivalentes: esta transformação possibilita colocar em uma única grandeza veículos de diferentes portes, traduzindo motos, caminhões, ônibus e demais tipos de veículos em uma quantidade equivalente a veículos de passeio. Para tal foi multiplicada a quantidade de veículos de cada categoria por seu número de eixos e dividida pela quantidade de eixos que tem um veículo de passeio;

- Soma de todas as séries já em veículos equivalentes: esta transformação foi feita objetivando o estudo do tráfego pedagiado total da rodovia;
- Deflacionamento da série de PIB: esta transformação foi feita com o intuito de somente estudar as variações no produto, decorrentes dos acréscimos na produção real, excluindo as variações nos preços; e
- Aplicação do logaritmo natural sobre as séries do PIB deflacionada e de tráfego em veículos equivalentes: esta transformação foi feita objetivando a redução de possível heterocedasticidade decorrente da diferença de unidades de medida, ou seja, variações sobre o logaritmo tendem a reproduzir variações percentuais de mesma escala.

### **3.4. Softwares Empregados na Modelagem**

Neste trabalho optou-se exclusivamente pelo uso de softwares de código fonte aberto e de reprodução gratuita, para que os resultados e todos os passos aqui descritos sejam facilmente reproduzíveis por qualquer pessoa interessada, de maneira a considerar os aspectos positivos de cada um dos pacotes estatísticos para determinadas tarefas. Primeiramente, constatou-se uma facilidade de uso maior do programa GRET (GNU Regression and Econometric Tool Library), devido à presença de interface gráfica amigável e compatibilidade tanto com arquivos de extensão “.CSV” (arquivo numérico separado por vírgulas) quanto arquivos de planilha eletrônica gerados pelo software Microsoft Excel.

Vale ressaltar que o GRET implementa os diversos testes estatísticos para verificação das relações de cointegração discutidas no capítulo anterior, no caso o teste de Johansen e o teste de Engle-Granger e já faz automaticamente as previsões fora da amostra.

No entanto, o GRETl não implementa rotinas de detecção da estrutura ótima para modelos ARIMA baseadas em critérios estatísticos, conforme discorrido no capítulo anterior, somente implementa aquelas para detecção das estruturas ótimas para modelos VAR e VEC.

Igual facilidade (e igual ausência de método para modelos ARIMA) foi encontrada no software JMulti, sendo que este ainda possui maiores facilidades para a elaboração de modelos mais sofisticados, como VAR Estrutural e VEC Estrutural (que não são objeto de interesse neste estudo).

No entanto, apesar da possibilidade de estimação neste pacote, optou-se por não se utilizar o Jmulti, devido ao fato de existir somente versões recentes para a plataforma Windows.

Assim, para estimar a estrutura ótima de modelos ARIMA e sua respectiva previsão, foi utilizado o software R, com o pacote “forecast” devidamente instalado. Este pacote, conforme pode ser visto em Hyndman et al. (2013), permite a seleção da estrutura ótima de *atrasos* de acordo com o critério de Schwarz, discutido no capítulo anterior.

### **3.5. Critérios estatísticos para escolha da estrutura ótima dos modelos**

As rotinas para escolha das melhores estruturas de *atrasos* consiste em:

- definição do universo de estruturas, o qual o algoritmo de pesquisa percorrerá;
- definição do critério estatístico que deverá ser utilizado para balizar a escolha da melhor estrutura (a ser definido pelo pesquisador);
- estimação de cada um dos modelos definidos no universo de possibilidades e aferição da soma do critério estatístico;

- comparação e ordenação dos modelos de acordo com o critério estatístico; e
- informação ao pesquisador do melhor modelo obtido e seus respectivos resultados.

Vale ressaltar que os programas de computador mencionados na seção anterior implementam, em larga medida, a sequência de tarefas descritas nesta seção, ressaltando as especificidades também mencionadas. Para maiores informações, consultar Morettin & Tolói (2006).

Estas técnicas visam substituir as técnicas tradicionais de identificação de sistemas, como por exemplo em Box & Jenkins (1976), que eram fortemente embasadas na interpretação/análise qualitativa do comportamento das funções de auto-correlação e auto-correlação parcial. Hamilton (1994) inclusive dedica a primeira parte de seu livro visando explicitar o comportamento qualitativo de equações em diferenças (por exemplo, o comportamento oscilatório em equações em diferenças de grau 2) para melhor orientar o pesquisador.

Morettin & Tolói (2006) e Box & Jenkins (1976) reproduzem um quadro (mostrado na Tabela 1), com base no comportamento qualitativo de equações de diferenças sobre as funções de auto-correlação e auto-correlação parcial, para as principais estruturas ARIMA.

**Quadro 1 – Quadro resumo do comportamento das funções de auto-correlação e auto-correlação parcial para cada tipo de modelo ARIMA até ordem 2.**

Ordem	ARIMA(1,d,0)	ARIMA (0,d,1)
Comportamento FAC	Decaimento Exponencial	$\rho_1 \neq 0$
Comportamento FACP	$\phi_{11} \neq 0$	Decaimento Exponencial
Região de Estabilidade	$-1 < \phi < 1$	$-1 < \theta < 1$
Ordem	ARIMA(2,d,0)	ARIMA (0,d,2)
Comportamento FAC	Exponenciais Amortecidades e/ou ondas sinusoidais	$\rho_1 \neq 0$ $\rho_2 \neq 0$
Comportamento FACP	$\phi_{11} \neq 0$ $\phi_{22} \neq 0$	Exponenciais Amortecidades e/ou ondas sinusoidais
Região de Estabilidade	$-1 < \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $\phi_2 + \phi_1 < 1$	$-1 < \theta_2 < 1$ $\theta_2 - \theta_1 < 1$ $\theta_2 + \theta_1 < 1$
Ordem	ARIMA(1,d,1)	
Comportamento FAC	Decaimento exponencial após o 1o lag	
Comportamento FACP	Decaimento exponencial após o 1o lag	
Região de Estabilidade	$-1 < \phi < 1$ $-1 < \theta < 1$	

Fonte: Adaptado – Box & Jenkins (1976) – página 176 e 177

### 3.6. Critérios estatísticos para escolha do melhor modelo

A fim de se escolher o melhor modelo, é feito o procedimento de *backtesting* que basicamente consiste no estabelecimento arbitrário de uma quantidade de observações que são deixadas fora da amostra usada na inferência dos parâmetros, e que posteriormente é comparada ante as previsões feitas pelo modelo.

Posteriormente, é calculado o critério de soma dos resíduos quadráticos, conforme discutido no capítulo anterior e é escolhido o melhor modelo, com base no menor valor calculado através deste critério.

Aqui não é discutida a formação de critérios de *score*, que contemplem tanto a aderência aos pontos fora da amostra (*out-of-sample*) quanto a estabilidade do modelo. A justificativa para tal se apoia no fato que, apesar de avaliada a estabilidade (e discutida no capítulo anterior), a escolha ou rejeição de um modelo instável/estável passa necessariamente pela contextualização do problema por quem o está estudando.

Em um horizonte temporal de até seis meses, que a janela de tempo para fins de *backtesting* escolhida neste trabalho, não deve haver grandes impactos em termos de erros de previsão decorrentes de eventual instabilidade do modelo, pois, provavelmente não haverá tempo suficiente para grandes divergências em função da absorção de choques.

Já em um horizonte de tempo grande, a instabilidade deve ter impactos consideráveis, gerando grandes discrepâncias em função da integração decorrente da propagação/acúmulo dos choques ao longo do tempo.



### 3.7. Conclusão

Neste capítulo foram apresentados os principais aspectos da metodologia empregada ao longo deste trabalho, de maneira a destacar:

- Metodologia de obtenção dos dados e principais fontes;
- Transformações matemáticas utilizadas sobre os dados e suas principais justificativas de emprego;
- *Softwares* utilizados para a modelagem;
- Os métodos empregados na identificação dos modelos de séries temporais (VAR, VEC e ARIMA) discutidos no capítulo anterior; e
- Métodos de avaliação de desempenho dos modelos.

Portanto, espera-se que o leitor tenha se familiarizado com o passo-a-passo utilizado na geração de modelos e na discussão de desempenho entre eles, para que finalmente se torne possível adentrar nos resultados obtidos e em sua interpretação/análise, com base no referencial teórico previamente exposto.

Tendo em vista isso, no próximo capítulo são apresentadas todas as saídas dos pacotes estatísticos, devidamente comentadas, fazendo a ponte entre o referencial teórico e o método prático de inferência, objetivando a elaboração de uma conclusão sobre o tema dentro desta área específica.

## 4. Resultados

Neste capítulo são apresentados os principais resultados referentes à identificação, testes de cointegração, aderência, *backtesting* e análise de estabilidade, de maneira a permitir a comparação e a escolha do melhor modelo.

Para fins de identificação, foram utilizados critérios estatísticos. O Software R, na função “*auto.arima()*” utiliza o critério de Akaike, que é basicamente uma pequena alteração em relação ao critério de Schwarz, penalizando menos a inserção de mais coeficientes. Entende-se que no caso de modelos univariados, a penalização excessiva de novos parâmetros poderia comprometer a aderência dos modelos, uma vez que nestes não foram inseridos regressores externos.

Já nos modelos multivariados, pelo fato de terem naturalmente excessivos parâmetros (mais de uma série temporal é modelada ao mesmo tempo, para se estabelecer relações de evolução conjunta), foi preferencialmente utilizado o critério de Schwarz, visando evitar a sobreparametrização.

Outro ponto que vale a pena ser melhor explicado (que foi supramencionado) é a questão da não inserção de regressores externos nos modelos ARIMA sazonais. O principal candidato a regressor externo nestes modelos é o PIB. Por outro lado, a fim de aumentar significativamente o número de graus de liberdade nos mesmos, optou-se por uma modelagem em média frequência (mensal).

