Centro de Estudos de Política e Economia do Setor Público (Cepesp – FGV)
Laboratório de Urbanismo da Metrópole (Lume – FAUUSP)

Desenvolvimento de software para a avaliação de impacto sócio-econômico de uma intervenção no sistema de transportes

Relatório de Pesquisa

São Paulo
Novembro
2006

Coordenador:	Ciro Biderman
Pesquisadores Principais:	Danilo Camargo Igliori Rodrigo de Losso
Pesquisadores:	Glauco Peres da Silva Guilherme Finkelfarb Lichand Henrique Eduardo F. Vinhais Marcos Felipe Mendes Lopes Marcelo Tyszler Rodrigo Mantovani Policano Vladimir Maciel
Estagiários:	Bruno Barbosa João Rafael Baena Casalecchi Michelle Lisboa Turu

Índice

Este modelo projeta o emprego para cada zona por setor de atividade, assumindo que as interações entre regiões podem ser captadas a partir da sua relação no espaço medida pelo tempo. Além disso, considera-se que as empresas e famílias tomam decisões de acordo com o que observam das características das regiões até o momento presente, sem abstraírem sobre o futuro.
Este texto visa explicitar o modelo de projeção do emprego total desenvolvido para o projeto DERSA. Para tanto, contém a metodologia de construção das variáveis, as regressões e projeções do modelo e a calibragem do resultado, evitando discrepâncias com o previsto pela PNAD/RAIS.

1. Construção das Variáveis

O banco de dados principal deste estudo é o Censo Demográfico do IBGE dos anos de 1991 e 2000.
Primeiramente, restringe-se o banco para: V0302 = 1 (considera somente o chefe do domicílio) e V0345 = 1 (considera aquele que trabalhou habitualmente, ou seja, pessoa que exerceu uma ocupação remunerada, mesmo que somente durante algumas horas diárias, semanais ou mensais como assalariado, conta-própria ou empregador, e a não remunerada que trabalhou habitualmente pelo menos 15 horas semanais). Desta forma, as outras observações são desconsideradas na leitura.
Em seguida, as observações são agrupadas por código do município (CodMun), código de atividade (V4461) e quintil de renda (Quintil_RT_CD).
Contudo trabalha-se com uma agregação de CodMun, chamada código AMC_pddt que corresponde à zona. Para tanto, há uma tabela que relaciona CodMun com AMC_pddt.

1.1- Variáveis Explicativas - Matriz X_{t-1}:

- Variável **ln(anos de estudo)$_{t-1}$** = X_ln_anos_estudo_91 (1 variável)
  Esta variável é proveniente do IPEADATA. Para construí-la, seria:
  É o logaritmo natural da média da variável V3241, por zona (AMC_pddt).
  V3241 = considera apenas números entre 0 e 17 (número de anos de estudo da pessoa).

- Variável **ln(área da zona)$_{t-1}$** = X_ln_area_91 (1 variável)
  É o logaritmo natural da soma da área dos municípios que constituem cada zona (Tabela retirada do IPEADATA).

- Variável **(Pop dos estratos)$_{t-1}$** = X_Pop91_0...5 (6 variáveis)
  É a agregação do número de domicílios por quintil de renda (estrato), por zona.

- Variável **ln(renda per capita domiciliar)$_{t-1}$** = X_ln_renda_pc_91 (1 variável)
  Esta variável é proveniente do IPEADATA. Para construí-la, seria:
  É o logaritmo natural da média de (renda total do domicilio dividida pelo número de pessoas moradoras no domicílio), por zona.
  A renda total do domicílio é o somatório da renda total (V3561) de todos os membros do domicílio e o número de pessoas moradoras no domicílio abrange pessoas em domicílios particulares exceto domésticas, pensionistas e com renda missing. Com isso, precisa-se considerar V0201 = 1 (situação do domicílio = particular permanente), V0202 = 1 a 6 (não considera cômodos e domicílios improvisados ou domicílios coletivos) e V0302 = 1 a 13, e 20 (condição da pessoa no

domicílio, desconsiderando pensionista, empregado(a) doméstico(a) e parente de empregado(a) doméstico(a)).
Além disso, o valor está representado em moeda de julho de 1991, precisando ser convertido para julho de 2000. Para isto, multiplica-se o valor por (IPCA do período) / 2750000.

