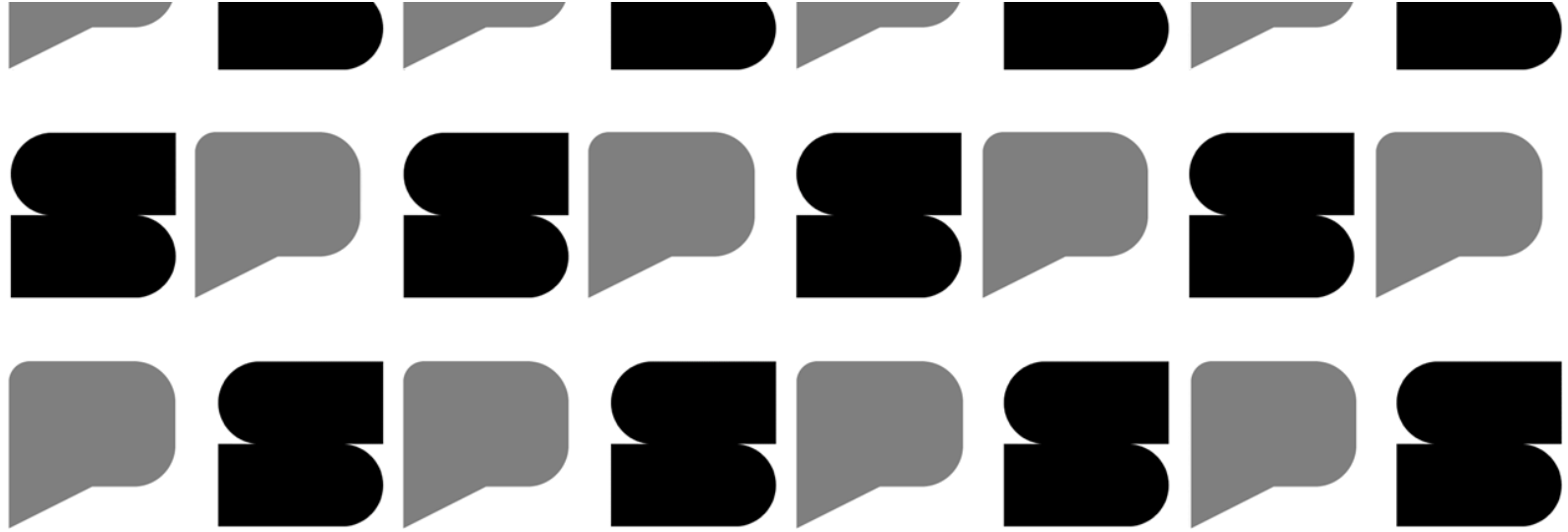




Projeto “Big Data”

Estimativa de matrizes de viagens rodoviárias
regionais no Estado de São Paulo



GOVERNO DO ESTADO DE SÃO PAULO

**Secretaria de Meio Ambiente,
Infraestrutura e Logística**

Equipe Técnica

Subsecretaria de Logística e Transportes

Denis Gerage Amorim
Subsecretário

André Nozawa Brito
Coordenador técnico

Karin Anne van de Bilt
Especialista em modelagem

Paulo Celso Pinheiro
Especialista em banco de dados

Mirian Miyuk Shirakawa
Analista

Atech – Grupo Embraer

Marco A. A. Fidos Junior
Gerente Técnico do Projeto

Fernando Antonialli
Gestor do Projeto

Bernardo Alvim
Especialista em Modelagem de Transportes

Carlos Magno Ribeiro Francisco
Analista de Sistemas

Denis Costa Garcia
Analista de Sistemas

Fabio Conn
Especialista em Modelagem de Transportes

Projeto “Big Data”

Estimativa de matrizes
de viagens rodoviárias
regionais no Estado de
São Paulo



GRUPO EMBRAER

Projeto “Big Data”

Estimativa de matrizes de viagens rodoviárias regionais no Estado de São Paulo

Sumário

1	INTRODUÇÃO	6
1.1	Objetivo do documento	6
2	DESCRIÇÃO DOS PRODUTOS	7
2.1	Elaboração de metodologia	7
2.1.1	Conceitos básicos	8
2.1.2	Conceitos Básicos de Carregamento em Rede	10
2.1.3	Ajustamento das Condições iniciais dos Dados de Telefonia	12
2.2	Uso de dados adicionais – aquisição e preparação de dados	13
2.3	Complementação e atualização da rede	14
3	MODELO E METODOLOGIA.....	18
3.1	Pressupostos teóricos.....	18
3.2	Procedimentos para aplicação do modelo teórico	20
3.2.1	Roteiro do Procedimento.....	20
3.3	Descrição de aplicação e indicadores para avaliação dos resultados estatísticos calibrados	22
3.3.1	Resultados para veículos de passeio	23
3.3.2	Resultados para veículos comerciais leves.....	26
3.3.3	Resultados para Veículos Comerciais Pesados:	28
4	RESULTADOS DO AJUSTAMENTO ESTATÍSTICO	31
4.1	Ajustamento Veículo Passeio	31
4.2	Ajustamento Veículo Comercial Leve	32
4.3	Ajustamento Veículo Comercial Pesado.....	33
4.4	Ajustamento Veículo Dia Útil	34
4.4.1	Veículo Passeio.....	34
4.4.2	Veículo Comercial Leve	35
4.4.3	Veículo Comercial Pesado	35
4.5	Ajustamento Fim de Semana.....	35
4.5.1	Veículo Passeio.....	36
4.5.2	Veículo Comercial Leve	36
4.5.3	Veículo Comercial Pesado	37

4.6	Ajustamento Feriado	37
4.6.1	Veículo Passeio	37
4.6.2	Veículo Comercial Leve	38
4.6.3	Veículo Comercial Pesado	38

5 OBTENÇÃO DAS MATRIZES.....39

6 CUIDADOS ESPECÍFICOS41

Lista de figuras:

FIGURA 2-1:	ESBOÇO DO CÁLCULO DO TRAÇO	12
FIGURA 2-2:	PEDÁGIOS FEDERAIS DO ESTADO DE SÃO PAULO	15
FIGURA 2-3:	CONCESSIONÁRIAS FEDERAIS E ESTADUAIS DO ESTADO DE SÃO PAULO	15
FIGURA 2-4:	PEDÁGIOS FORA DO ESTADO DE SÃO PAULO.....	16
FIGURA 2-5:	EXEMPLO DE FLOWBUNDLE	16
FIGURA 3-1 -	PROCEDIMENTO DESENVOLVIDO.....	21
FIGURA 3-2:	LOTES DE PRAÇAS DE PEDÁGIO	24
FIGURA 4-1:	AJUSTE VEÍCULOS DE PASSEIO	31
FIGURA 4-2:	CALIBRAÇÃO VEÍCULOS DE PASSEIO	32
FIGURA 4-3:	AJUSTE VEÍCULOS COMERCIAIS LEVES	32
FIGURA 4-4:	CALIBRAÇÃO VEÍCULOS COMERCIAIS LEVES	33
FIGURA 4-5:	AJUSTE VEÍCULOS COMERCIAIS PESADOS	33
FIGURA 4-6:	CALIBRAÇÃO VEÍCULOS COMERCIAIS PESADOS.....	34
FIGURA 4-7:	CALIBRAÇÃO VEÍCULOS DE PASSEIO – DIA ÚTIL.....	34
FIGURA 4-8:	CALIBRAÇÃO VEÍCULOS COMERCIAIS LEVES – DIA ÚTIL.....	35
FIGURA 4-9:	CALIBRAÇÃO VEÍCULOS COMERCIAIS PESADOS – DIA ÚTIL	35
FIGURA 4-10:	CALIBRAÇÃO VEÍCULOS DE PASSEIO – FINAL DE SEMANA	36
FIGURA 4-11:	CALIBRAÇÃO VEÍCULOS COMERCIAIS LEVES – FINAL DE SEMANA.....	36
FIGURA 4-12:	CALIBRAÇÃO VEÍCULOS COMERCIAIS PESADOS – FINAL DE SEMANA	37
FIGURA 4-13:	CALIBRAÇÃO VEÍCULOS DE PASSEIO – FERIADO	37
FIGURA 4-14:	CALIBRAÇÃO VEÍCULOS COMERCIAL LEVE – FERIADO.....	38
FIGURA 4-15:	CALIBRAÇÃO VEÍCULOS COMERCIAIS PESADOS – FERIADO.....	38
FIGURA 6-1:	EXEMPLO DE VIÁRIO COMPLEMENTADO	41
FIGURA 6-2:	EXEMPLO DE VIÁRIO A SER COMPLEMENTADO	42
FIGURA 6-3:	EXEMPLO DE VIÁRIO ESPECÍFICO.....	42

Lista de tabelas:

TABELA 2-1: EXTRATO MATRIZ DE LINK SELECIONADO DO PEDÁGIO.....	17
TABELA 3-1: AJUSTAMENTO VEÍCULOS DE PASSEIO.....	23
TABELA 3-2: TESTE DE ROBUSTEZ VEÍCULOS DE PASSEIO	24
TABELA 3-3: TESTE DE ENDOGENEIDADE VEÍCULOS DE PASSEIO.....	25
TABELA 3-4: TESTE DE SOBRE IDENTIFICAÇÃO VEÍCULOS DE PASSEIO	25
TABELA 3-5: TESTE VALIDADE DA EQUAÇÃO DE PRIMEIRO ESTÁGIO VEÍCULOS DE PASSEIO	25
TABELA 3-6: AJUSTAMENTO VEÍCULOS COMERCIAIS LEVES	26
TABELA 3-7: TESTE DE ROBUSTEZ VEÍCULOS DE COMERCIAIS LEVES.....	27
TABELA 3-8: TESTE DE ENDOGENEIDADE VEÍCULOS COMERCIAIS LEVES	27
TABELA 3-9: TESTE DE SOBRE IDENTIFICAÇÃO VEÍCULOS COMERCIAIS LEVES.....	27
TABELA 3-10: TESTE DE PRIMEIRO ESTÁGIO VEÍCULOS COMERCIAIS LEVES	28
TABELA 3-11: AJUSTAMENTO VEÍCULOS COMERCIAIS PESADOS.....	28
TABELA 3-12: TESTE DE ROBUSTEZ VEÍCULOS DE COMERCIAIS PESADOS	29
TABELA 3-13: TESTE DE ENDOGENEIDADE VEÍCULOS COMERCIAIS PESADOS.....	30
TABELA 3-14: TESTE DE SOBRE IDENTIFICAÇÃO VEÍCULOS COMERCIAIS PESADOS	30
TABELA 3-15: TESTE DE PRIMEIRO ESTÁGIO VEÍCULOS COMERCIAIS PESADOS.....	30
TABELA 5-1: TIPOS DE DIAS AO LONGO DO PERÍODO EM ESTUDO	39

O projeto denominado “Big Data” teve como um de seus principais objetivos a estimativa de matrizes de viagens rodoviárias regionais no Estado de São Paulo.

Os principais insumos para a obtenção dessas matrizes foram dados de rastreamento de movimentação de telefonia celular, para o período entre março de 2019 e fevereiro de 2020.

O período de amostragem foi definido em 12 meses, tendo-se escolhido a janela temporal mais recente sem o efeito da pandemia de Covid-19.

As matrizes estimadas neste projeto serão importante fonte de informação para as modelagens a serem desenvolvidas no Plano de Logística e Investimentos, em estruturação pela Secretaria de Meio Ambiente, Infraestrutura e Logística.

1.1 Objetivo do documento

Este documento apresenta o relatório final de metodologia de calibração das Matrizes Origem-Destino de viagens regionais de veículos que utilizam as rodovias do Estado de São Paulo, estimadas no escopo do projeto “Big Data”. O projeto tem como objetivo consolidar e registrar procedimentos metodológicos utilizados para desenvolvimento das atividades de calibração de matrizes origem-destino, utilizando tecnologias de big data e bases de dados de grande volume, em particular as bases de dados de movimentação de telefonia celular.

Além do processamento de dados, outras atividades adicionais foram conduzidas, como por exemplo:

- Avaliação de insumos para calibrações de matrizes Origem-Destino;
- Definição de parâmetros de avaliação – qualitativos e quantitativos – baseado em modelos anteriores de matrizes Origem-Destino existentes;
- Calibração de modelo de transportes utilizando dados processados pela ATECH em software de simulação de transportes (PTV-Visum).

2

Descrição dos produtos

A descrição da elaboração do produto deste relatório está estruturada conforme as etapas apresentadas a seguir:

- Reuniões Realizadas
- Elaboração de Metodologia
 - Levantamento de dados
 - Complementação e atualização da rede
 - Modelo e metodologia final de cálculo de viagens por intermédio de deslocamento de celulares



Escopo

Foram utilizados os seguintes conjuntos de informação que serviram como a base principal para a calibração das matrizes:

- a) Rede de simulação e modelagem da Secretaria de Logística e Transportes – em formato PTV-Visum;
- b) Dados de transações de telefonia (março de 2019);
- c) Dados de volume de tráfego das praças de pedágios, fornecidos pela ARTESP (março de 2019);
- d) Dados externos: população (2019), empregos (2019) e dados ANTT (volumes nas praças de concessões federais no estado de São Paulo, março de 2019).

2.1 Elaboração de metodologia

Este item apresenta a metodologia de Geração de Matriz OD a partir de deslocamentos de celulares detectados por informações de CDR – Registro Detalhado de Chamadas (Call Detail Record).

Para que se compreenda esta metodologia é necessário destacar alguns conceitos básicos apresentados a seguir.

2.1.1 Conceitos básicos

2.1.1.1 Viagem de Veículo

Viagem é uma transação realizada por uma ou mais pessoas em um veículo, que possui origem, destino, tempo, distância e motivos definidos, em que são transportados pessoas ou cargas, identificados tradicionalmente por entrevistas ou rastreo, associado a informações anexas transmitidas on-line.

O conjunto de viagens entre uma origem e destino consiste em fluxo que, somado a outros que compartilham um ponto comum em seus caminhos mínimos, determina o volume que solicita um trecho rodoviário ou link.

2.1.1.2 Deslocamento de celular por CDR – Registro Detalhado de Chamadas

É a transação realizada por um ou mais aparelhos celulares móveis em posse de uma pessoa, com origem e destino identificados por área de transmissão de torres de telefonia móvel, a cada intervalo entre pausas de movimentos maiores que 30 minutos.

A qualificação funcional da origem e destino é definida por inferência em função do tempo e horário e dispendido em cada local e as características dos usuários.

A informação obtida de operadora de telefonia móvel sobre estas transações implica em confusão de dados a fim de preservar o sigilo de informações específicas e busca homogeneizar os dados fornecidos com os dados de outros usuários que tenham o mesmo comportamento.

2.1.1.3 Rede Rodoviária

Rede Rodoviária é aquela utilizada em planejamento estratégico da DERSA com detalhamento (granularidade) compatível com a identificação dos sistemas principal, secundário e vicinal estadual.