Assim, o PIB como um simples regressor externo (variável exógena), torna impraticável obter externamente previsões em frequência mensal necessárias à previsão mensal do tráfego, com viés econômico e que não somente se atenha à análise de séries temporais. Quando contextualizada a previsão do PIB dentro de modelos vetoriais auto-

regressivos, esta passa a ter um viés econômico, pelos motivos discutidos no capítulo anterior (uma série tender a ser um indicador precedente da outra).

#### **4.1. Modelos ARIMA Sazonais**

Conforme Morettin & Tolói (2006), modelos ARIMA sazonais basicamente consistem na extensão dos modelos ARIMA regulares, de maneira a multiplicar o polinômio ARIMA regular por um outro polinômio ARIMA, que descreva a evolução estocástica da componente sazonal, de maneira a se obter um único polinômio capaz de descrever a evolução da série no curto prazo, em função de sua inércia/nível, e de médio/longo prazo, de maneira a capturar as periodicidades.

Comumente, como pode ser verificado em Box & Jenkins (1976), a parte sazonal de um modelo ARIMA pode ter duas formas:

- Diferença sazonal mais uma componente de média móvel na periodicidade dos dados; ou
- Uma componente auto-regressiva direta com um retardo na periodicidade dos dados.

Ainda de acordo com esses autores, modelos sazonais do primeiro tipo têm por característica possuir memória sazonal mais longa que os segundos, pois no primeiro tipo há um termo que reproduz integralmente a observação em igual período passado e um termo de absorção dos choques sobre a série, enquanto que no segundo tipo há somente um termo cujo valor tem módulo inferior a um, fazendo com que a participação das observações passadas na observação atual decaiam exponencialmente.

Nesta seção foi estabelecido, convenientemente, que o modelo com a diferença sazonal mais a componente de média móvel sazonal seria denominado “SARIMA1” e o outro modelo como “SARIMA2”.

#### **4.2. Estratégia de Identificação de modelos ARIMA Sazonais**

Os modelos ARIMA foram identificados em dois estágios, uma identificação referente à parte regular e depois uma identificação empírica para a parte de evolução estocástica sazonal.

Para a identificação da parte regular, foi feita uma estimação via análise combinatória das principais possibilidades de modelos, conforme Tabela 1, de maneira a possibilitar o cálculo dos critérios estatísticos de seleção.

A segunda coluna da Tabela 2 reflete o critério de informação de Akaike, que é uma pequena modificação do critério de Schwarz (também conhecido por critério de informação Bayesiano), onde a adição de novos parâmetros sofre uma penalização menor do que no segundo critério mencionado. Logo, o programa atribui um peso maior para o ajuste do modelo do que para o número de parâmetros utilizados.

Como a estimação destes modelos não demanda grande poderio computacional, foi estimada uma grande quantidade de diferentes modelos, a fim de assegurar a obtenção da melhor estrutura.

Tendo estimado a melhor estrutura, se adicionam duas possibilidades de modelagem das componentes sazonais, que posteriormente são comparadas empiricamente (via *backtesting*).

**Tabela 2 – Seleção da Parte Regular do Modelo ARIMA Sazonal.**

```
> fitArima <- auto.arima(Panel$Trafego, trace=TRUE, stepwise=FALSE)
```

ARIMA(0,1,0) : 5011.375

ARIMA(0,1,0) with drift : 5012.396

ARIMA(0,1,1) : 5006.747

ARIMA(0,1,1) with drift : 5007.378

ARIMA(0,1,2) : 5007.21

ARIMA(0,1,2) with drift : 5007.656

ARIMA(0,1,3) : 5008.44

ARIMA(0,1,3) with drift : 5009.086

ARIMA(0,1,4) : 5006.685

ARIMA(0,1,4) with drift : 5006.805

ARIMA(0,1,5) : 5007.531

ARIMA(0,1,5) with drift : 5007.395

ARIMA(1,1,0) : 4956.876

ARIMA(1,1,0) with drift : 4958.33

ARIMA(1,1,1) : 4934.325

ARIMA(1,1,1) with drift : 4932.503

ARIMA(1,1,2) : 4928.932

ARIMA(1,1,2) with drift : 4923.81

ARIMA(1,1,3) : 4930.94

ARIMA(1,1,3) with drift : 4925.679

ARIMA(1,1,4)	: 4924.399
<u>ARIMA(1,1,4) with drift</u>	<u>: 4918.72</u>
ARIMA(2,1,0)	: 4946.043
ARIMA(2,1,0) with drift	: 4947.273
ARIMA(2,1,1)	: 4928.706
ARIMA(2,1,1) with drift	: 4923.471
ARIMA(2,1,2)	: 4930.807
ARIMA(2,1,2) with drift	: 4925.338
ARIMA(2,1,3)	: 4932.903
ARIMA(2,1,3) with drift	: 4927.486
ARIMA(3,1,0)	: 4945.451
ARIMA(3,1,0) with drift	: 4946.967
ARIMA(3,1,1)	: 4947.226
ARIMA(3,1,1) with drift	: 4948.737
ARIMA(3,1,2)	: 4931.684
ARIMA(3,1,2) with drift	: 4927.122
ARIMA(4,1,0)	: 4946.405
ARIMA(4,1,0) with drift	: 4947.711
ARIMA(4,1,1)	: 4929.831
ARIMA(4,1,1) with drift	: 4922.383
ARIMA(5,1,0)	: 4939.498
ARIMA(5,1,0) with drift	: 4940.274

Fonte: Autor obtido através do software estatístico R

Como pode ser visto na Tabela 2, o melhor modelo, que foi sublinhado pelo autor, tem a estrutura ARIMA(1,1,4), ou seja, um componente auto regressivo, uma diferença na série e quatro componentes de média móvel.

Por conseguinte, conforme já mencionado, são obtidos os dois modelos sazonais completos, conforme as Tabelas 3 e 4.

**Tabela 3 – Estimação do Modelo “SARIMA1”**

summary(fitSARIMA1)							
Series: Panel\$Trafego							
ARIMA(1,1,4)(0,1,1)[12] with drift							
Coefficients:							
ar1	ma1	ma2	ma3	ma4	sma1	drift	
-0.4864	0.0447	-0.2215	-0.0978	-0.5514	-0.6753	34710.72	
s.e.	0.2658	0.2187	0.1892	0.1085	0.1450	0.1009	NaN
sigma^2 estimated as 5.967e+10: log likelihood=-2272.78							
AIC=4561.55 AICc=4562.44 BIC=4586.69							
Training set error measures:							
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	
Training set	-29832.33	235137.9	147069.3	-1.109765	4.154517	0.5490204	0.1789293

Fonte: Autor obtido através do software estatístico R

É possível ver no Tabela 3 os valores de cada um dos coeficientes na primeira linha abaixo da descrição do termos do polinômio ARIMA e na segunda linha os valores do desvio padrão do valor destes coeficientes.

A Tabela 4, a seguir, tem a mesma estrutura da Tabela 3.

**Tabela 4 – Estimação do Modelo “SARIMA2”.**

```
summary(fitSARIMA2)

Series: Panel$Trafego

ARIMA(1,1,4)(1,0,0)[12] with drift

Coefficients:

      ar1    ma1    ma2    ma3    ma4   sar1   drift
-0.5280 -0.0231 -0.2902 -0.1227 -0.5641 0.8141 12679.039

s.e. 0.2497 0.1804 0.2034 0.0990 0.1425 0.0602 2248.984

sigma^2 estimated as 6.75e+10: log likelihood=-2451.21

AIC=4918.42 AICc=4919.25 BIC=4944.1

Training set error measures:

      ME    RMSE    MAE    MPE    MAPE    MASE    ACF1
Training set 13323.26 259065.2 173210.1 0.2046363 4.916068 0.6466058 0.08062521
```

Fonte: Autor obtido através do software estatístico R

### **4.3. Identificação de Estrutura de Retardos de Modelos VAR**

A fim de identificar a melhor estrutura para o modelo VAR, foi feita, de maneira análoga à identificação dos modelos ARIMA, a identificação através da comparação de critério estatístico de seleção, conforme Tabela 7.

Porém, antes disso, foram feitos os testes de cointegração de Engle-Granger e o teste de Johansen, conforme as Tabelas 5 e 6, respectivamente.



Estes testes foram feitos pois, se as variáveis não forem estacionárias e nem cointegradas, deverá ser estimado um modelo VAR até a diferença da série em que ela se torna estacionária; se as variáveis forem não estacionárias e cointegradas, deve ser estimado um modelo VEC; e se as variáveis forem estacionárias e não cointegradas, um VAR em nível (sobre as séries originais).