○ Variável **ln(Salário Formal Total)$_{t-1}$** = Sal_Tot (1 variável)
É o logaritmo natural de (massa salarial formal dividida por emprego formal), por zona.
Inicialmente, a massa salarial e o emprego estão abertos em 42 setores com base na RAIS.
O salário formal médio de todos os setores em t-1 é calculado através de:

$$Sal\_Tot_{t-1} = \frac{\sum_{1}^{43} massa\_salarial\_formal_{t-1}}{\sum_{1}^{43} emprego\_formal_{t-1}}$$

○ Variável **Dummy se município no litoral** = X_costeira (1 variável)
Há uma tabela (fonte IBGE) que indica se o município está no litoral ou não. Na agregação por zona, X_costeira = 1 se pelo menos um dos municípios, que constituem a zona, está no litoral, caso contrário X_costeira = 0.

○ Variável **ln(Salário Formal do setor r)$_{t-1}$** = Sal_i (1 variável)
É o logaritmo natural de (massa salarial dividida por emprego), por setor e zona.
Inicialmente, a massa salarial está aberta em 42 setores com base na RAIS. Como as regressões são feitas por setor, esta variável é o logaritmo natural do salário apenas do setor considerado pela variável dependente (Emprego do setor r)$_{t-1}$.
Com isso, chega-se ao salário formal total em t para o setor r, calculado através de:

$$Sal\_r_{t-1} = \frac{massa\_salarial\_formal_{t-1}}{emprego\_formal_{t-1}}$$

○ Variável **(% pessoas em domicílio com banheiro e água encanada)$_{t-1}$** = X_struc_91 (1 variável)
Para a população total, precisa-se considerar V0201 = 1 (situação do domicílio = particular permanente), V0202 = 1 a 6 (tipo de domicílio, não considera cômodos e domicílios improvisados ou domicílios coletivos) e V0302 = 1 a 13, e 20 (condição da pessoa no domicílio, desconsiderando pensionista, empregado(a) doméstico(a) e parente de empregado(a) doméstico(a)).
Para o número de pessoas que moram em domicílio com banheiro e água encanada, considera-se V0206 = 1 a 3 (tem instalação sanitária do tipo rede geral, fossa séptica ligada à rede pluvial ou fossa séptica sem escoadouro), V0207 = 1 (uso da instalação sanitária somente pelo domicílio), V0213 > 0 (há pelo menos um banheiro no domicílio) e V0205 = 1 a 3 (água encanada = rede geral com canalização interna, poço ou nascente com canalização interna ou outra forma com canalização interna).

○ Variável **(Demanda do setor r)$_{t-1}$** = c_i (1 variável)
É dada pela fórmula:

$$c_{s,z,t-1} \equiv \sum_{j \neq s} m_{s,j} \ln(empr_{j,z,t-1})$$

Onde $c_{s,z,t-1}$ é a demanda no setor, $m_{s,j}$ representa as unidades de insumos produzidos pelo setor s usados para produzir uma unidade do produto j (coeficiente técnico intersetorial) estimado a partir da matriz de insumo-produto (V), e $empr_{j,z,t-1}$ é o logaritmo natural do emprego do setor.
Assim, introduzem-se todos os setores em uma variável sintética ponderando-os pelo comércio com o setor que está sendo analisado.
Na forma matricial, considerando 31 setores, tem-se:
$empr_{(788x31)} * X'_{(31x31)} = C_{(788x31)}$
Onde empr é a matriz ln(emprego) com 788 zonas e 31 setores; X' é a matriz transposta de coeficientes técnicos diretos por atividade agregada para 31 setores; e C é a matriz de demanda.
Definem-se ainda as seguintes matrizes:
V – matriz insumo-produto obtida no IBGE – Contas Nacionais – Recursos – Tabela 1 – Produção das Atividades.
Un – matriz obtida no IBGE – Contas Nacionais – Recursos – Tabela 2 – Consumo Intermediário Nacional. É a matriz de consumo intermediário nacional, com o valor consumido de produtos (de origem interna) por atividade;
Fn – vetor obtido no IBGE – Contas Nacionais – Recursos – Tabela 2 – Demanda Final. É o vetor com o valor consumido de produtos por atividade. Deve-se usar a última coluna da tabela que corresponde à soma das demandas de todas as categorias apresentadas.
Para as matrizes e vetores acima, inicialmente, há a abertura de 80 produtos que precisam ser agrupados em 31 setores. Para tanto, ao agrupar cada matriz, soma-se os valores dos elementos da matriz 80x31 de acordo com a classificação dos produtos pelos setores, resultando numa matriz 31x31. Já o vetor Fn terá dimensão 31x1.
Além das matrizes acima, há:
g – vetor com os valores da última coluna da Tabela 1, com o valor bruto da produção total.
q – vetor com os valores da última linha da Tabela 1, com o valor bruto da produção total.
Por construção, o somatório dos elementos do vetor g deve ser igual ao somatório dos elementos do vetor q.
Com estas matrizes e vetores, é possível encontrar a matriz de coeficientes técnicos diretos por atividade (X = D.Bn).
Pelas notas metodológicas do IBGE sobre a matriz insumo-produto, define-se:
$Bn = Un.\langle g \rangle^{-1}$
Onde $\langle g \rangle$ é uma matriz quadrada com os valores de g na diagonal principal e zero para os outros elementos.
Sabe-se que: $q = Un.i + Fn$
Onde i é um vetor unitário.
Substituindo Un (da equação de Bn) na equação de q, tem-se:
$q = Bn.\langle g \rangle.i + Fn$
$q = Bn.g + Fn$
Sabe-se também que $D = V.\langle q \rangle^{-1}$
Multiplicando ambos os lados da equação por i e invertendo, tem-se:
$V.i = D.\langle q \rangle.i$
Como $g = V.i$, então:
$g = D.\langle q \rangle.i = D.q$
Portanto $D = g.q^{-1}$.
Uma vez calculado Bn e D, encontra-se X = D.Bn e conseqüentemente C:
$C_{(788x31)} = empr_{(788x31)} * X'_{(31x31)}$
Vale ressaltar que para cada regressão por setor, a variável (Demanda do setor $r)_t$ é a coluna da matriz C do respectivo setor.

○ Variável **(Oferta do setor r)$_{t-1}$** = f_i (1 variável)

É dada pela fórmula:

$$f_{s,z,t-1} \equiv \sum_{j \neq s} m_{j,s} \ln(empr_{j,z,t-1})$$

Onde $f_{s,z,t-1}$ é a oferta no setor, $m_{j,s}$ representa as unidades de insumos produzidos pelo setor s usados para produzir uma unidade do produto j (coeficiente técnico intersetorial) estimado a partir da matriz de insumo-produto (V), e $empr_{j,z,t-1}$ é o logaritmo natural do emprego do setor.

Assim, introduzem-se todos os setores em uma variável sintética ponderando-os pelo comércio com o setor que está sendo analisado.

Na forma matricial, considerando 31 setores, tem-se:

$empr_{(788x31)} * X_{(31x31)} = C_{(788x31)}$

Onde empr é a matriz ln(emprego) com 788 zonas e 31 setores; X é a matriz de coeficientes técnicos diretos por atividade agregada para 31 setores; e C é a matriz de oferta.

O cálculo da matriz X segue a metodologia citada na variável anterior. Vale ressaltar que para cada regressão por setor, a variável (Oferta do setor r)$_t$ é a coluna da matriz C do respectivo setor.

○ Variável **(Tempo de deslocamento da zona z para o 1° porto mais próximo)$_{t-1}$** = DP_1 (1 variável)

É o logaritmo natural do tempo representativo da distância da zona para o 1° porto mais próximo. Considera-se a média entre os tempos de chegada ao porto a partir da zona e de partida do porto para a zona. Estes tempos são disponibilizados pela DERSA juntamente com a matriz de tempos.