Definição de links, nós e zonas:

- A rede é composta por nós e links que representam a rede viária de transportes;
- Os links são as representações das vias com a caracterização de cada uma delas de modo a representar os tempos de viagens e custos quando da sua utilização;
- Já os nós são as representações de entroncamentos entre as vias;

- As zonas são as representações da área espacial em que as pessoas têm sua origem ou seu destino da viagem. Suas características socioeconômicas (população, empregos etc.) são importantes para a determinação da atração e produção de viagens. As zonas são representadas por centroides na rede de transportes que são conectados à rede viária através de conectores, que fazem a ligação da demanda (matriz) com a rede;
- No presente estudo foram utilizadas 910 zonas, definidas pela equipe da SLT.

2.1.1.4 Diferenças entre Transações e Viagens de Pessoas

Os dados de transações de deslocamento de celular não têm correspondência imediata com viagens, porque uma viagem de pessoa pode corresponder a mais de uma transação se houver uma ou mais pausas estacionárias superiores a 30 minutos durante o mesmo trajeto. Portanto é necessário proceder ajustes ou correções de forma a aproximar os dados de transações aos dados de viagem.

Essas correções serão feitas pela identificação de locais geradores que sejam “outliers” de produção e atração de viagens em relação à característica do Local Gerador, casa (C), trabalho (T) e outro (O). As colunas simétricas serão corrigidas de forma análoga.

2.1.1.5 Finalidade/Motivo da Origem e do Destino

No estudo foram identificadas as atividades em relação ao Local Gerador relativas à casa, trabalho e outros. O Local Gerador influencia na geração de viagens e tem tratamento diferenciado quando da análise dos dados de telefonia móvel para a calibração da matriz OD.

2.1.1.6 Regime de Atividade

No estudo foram identificados os diferentes tipos de dias em útil, finais de semana e feriados. Cada um dos tipos de dias influencia na geração de viagens e tem tratamento diferenciado quando da análise dos dados de telefonia móvel para a calibração da matriz OD.

2.1.1.7 Classe de veículo

Os veículos foram divididos em três classes:

- Passeio: contendo os automóveis, caminhonetes, utilitários e motos;
- Caminhões leves: contendo os caminhões de 2 eixos e 3 eixos;
- Caminhões pesados: contendo os veículos de 4 eixos ou mais.

2.1.1.8 Praças de Pedágios

As praças de pedágios são um elemento importante para a calibração da rede de transporte e da matriz OD, pois têm peso nos custos da rede de transporte. Devido a sua necessidade de controle (arrecadação das concessionárias), seus dados volumétricos são extremamente confiáveis. Os volumes das praças de pedágio serão utilizados como balizador e aferição do modelo.

2.1.2 Conceitos Básicos de Carregamento em Rede

O carregamento de uma rede de transporte é obtido com o cálculo de todos os fluxos de cada um dos links que representam as vias através da alocação das matrizes de transporte nesta rede.

O fluxo (ou volume) é definido pelo número de veículos que passa por uma determinada seção (ou faixa) de uma via por unidade de tempo de observação, sendo usualmente expresso em veículos por hora ou por dia. Neste projeto os cálculos de carregamento são diários referentes ao volume de cada seção da via.

Para a alocação na rede de transporte, foi utilizado o princípio do custo generalizado, onde a impedância é definida pelos tempos gastos em cada link da rede e seus custos que são transformados em tempo.

Para a alocação usou-se o módulo "Tribut" do PT-Visum, utilizado na elaboração do PAM-TL (Plano de Ação de Transporte e Logística para a Macrometrópole Paulista) que através dos valores do tempo, transforma os custos de transporte em reais, em impedâncias de tempos. A grande vantagem desse método é que ele considera uma curva de distribuição do valor do tempo, contemplando o valor médio do tempo e sua dispersão.

Foram considerados os custos quilométricos, relativos a custos de gasolina, óleos, pneus e etc. e os valores do tempo obtidos no estudo do PAMTL.

Foi considerado o valor médio do tempo para veículos da classe C para representar os veículos de telefonia. Como se trata de um valor médio de todas as classes esse valor foi atualizado e aplicado para caminhões leves e realizada a estimativa para caminhões pesados. Foi considerado valor de R\$ 44,18 com dispersão de R\$0,14 para automóveis e caminhões leves, e para caminhões pesados, R\$ 88,36 com dispersão de R\$0,14.

Os valores quilométricos do PAM-TL, para representação dos custos de combustível, óleo, pneus e demais custos que variam conforme a extensão do trajeto, foram atualizados para o presente projeto, utilizando R\$0,50 para automóveis, R\$ 1,00 para caminhões leves e R\$ 2,00 para caminhões pesados.

O consumo de insumos para rodagem dos veículos sofre uma variação nos diversos tipos de rodovias, levando em consideração suas características, pista simples ou dupla e vias rurais e urbanas. Para representar esta variação foi inserido uma componente de fator de custo (FAT_CUST).

Custo (R\$) por Link = [PEDAGIO_TAR2019]+CustKM*[LENGTH]*[FAT_CUST]

Sendo:

Pedágio_TAR2019, tarifa de pedágio 2019;

CustKM = Custo quilométrico para cada uma das categorias;

LENGTH = Extensão do link; e

FAT_CUST= Fator de custo para diferentes vias. Adotando 1,00 para Freeway; 1,05 para Pista Dupla; 1,20 para Pista Simples; e 1,30 para Urbanas e não pavimentadas.

O custo generalizado é representado por:

Custo Generalizado (min) = Tempo Total(min) + Custo (R\$) transformado em tempo (min)

2.1.2.1 Fórmula utilizada para determinação do volume de tráfego de um determinado link da rede de transportes:

Volume de tráfego de determinado link da rede

Fórmula:

$$VV_{P, CV, RA, ANO, MES, PD} = Tr [ML_P, QT_{RA, ANO, MES, PD, CR, MOT_OD},] \cdot CA_{CR, MOT_OD}$$

Onde:

- $VV_{P, RA, ANO, MES, PD}$ é o volume de classes de veículo (CV) que passa pelo pedágio (P) em um determinado período de tempo (ANO, MES, PD) para determinado regime de atividade diária (RA), podendo ser RA dia útil, final de semana e feriado.
- $Tr [.]$ traço do produto entre a matriz de link selecionado do pedágio e a matriz de transações de deslocamento de celular;
- ML_P é a matriz de link selecionado correspondente ao pedágio (P);

- $QT_{RA, ANO, MES, PD, CR, MOT_OD}$ são as matrizes de transações de deslocamentos de celular desagregadas por regime de atividade, ano, mês, período diário, classe de renda, motivo origem e destino;
- $[CA_{CR, MOT_OD}]$ são os vetores de coeficientes estimados estatisticamente por regressão aplicada sobre a amostra aleatória de dados correspondente a 4% do volume mensal, ou seja, um dia útil equivalente e aplicada quanto ao método de variáveis instrumentais, conforme será explicado adiante.

A Figura a seguir mostra o cálculo para se obter o traço do produto entre a matriz de link selecionado do pedágio e a matriz de transações de deslocamento de celular.

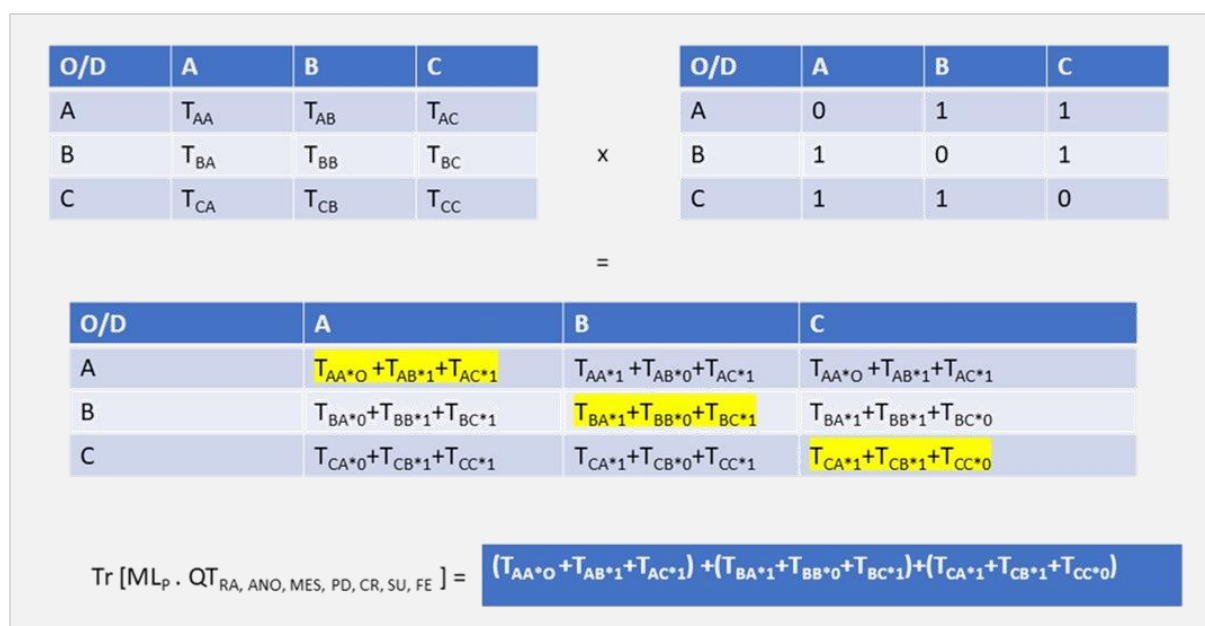


Figura 2-1: Esboço do Cálculo do Traço

2.1.3 Ajustamento das Condições iniciais dos Dados de Telefonia

Nesta fase do trabalho foram ajustadas as transações de deslocamento de aparelhos móveis para refletir fluxos de veículos, e para que se possa gerar a matriz origem e destino ao final dos trabalhos.

- Determinação do tráfego de veículos nos pedágios por Regime de Atividade e classe de veículo ($VV_P, CV, RA, ANO, MES, PD$);
- Determinação das matrizes de links selecionados de pedágio (ML_P);
- Determinação das quantidades de transações por atributos de usuários ($QT_{RA, ANO, MES, PD, CR, MOT_OD}$);

- d) Processamento das transações de deslocamento de celulares e matrizes de links selecionados de cada pedágio para identificação do Traço da Matriz (Tr [ML_p, QT_{RA}, ANO, MES, PD, CR, MOT_OD]) referente a cada período, regime de atividade diária;
- e) Identificação das variáveis dependentes (volumes de pedágio) e independentes, volumes de transações por atributo explicativo de usuário;
- f) Aplicação de regressão sobre a amostra de março e obtenção do vetor de coeficientes (CA_{CR, MOT_OD});
- g) Treinamento da regressão para o universo de dados dos demais meses de e obtenção dos vetores finais de coeficientes (CF_{CR, MOT_OD});
- h) Aplicação dos vetores CF_s sobre os dados totais de telefonia de cada segmento de transações para geração das matrizes de veículos por segmento;
- i) Obtenção das Matrizes por Classe de Veículos.

2.2 Uso de dados adicionais – aquisição e preparação de dados

Para ajustes do modelo de simulação de transportes foi necessário realizar atividades adicionais para a obtenção de dados extras, complementares aos dados existentes, e processamento destes dados.

Foram necessários os seguintes tipos de dados:

- Emprego;
- População;
- Pedágios ANTT do Estado de São Paulo, fora da concessão da ARTESP.

Os dados de população foram obtidos no site do IBGE e foram utilizados os dados da população estimada para 2019, obtidos no link:

- <https://www.ibge.gov.br/#>
- Gerando o arquivo 4 - Dados Externos\estimativa_dou_2019.xls:

Os dados de empregos foram obtidos no site do ministério do trabalho e foram utilizados os dados de empregos formais por município em 2019, obtidos no link:

- <https://bi.mte.gov.br/scripts10/dardoweb.cgi>
- Gerando o arquivo 4 - Dados Externos\consulta31939332.csv:

Os dados de empregos e população foram tabulados para a correspondência das zonas.

- Gerando o arquivo 4 - Dados Externos\Consolidado_Emp_pop2019.xlsx:

Foram obtidos os dados de volume de veículos das praças de pedágio federais (ANTT) instaladas no Estado de São Paulo de 2019, obtidos no link:

- <https://dados.antt.gov.br/dataset/volume-trafego-praca-pedagio>
- Gerando o arquivo 4 – Dados Externos\ **volume_praca_2019.csv**.

Foram pesquisadas as tarifas das praças de Pedágio para 2019. As tarifas das praças da ARTESP estão contidas no arquivo "TARIFAS DE PEDAGIO-2019.pdf", obtido no SIC da ARTESP, no link:

- <http://www.artesp.sp.gov.br/Shared%20Documents/TARIFAS%20DE%20PEDAGIO-2019.pdf>
- O arquivo está na pasta do arquivo 4 – Dados Externos\ **TARIFAS DE PEDAGIO-2019.pdf**.

2.3 Complementação e atualização da rede

Foi realizada a revisão da rede de transportes, em uso pela SLT, utilizando o software PTV-Visum com ajustes de pares de ligação de centroide, ajuste das tarifas de pedágio 2019 e revisão do viário principal. Esta rede foi complementada com as praças de pedágio fora da jurisdição da ARTESP, tanto dentro do estado quanto fora dele.

As seguintes praças das concessionárias federais presentes no estado foram posicionadas na rede:

- Autopista Fernão Dias;
- Autopista Régis Bittencourt;
- NovaDutra;
- Transbrasiliana.

A complementação ocorreu através do uso das funções de edição de links e nós do PTV-Visum. As praças de pedágio dentro do Estado de São Paulo já possuíam as posições corretas, necessitando o valor da tarifa.

A Figura **2-2** mostra a localização das praças na rede do PTV-Visum das quatro concessionárias federais, sob jurisdição da ANTT, presentes no estado de São Paulo.