**Tabela 5 – Teste de Cointegração de Engle-Granger**

<p>Step 1: testing for a unit root in PIB</p> <p>Augmented Dickey-Fuller test for PIB</p> <p>including 12 lags of (1-L)PIB</p> <p>(max was 12, criterion modified AIC)</p> <p>sample size 164</p> <p>unit-root null hypothesis: <math>a = 1</math></p> <p>test with constant</p> <p>model: <math>(1-L)y = b_0 + (a-1)y(-1) + \dots + e</math></p> <p>1st-order autocorrelation coeff. for e: 0.007</p> <p>lagged differences: <math>F(12, 150) = 15.656 [0.0000]</math></p> <p>estimated value of <math>(a - 1)</math>: 0.00602697</p> <p>test statistic: <math>\tau_c(1) = 0.565415</math></p> <p>asymptotic p-value 0.9888</p> <p>Step 2: testing for a unit root in Trafego</p> <p>Augmented Dickey-Fuller test for Trafego</p> <p>including 8 lags of (1-L)Trafego</p> <p>(max was 12, criterion modified AIC)</p> <p>sample size 164</p>
--

unit-root null hypothesis:  $a = 1$

test with constant

model:  $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$

1st-order autocorrelation coeff. for e: -0.050

lagged differences:  $F(8, 154) = 7.537$  [0.0000]

estimated value of  $(a - 1)$ : 0.0138432

test statistic:  $\tau_c(1) = 0.328443$

asymptotic p-value 0.9798

Step 3: cointegrating regression

Cointegrating regression -

OLS, using observations 1998:05-2013:01 (T = 177)

Dependent variable: PIB

	coefficient	std. error	t-ratio	p-value
--	-------------	------------	---------	---------

const	10474.7	10153.2	1.032	0.3037
-------	---------	---------	-------	--------

Trafego	0.0497392	0.00264195	18.83	6.27e-044 ***
---------	-----------	------------	-------	---------------

Mean dependent var 198378.0 S.D. dependent var 43001.06

Sum squared resid 1.08e+11 S.E. of regression 24792.81

R-squared 0.669464 Adjusted R-squared 0.667575

Log-likelihood -2041.087 Akaike criterion 4086.174

Schwarz criterion 4092.527 Hannan-Quinn 4088.750

rho 0.587672 Durbin-Watson 0.684059

A26 Definir clíiquer  
Adriana, 11/9/2011

Step 4: testing for a unit root in uhat

Augmented Dickey-Fuller test for uhat

including 12 lags of  $(1-L)uhat$

(max was 12, criterion modified AIC)

sample size 164

unit-root null hypothesis:  $a = 1$

model:  $(1-L)y = (a-1)*y(-1) + \dots + e$

1st-order autocorrelation coeff. for e: 0.103

lagged differences:  $F(12, 151) = 24.142 [0.0000]$

estimated value of  $(a - 1)$ : -0.137823

test statistic:  $\tau_c(2) = -2.58148$

asymptotic p-value 0.2449

There is evidence for a cointegrating relationship if:

- (a) The unit-root hypothesis is not rejected for the individual variables.
- (b) The unit-root hypothesis is rejected for the residuals (uhat) from the cointegrating regression.

Fonte: Autor obtido através do software estatístico GRETL

Com base na Tabela 5, é possível ver que, segundo Engle-Granger (1987), as variáveis não são cointegradas, pois são estacionárias e logo, não são integradas e conseqüentemente, pela definição de cointegração, não podem ser cointegradas.

No entanto, mesmo assim, foi feito o teste de Johansen, conforme Tabela 6, que verifica o traço da matriz sobre o vetor de correção de erros.

Johansen test:

Number of equations = 2

Lag order = 12

Estimation period: 1999:05 - 2013:01 (T = 165)

Case 3: Unrestricted constant

Log-likelihood = -3264.91 (including constant term: -3733.15)

Rank Eigenvalue Trace test p-value Lmax test p-value

0 0.096032 16.809 [0.0298] 16.659 [0.0185]

1 0.00090969 0.15017 [0.6984] 0.15017 [0.6984]

Corrected for sample size (df = 129)

Rank Trace test p-value

0 16.809 [0.0314]

1 0.15017 [0.7017]

eigenvalue 0.096032 0.00090969

beta (cointegrating vectors)

PIB 0.00011690 -9.3027e-005

Trafego -7.1235e-006 8.4543e-006

alpha (adjustment vectors)

A25 1- preço de imóveis para outros itens  
Adriana, 12/8/2011

Trafego	34988.	359.52
renormalized beta		
PIB	1.0000	-11.003
Trafego	-0.060937	1.0000
renormalized alpha		
PIB	0.019014	0.00094622
Trafego	4.0900	0.0030395
long-run matrix (alpha * beta')		
	PIB	Trafego
PIB	0.0086026	-0.00021246
Trafego	4.0566	-0.24619

Fonte: Autor obtido através do software estatístico GRET

Na Tabela 6 é possível visualizar que as variáveis também não são cointegradas, segundo o teste de Johansen.

Por conseguinte, apesar da literatura apontar para a identificação de um modelo VAR sem o termo de correção de erros, neste trabalho procedeu-se tanto com a identificação de um modelo VAR quanto de um modelo VEC, a fim de testar a assertividade de modelos de séries temporais, tal que o número de retardos tanto para o modelo VAR quanto para o modelo VEC foi estimado através da seleção a partir do critério de Schwarz, conforme Tabela 7.

# A26 Definição de Tabela 7 – Identificação da Estrutura do Modelo VAR.

Adriana, 11/9/2011

VAR system, maximum lag order 20

The asterisks below indicate the best (that is, minimized) values

of the respective information criteria, AIC = Akaike criterion,

BIC = Schwarz Bayesian criterion and HQC = Hannan-Quinn criterion.

lags	loglik	p(LR)	AIC	BIC	HQC
1	-3505.24092		46.186156	46.740746*	46.411439*
2	-3500.56780	0.05300	46.177357	46.811174	46.434824
3	-3494.89751	0.02299	46.155523	46.868567	46.445173
4	-3491.70968	0.17280	46.166140	46.958411	46.487973
5	-3482.55212	0.00107	46.098720*	46.970219	46.452738
6	-3479.88859	0.25535	46.116191	47.066916	46.502391
7	-3476.34601	0.13146	46.122170	47.152123	46.540554
8	-3473.92017	0.30286	46.142747	47.251927	46.593315
9	-3470.39683	0.13344	46.148978	47.337385	46.631729
10	-3469.46243	0.75988	46.189051	47.456686	46.703985
11	-3467.46708	0.40726	46.215256	47.562117	46.762373
12	-3463.31299	0.08092	46.213242	47.639330	46.792542
13	-3463.02327	0.96532	46.261742	47.767058	46.873226
14	-3458.73090	0.07236	46.257920	47.842463	46.901588
15	-3454.89888	0.10469	46.260116	47.923886	46.935967
16	-3452.51731	0.31248	46.281272	48.024269	46.989306
17	-3446.45869	0.01650	46.254362	48.076586	46.994580
18	-3445.14896	0.62338	46.289529	48.190980	47.061930
19	-3442.51838	0.26153	46.307430	48.288108	47.112014
20	-3440.83705	0.49907	46.337739	48.397645	47.174507

Fonte: Autor obtido através do software estatístico GRETL

425 1- prego de conforme a Tabela 8, optou-se por obter modelos VAR e VEC de  
Adriana, 12/8/2011  
ordem 1. Assim, foram estimados os modelos, conforme tabelas 8 e 9.

**Tabela 8 – Modelo VAR(1).**

VAR system, lag order 1					
OLS estimates, observations 1998:06-2013:01 (T = 176)					
Log-likelihood = -4067.0988					
Determinant of covariance matrix = 4.0443916e+017					
AIC = 46.5352					
BIC = 47.0396					
HQC = 46.7398					
Portmanteau test: LB(44) = 412.361, df = 172 [0.0000]					
Equation 1: PIB					

	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-ratio</i>	<i>p-value</i>	
const	-5468.02	2444.07	-2.2373	0.02663	**
PIB_1	0.973596	0.0201256	48.3761	<0.00001	***
Trafego_1	0.00191621	0.00136612	1.4027	0.16263	
S1	-7270.76	1731.01	-4.2003	0.00004	***
S2	-4400.54	2015.09	-2.1838	0.03042	**
S3	12191.9	1656.96	7.3580	<0.00001	***
S4	4863.24	1651.99	2.9439	0.00372	***
S5	12545.8	1644.09	7.6308	<0.00001	***
S6	5743.71	1718.28	3.3427	0.00103	***
S7	2613.97	1686.67	1.5498	0.12315	

S8	5324.9	1614.41	3.2984	0.00120	***
S9	-1374.43	1615.39	-0.8508	0.39612	
S10	13806.3	1614.46	8.5517	<0.00001	***
S11	6612.35	1615.68	4.0926	0.00007	***

Mean dependent var	198673.5	S.D. dependent var	42943.20
Sum squared resid	3.15e+09	S.E. of regression	4411.386
R-squared	0.990231	Adjusted R-squared	0.989447
F(13, 162)	1263.194	P-value(F)	1.9e-155
rho	-0.191314	Durbin-Watson	2.375513

F-tests of zero restrictions:

All lags of PIB       $F(1, 162) = 2340.2 [0.0000]$

All lags of Trafego       $F(1, 162) = 1.9675 [0.1626]$

Equation 2: Trafego

	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-ratio</i>	<i>p-value</i>	
const	1.25978e+06	88571.7	14.2232	<0.00001	***
PIB_1	9.75201	0.72934	13.3710	<0.00001	***
Trafego_1	0.252305	0.0495074	5.0963	<0.00001	***
S1	181443	62730.9	2.8924	0.00435	***
S2	-623254	73025.9	-8.5347	<0.00001	***
S3	-214817	60047.5	-3.5774	0.00046	***



S4	-460430	59867.2	-7.6909	<0.00001	***
S5	-560794	59581.1	-9.4123	<0.00001	***
S6	-653837	62269.8	-10.5001	<0.00001	***
S7	-296413	61124.1	-4.8494	<0.00001	***
S8	-481878	58505.5	-8.2365	<0.00001	***
S9	-451521	58541	-7.7129	<0.00001	***
S10	-195290	58507.3	-3.3379	0.00105	***
S11	-395400	58551.5	-6.7530	<0.00001	***

Mean dependent var	3797779	S.D. dependent var	657247.4
Sum squared resid	4.14e+12	S.E. of regression	159866.5
R-squared	0.945231	Adjusted R-squared	0.940836
F(13, 162)	215.0680	P-value(F)	6.60e-95
rho	0.307772	Durbin-Watson	1.255931

F-tests of zero restrictions:

All lags of PIB      F(1, 162) = 178.78 [0.0000]

All lags of Trafego      F(1, 162) = 25.972 [0.0000]

Fonte: Autor obtido através do software estatístico GRETL

Conforme pode ser visto na Tabela 8, as equações para o Tráfego (equação 2) e para o PIB (equação 1), possuem altíssimos  $R^2$ , 0,99 e 0,94, respectivamente e com significância superior a 95% para quase todos os termos.