○ Variável **(Tempo de deslocamento da zona z para o 2° porto mais próximo)$_{t-1}$** = DP_2 (1 variável)

É o logaritmo natural do tempo representativo da distância da zona para o 2° porto mais próximo. Considera-se a média entre os tempos de chegada ao porto a partir da zona e de partida do porto para a zona. Estes tempos são disponibilizados pela DERSA juntamente com a matriz de tempos.

○ Variável **(Tempo de deslocamento da zona z para o 3° porto mais próximo)$_{t-1}$** = DP_3 (1 variável)

É o logaritmo natural do tempo representativo da distância da zona para o 3° porto mais próximo. Considera-se a média entre os tempos de chegada ao porto a partir da zona e de partida do porto para a zona. Estes tempos são disponibilizados pela DERSA juntamente com a matriz de tempos.

Com isto, tem-se todas as variáveis explicativas que constituem a matriz X_{t-1}. Para a construção desta matriz, basta posicionar cada variável ao lado da outra, criando a matriz com o número de zonas como linhas e as 18 variáveis supracitadas como colunas.

1.2- Matriz de Tempos - W_{t-1}:

É preciso transformar a informação disponibilizada pela DERSA (3 colunas: origem, destino e tempo ponderado) em uma matriz quadrada, com as linhas significando a origem e as colunas significando os destinos.

1.3- Unidades da Federação - _IUF_11...53:

O modelo possui as 27 dummies de unidade da federação (UF), buscando captar o efeito da UF, que é fixo no tempo, sobre a variável dependente. Na programação, para cada observação, calculam-se as 27 dummies de unidade da federação, da seguinte maneira:
Se V1101=11 então _IUF_11=1, caso contrário, _IUF_11=0;
Se V1101=12 então _IUF_12=1, caso contrário, _IUF_12=0;
Se V1101=13 então _IUF_13=1, caso contrário, _IUF_13=0;
Se V1101=14 então _IUF_14=1, caso contrário, _IUF_14=0;
Se V1101=15 então _IUF_15=1, caso contrário, _IUF_15=0;
Se V1101=16 então _IUF_16=1, caso contrário, _IUF_16=0;
Se V1101=17 então _IUF_17=1, caso contrário, _IUF_17=0;
Se V1101=21 então _IUF_21=1, caso contrário, _IUF_21=0;
Se V1101=22 então _IUF_22=1, caso contrário, _IUF_22=0;
Se V1101=23 então _IUF_23=1, caso contrário, _IUF_23=0;
Se V1101=24 então _IUF_24=1, caso contrário, _IUF_24=0;
Se V1101=25 então _IUF_25=1, caso contrário, _IUF_25=0;
Se V1101=26 então _IUF_26=1, caso contrário, _IUF_26=0;
Se V1101=27 então _IUF_27=1, caso contrário, _IUF_27=0;
Se V1101=28 então _IUF_28=1, caso contrário, _IUF_28=0;
Se V1101=29 então _IUF_29=1, caso contrário, _IUF_29=0;
Se V1101=31 então _IUF_31=1, caso contrário, _IUF_31=0;
Se V1101=32 então _IUF_32=1, caso contrário, _IUF_32=0;
Se V1101=33 então _IUF_33=1, caso contrário, _IUF_33=0;
Se V1101=35 então _IUF_35=1, caso contrário, _IUF_35=0;
Se V1101=41 então _IUF_41=1, caso contrário, _IUF_41=0;
Se V1101=42 então _IUF_42=1, caso contrário, _IUF_42=0;
Se V1101=43 então _IUF_43=1, caso contrário, _IUF_43=0;
Se V1101=50 então _IUF_50=1, caso contrário, _IUF_50=0;
Se V1101=51 então _IUF_51=1, caso contrário, _IUF_51=0;
Se V1101=52 então _IUF_52=1, caso contrário, _IUF_52=0;
Se V1101=53 então _IUF_53=1, caso contrário, _IUF_53=0.