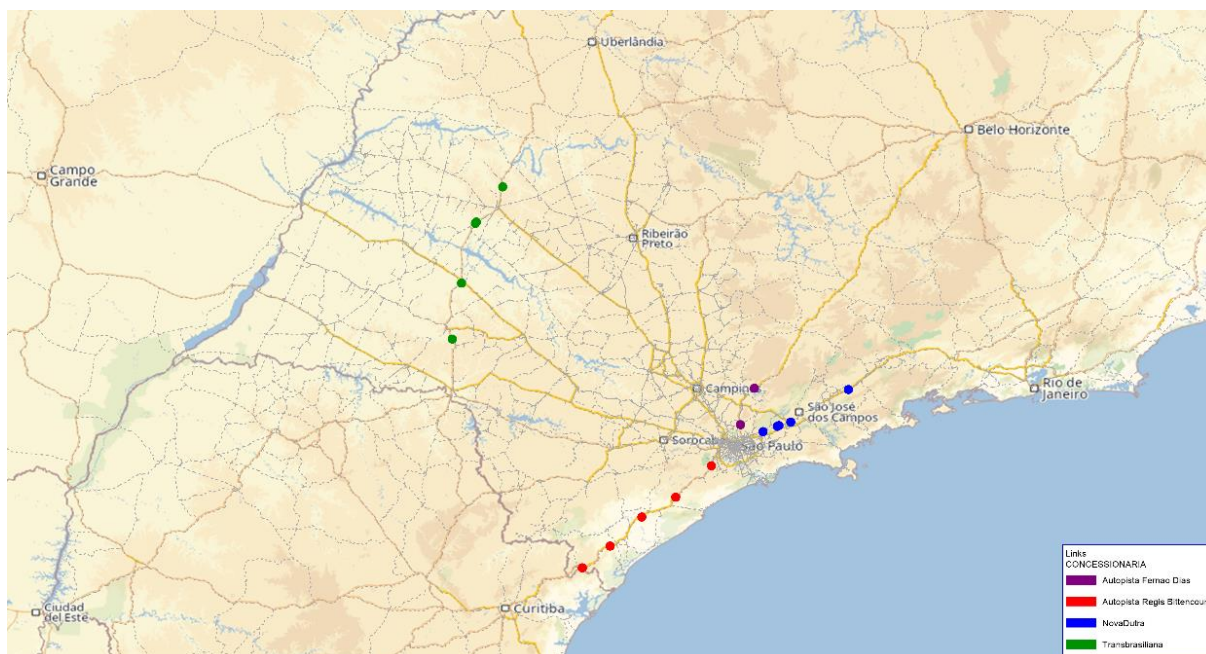


Figura 2-2: Pedágios Federais do Estado de São Paulo

Na Figura **2-3** a seguir pode-se visualizar as concessionárias federais em contraste com as estaduais dentro do estado de São Paulo.

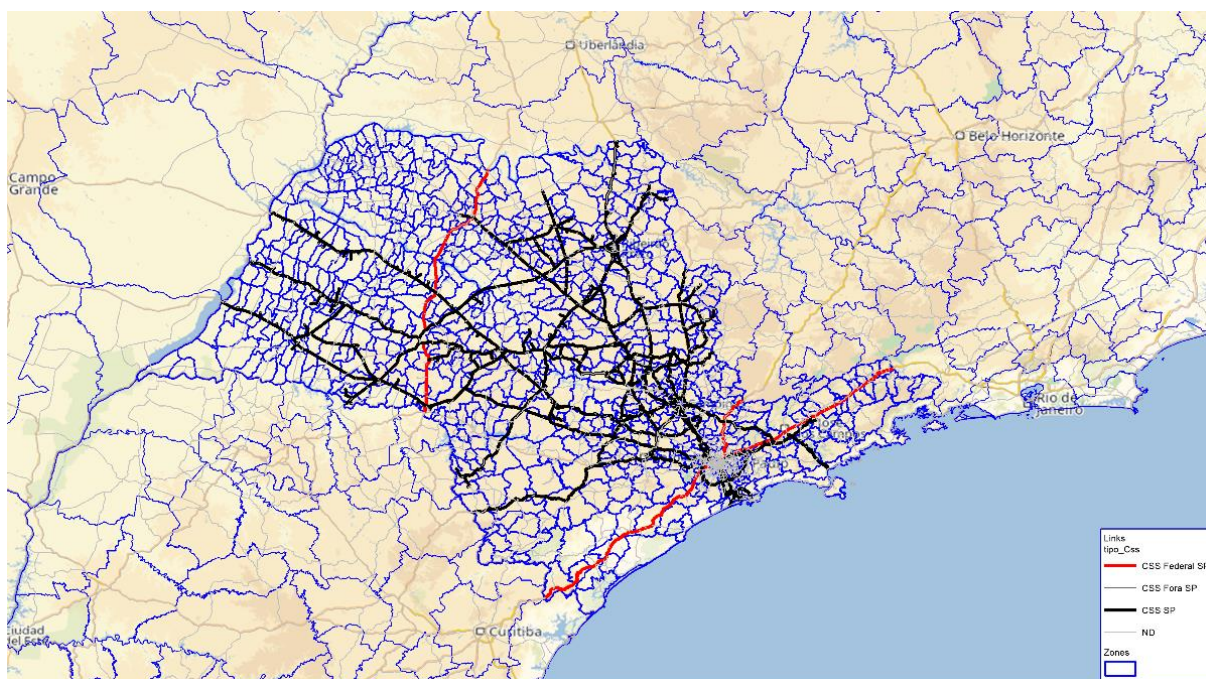


Figura 2-3: Concessionárias Federais e Estaduais do Estado de São Paulo

Foram acrescentadas as praças de pedágio fora do estado de São Paulo, com suas respectivas tarifas. A Figura **2-4** mostra Pedágios Fora do Estado de São Paulo.

Com a junção de diversas matrizes foi elaborada uma planilha com os pares OD que passam nas praças e comparada com os volumes classificados das praças. Esta é uma das entradas do modelo e quando da sua finalização será montado o método para replicar esta tarefa.

Tabela 2-1: Extrato matriz de link selecionado do pedágio

	P e d a g i o s		Lote_sele	L1	L1	L1	L1	L1	L1	L1	L1	L1	
		Concessão	CSS	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban
		Pedagio	Praça	SP 330-26	SP 330-26	SP 330-82	SP 330-82	SP 330-118	SP 330-118	SP 330-152	SP 330-152	SP 348-36	SP 348-36
		Sentido	Sentido	2	1	2	1	2	1	2	1	2	2
			cod_praça	SP 330-26	SP 330-26	SP 330-82	SP 330-82	SP 330-118	SP 330-118	SP 330-152	SP 330-152	SP 348-36	SP 348-36
			codPraça2	1001	1002	1003	1004	1005	1006	1007	1008	1009	1009
O	D	Key	Distancia	Ped1001	Ped1002	Ped1003	Ped1004	Ped1005	Ped1006	Ped1007	Ped1008	Ped1009	
11	393	11-393	77,797	0	0	0	0	0	0	0	0	0	
142	393	142-393	154,457	0	0	0	0	0	0	0	0	0	
211	1	211-1	326,229	0	0	0	0	0	0	0	0	0	
211	2	211-2	164,568	0	0	0	0	0	0	0	0	0	
211	10	211-10	393,926	0	0	0	0	0	0	0	0	0	
211	13	211-13	220,991	0	0	0	0	0	0	0	0	0	
211	16	211-16	377,802	0	0	0	0	0	0	0	0	0	
211	17	211-17	277,822	0	0	0	0	0	0	0	0	0	
211	18	211-18	334,21	0	0	0	0	0	0	0	0	0	
211	24	211-24	333,996	0	0	0	0	0	0	0	0	0	
211	27	211-27	380,916	0	0	0	0	0	0	0	0	0	
211	32	211-32	262,332	0	0	0	0	0	0	0	0	0	

a seguir é apresentado um extrato da planilha de passagem nas praças por par Origem e Destino.

Tabela 2-1: Extrato matriz de link selecionado do pedágio

	P e d a g i o s		Lote_sele	L1	L1	L1	L1	L1	L1	L1	L1	L1	
		Concessão	CSS	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban	Autoban
		Pedagio	Praça	SP 330-26	SP 330-26	SP 330-82	SP 330-82	SP 330-118	SP 330-118	SP 330-152	SP 330-152	SP 348-36	SP 348-36
		Sentido	Sentido	2	1	2	1	2	1	2	1	2	2
			cod_praça	SP 330-26	SP 330-26	SP 330-82	SP 330-82	SP 330-118	SP 330-118	SP 330-152	SP 330-152	SP 348-36	SP 348-36
			codPraça2	1001	1002	1003	1004	1005	1006	1007	1008	1009	1009
O	D	Key	Distancia	Ped1001	Ped1002	Ped1003	Ped1004	Ped1005	Ped1006	Ped1007	Ped1008	Ped1009	
11	393	11-393	77,797	0	0	0	0	0	0	0	0	0	
142	393	142-393	154,457	0	0	0	0	0	0	0	0	0	
211	1	211-1	326,229	0	0	0	0	0	0	0	0	0	
211	2	211-2	164,568	0	0	0	0	0	0	0	0	0	
211	10	211-10	393,926	0	0	0	0	0	0	0	0	0	
211	13	211-13	220,991	0	0	0	0	0	0	0	0	0	
211	16	211-16	377,802	0	0	0	0	0	0	0	0	0	
211	17	211-17	277,822	0	0	0	0	0	0	0	0	0	
211	18	211-18	334,21	0	0	0	0	0	0	0	0	0	
211	24	211-24	333,996	0	0	0	0	0	0	0	0	0	
211	27	211-27	380,916	0	0	0	0	0	0	0	0	0	
211	32	211-32	262,332	0	0	0	0	0	0	0	0	0	

Modelo e metodologia de cálculo de viagens por intermédio de deslocamento de celulares

3.1 Pressupostos teóricos

A escolha do modelo a ser perseguido para determinação da relação entre transações de celulares envolve a definição do método para cálculo do vetor de coeficientes. Esta relação pode ser definida pela fórmula abaixo:

$$VV_{P, CV, RA, ANO, MES, PD} = Tr [ML_P \cdot QT_{RA, ANO, MES, PD, CR, MOT_OD}] \cdot CA_{CR, MOT_OD}$$

Onde:

- $VV_{P, RA, ANO, MES, PD}$ é o volume por classes de veículo (CV) que passa pelo pedágio (P) em um determinado período de tempo (ANO, MES, PD) para determinado regime de atividade diária (RA);
- $Tr [\cdot]$ traço do produto entre a matriz de link selecionado do pedágio e a matriz de transações de deslocamento de celular;
- ML_P é a matriz de link selecionado correspondente ao pedágio(P);
- $QT_{RA, ANO, MES, PD, CR, MOT_OD}$ é a matriz de quantidade transações que passa pelo pedágio (P) em um determinado período de tempo (ANO, MES, PD) para determinado regime de atividade diária (RA).

Para proceder este cálculo considerou-se uma equação da forma:

$$y_1 = \alpha_1 \cdot y_2 + \delta_0 + \delta_1 \cdot x_1 + \dots + \delta_k \cdot x_k + \partial_1 \cdot z_1 + \dots + \partial_n \cdot z_n + u$$

Onde:

- $y_1 = VV_P, MÊS, PD$ (corresponde a uma amostra de 4% ou um dia útil equivalente médio por mês;
- $y_2 = QT_{MES, PD, CR, MOT_OD}$ considera-se y_2 variável endógena;
- $z_n =$ variáveis explicativas independentes;
- $y_2 = F(x_1, \dots, x_k, z_1, \dots, z_n) =$ variável endógena representada como função $F(.)$ de variáveis gravitacionais referentes a distância entre origem e destino de transações e intensidade de população e emprego nas zonas de origem e destino de transações de celular;
- $\partial_1.z_1 + \dots + \partial_n.z_n =$ variáveis dummy características dos pedágios e dos meses considerados na regressão.

Nesta metodologia de variáveis instrumentais considera-se que QT está significativamente correlacionada com a parcela de erro u que é um elemento desconhecido. Isto assim procede porque tanto o volume de pedágio quanto a quantidade de transações que passam pelo pedágio acontecem simultaneamente. A variável QT é então considerada uma variável endógena.

Para eliminar este problema pode-se aplicar variáveis instrumentais, porque este método elimina a correlação entre o termo de erro u e as variáveis independentes, de fato a variável endógena passa a ser representada por uma função F com erro $v \neq u$. Para realizar estimativas de variáveis instrumentais, precisamos ter instrumentos que são (a) não correlacionados com u mas (b) parcialmente e suficientemente fortemente correlacionados com y_2 uma vez que as outras variáveis independentes são controladas.

Na prática, podemos testar o segundo requisito (b), mas não podemos testar o primeiro requisito (a) porque u não é observável. Para testar o segundo requisito (b), precisamos expressar uma equação de forma reduzida de y_2 com todas as variáveis exógenas. As variáveis exógenas incluem todas as variáveis independentes que não estão correlacionadas com o termo de erro e a variável instrumental, z . A equação da forma reduzida para y_2 é:

$$y_2 = \delta_0 + \delta_1.z_1 + \dots + \delta_k.z_k + \partial_1.z_1 + \dots + \partial_n.z_n + v$$

Sendo que o requisito (b) será atendido se δ_z e ∂_n se apresentarem como estatisticamente significativos.

Este método foi aplicado aos dados obtidos de amostra aleatória de 4% de transações de deslocamento de aparelhos celulares sobre 4% dos volumes de veículos observados em 2455 estatísticas mensais de volume pedágios apurados no ano de 2019, abrangendo os sistemas estadual e federal de concessão.

O Software utilizado foi o STATA/BE 17.0 por intermédio da função `ivregress 2SLS` que consiste em Regressão Linear em 2 estágios.

Os resultados se mostraram promissores para aplicação no universo de pedágios do estado de São Paulo e para o período de março a dezembro de 2019. Os scripts são apresentados em anexo e já foram encaminhados à SLT.

3.2 Procedimentos para aplicação do modelo teórico

O roteiro dos procedimentos para Cálculo da Matriz O/D de veículos a partir de dados de telefonia foi construído conforme apresentado no algoritmo abaixo. Os títulos entre aspas (".") referem-se a exemplos fornecidos em relatórios anteriores.

3.2.1 Roteiro do Procedimento

- a) Variando mês de março de 2019 a dezembro de 2019
 - i. Os arquivos de telefonia "03_201903_telefonia_consolidado.xlsx" são importados pelo software STATA onde se procede a divisão em 25 amostras iguais com a função "splittsample" gerando o arquivo "f2019_03_004.xlsx".
 - ii. O arquivo de amostra de transações de telefone "f2019_#mes#_004.xlsx" gerado no passo anterior é colado na planilha 'mar' nas colunas A - K no arquivo "f2019_03_004_aux_mar19_L1.xlsx".
- b) Variando Lote de pedágio de 1 a 10:
 - i. O arquivo Flow_L1.xlsx é o select link analysis do lote1 é colado na planilha 'Lote1' no arquivo "f2019_03_004_aux_mar19_L1.xlsx";
 - ii. A planilha 'mar' do arquivo "f2019_03_004_aux_mar19_L1.xlsx" são selecionados os pares O/D que passam pelo Lote 1 e possuem transações de celular que são colados como valores na planilha 'base' do arquivo "f2019_03_004_aux1_L3.xlsx";
 - iii. Os valores totais de variáveis dependentes e independentes do Lote1 são calculados e consolidados na planilha 'mar_2019_L03' do arquivo "f2019_03_004_aux1_L3.xlsx".
 - iv. Os valores gerados na planilha 'mar_2019_L03' do arquivo "f2019_03_004_aux1_L3.xlsx" são consolidados no arquivo "InputStata_final.dta.xlsx" onde é calculado o modelo de regressão de variáveis instrumentais para cada tipo de veículo, gerando resultados em STATA.