**Tabela 9 – Modelo VEC(1).**

VECM system, lag order 1			
Maximum likelihood estimates, observations 1998:06-2013:01 (T = 176)			
Cointegration rank = 1			
Case 3: Unrestricted constant			
beta (cointegrating vectors, standard errors in parentheses)			
PIB	1.0000		
	(0.0000)		
Trafego	-0.076559		
	(0.0018993)		
alpha (adjustment vectors)			
PIB	-0.024768		
Trafego	9.7690		
Log-likelihood = -4067.1153			
Determinant of covariance matrix = 4.0451489e+017			
AIC = 46.5354			
BIC = 47.0398			
HQC = 46.7400			
Equation 1: d_PIB			

<i>Coefficient</i>	<i>Std. Error</i>	<i>t-ratio</i>	<i>p-value</i>
--------------------	-------------------	----------------	----------------

const	-1494.86	1643.89	-0.9093	0.36451	
S1	-7256.59	1723.93	-4.2093	0.00004	***
S2	-4357.79	1994	-2.1854	0.03028	**
S3	12228.3	1638.71	7.4621	<0.00001	***
S4	4887.98	1640.92	2.9788	0.00334	***
S5	12565.6	1635.25	7.6842	<0.00001	***
S6	5748.41	1712.95	3.3559	0.00098	***
S7	2618.3	1681.46	1.5572	0.12137	
S8	5339.82	1607.32	3.3222	0.00110	***
S9	-1364.57	1609.58	-0.8478	0.39780	
S10	13825	1606.07	8.6080	<0.00001	***
S11	6618.5	1610.48	4.1096	0.00006	***
EC1	-0.0247683	0.0177276	-1.3972	0.16426	

Mean dependent var	698.9404	S.D. dependent var	7490.692
Sum squared resid	3.15e+09	S.E. of regression	4398.241
R-squared	0.678883	Adjusted R-squared	0.655242
rho	-0.192825	Durbin-Watson	2.378642

Equation 2: d\_Trafego

	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-ratio</i>	<i>p-value</i>	
const	911177	59568.7	15.2962	<0.00001	***
S1	181590	62469	2.9069	0.00416	***

S2	-622809	72255.6	-8.6195	<0.00001	***
S3	-214438	59381	-3.6112	0.00041	***
S4	-460173	59461	-7.7391	<0.00001	***
S5	-560589	59255.5	-9.4605	<0.00001	***
S6	-653788	62071.2	-10.5329	<0.00001	***
S7	-296368	60930.1	-4.8641	<0.00001	***
S8	-481723	58243.4	-8.2709	<0.00001	***
S9	-451418	58325.4	-7.7397	<0.00001	***
S10	-195095	58198.2	-3.3523	0.00100	***
S11	-395336	58358	-6.7743	<0.00001	***
EC1	9.76902	0.642384	15.2075	<0.00001	***

Mean dependent var	29838.26	S.D. dependent var	394098.9
Sum squared resid	4.14e+12	S.E. of regression	159376.5
R-squared	0.847669	Adjusted R-squared	0.836455
rho	0.307758	Durbin-Watson	1.255872

Cross-equation covariance matrix:

	PIB	Trafego
PIB	1.7916e+007	1.3017e+008
Trafego	1.3017e+008	2.3525e+010

determinant = 4.04515e+017

Fonte: Autor obtido através do software estatístico GRETL

Conforme pode ser visto na tabela 8, as equações para o Tráfego (equação 2) e para o PIB (equação 1), não possuem altíssimos  $R^2$  como

os modelos VAR, 0,68 e 0,85, respectivamente e com significância superior a 95% para quase todos os termos, ou seja, em termos de ajuste, o modelo VEC é consideravelmente pior que o VAR.

#### 4.4. Comparativo de Desempenho Empírico

A fim de medir empiricamente o desempenho dos modelos, foi feito o procedimento de *backtesting*, que consiste em analisar a capacidade preditiva dos modelos fora dos pontos utilizados na inferência dos modelos, conforme mostrado nas tabelas 10 a 13.

**Tabela 10 – Erros de Previsão do Modelo SARIMA1.**

Tabela Resumo	Previsto	Observado	Erro Absoluto (%)	Erro Quadrático
fev/13	5,074,452	4,553,970	11.43%	2.71E+11
mar/13	5,226,206	5,020,136	4.10%	4.25E+10
abr/13	5,210,780	4,731,025	10.14%	2.30E+11
mai/13	5,584,586	4,831,451	15.59%	5.67E+11
jun/13	5,748,151	4,626,837	24.24%	1.26E+12
jul/13	5,071,682	4,962,194	2.21%	1.20E+10
SQR	2.38E+12			
Erro Absoluto Médio (%)	11.28%			

Fonte: Autor

**Tabela 11 – Erros de Previsão do Modelo SARIMA2.**

Tabela Resumo	Previsto	Observado	Erro Absoluto (%)	Erro Quadrático
<b>fev/13</b>	5,176,306	4,553,970	13.67%	3.87E+11
<b>mar/13</b>	5,182,009	5,020,136	3.22%	2.62E+10
<b>abr/13</b>	5,188,096	4,731,025	9.66%	2.09E+11
<b>mai/13</b>	5,386,889	4,831,451	11.50%	3.09E+11
<b>jun/13</b>	5,552,947	4,626,837	20.02%	8.58E+11
<b>jul/13</b>	5,265,990	4,962,194	6.12%	9.23E+10
<b>SQR</b>	1.88E+12			
<b>Erro Absoluto Médio (%)</b>	10.70%			

Fonte: Autor

**Tabela 12 – Erros de Previsão do Modelo VAR(1).**

Tabela Resumo	Previsto	Observado	Erro Absoluto (%)	Erro Quadrático
<b>fev/13</b>	4,653,367	4,553,970	2.18%	9.88E+09
<b>mar/13</b>	4,783,466	5,020,136	4.71%	5.60E+10
<b>abr/13</b>	4,655,493	4,731,025	1.60%	5.71E+09
<b>mai/13</b>	4,536,380	4,831,451	6.11%	8.71E+10
<b>jun/13</b>	4,498,995	4,626,837	2.76%	1.63E+10
<b>jul/13</b>	4,861,875	4,962,194	2.02%	1.01E+10
<b>SQR</b>	1.85E+11			
<b>Erro Absoluto Médio (%)</b>	3.23%			

Fonte: Autor

**Tabela 13 – Erros de Previsão do Modelo VEC(1).**

<b>Tabela Resumo</b>	<b>Previsto</b>	<b>Observado</b>	<b>Erro Absoluto (%)</b>	<b>Erro Quadrático</b>
<b>fev/13</b>	4,654,499	4,553,970	2.21%	1.01E+10
<b>mar/13</b>	4,785,947	5,020,136	4.66%	5.48E+10
<b>abr/13</b>	4,659,374	4,731,025	1.51%	5.13E+09
<b>mai/13</b>	4,541,670	4,831,451	6.00%	8.40E+10
<b>jun/13</b>	4,505,714	4,626,837	2.62%	1.47E+10
<b>jul/13</b>	4,870,056	4,962,194	1.86%	8.49E+09
<b>SQR</b>	1.77E+11			
<b>Erro Absoluto Médio (%)</b>	3.14%			

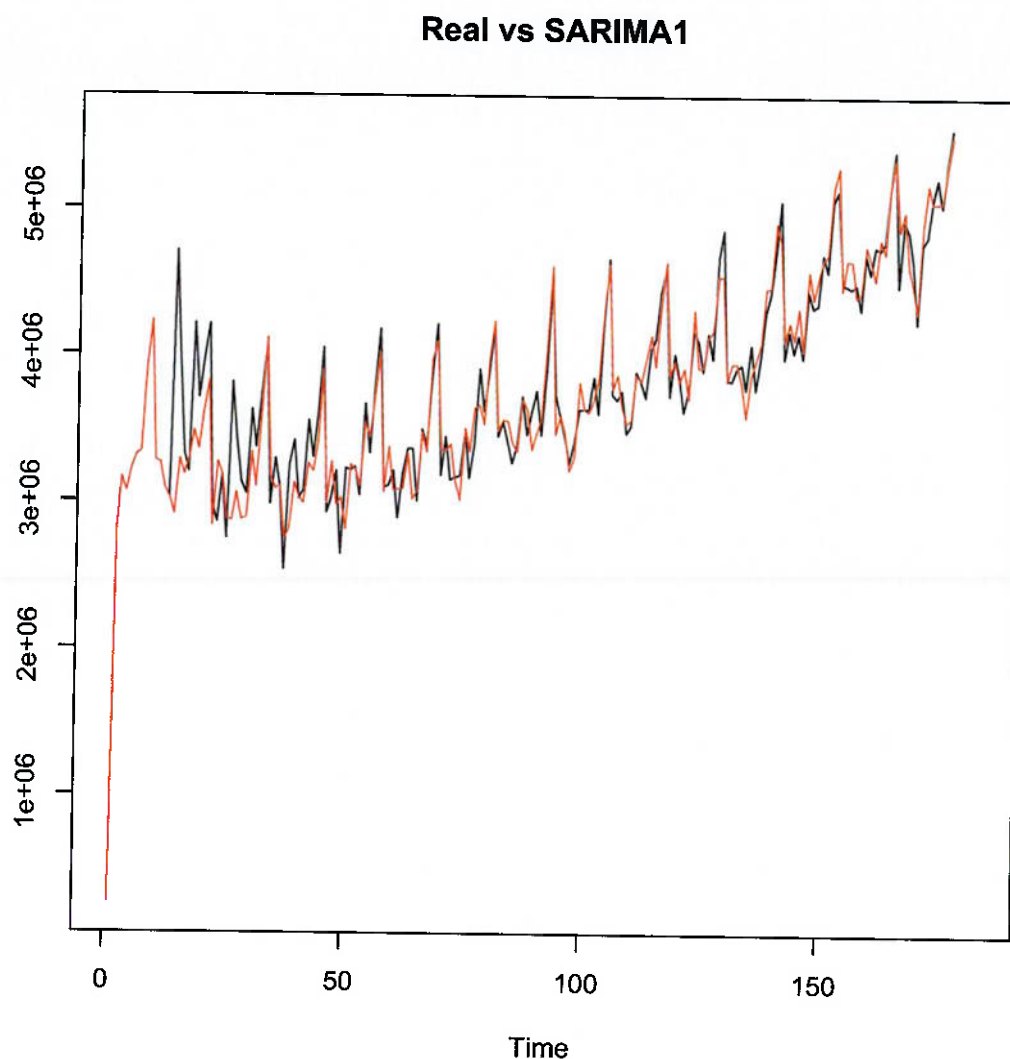
Fonte: Autor

Com base nas Tabelas 10 a 13, é possível verificar que, contra a intuição, o melhor modelo foi justamente o modelo VEC(1) que, de acordo com a literatura existente, sequer deveria ser estimado, muito embora também vale salientar que não há uma diferença muito significativa entre o desempenho dos modelos VAR e VEC.