1.4- Emprego Inicial = Empr_91_1...43 (só do setor r):

É a agregação do número do total de empregos (formais e informais) por setor e zona, para o ano de 1991. Considera somente o emprego do setor em questão na regressão.

1.5- Emprego Final = Empr_00_1...43 (só do setor r):

É a agregação do número do total de empregos (formais e informais) por setor e zona, para o ano de 2000. Considera somente o emprego do setor em questão na regressão.

2. Regressões

Para a obtenção do modelo de projeção do emprego, inicialmente é necessário utilizar um modelo de regressão com variáveis instrumentais em 2 estágios.
Vale destacar que a agregação de zonas e setores é diferente entre as etapas de regressão e projeção. Para a regressão, utiliza-se 788 AMC_pddt e 31 setores, devido a esta ser a máxima abertura possível pelo Censo de 1991. Já para a projeção, trabalha-se com 899 AMC_pddt e 42 setores, uma vez que as informações utilizadas são de 2000 em diante.

○ 1° Estágio:

$$W_{t-1}\ln(empr_{r,t-1}) = \gamma_0 W_{t-1}(X_{t-1}) + \gamma_1 \ln(empr_{r,t-1}) + \gamma_2 X_{t-1} + \gamma_3 \_IUF\_$$

○ 2° Estágio:

$$\ln(empr_{r,t}) = \beta_0 + \beta_1 \ln(empr_{r,t-1}) + \beta_2 X_{t-1} + \beta_3 \_IUF\_ + \beta_4 [W_{t-1} \ln(\widehat{empr}_{z,r,t-1})]$$

Exemplo: Se r=1, então tem-se:

1° Estágio:

- Variável Dependente: W_{t-1}.Emprego Inicial = W_{t-1}.ln(empr $_{1,t-1}$) = W * Empr_91_1
- Variáveis Independentes:

W_{t-1}.Variáveis Matriz X = W_{t-1}.ln(X_{t-1}):
W * X_ln_anos_estudo_91; W * X_ln_area_91; W * X_ln_renda_pc_91; W * X_costeira; W * X_struc_91; W * X_Pop_91_0; W * X_Pop_91_1; W * X_Pop_91_2; W * X_Pop_91_3; W * X_Pop_91_4; W * X_Pop_91_5; W * DP_1; W * DP_2; W * DP_3; W * Sal_Tot; W * Sal_1; W * c_1; W * f_1.
Emprego Inicial: ln(empr $_{1,t-1}$) = Empr_91_1.
Variáveis Matriz X = X_{t-1}:
X_ln_anos_estudo_91; X_ln_area_91; X_ln_renda_pc_91; X_costeira; X_struc_91; X_Pop_91_0; X_Pop_91_1; X_Pop_91_2; X_Pop_91_3; X_Pop_91_4; X_Pop_91_5; Sal_Tot; Sal_1; c_1; f_1; DP_1; DP_2; DP_3.
Unidades da Federação (sem dummy SP): _IUF_11; _IUF_12; _IUF_13; _IUF_14; _IUF_15; _IUF_16; _IUF_17; _IUF_21; _IUF_22; _IUF_23; _IUF_24; _IUF_25; _IUF_26; _IUF_27; _IUF_28; _IUF_29; _IUF_31; _IUF_32; _IUF_33; _IUF_41; _IUF_42; _IUF_43; _IUF_50; _IUF_51; _IUF_52; _IUF_53.