- c) Aplicação dos fatores da equação de conversão sobre os dados de transação de telefones celulares EXCEL. "sub_matriz_pedagio_anual_Mar.xlsx"
- d) Correção das matrizes origem e destino para variáveis dummy de lotes e meses EXCEL/VISUM. Obtenção das matrizes preliminares de dia útil, final de semana e feriado mensal "sub_matriz_pedagio_anual_Mar.xlsx"
- e) Obtenção das matrizes mensais corrigidas finais, por tipo de veículo EXCEL. Aplicação do peso de dia útil, feriado e final de semana em cada mês para obtenção da matriz média anual de cada uma das categorias diárias.
- f) Obtenção das matrizes B por tipo de veículo e matrizes A gerais por regime de atividade.
- g) Alocação em rede VISUM.

O Fluxograma a seguir apresenta o procedimento desenvolvido:

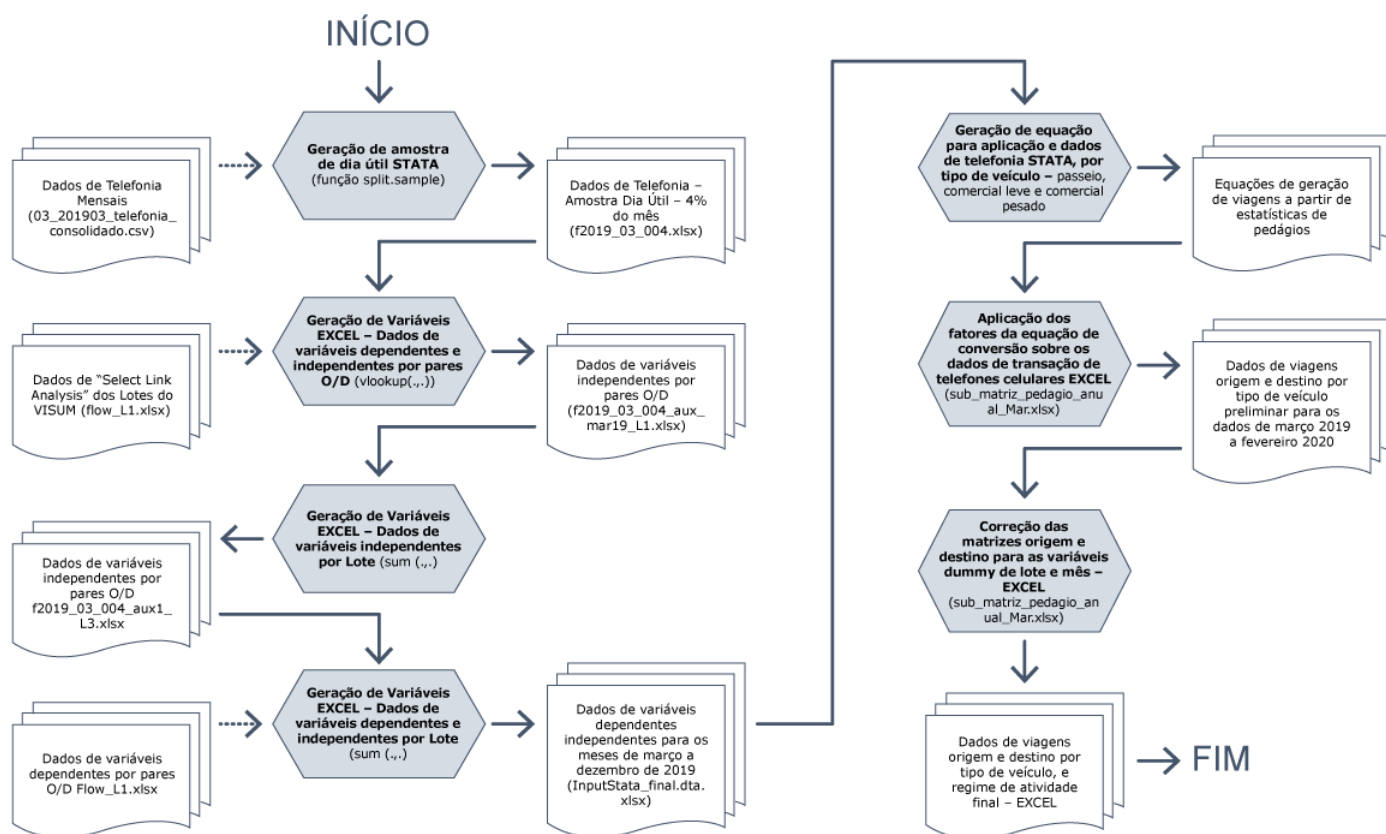


Figura 3-1 - procedimento desenvolvido

3.3 Descrição de aplicação e indicadores para avaliação dos resultados estatísticos calibrados

a) Sumário de Aplicação:

As transações de celulares são estimadas em um primeiro estágio em uma regressão linear múltipla de mínimos quadrados ordinários.

As variáveis independentes instrumentais deste primeiro estágio foram construídas a partir de dados da geografia e socio economia e não têm o mesmo erro compartilhado por viagens de veículo e transações de celular.

A variável endógena de transações de celular é então estimada neste primeiro estágio, obtendo um erro diferente daquele erro primariamente compartilhado com as viagens de veículos.

A variável endógena junta-se a outras variáveis independentes de caráter idiossincrático (dummies) especificadas pelo lote e mês em que se realizaram viagens e transações.

b) Indicadores de Calibração:

A equação de segundo estágio é então calculada fornecendo informações sobre o ajustamento, robustez dos coeficientes estimados, e testes de endogeneidade e viés de estimativa no primeiro estágio.

- Os parâmetros gerais de calibração visam obter uma avaliação geral do ajustamento pelas estatísticas "F" e "R²" e da robustez dos coeficientes obtidos;
- O teste de Bootstrap indica se a amostra coletada de 4% das transações de celulares foi suficientemente representativa, de forma a fornecer coeficientes robustos e significativos para um total de 2000 replicações de ajustamento para amostragem com reposição;
- O teste de endogeneidade identifica se há evidências para se considerar que as variáveis de transações de celulares apresentam endogeneidade em relação a viagens de veículos de passeio, comerciais leves e comerciais pesados por meio dos testes de Durbin (chi²) e Wu-Hausman F;
- O teste sobre identificação visa garantir que as variáveis instrumentais não estejam correlacionadas com o erro estrutural, por meio dos testes de Sargan (chi²), Basmann (chi²);
- O teste de Primeiro Estágio visa identificar se as variáveis instrumentais estão correlacionadas com a variável endógena ou se fornecem uma estimativa viesada desta.

3.3.1 Resultados para veículos de passeio

a) Ajustamento do Modelo:

A seguir é apresentada a **Tabela 3-1** com o resultado obtido através do software STATA para os veículos de passeio.

Tabela 3-1: Ajustamento Veículos de Passeio

Instrumental variables 2SLS regression				Number of obs	=	2,455
				Wald chi2(18)	=	2618.57
				Prob > chi2	=	0.0000
				R-squared	=	0.5391
				Root MSE	=	9964.6
VP	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
CR_ABCD	1.483492	.060319	24.59	0.000	1.365269	1.601715
dum_m4	-2196.065	853.6307	-2.57	0.010	-3869.15	-522.9795
dum_m5	773.3994	852.5416	0.91	0.364	-897.5515	2444.35
dum_m6	-2595.545	854.1949	-3.04	0.002	-4269.737	-921.3541
dum_m7	-1321.301	868.3754	-1.52	0.128	-3023.286	380.6835
dum_m8	1350.245	851.7935	1.59	0.113	-319.2399	3019.729
dum_m9	1548.066	852.7454	1.82	0.069	-123.2842	3219.416
dum_m10	2482.149	854.2119	2.91	0.004	807.9243	4156.373
dum_m11	-385.8297	850.3482	-0.45	0.650	-2052.482	1280.822
dum_l2	-10672.81	1070.217	-9.97	0.000	-12770.4	-8575.224
dum_l3	-5493.408	1004.011	-5.47	0.000	-7461.233	-3525.583
dum_l4	-11005.73	954.6917	-11.53	0.000	-12876.89	-9134.565
dum_l5	-12617.56	1006.99	-12.53	0.000	-14591.22	-10643.9
dum_l6	-10606.13	1076.069	-9.86	0.000	-12715.19	-8497.075
dum_l7	-12557.03	1053.942	-11.91	0.000	-14622.72	-10491.34
dum_l8	-12446.1	1041.112	-11.95	0.000	-14486.64	-10405.56
dum_l9	-13235.61	1040.616	-12.72	0.000	-15275.18	-11196.04
dum_l10	-3966.132	1081.442	-3.67	0.000	-6085.72	-1846.544
_cons	11475.83	1144.011	10.03	0.000	9233.608	13718.05
Instrumented: CR_ABCD						
Instruments: dum_m4 dum_m5 dum_m6 dum_m7 dum_m8 dum_m9 dum_m10 dum_m11						
dum_l2 dum_l3 dum_l4 dum_l5 dum_l6 dum_l7 dum_l8 dum_l9						
dum_l10 lgrv_pe ldist						

Considerações:

- y_1 = VVP, ANO, MES, = VP = 4% volumes passeio passando pelo pedágio no mês
- y_2 = QTP, ANO, MES, = CR_ABCD = transações de celulares de usuários das classes A, B, C, D
- x_n = variáveis explicativas independentes ldist, lgrv_ep, lgrv_pe

Onde:

- $ldist = \sum (\ln (dist_{ij}), ij \in MLP$
- $lgrv_pe = \sum (\ln (população_i \cdot emprego_j), ij \in MLP$
- dum_m4, ..., dum_m11 = variáveis dummy indicativas dos meses abril a novembro de 2019
- dum_l2, ..., dum_l10 = variáveis dummy indicativas dos lotes 2 a 10 de pedágios

Os lotes de processamento são apresentados na figura a seguir.

c) Teste Endogeneidade:

Abaixo são apresentados os resultados obtidos para rejeição da hipótese de que a variável CR_ABCD seja exógena.

Tabela 3-3: Teste de Endogeneidade Veículos de Passeio

```
estat endogenous CR_ABCD

Tests of endogeneity
H0: Variables are exogenous

Durbin (score) chi2(1)          = 35.5916 (p = 0.0000)
Wu-Hausman F(1,2435)           = 35.8209 (p = 0.0000)
```

d) Teste de sobre Identificação

Abaixo são apresentados os resultados obtidos para rejeição da hipótese de que as variáveis instrumentais estejam estruturalmente correlacionadas com as variáveis dependentes.

Tabela 3-4: Teste de sobre Identificação Veículos de Passeio

```
estat overid

Tests of overidentifying restrictions:

Sargan (score) chi2(1) = 12.8454 (p = 0.0003)
Basman chi2(1)         = 12.8077 (p = 0.0003)
```

e) Validade da Equação de Primeiro Estágio

Abaixo são apresentados os resultados obtidos para rejeição das hipóteses de que as variáveis instrumentais e independentes não sejam correlacionadas com a variável endógena ou representem risco de viés nos resultados.

Tabela 3-5: Teste Validade da equação de Primeiro Estágio Veículos de Passeio

estat firststage					
First-stage regression summary statistics					
Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(2,2435)	Prob > F
CR_ABCD	0.6497	0.6470	0.4546	1014.78	0.0000
Minimum eigenvalue statistic = 1014.78					
Critical Values		# of endogenous regressors:		1	
H0: Instruments are weak		# of excluded instruments:		2	
2SLS relative bias			5%	10%	20%
			(not available)		
2SLS size of nominal 5% Wald test			10%	15%	20%
LIML size of nominal 5% Wald test			19.93	11.59	8.75
			8.68	5.33	4.42
			25%		
			7.25		
			3.92		

3.3.2 Resultados para veículos comerciais leves

a) Ajustamento do Modelo:

A seguir é apresentada a Tabela 3-6 com o resultado obtido através do software STATA para veículos comerciais leves.

Tabela 3-6: Ajustamento Veículos Comerciais Leves

mental variables 2SLS regression					Number of obs = 2,455	
					Wald chi2(18) = 3720.95	
					Prob > chi2 = 0.0000	
					R-squared = 0.5900	
					Root MSE = 1101.7	
VCL	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
CR_BCDE	.1867095	.0078131	23.90	0.000	.1713961	.2020228
dum_m4	-250.6005	98.17285	-2.55	0.011	-443.0157	-58.18521
dum_m5	268.8604	100.1409	2.68	0.007	72.58787	465.133
dum_m6	-392.7238	95.12263	-4.13	0.000	-579.1607	-206.2868
dum_m7	-140.98	104.9054	-1.34	0.179	-346.5908	64.63075
dum_m8	360.7016	95.64084	3.77	0.000	173.249	548.1542
dum_m9	357.4073	92.09766	3.88	0.000	176.8992	537.9154
dum_m10	495.7752	94.93466	5.22	0.000	309.7067	681.8437
dum_m11	7.738792	101.7651	0.08	0.939	-191.7171	207.1947
dum_l2	-969.0578	145.933	-6.64	0.000	-1255.081	-683.0344
dum_l3	-856.3955	152.9164	-5.60	0.000	-1156.106	-556.6849
dum_l4	-1112.516	138.0633	-8.06	0.000	-1383.115	-841.9173
dum_l5	-1104.118	152.2558	-7.25	0.000	-1402.534	-805.7016
dum_l6	-1052.383	145.4583	-7.23	0.000	-1337.476	-767.2902
dum_l7	-1303.668	148.3621	-8.79	0.000	-1594.453	-1012.884
dum_l8	-1243.892	141.2186	-8.81	0.000	-1520.675	-967.1083
dum_l9	-1310.607	140.5792	-9.32	0.000	-1586.137	-1035.077
dum_l10	239.0653	268.8431	0.89	0.374	-287.8576	765.9881
_cons	1187.576	163.4717	7.26	0.000	867.1777	1507.975

Considerações:

- y_1 = VVP, ANO, MES, = VCL=4% volumes de veículos comerciais leves passando pelo pedágio no mês
- y_2 = QTP, ANO, MES, = CR_BCDE= transações de celulares de usuários das classes A, B, C, D.
- x_n = variáveis explicativas independentes ldist, lgrv_ep, lgrv_pe

Onde:

- $ldist = \sum (\ln (dist_{ij}), ij \in MLP$
- $lgrv_{ee} = \sum (\ln (emprego_i.emprego_j), ij \in MLP$
- dum_m4, ..., dum_m11 = variáveis dummy indicativas dos meses abril a novembro de 2019.
- dum_l2, ..., dum_l10 = variáveis dummy indicativas dos lotes 2 a 10 de pedágios.

b) Teste de Robustez dos Coeficientes:

Abaixo são apresentados os resultados obtidos após a aplicação de 2000 replicações do modelo pelo método Bootstrap. Observa-se que o coeficiente CR_BCDE e a maioria dos coeficientes dummy se mostram significativos com $P > |z|$ inferior a 5%.