Também vale ressaltar o desempenho muito pior dos modelos univariados, em relação aos modelos multivariados, cuja causa possivelmente se deve à ausência dos fatores explicativos macroeconômicos.

Não obstante, no caso dos modelos multivariados, uma curiosidade é que não ocorreu uma difusão sobre a propagação dos erros, tal que o erro empírico se mostrou inferior ao erro teórico.

A seguir, nas Figuras 7 a 10, são mostradas as curvas ajustadas de cada um dos modelos.

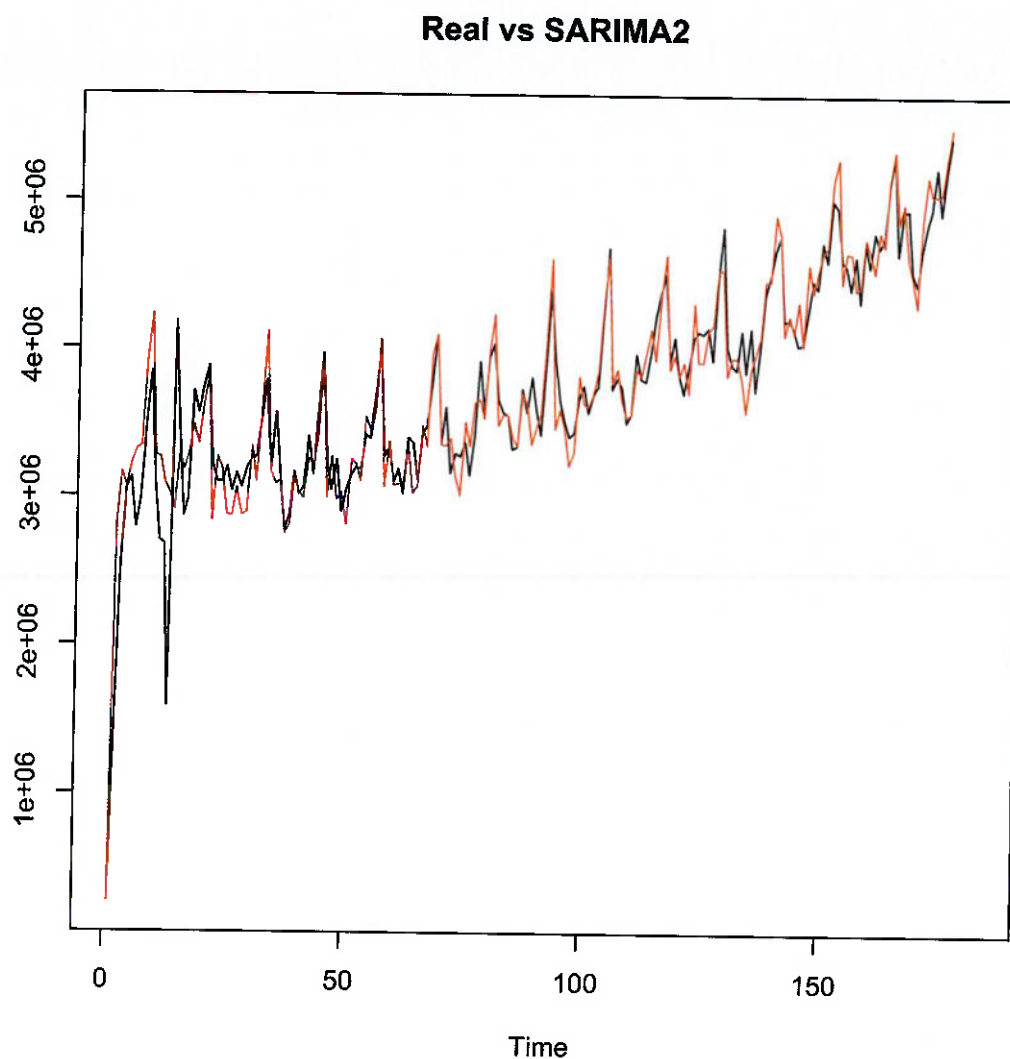


**Figura 7 – Ajuste do modelo SARIMA1.**

Fonte: Autor – obtido através do programa R

Na Figura 7 é possível visualizar que o ajuste (linha vermelha) do modelo é alto, sendo capaz de descrever as evoluções sazonais ocorridas na série, apesar de um descolamento grande no início da série. No eixo vertical, a quantidade de veículos em eixos equivalentes e no eixo horizontal, a disposição temporal, tal que o período 0 é o primeiro mês observado.



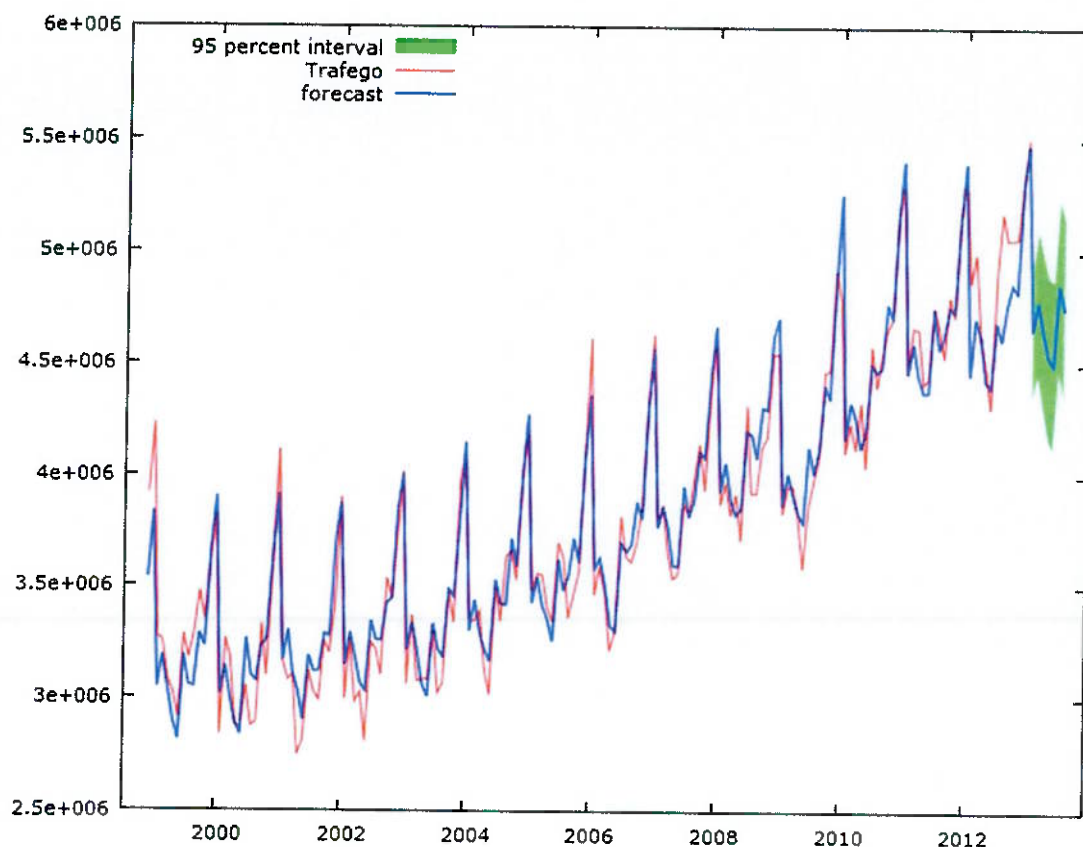


**Figura 8 – Ajuste do modelo SARIMA2.**

Fonte: Autor – obtido através do programa R

Na Figura 8 é possível visualizar que o ajuste (linha vermelha) do modelo também é alto, sendo capaz de descrever as evoluções sazonais ocorridas na série sem o grande descolamento havido no começo da série como no outro modelo.