2° Estágio:

- Variável Dependente: Emprego Final = ln(empr $_{1,t}$) = Empr_91_1
- Variáveis Independentes:

Emprego Inicial: ln(empr $_{1,t-1}$) = Empr_91_1.
Variáveis Matriz X = X_{t-1}:
X_ln_anos_estudo_91; X_ln_area_91; X_ln_renda_pc_91; X_costeira; X_struc_91; X_Pop_91_0; X_Pop_91_1; X_Pop_91_2; X_Pop_91_3; X_Pop_91_4; X_Pop_91_5; Sal_Tot; Sal_1; c_1; f_1; DP_1; DP_2; DP_3.
Unidades da Federação (sem dummy SP): _IUF_11; _IUF_12; _IUF_13; _IUF_14; _IUF_15; _IUF_16; _IUF_17; _IUF_21; _IUF_22; _IUF_23; _IUF_24; _IUF_25; _IUF_26; _IUF_27; _IUF_28; _IUF_29; _IUF_31; _IUF_32; _IUF_33; _IUF_41; _IUF_42; _IUF_43; _IUF_50; _IUF_51; _IUF_52; _IUF_53.
Variável dependente prevista no 1° Estágio: {W_{t-1}.ln(empr $_{1,t-1}$)}.

3. Projeções

Com o modelo especificado acima, projeta-se o emprego por setor e por zona. Em seguida, compara-se a projeção feita com a variável **(Emprego dos setores r)$_t$** calculada no modelo de população.
Para cada ano, calcula-se:

$$\ln(empr_{r,t+1}) = \hat{\beta}_0 + \hat{\beta}_1 \ln(empr_{r,t}) + \hat{\beta}_2 X_t + \hat{\beta}_3 \_IUF\_ + \hat{\beta}_4 [W_t \ln(\widehat{empr}_{z,r,t})]$$

Se t=2000, então a variável considerada é a observada. Caso contrário, se t ≥ 2005, então a variável é a projetada.
No caso do emprego, a variável foi anteriormente projetada, pela equação acima. Assim, para se projetar a população em t+1, utiliza-se como variável explicativa o

emprego em t, que foi projetado utilizando as variáveis explicativas em t-1, ou seja, a variável emprego é endógena.
Pela equação acima, tem-se:
$\ln(empr_{i,t})$ é o emprego no setor r e no tempo anterior (t);
X_t é a matriz de variáveis explicativas em t;
IUF representa as dummies de unidade da federação;
$W_t.\ln(\hat{empr}_{z,i,t})$ é a previsão do produto de matriz de tempos W_t com o logaritmo natural do emprego no setor r e no tempo anterior (t) , calculado a partir da equação do 1° estágio.
Dois pontos precisam ser frisados na equação em questão. Primeiramente, com relação a esta previsão do termo $W_t.\ln(empr_{z,i,t})$, utiliza-se os coeficientes estimados no 1° estágio, com as variáveis independentes em t. Ou seja, no 1° estágio a equação é:

$$W_t \ln(\widehat{empr}_{z,i,t}) = \hat{\gamma}_0 + \hat{\gamma}_1 W_t(X_t) + \hat{\gamma}_2 \ln(empr_{i,t}) + \hat{\gamma}_3 X_t + \hat{\gamma}_4 \_IUF\_$$

O segundo ponto relevante na projeção do emprego é a projeção das variáveis independentes de ambas as equações acima.

3.1- Projeção das Variáveis Independentes:

○ Variável **ln(anos de estudo)$_t$** = $3{,}1.[\ln(\text{anos de estudo})_{t-1}]^{3/2}$

○ Variável **ln(área da zona)$_t$** = $\ln(\text{área da zona})_{t-1}$
É constante no tempo.

○ Variável **(Pop dos estratos i)$_t$**
É a projeção do número de domicílios por quintil de renda e por zona para o período anterior. Vale lembrar que se t=2000, então esta variável é a observada com base no Censo 2000. Caso contrário, se t ≥ 2005, então esta variável é a projetada. Neste modelo, é utilizada esta variável projetada pelo modelo de população.