Tabela 3-7: Teste de Robustez Veículos de Comerciais Leves

VCL	Observed coefficient	Bootstrap std. err.	z	P> z	Normal-based [95% conf. interval]	
CR_BCDE	.1867095	.0079556	23.47	0.000	.1711167	.2023022
dum_m4	-250.6005	94.48552	-2.65	0.008	-435.7887	-65.41226
dum_m5	268.8604	98.84851	2.72	0.007	75.12091	462.6
dum_m6	-392.7238	89.09466	-4.41	0.000	-567.3461	-218.1014
dum_m7	-140.98	101.9912	-1.38	0.167	-340.879	58.91898
dum_m8	360.7016	93.35933	3.86	0.000	177.7207	543.6826
dum_m9	357.4073	88.66549	4.03	0.000	183.6261	531.1884
dum_m10	495.7752	89.64081	5.53	0.000	320.0824	671.4679
dum_m11	7.738792	99.45465	0.08	0.938	-187.1887	202.6663
dum_m12	-969.0578	142.5517	-6.80	0.000	-1248.454	-689.6615
dum_m13	-856.3955	152.4669	-5.62	0.000	-1155.225	-557.5658
dum_m14	-1112.516	139.0489	-8.00	0.000	-1385.047	-839.9856
dum_m15	-1104.118	147.9576	-7.46	0.000	-1394.109	-814.126
dum_m16	-1052.383	145.048	-7.26	0.000	-1336.672	-768.0942
dum_m17	-1303.668	145.9984	-8.93	0.000	-1589.82	-1017.517
dum_m18	-1243.892	141.0739	-8.82	0.000	-1520.391	-967.3919
dum_m19	-1310.607	140.2837	-9.34	0.000	-1585.558	-1035.656
dum_m10	239.0653	266.3414	0.90	0.369	-282.9544	761.0849
_cons	1187.576	164.4088	7.22	0.000	865.3411	1509.812

c) Teste Endogeneidade:

Abaixo são apresentados os resultados obtidos para rejeição da hipótese de que a variável CR_BCDE seja exógena.

Tabela 3-8: Teste de Endogeneidade Veículos Comerciais Leves

dum_m17	-1303.668	148.3621	-8.79	0.000	-1594.453
dum_m18	-1243.892	141.2186	-8.81	0.000	-1520.675
dum_m19	-1310.607	140.5792	-9.32	0.000	-1586.137
dum_m10	239.0653	268.8431	0.89	0.374	-287.8576
_cons	1187.576	163.4717	7.26	0.000	867.1777

d) Teste de sobre Identificação:

Abaixo são apresentados os resultados obtidos para rejeição da hipótese de que as variáveis instrumentais estejam estruturalmente correlacionadas com as variáveis dependentes.

Tabela 3-9: Teste de Sobre Identificação Veículos Comerciais Leves

```
Instrumented: CR_BCDE
Instruments: dum_m4 dum_m5 dum_m6 dum_m7 dum_m8 dum_m9 dum_m10 dum_m
dum_m12 dum_m13 dum_m14 dum_m15 dum_m16 dum_m17 dum_m18 dum_m19
dum_m10 lgrv_ee ldist
```

e) Validade da Equação de Primeiro Estágio:

Abaixo são apresentados os resultados obtidos para rejeição das hipóteses de que as variáveis instrumentais e independentes não sejam correlacionadas com a variável endógena ou representem risco de vies nos resultados.

Tabela 3-10: Teste de Primeiro Estágio Veículos Comerciais Leves

estat firststage

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(2,2435)	Prob > F
CR_BCDE	0.6627	0.6601	0.4710	1084.14	0.0000

Minimum eigenvalue statistic = 1084.14

Critical Values

of endogenous regressors: 1

H0: Instruments are weak

of excluded instruments: 2

2SLS relative bias	5%	10%	20%	30%
	(not available)			
2SLS size of nominal 5% Wald test	19.93	11.59	8.75	7.25
LIML size of nominal 5% Wald test	8.68	5.33	4.42	3.92

3.3.3 Resultados para Veículos Comerciais Pesados:

a) Ajustamento do Modelo:

A seguir é apresentada a Tabela 3-11 com o resultado obtido através do software STATA para veículos comerciais pesados.

Tabela 3-11: Ajustamento Veículos Comerciais Pesados

Instrumental variables 2SLS regression					
Number of obs = 2,451					
Wald chi2(18) = 1448.38					
Prob > chi2 = 0.0000					
R-squared = 0.3650					
Root MSE = 1430.8					
VCP	Coefficient	Std. err.	z	P> z	[95% conf. interval]
CTTCOO	.2913376	.0164044	17.76	0.000	.2591856 .3234895
dum_m4	-.298.101	123.1994	-2.42	0.016	-539.5673 -56.63463
dum_m5	91.8477	122.8259	0.75	0.455	-148.8866 332.582
dum_m6	-356.9369	123.2476	-2.90	0.004	-598.4978 -115.3761
dum_m7	-140.5118	125.0495	-1.12	0.261	-385.6043 104.5806
dum_m8	186.8309	122.7063	1.52	0.128	-53.66899 427.3309
dum_m9	123.4989	122.6642	1.01	0.314	-116.9185 363.9164
dum_m10	276.1936	122.8811	2.25	0.025	35.35101 517.0362
dum_m11	-163.7145	122.8458	-1.33	0.183	-404.4877 77.05882
dum_12	1148.661	146.5595	7.84	0.000	861.4099 1435.912
dum_13	-854.4071	140.1317	-6.10	0.000	-1129.06 -579.7541
dum_14	87.81113	134.8617	0.65	0.515	-176.513 352.1353
dum_15	-940.1957	140.253	-6.70	0.000	-1215.086 -665.305
dum_16	-665.3305	147.1991	-4.52	0.000	-953.8354 -376.8257
dum_17	-1094.216	146.1031	-7.49	0.000	-1380.572 -807.859
dum_18	-947.9392	145.2454	-6.53	0.000	-1232.615 -663.2634
dum_19	-1041.738	146.0792	-7.13	0.000	-1328.048 -755.4284
dum_110	282.5717	155.256	1.82	0.069	-21.72442 586.8679
_cons	1085.661	157.3456	6.90	0.000	777.2697 1394.053
Instrumented: CTTCOO					
Instruments: dum_m4 dum_m5 dum_m6 dum_m7 dum_m8 dum_m9 dum_m10 dum_m11 dum_12 dum_13 dum_14 dum_15 dum_16 dum_17 dum_18 dum_19 dum_110 lgrv_ee ldist					

Considerações:

- y_1 = VVP, ANO, MES, = VCP=4% volumes de veículo comerciais pesados passando pelo pedágio no mês
- y_2 = QTP, ANO, MES, = CTTCOO= transações de celulares de usuários nos motivos casa-trabalho, trabalho-casa e outro-outro
- x_n = variáveis explicativas independentes ldist, lgrv_ep, lgrv_pe

Onde:

- $ldist = \sum(\ln(dist_{ij}), ij \in MLP$
- $lgrv_{ee} = \sum(\ln(emprego_i.emprego_j), ij \in MLP$
- dum_m4, ..., dum_m11 = variáveis dummy indicativas dos meses abril a novembro de 2019
- dum_l2, ..., dum_l10 = variáveis dummy indicativas dos lotes 2 a 10 de pedágios

b) Teste de Robustez dos Coeficientes:

Abaixo são apresentados os resultados obtidos após a aplicação de 2000 replicações do modelo pelo método Bootstrap. Observa-se que o coeficiente CTTCOO e a maioria dos coeficientes dummy se mostram significativos com $P > |z|$ inferior a 5%.

Tabela 3-12: Teste de Robustez Veículos de Comerciais Pesados

VCP	Observed coefficient	Bootstrap std. err.	z	P> z	Normal-based [95% conf. interval]	
CTTCOO	.2913376	.0175443	16.61	0.000	.2569514	.3257237
dum_l10	282.5717	224.9075	1.26	0.209	-158.2389	723.3823
dum_l2	1148.661	220.5137	5.21	0.000	716.4623	1580.86
dum_l3	-854.4071	139.9035	-6.11	0.000	-1128.613	-580.2013
dum_l4	87.81113	179.6216	0.49	0.625	-264.2407	439.863
dum_l5	-940.1957	140.4401	-6.69	0.000	-1215.453	-664.9382
dum_l6	-665.3305	142.7253	-4.66	0.000	-945.0669	-385.5941
dum_l7	-1094.216	141.4158	-7.74	0.000	-1371.386	-817.0458
dum_l8	-947.9392	141.2894	-6.71	0.000	-1224.861	-671.0171
dum_l9	-1041.738	138.6943	-7.51	0.000	-1313.574	-769.9025
dum_m4	-298.101	123.0192	-2.42	0.015	-539.2142	-56.98773
dum_m5	91.8477	122.6332	0.75	0.454	-148.509	332.2044
dum_m6	-356.9369	118.5796	-3.01	0.003	-589.3486	-124.5252
dum_m7	-140.5118	129.8805	-1.08	0.279	-395.073	114.0494
dum_m8	186.8309	122.9945	1.52	0.129	-54.23377	427.8957
dum_m9	123.4989	123.5749	1.00	0.318	-118.7035	365.7013
dum_m10	276.1936	127.2695	2.17	0.030	26.75002	525.6372
dum_m11	-163.7145	126.0715	-1.30	0.194	-410.81	83.38107
_cons	1085.661	171.4505	6.33	0.000	749.6245	1421.698

c) Teste Endogeneidade:

Abaixo são apresentados os resultados obtidos para rejeição da hipótese de que a variável CR_BCDE seja exógena.

Tabela 3-13: Teste de Endogeneidade Veículos Comerciais Pesados

```
estat endogenous CTTC00

Tests of endogeneity
H0: Variables are exogenous

Durbin (score) chi2(1)      =  50.9709  (p = 0.0000)
Wu-Hausman F(1,2431)       =  51.6286  (p = 0.0000)
```

d) Teste de sobre Identificação:

Abaixo são apresentados os resultados obtidos para rejeição da hipótese de que as variáveis instrumentais estejam estruturalmente correlacionadas com as variáveis dependentes.

Tabela 3-14: Teste de Sobre Identificação Veículos Comerciais Pesados

```
estat overid

Tests of overidentifying restrictions:

Sargan (score) chi2(1) =  7.03109  (p = 0.0080)
Basmann chi2(1)      =  6.99378  (p = 0.0082)
```

e) Validade da equação de Primeiro Estágio:

Abaixo são apresentados os resultados obtidos para rejeição das hipóteses de que as variáveis instrumentais e independentes não sejam correlacionadas com a variável endógena ou representem risco de viés nos resultados.

Tabela 3-15: Teste de Primeiro Estágio Veículos Comerciais Pesados

estat firststage

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(2,2431)	Prob > F
CTTC00	0.6614	0.6587	0.5007	1218.81	0.0000

Minimum eigenvalue statistic = 1218.81

Critical Values

H0: Instruments are weak

of endogenous regressors: 1

of excluded instruments: 2

2SLS relative bias	5%	10%	20%	30%
	(not available)			
2SLS size of nominal 5% Wald test	10%	15%	20%	25%
LIML size of nominal 5% Wald test	19.93	11.59	8.75	7.25
	8.68	5.33	4.42	3.92

4

Resultados do ajustamento

Resultados do ajustamento estatístico e na alocação de rede.

4.1 Ajustamento Veículo Passeio

Conforme apresentado na figura abaixo, os dados de veículo passeio apresentam significativa redução da dispersão após a aplicação do modelo.

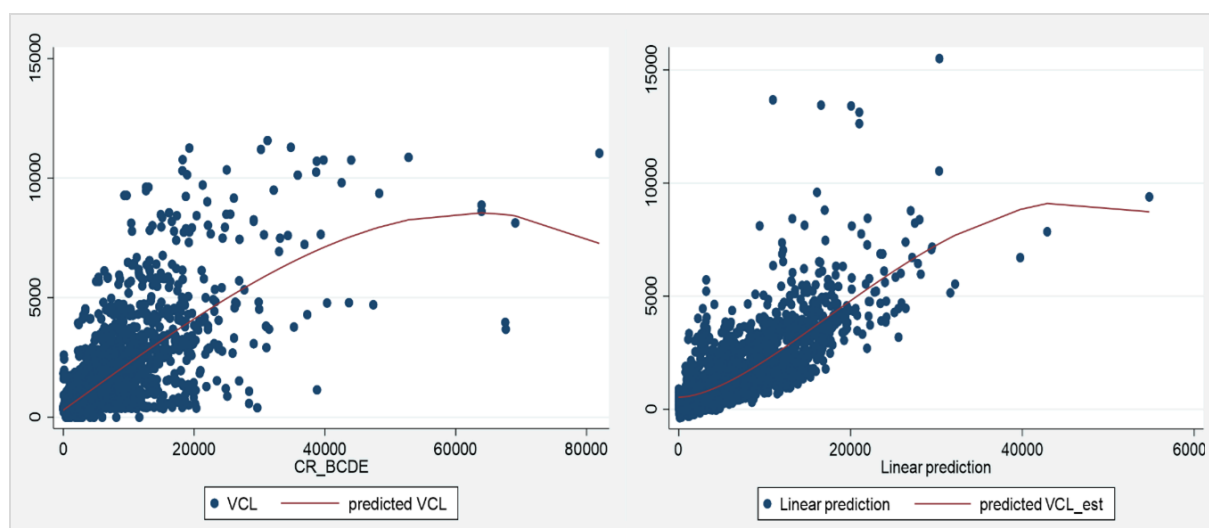


Figura 4-1: Ajuste Veículos de Passeio

A aplicação dos coeficientes encontrados em regressão sobre as matrizes de transações produz os resultados de alocação de pedágio em rede de simulação com os resultados de correlação entre observado em pedágio e simulado $\text{pwcorr} = .7955$, considerado satisfatório.