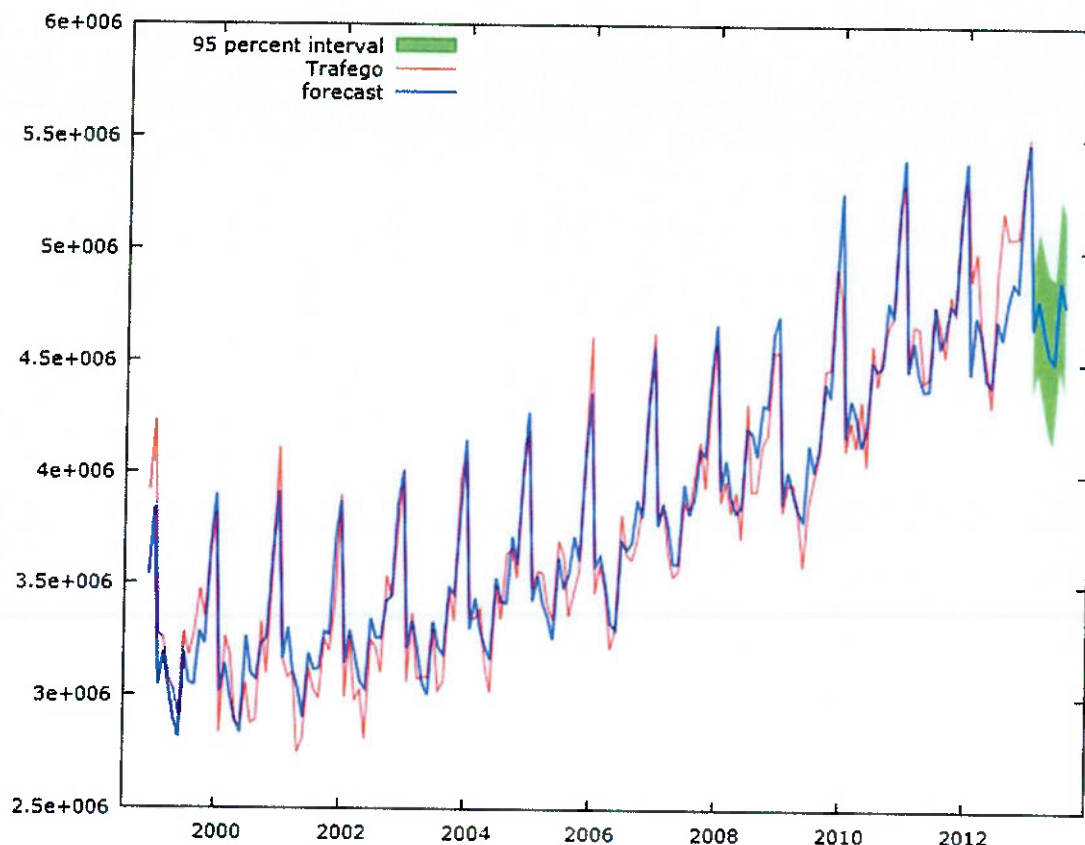
Vale a pena ressaltar também que, em decorrência da não aplicação do fator sazonal, a série na Figura 8 se tornou mais longa do que a série utilizada na Figura 7.



**Figura 9 – Ajuste do modelo VAR(1).**

Fonte: Autor – obtido através do programa GRET

Na Figura 9 é possível visualizar que o ajuste (linha azul) do modelo é alto, sendo capaz de descrever as evoluções sazonais ocorridas na série, assim como nos modelos univariados, com um pequeno descolamento do ajustado versus observado no final do ano de 2012.



**Figura 10 – Ajuste do modelo VEC(1).**

Fonte: Autor – obtido através do programa GRETL

Na Figura 10 é possível visualizar que o ajuste (linha azul) do modelo é alto como no modelo VAR, sendo capaz de descrever as evoluções sazonais ocorridas na série, também com um pequeno descolamento do ajustado versus observado no final do ano de 2012.

#### **4.5. Análise de Estabilidade**

Para analisar a estabilidade dos modelos, recorre-se à representação deles como um VAR(1), conforme disposto no capítulo anterior.

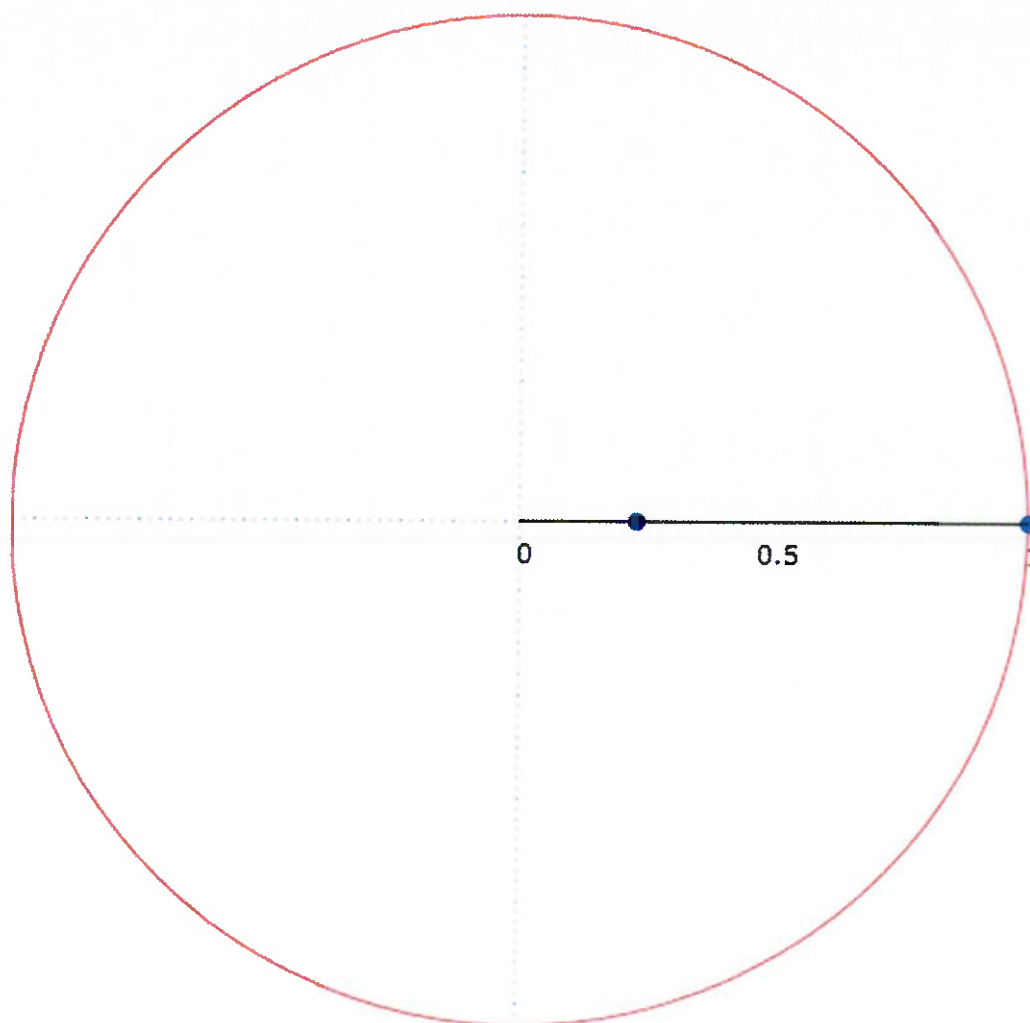
Depois, são calculados os autovalores de cada um dos modelos, de acordo com essas representações e de acordo com o módulo deles, é possível classificá-los como instáveis ou estáveis.

Os modelos univariados possuem todos os coeficientes auto regressivos com módulo menor do que um. Portanto, sem maiores transformações, é possível concluir que ambos são estáveis.

Já o modelo VAR(1) está na representação adequada para análise, cabendo somente ao autor (através do emprego do software GRETL) calcular os autovalores da matriz de termos auto regressivos, conforme indicado na Figura 11.

No modelo VEC(1) sua representação é automaticamente convertida para VAR(1) e suas raízes inversas calculadas, como mostrado na Figura 12.

VAR inverse roots in relation to the unit circle

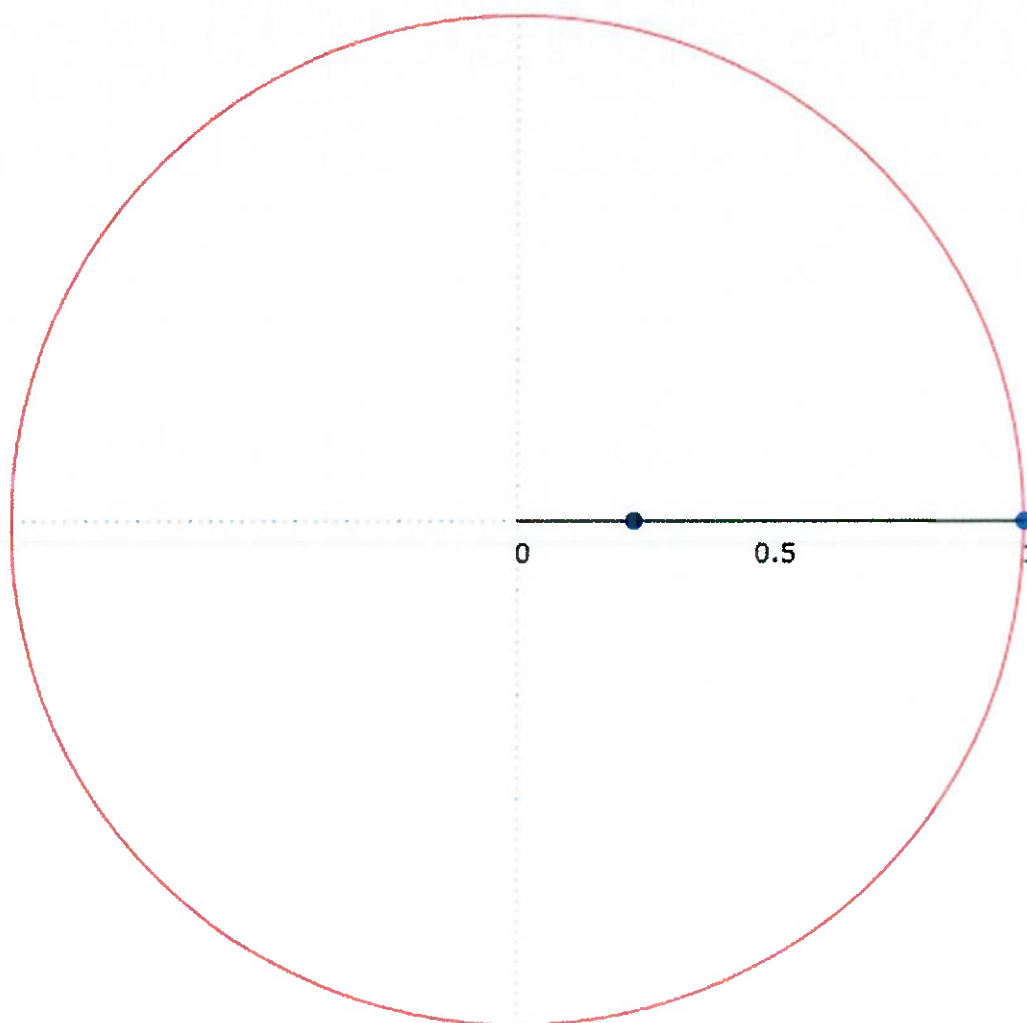


**Figura 11 – Autovalores do modelo VAR(1).**

Fonte: Autor – obtido através do programa GRETL

Na Figura 11 é possível verificar que o modelo VAR(1) é instável, pois possui um autovalor justamente em cima do círculo unitário, tal que eventuais choques que ocorram sobre o modelo não serão amortecidos e serão propagados indefinidamente ao longo das previsões, acarretando em instabilidade.

VAR inverse roots in relation to the unit circle



**Figura 12 – Autovalores do modelo VEC(1).**

Fonte: Autor – obtido através do programa GRETL

Na Figura 12 é possível verificar que o modelo VEC(1) também é instável, pois possui um autovalor justamente em cima do círculo unitário, tal que eventuais choques que ocorram sobre o modelo não serão amortecidos e serão propagados indefinidamente ao longo das previsões, acarretando em instabilidade.

Outro fato curioso é que os autovalores dos modelos VAR(1) e VEC(1) são bastante próximos entre si, o que em tese, deve acarretar em uma

propagação de instabilidade bastante semelhante entre os dois modelos ao longo do tempo.

#### **4.6. Conclusão**

Neste capítulo foi possível verificar que os modelos multifatoriais tiveram um desempenho substancialmente melhor em termos de capacidade preditiva sobre pontos fora da amostra em relação aos modelos univariados, muito embora os modelos univariados são estáveis e os modelos multivariados instáveis.

Não obstante, também vale ressaltar que, indo na contramão da literatura técnica existente, o modelo VEC(1), que nem deveria sequer ser estimado, foi o que teve melhor desempenho na comparação entre os quatro modelos testados, embora seu desempenho não tenha sido significativamente superior ao do modelo VAR(1).

## 5. Conclusão do Trabalho

Ao longo deste trabalho foi discutido um amplo referencial teórico sobre análise de séries temporais e sistemas dinâmicos, que englobou desde modelos univariados de séries temporais até modelos multivariados.

Não obstante, foram discutidos os principais aspectos da importância da realização de previsões de demanda no planejamento de execução de um serviço de utilidade pública regulada, tal que, a partir destas é possível planejar as linhas de custos, despesas, insumos entre outros, necessários às melhores práticas de gestão.

Foi ressaltada a importância da geração de fluxo de caixa próprio dentro do contexto da administração financeira e que embora atualmente o Brasil tenha se estabilizado e que tradicionalmente o custo de capital de terceiros seja inferior ao custo de capital próprio, o país historicamente teve grandes incertezas no cenário macroeconômico, que por consequência, desencorajou os gestores a buscarem recursos de terceiros e darem importância ainda maior à questão da geração própria de caixa, como fonte de recursos para a empresa.

Ao longo das discussões sobre séries temporais, se enfatizou a necessidade da busca de um modelo de previsão híbrido, conforme Morettin & Tolói (2004), em detrimento de modelos puramente econométricos ou puramente estatísticos (embasados em análise de séries temporais). Este ponto teórico se tornou muito interessante ao passo que foi possível observar que modelos puramente estatísticos (modelos univariados) tiveram um desempenho muito pior do que os modelos híbridos, no caso multivariados. Estes últimos foram considerados em decorrência de fortes expectativas em relação ao potencial explicativo, que variáveis macroeconômicas, como o PIB teriam sobre a evolução do tráfego.



Também foi estabelecida metodologia objetiva, embasada em referencial teórico existente, para a identificação, inferência e comparação de diferentes modelos, tal que espera-se que esta tenha sido a maior contribuição dentro do campo de conhecimento aplicado à Engenharia Financeira de projetos de concessão de serviços de utilidade pública regulada.

Dentro destes passos, identificou-se a estrutura de atrasos de cada um dos tipos de modelos propostos, foram realizadas as inferências e depois o processo de *backtesting*, que consiste na comparação direta entre previsto e observado para pontos fora da amostra.

Por fim, foi analisada a estabilidade de cada um dos modelos, para efeitos qualitativos. Este ponto se torna subjetivo na análise do usuário destes modelos, visto que, para efeitos de previsão de curto e médio prazo, não se torna crítico o não amortecimento de erros dos mesmos.

Ainda cabe ressaltar que o processo natural de difusão que ocorre em previsões fora da amostra não ocorreu nos modelos multivariados testados, tal que o erro empírico foi menor que o erro teórico e sendo assim, a instabilidade destes não deve se tornar um grande problema para previsões de longo prazo.

Por fim, como propostas de estudos futuros para próximos trabalhos e extensões à presente contribuição, sugerem-se os seguintes pontos:

- Comparação entre o Critério de Schwarz com outras métricas de critério de informação (Akaike, Hannan-Quinn, entre outros) para a seleção de estruturas de modelos contra a utilizada neste trabalho, e verificação do impacto das mesmas no desempenho dos modelos;
- Verificação do fator instabilidade em termos de previsão de longo prazo;

- Comparação com outras métricas de previsão em séries temporais, como por exemplo abordagens não lineares, como redes neurais; e
- Extrapolação deste trabalho para outros serviços de utilidade pública concessionados à iniciativa privada pelo Estado.

## Referências

- ARTESP, no sítio <http://www.artesp.sp.gov.br>, visitado em 08/09/2013.
- ASHLEY, R.A., GRANGER, C.W.J. – Time series analysis of residuals from St. Louis model. Em **Journal of Macroeconomics**, 1, 373-394, 1979.
- BRIGHAM, Eugene F., EHRHARDT, Michael C. **Financial Management: Theory and Practice**, 11ª Edição, Thomson South-Western, 2005.
- BHATTACHARYYA, M. N. Comparison of Box-Jenkins and Bonn Model prediction performance. Em **Lecture Notes in Economics and Mathematical Sciences**, n. 178. New York – Springer, 1980.
- BOX, G.E.P., JENKINS, G.M. **Times Series Analysis: Forecasting and Control**. 1ª Edição, San Francisco – Holden Day, 1976.
- CHOPRA, Sunil, MEINDL, Peter. – **Gerenciamento da Cadeia de Suprimentos** (“Supply Chain Management: strategy, planning and operation”), 1ª Edição, São Paulo – Pearson Prentice Hall, 2003.
- DICKEY, D.A., FULLER, W.A. – Distribution of the estimators for autoregressive time series with a unit root. Em **European Journal of Finance**, vol. 15, p. 619-637, 1979.
- ECORODOVIAS, no sítio <http://www.ecorodovias.com.br>, visitado em 08/09/2013.
- ENGLE, R.F., GRANGER, C.W.J. – Cointegration and error correction: Representation, estimation and testing. Em **Econometrica**, vol. 55, 251-276, 1987.
- FREDRICKSON, J. W., LAQUINTO, A. Inertia and creeping rationality in strategic decision processes. Em **Academy of Management Journal**, v. 32, p. 516-542, 1989.
- FROOT, Kenneth A., SCHARFSTEIN, David S., STEIN, Jeremy C. A framework for Risk Management, Em **Harvard Business Review**, 1994.
- GREENE, William H. – **Econometric Analysis**, 5ª Edição, Upper Saddle River, New Jersey – Prentice Hall, 2002.
- GUJARATI, Damodar N., PORTER, Dawn C. – **Econometria Básica**, Editora Bookman, São Paulo, 2011.
- HAMILTON, James D. – **Time Series Analysis**, 1ª Edição, Princeton, New Jersey – Princeton University Press, 1994.
- IPEADATA, no sítio <http://www.ipeadata.gov.br>, visitado em 01/11/2012.
- HEER, F. , **História das Civilizações**, v.3, Lisboa – Arcádia, 1968.
- HENDRY, David F., JUSELIUS, Katarina, 2000. Explaining Cointegration Analysis: Part 1. Em **The Energy Journal, International Association for Energy Economics**, vol. 0(Number 1), 1-42, 2000.