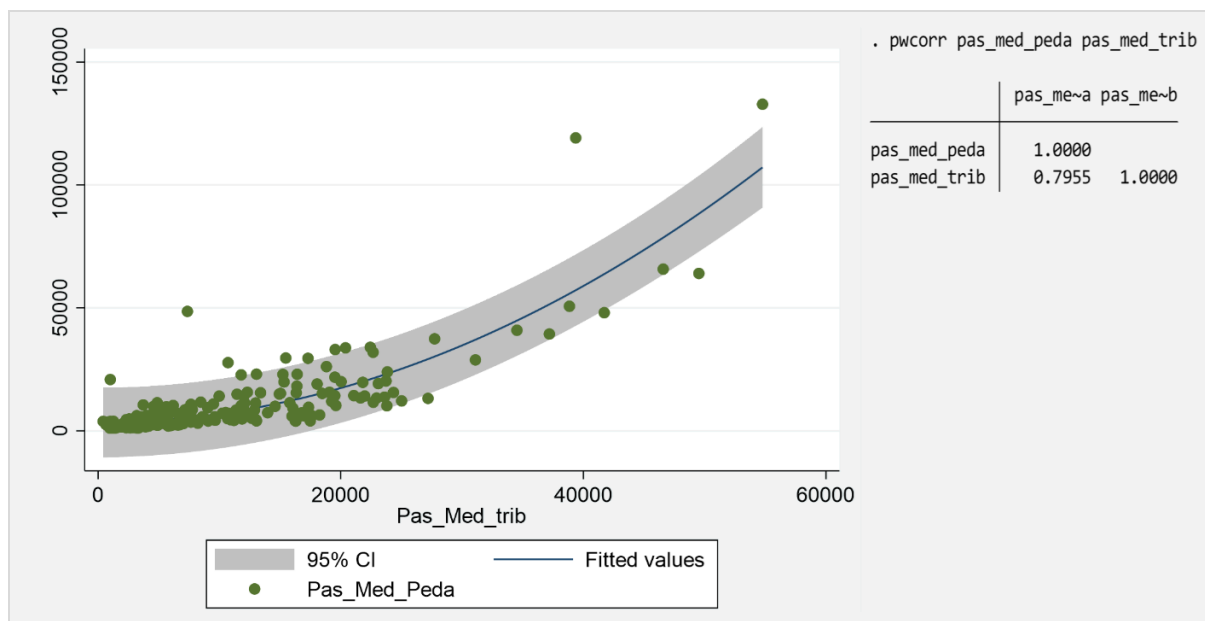


Figura 4-2: Calibração Veículos de Passeio

4.2 Ajustamento Veículo Comercial Leve

Conforme apresentado na figura abaixo, os dados de veículo passeio apresentam significativa redução da dispersão após a aplicação do modelo.

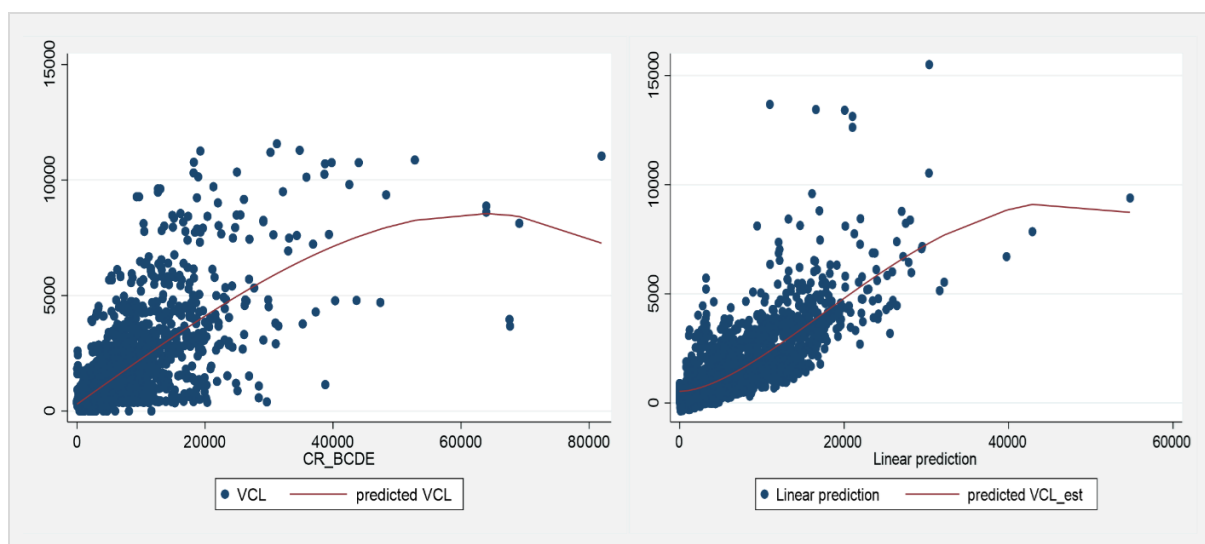


Figura 4-3: Ajuste Veículos Comerciais Leves

A aplicação dos coeficientes encontrados em regressão sobre as matrizes de transações produz os resultados de alocação de pedágio em rede de simulação com os resultados de correlação entre observado em pedágio e simulado $pwcorr = .5139$, considerado satisfatório.

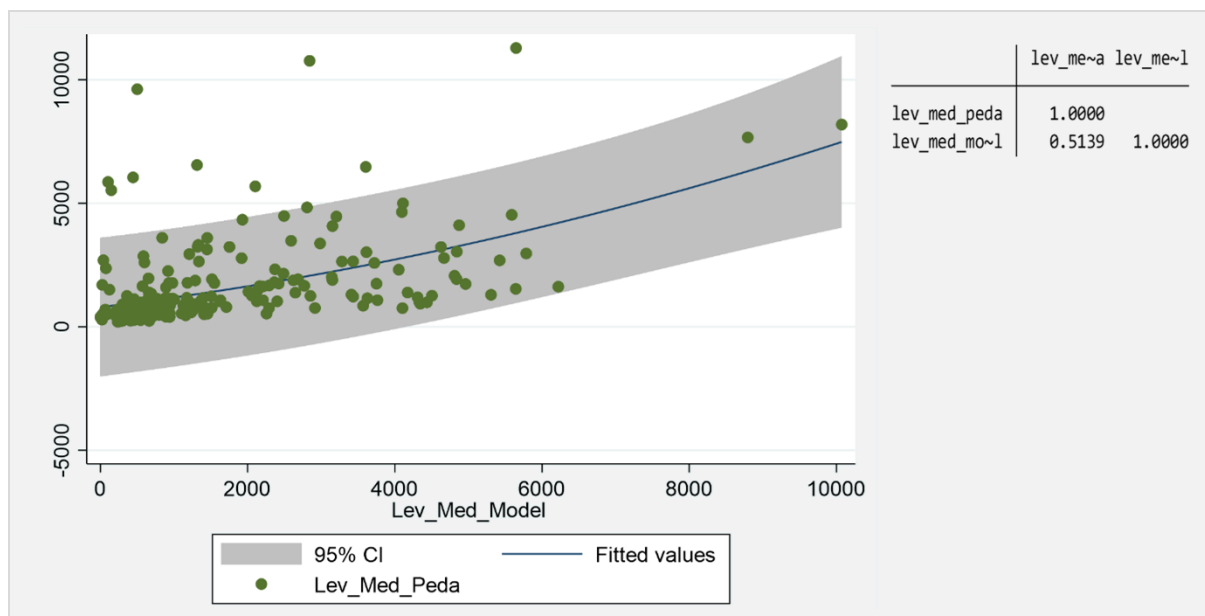


Figura 4-4: Calibração Veículos Comerciais Leves

4.3 Ajustamento Veículo Comercial Pesado

Conforme apresentado na figura abaixo, os dados de veículo passeio apresentam significativa redução da dispersão após a aplicação do modelo.

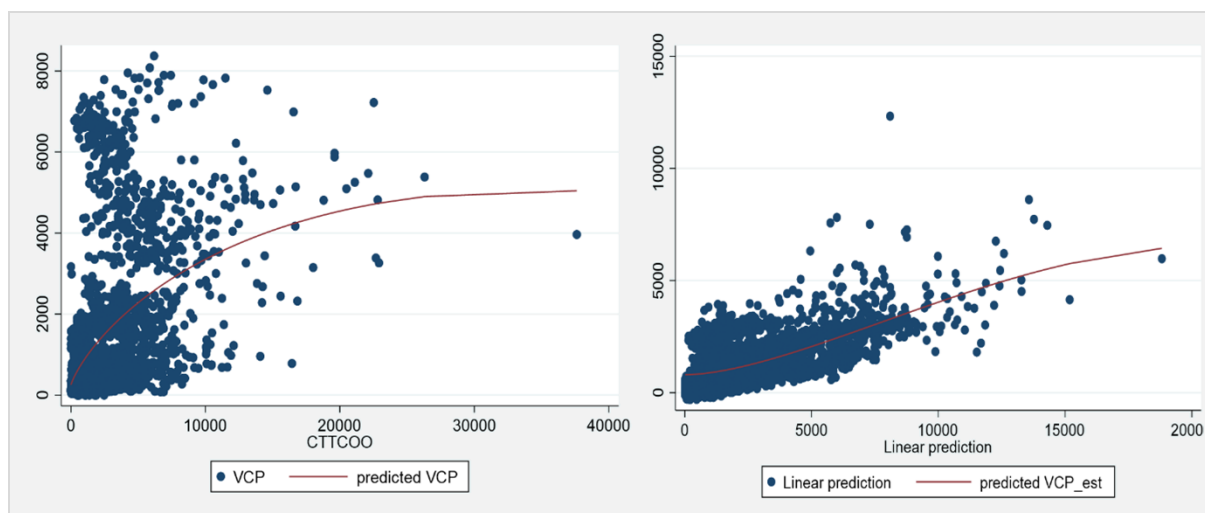


Figura 4-5: Ajuste Veículos Comerciais Pesados

A aplicação dos coeficientes encontrados em regressão sobre as matrizes de transações produz os resultados de alocação de pedágio em rede de simulação com os resultados de correlação entre observado em pedágio e simulado $\text{pwcorr} = .6709$, considerado satisfatório.

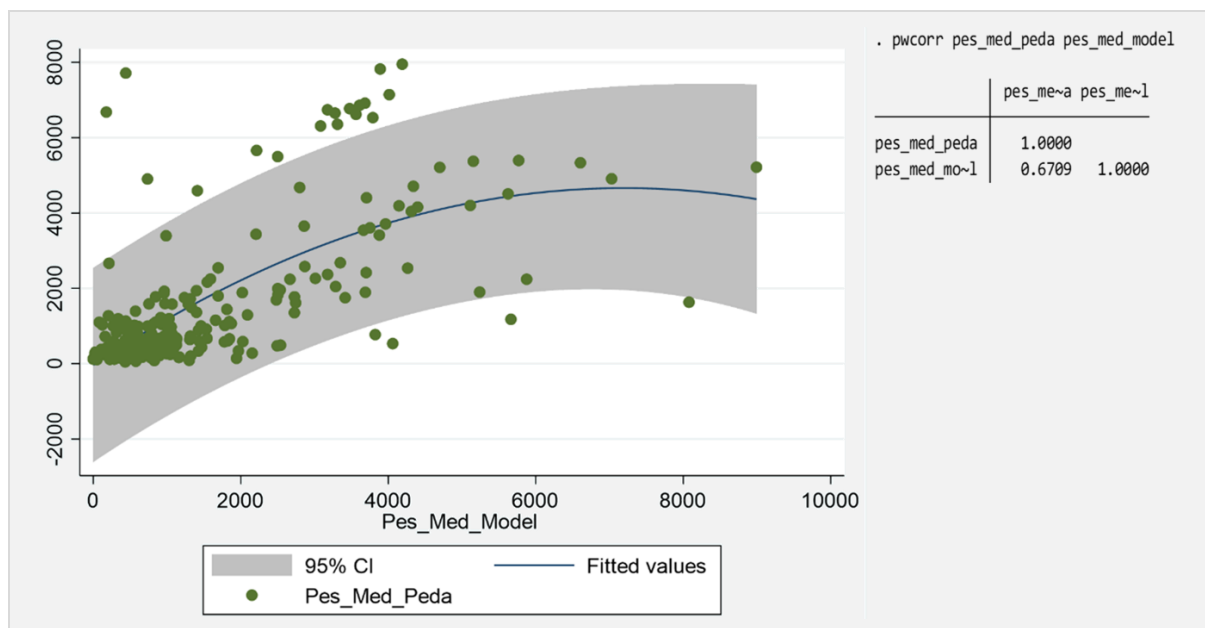


Figura 4-6: Calibração Veículos Comerciais Pesados

4.4 Ajustamento Veículo Dia Útil

Os resultados de ajustamento quando expandidos para o período anual por regime de atividade em dia útil apresentaram melhoria na calibração de rede com coeficientes correlação pareados de 0,8755, 0,6772 e 0,7742 para passeio, comercial leve e pesado respectivamente.

4.4.1 Veículo Passeio

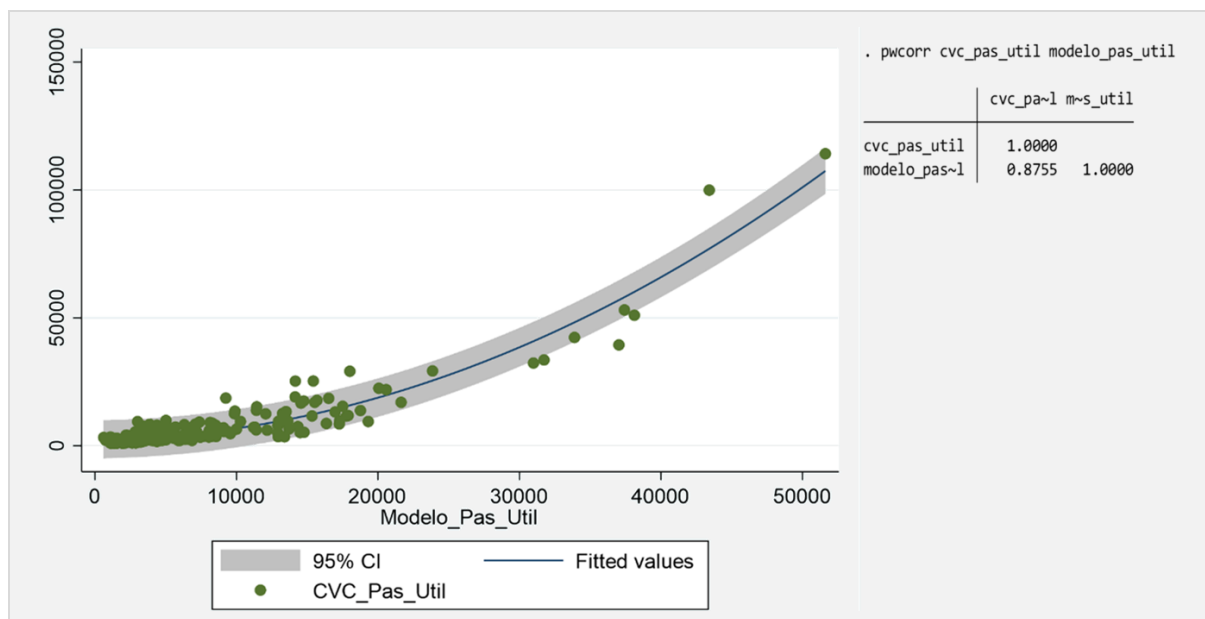


Figura 4-7: Calibração Veículos de Passeio – Dia Útil

4.4.2 Veículo Comercial Leve

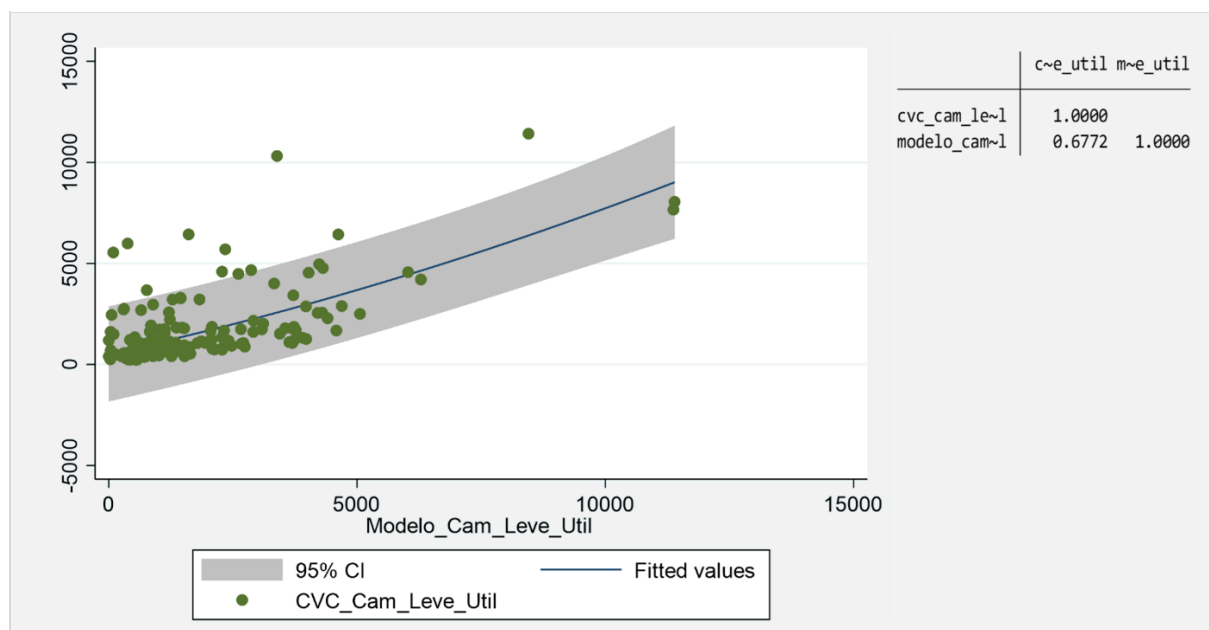


Figura 4-8: Calibração Veículos Comerciais Leves – Dia Útil

4.4.3 Veículo Comercial Pesado

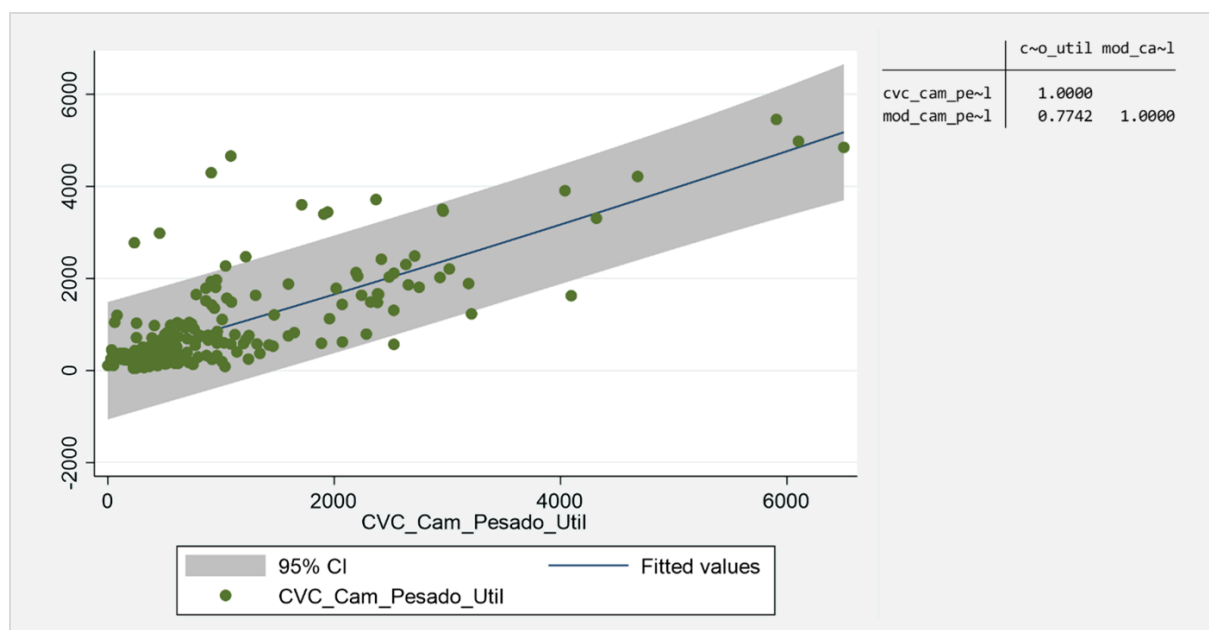


Figura 4-9: Calibração Veículos Comerciais Pesados – Dia Útil

4.5 Ajustamento Fim de Semana

Os resultados de ajustamento quando expandidos para o período anual por regime de atividade em dia de fim de semana apresentaram melhoria na calibração de rede com coeficientes

correlação pareados de 0,8516, 0,6592 e 0,7068 para passeio, comercial leve e pesado respectivamente.

4.5.1 Veículo Passeio

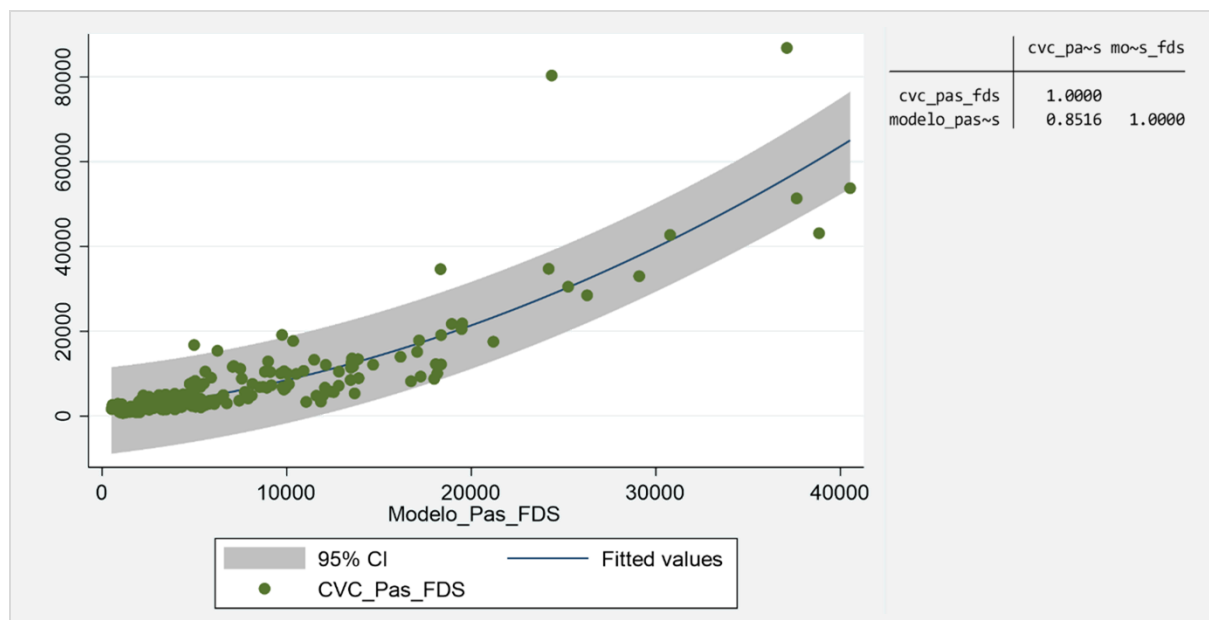


Figura 4-10: Calibração Veículos de Passeio – Final de Semana

4.5.2 Veículo Comercial Leve

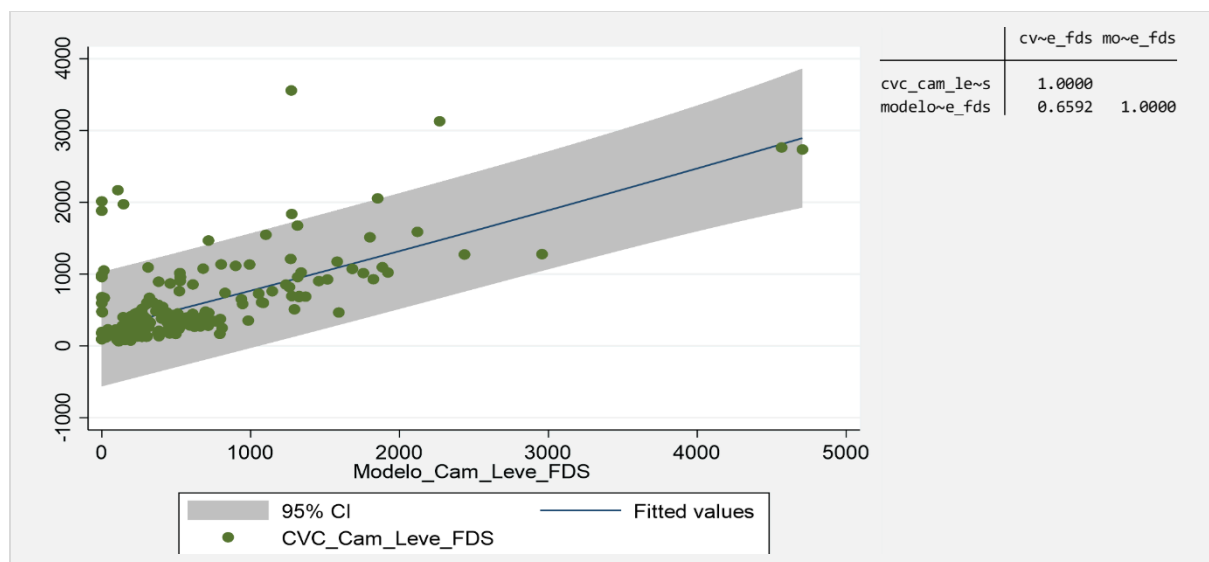


Figura 4-11: Calibração Veículos Comerciais Leves – Final de Semana

4.5.3 Veículo Comercial Pesado

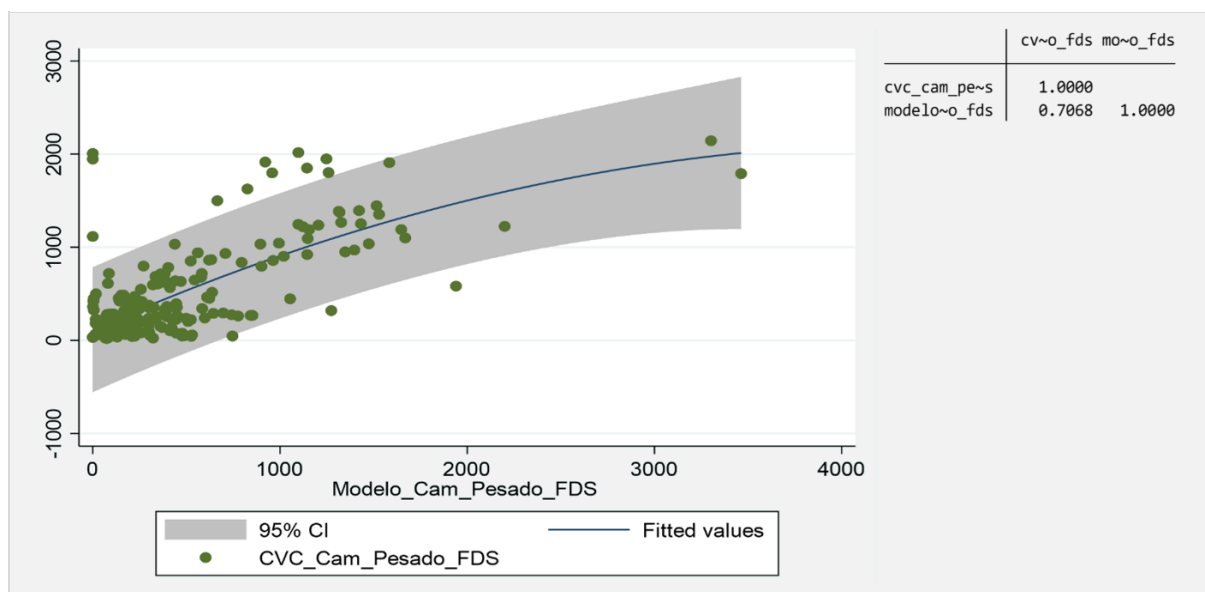


Figura 4-12: Calibração Veículos Comerciais Pesados – Final de Semana

4.6 Ajustamento Feriado

Os resultados de ajustamento quando expandidos para o período anual por regime de atividade em dia de feriado apresentaram melhoria na calibração de rede com coeficientes correlação pareados de 0,8857, 0,6381 e 0,7022 para passeio, comercial leve e pesado respectivamente.