- HENDRY, David F., JUSELIUS, Katarina, 2000. Explaining Cointegration Analysis: Part 2. Em **The Energy Journal**, International Association for Energy Economics, vol. 0(Number 1), 75-120, 2001.
- JOHANSEN, S. – Statistical Analysis of cointegration vectors. Em **Journal of Economic Dynamics and Control**, vol. 12, 231-254, 1988.
- LOPES, Simone Saisse, CARDOSO, Marcelo Porteiro, PICCININI, Maurício Serrão – O Transporte Rodoviário de Carga e o Papel do BNDES. Em **Revista do BNDES**, vol. 14, N. 29, p. 35-60, Rio de Janeiro, 2008.
- LÜTKEPOHL, Helmuth – **Applied Time Series Econometrics**, 1ª Edição, New York – Cambridge University Press, 2004.
- LÜTKEPOHL, Helmuth – **Introduction to Multiple Time Series Analysis**, Heidelberg – Springer Verlag, 1991.
- MATARAZZO, Dante C. **Análise Financeira de Balanços**: abordagem básica. Volume 1, 6ª Edição, São Paulo - Editora Atlas, 2006.
- MCCLAVE, James T., BENSON, George P., SINSICH, Terry. **Statistics for Business and Economics**. 10ª Edição, Upper Saddle River – Pearson, Prentice Hall, 2008.
- MORETTIN, A. Pedro, TOLOI, M.C. Clélia – **Análise de Séries Temporais**, 2ª Edição, São Paulo – Editora Edgar Blücher, 2006.
- MORETTIN, A. Pedro – **Econometria Financeira**: Um Curso em Séries Temporais Financeiras, 1ª Edição, São Paulo – Editora Edgar Blücher, 2011.
- NOTINI, Hilton – **Ensaio Sobre Ciclos de Negócios**, Tese de Doutorado, Fundação Getúlio Vargas, Rio de Janeiro, 2009.
- PIERCE, D.A. Seasonal adjustment when both deterministic and stochastic seasonality are present. Em **Seasonal Analysis of Economic Time Series** (Arnold Zellner, ed.), p. 242-269. Washington D.C. – U.S. Department of Commerce, Bureau of the Census, 1979.
- PILÃO, Nivaldo E., HUMMEL, Paulo Roberto Vampre. **Matemática Financeira e Engenharia Econômica**, 1ª Edição, São Paulo – Pioneira Thomson Learning, 2003.
- PRIESTLEY, B.M. Discussion of the paper by Professor Makridakis and Dr. Hibon. Em **Journal of the Royal Statistical Society**, Series A, 142, 127-128, 1979.
- SAMUELSON, Paul Anthony – A Synthesis of the Principle of Acceleration and the Multiplier. Em **Journal of Political Economy** – University of Chicago Press, vol. 47, p. 786, Chicago, 1939.
- SCHWARZ, Gideon – Estimating the dimension of a model. Em **The Annals of Statistics**, vol. 6, 461-464, 1978.
- STULZ, René M. **Risk Management and Derivatives**, 1ª Edição, Thomson South-Western, 2003.



## Apêndice A – Derivação do Método dos Mínimos Quadrados

Os problemas em econometria geralmente se baseiam no estabelecimento de uma relação empírica entre conjuntos de dados, sob a forma de alguma função matemática, como por exemplo  $y(\beta) = f(\beta)$ , onde  $\beta$  é um vetor de parâmetros e  $y$  é a variável a ser explicada em termos deste vetor de parâmetros  $\beta$ .

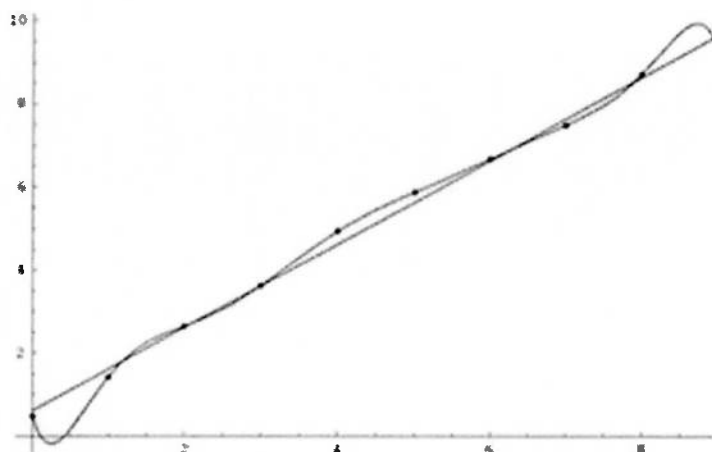
Assim, o problema acima poderia ser re-enunciado como um problema de ajuste de curva dado um conjunto de pontos, onde tal curva pode ser de qualquer forma, com esta forma determinada por uma **matriz de desenho**.

Consequentemente, de acordo com esta **matriz de desenho**, é possível obter uma função de linha reta, um polinômio de qualquer grau, uma relação linear múltipla.

Enunciando o problema (em notação matricial):

$$xM = y$$

Seja  $M$  a **matriz de desenho**,  $x$  é a matriz de variáveis explicativas, e  $y$  é o vetor da variável a ser explicada. É possível obter uma função que se ajuste aos pontos dados com um grau de complexidade / poder de extrapolação desejados, conforme indicado na figura 13.



**Figura 13 – Descrição da Sobre Parametrização.**

Esta imagem foi copiada da Wikipedia.

Assim, a ideia principal é encontrar os coeficientes da matriz de desenho que minimizem a distância dos pontos observados versus a função obtida.

A fim de entender as demonstrações e manipulações matemáticas que se seguem, é necessário algum conhecimento em cálculo diferencial e álgebra linear. Com isso, o leitor atento será capaz de replicar parte do processo de estimação, que é fundamentado no estimador de mínimos quadrados ordinário.

Como dito anteriormente, seja o problema escrito pela seguinte equação:

$$xM = y$$

E, desta expressão, o objetivo principal do estimador econométrico é obter uma matriz de desenho ( $M$ ) que minimize a distância entre a função estabelecida no lado esquerdo da equação e as observações à direita da equação. Assim é possível pensar como minimizar a distância entre ambos os lados da equação através da matriz ( $M$ ). Seja a distância quadrática dos dois lados da equação dada por:

$$D = (y - xM)^T(y - xM)$$

Em outras palavras, a distância proposta acima é a soma do quadrado da diferença entre cada observação e cada previsão.

Fazendo uso das seguintes relações do cálculo de matrizes, que são basicamente uma extensão do cálculo tradicional escalar, tem-se:

$$\frac{\partial x^T a}{\partial x} = \frac{\partial a^T x}{\partial x} = a$$

$$\partial(XY) = \partial(X)Y + \partial(Y)X$$

$$\partial A = 0$$

\*A é uma constante.

Assim, é possível verificar que:

$$\frac{\partial (Bx + b)^T C (Dx + d)}{\partial x} = B^T C (Dx + d) + C^T D^T (Bx + b)$$

Por conseguinte, é possível minimizar a distância acima através do uso da derivada da relação da distância e igualar a zero (que será um candidato a mínimo da função da distância):

$$\begin{aligned} \frac{\partial D}{\partial M} &= \frac{\partial (y - xM)^T (y - xM)}{\partial M} = -x^T (y - xM) - x^T (y - xM) \\ &= -2x^T (y - xM) \end{aligned}$$



E, fazendo-se:

$$\frac{\partial D}{\partial M} = 0$$

Então, segue que:

$$-2x^T(y - xM) = 0$$

$$-x^T(y - xM) = 0$$

$$-x^T y + x^T x M = 0$$

$$x^T x M = x^T y \Rightarrow M = (x^T x)^{-1} (x^T y)$$

Assim, é possível reproduzir e entender o processo de estimação mais usual em econometria, no qual os modelos multifatoriais se apoiam, que é a regressão linear múltipla por mínimos quadrados.

## Apêndice B – Código Fonte (R) – Modelos Univariados

```
# Lê o arquivo CSV

Panel <-
read.csv(file="C:/Users/Fernando/Desktop/modelagem_POLI_Ecovias/pa
inel_ecovias.csv", head=TRUE, sep=";")

# Instala o pacote Forecast

require(forecast)

# Ajusta o melhor modelo ARIMA (parte regular) sobre a série de dados

fitArima <- auto.arima(Panel$Trafego, trace=TRUE, stepwise=FALSE)

# Lê qual a estrutura regular detectada

AR_order <- fitArima$arma[1]

MA_order <- fitArima$arma[2]

Diff_order <- fitArima$arma[6]

# Verifica se o modelo detectado tem ou não "drift"

Total_ARMA_O <- AR_order + MA_order

hasDrift <- FALSE

if (length(fitArima$coef) > Total_ARMA_O)

  hasDrift <- TRUE

# Ajusta o melhor um Modelo ARIMA com Sazonalidade (0,1,12)

fitSARIMA1 <- Arima(Panel$Trafego,
order=c(AR_order,Diff_order,MA_order), seasonal=list(order=c(0,1,1),
period=12), include.drift=hasDrift)
```

```
# Ajusta o melhor um Modelo ARIMA com Sazonalidade (1,0,0)

fitSARIMA2 <- Arima(Panel$Trafego,
order=c(AR_order,Diff_order,MA_order), seasonal=list(order=c(1,0,0),
period=12), include.drift=hasDrift)

# Verifica a saída dos Modelos

summary(fitSARIMA1)

summary(fitSARIMA2)

# Realiza a projeção para 6 Meses adiante

forecast.Arima(fitSARIMA1, h=6)

forecast.Arima(fitSARIMA2, h=6)

# Retorna os Valores Ajustados em Figura, comparando com as
observações

ts.plot(ts(fitted.values(fitSARIMA1)), ts(Panel$Trafego), col=1:2,
main="Real vs SARIMA1")

ts.plot(ts(fitted.values(fitSARIMA2)), ts(Panel$Trafego), col=1:2,
main="Real vs SARIMA2")
```

## **Apêndice C – Código Fonte (GRET) – Modelo VAR(1)**

open

C:\Users\Fernando\Desktop\modelagem\_POLI\_Ecovias\painel\_ecovias.xl

sx

setobs 12 1998:05 --time-series

coint 12 PIB Trafego --test-down

coint 12 PIB Trafego --test-down

coint2 12 PIB Trafego --seasonals

var 24 PIB Trafego --seasonals --lagselect

var 1 PIB Trafego --seasonals

## **Apêndice D – Código Fonte (GRET) – Modelo VEC(1)**

open

C:\Users\Fernando\Desktop\modelagem\_POLI\_Ecovias\painel\_ecovias.g

dt

vecm 1 1 PIB Trafego --seasonals