4.6.1 Veículo Passeio

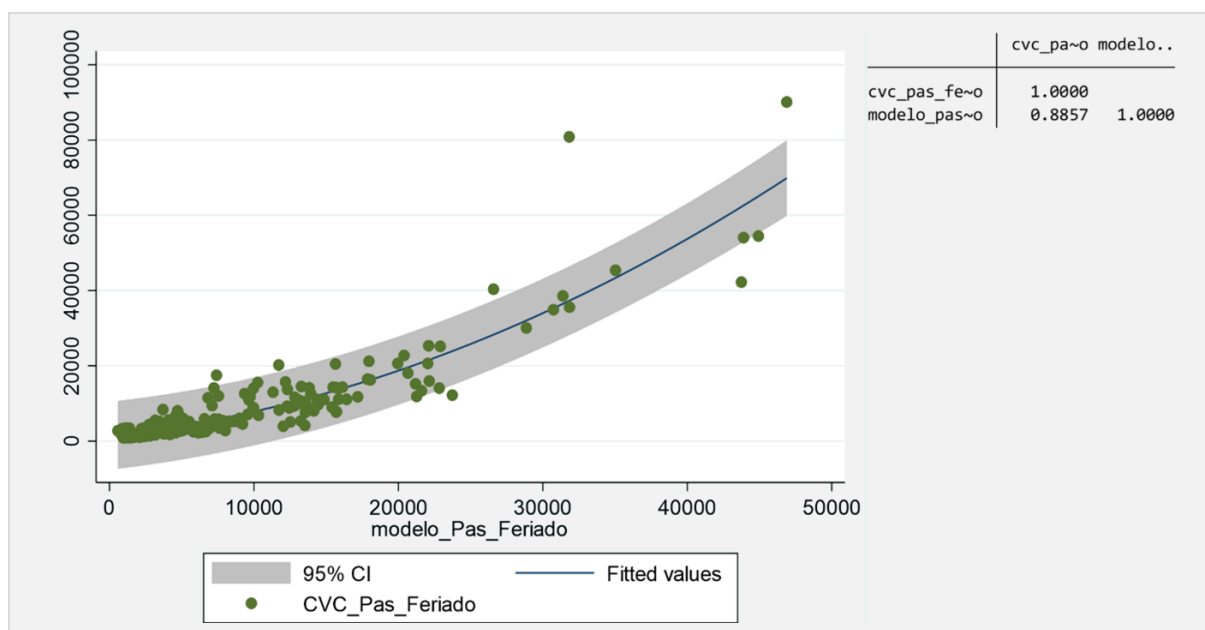


Figura 4-13: Calibração Veículos de Passeio – Feriado

4.6.2 Veículo Comercial Leve

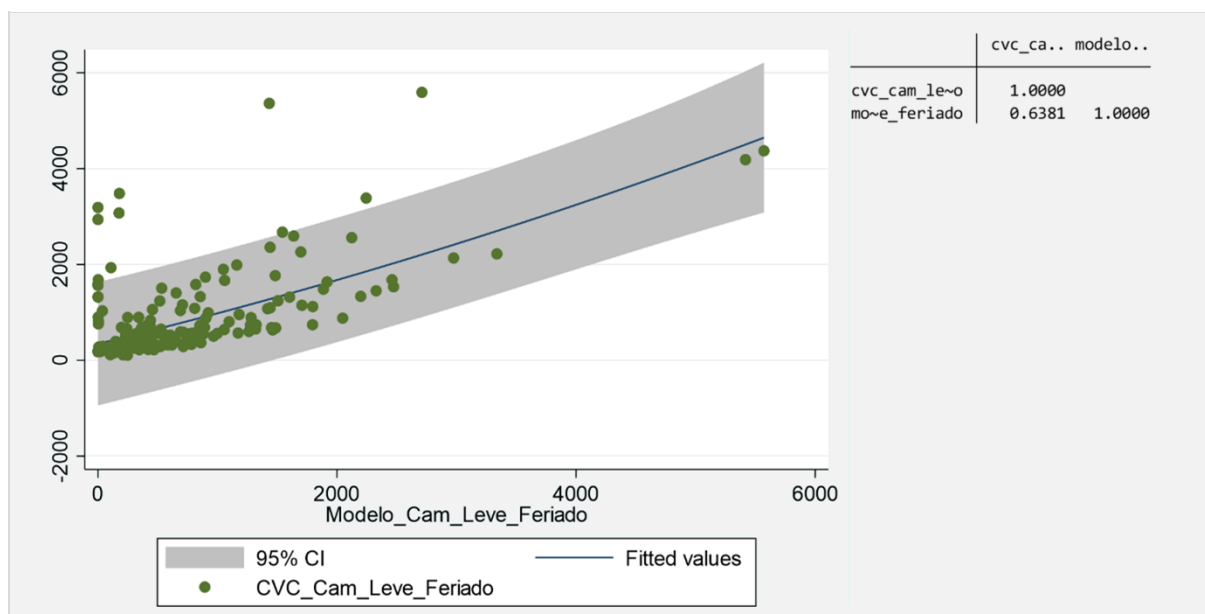


Figura 4-14: Calibração Veículos Comercial Leve – Feriado

4.6.3 Veículo Comercial Pesado

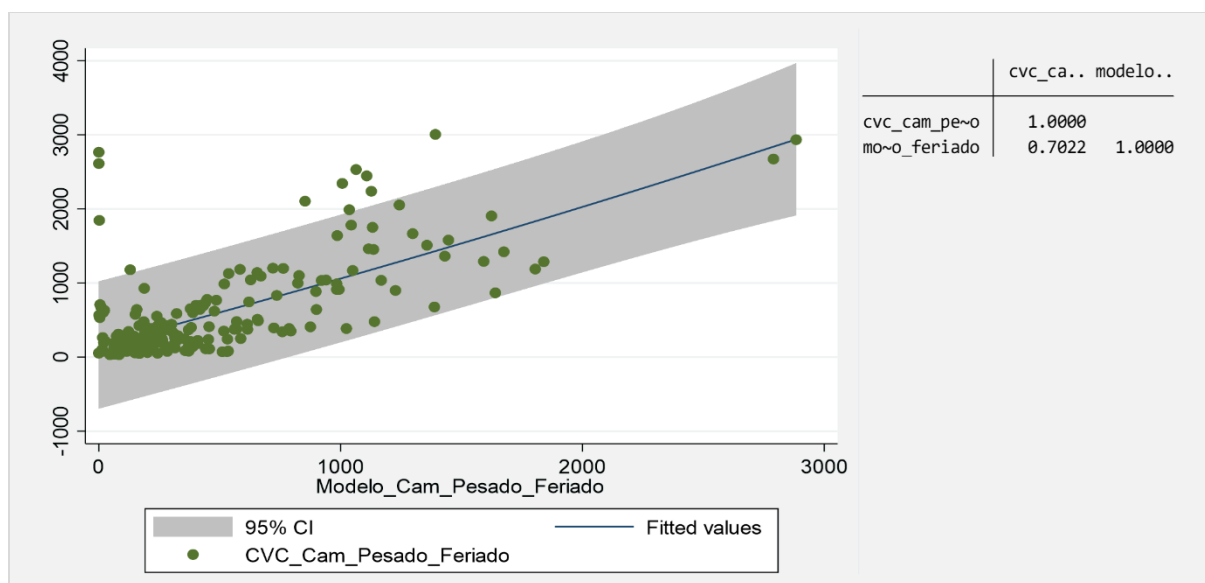


Figura 4-15: CALIBRAÇÃO VEÍCULOS COMERCIAIS PESADOS – FERIADO

5

Obtenção das Matrizes

A construção das matrizes de veículos é o resultado da aplicação do modelo de calibração da matriz de telefonia.

Optou-se por realizar primeiramente a obtenção das matrizes desagregadas por tipo de veículo (Calibração B) e depois, através da soma, obter a matriz agregada (Calibração A).

Foram aplicados os fatores obtidos no modelo para a obtenção das matrizes para cada um dos tipos de dia (útil; final de semana e feriado) para cada mês.

Foi levado em consideração a ponderação de cada tipo de dia no mês para que se obtenha o valor anual médio de cada um dos tipos de dia

A Tabela a seguir mostra o peso de cada um dos tipos dia ao longo do ano.

Tabela 5-1: Tipos de dias ao longo do Período em Estudo

Mês	Ano	DU	FDS	F	Total
3	2019	17	8	6	31
4	2019	20	6	4	30
5	2019	22	8	1	31
6	2019	17	8	5	30
7	2019	20	6	5	31
8	2019	22	9	0	31
9	2019	21	9	0	30
10	2019	23	8	0	31
11	2019	19	7	4	30
12	2019	18	7	6	31
1	2020	22	8	1	31
2	2020	16	7	6	29
Total		237	91	38	366

As matrizes resultantes são entregues através de meio digital, tanto em formato de banco de dados como contidas no modelo do software PT-Visum.

a) Matrizes Calibração A:	▪ Mat_DU. Matriz dia útil; ▪ Mat_FDS, Matriz final de semana; e ▪ Mat_F, Matriz feriado.
b) Matrizes Calibração B:	▪ Mat_pas_DU, Matriz veículos de passeio dia útil; ▪ Mat_Cam_Lev_DU, Matriz de caminhões leves dia útil; ▪ Mat_Cam_Pes_DU Matriz de caminhões pesados dia útil; ▪ Mat_pas_FDS, Matriz veículos de passeio final de semana; ▪ Mat_Cam_Lev_FDS, Matriz de caminhões leves final de semana;

6 Cuidados específicos

A modelagem de transportes retrata as viagens através do uso da infraestrutura por seus usuários. Habitualmente é possível construir matrizes de demanda através das pesquisas origem-destino em pontos específicos da rede de transporte.

Desta maneira existe uma concentração das viagens no eixo principal a ser estudado. No caso da utilização de matrizes provenientes da telefonia móvel o universo é maior, abrangendo-se praticamente todo o universo de viagens de uma determinada área, sejam elas rodoviária, urbanas, e de qualquer modo de transporte. Com a ampliação das matrizes, a rede deve ter a representação condizente com este maior fluxo de viagens.

As diferenças básicas entre matrizes estimadas por pesquisas origem-destino e a partir de dados de telefonia são:

- a) A Telefonia abrange um maior universo; e**
- b) Existe a necessidade de uma rede detalhada e a adequação da rede ao maior fluxo e caminhos inexistentes.**

Assim, há a necessidade de uma rede detalhada, já que os fluxos não contemplados usualmente são abrangidos neste caso. Com isso, sem um detalhamento da rede, o fluxo na rede simplificada pode ser superestimado.

Na Figura 6-1, podemos exemplificar a necessidade de detalhamento. Entre os municípios de Cordeirópolis e Limeira existe um viário urbano que realiza esta ligação. No caso da não representação deste viário, as viagens se concentram no viário rodoviário, que teria uma quantidade além da realidade.

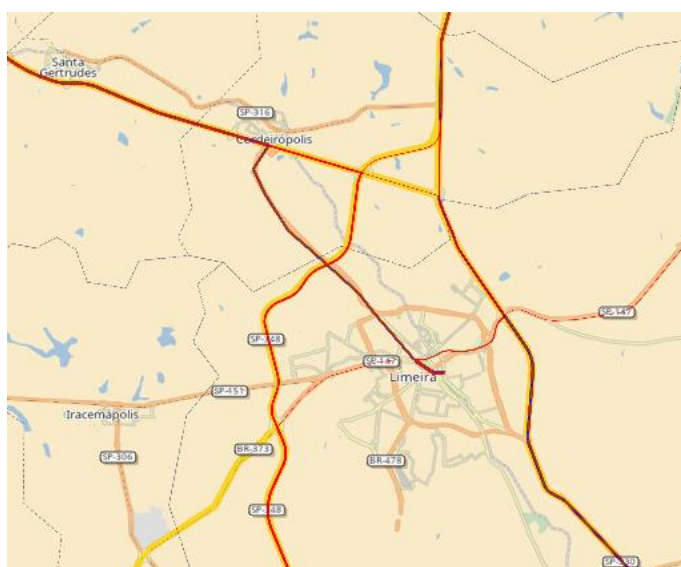


Figura 6-1: Exemplo de Viário Complementado

No caso de uma OD Rodoviária tradicional, as viagens intermunicipais que utilizam as vias urbanas não estariam contempladas na matriz.

Na Figura 6-2, temos um exemplo onde não foi detalhado o caminho entre os municípios de Guararema e Jacareí.

As viagens que utilizariam o caminho por vias urbanas, pela falta das mesmas na rede, acabam utilizando as rodovias e assim apresentando um volume acima do esperado.

Nestes casos existe uma necessidade de complementação do sistema viário urbano.



Figura 6-2: Exemplo de Viário a ser Complementado

Alguns casos mais específicos acontecem no estado, como é o caso do Rodoanel. Devido ao zoneamento utilizado, muitas das viagens da região metropolitana são consideradas intrazonais e não carregam o viário principal. Assim a atração das viagens para o Rodoanel é prejudicada, sendo necessário um detalhamento específico, ajustando as variáveis para representarem o nível de saturação com a matriz existente.

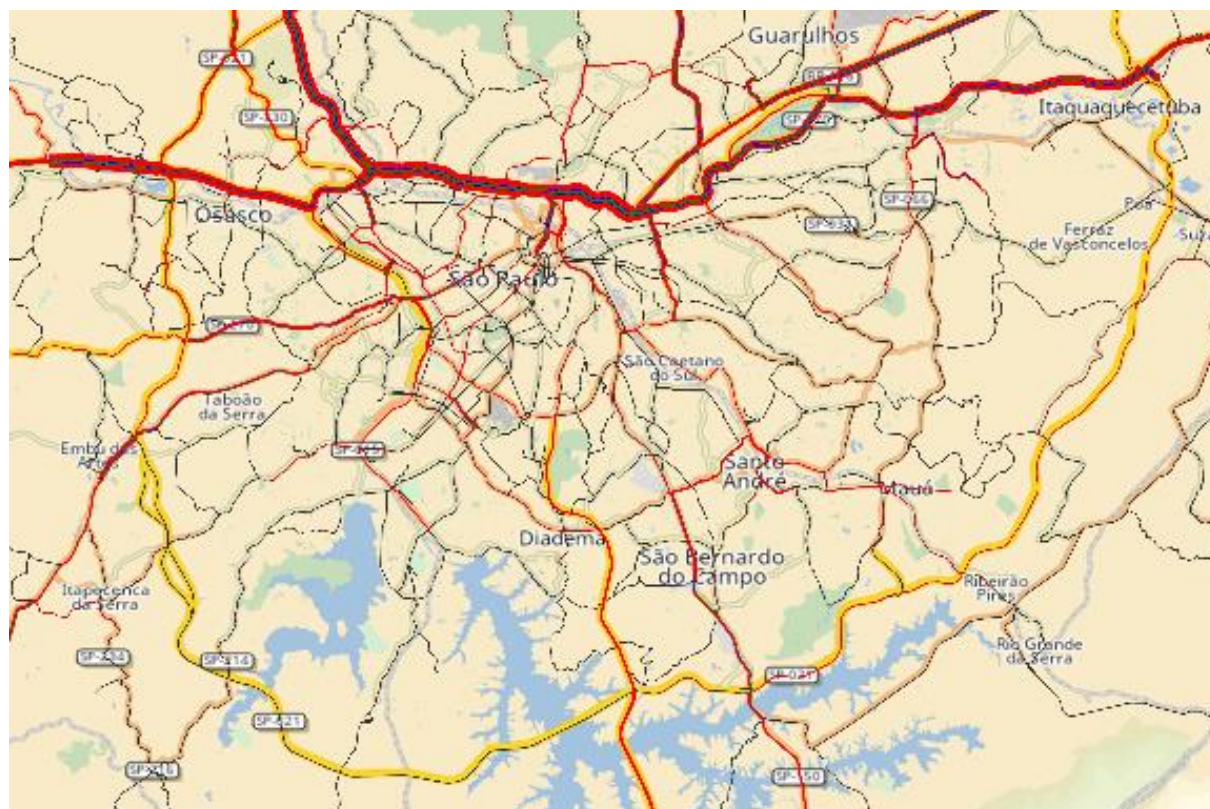


Figura 6-3: Exemplo de Viário Específico