○Variável **ln(renda per capita domiciliar)$_t$** = ln { (renda per capita domiciliar)$_{t-1}$. (salário total)$_t$ / [(salário total)$_{t-1}$. (1 + taxa de crescimento da população da UF)] }
Onde salário total é calculado a partir da massa salarial e emprego, em t.
Para tanto, calcula-se a massa salarial da seguinte maneira:
(massa salarial)$_t$ = (massa salarial)$_{t-1}$. (1 + taxa de crescimento do salário), com taxa de crescimento do salário = 0,144552314 (fonte FIESP).
Inicialmente, a massa salarial e o emprego estão abertos em 42 setores com base na RAIS.
Para o cálculo do salário total em t, utiliza-se a expressão:

$$Sal\_Tot_t = \frac{\sum_{1}^{43} massa\_salarial\_formal_t}{\sum_{1}^{43} emprego\_formal_t}$$

Para o cálculo da taxa de crescimento da população da UF, é utilizada a relação entre as populações em t e t-1, por UF, com base nas previsões populacionais do IBGE.

○ Variável **ln(Salário Formal Total)$_t$** = Sal_Tot (1 variável)
É o logaritmo natural de (massa salarial dividida por emprego), por zona, de acordo com o que foi explicitado na variável anterior.

○ Variável **Dummy se município no litoral** = X_costeira (1 variável)
É constante no tempo.

○ Variável **ln(Salário Formal do setor r)$_t$** = Sal_i (1 variável)
É o logaritmo natural de (massa salarial dividida por emprego), por setor e zona. Inicialmente, a massa salarial está aberta em 42 setores com base na RAIS, para a regressão e projeção. Como as regressões são feitas por setor, esta variável é o logaritmo natural do salário apenas do setor r considerado pela variável dependente (Emprego do setor r)$_t$.
Com isso, chega-se ao salário total em t para o setor r, calculado através de:

$$Sal\_r_t = \frac{massa\_salarial\_formal_t}{emprego\_formal_t}$$

○ Variável **(% pessoas em domicílio com banheiro e água encanada)$_t$** = 1,44 . ln [1 + (% pessoas em domicílio com banheiro e água encanada)$_{t-1}$] (1 variável)

○ Variável **(Demanda do setor r)$_t$** = c_i (1 variável)
É dada pela fórmula:

$$c_{s,z,t} \equiv \sum_{j \neq s} m_{s,j} \ln(empr_{j,z,t})$$

Onde $c_{s,z,t}$ é a demanda no setor, $m_{s,j}$ representa as unidades de insumos produzidos pelo setor s usados para produzir uma unidade do produto j (coeficiente técnico intersetorial) estimado a partir da matriz de insumo-produto (V), e $empr_{j,z,t}$ é o logaritmo natural do emprego do setor.
Assim, introduzem-se todos os setores em uma variável sintética ponderando-os pelo comércio com o setor que está sendo analisado.
Na forma matricial, considerando 42 setores, tem-se:
$empr_{(899x42)} * X'_{(42x42)} = C_{(899x42)}$
Onde empr é a matriz ln(emprego) com 899 zonas e 42 setores; X' é a matriz transposta de coeficientes técnicos diretos por atividade agregada para 42 setores; e C é a matriz de demanda.
Definem-se ainda as seguintes matrizes:
V – matriz insumo-produto obtida no IBGE – Contas Nacionais – Recursos – Tabela 1 – Produção das Atividades.
Un – matriz obtida no IBGE – Contas Nacionais – Recursos – Tabela 2 – Consumo Intermediário Nacional. É a matriz de consumo intermediário nacional, com o valor consumido de produtos (de origem interna) por atividade;
Fn – vetor obtido no IBGE – Contas Nacionais – Recursos – Tabela 2 – Demanda Final. É o vetor com o valor consumido de produtos por atividade. Deve-se usar a última coluna da tabela que corresponde à soma das demandas de todas as categorias apresentadas.
Para as matrizes e vetores acima, inicialmente, há a abertura de 80 produtos que precisam ser agrupados em 42 setores. Para tanto, ao agrupar cada matriz, soma-se os valores dos elementos da matriz 80x42 de acordo com a classificação dos produtos pelos setores, resultando numa matriz 42x42. Já o vetor Fn terá dimensão 42x1.
Além das matrizes acima, há:
g – vetor com os valores da última coluna da Tabela 1, com o valor bruto da produção total.
q – vetor com os valores da última linha da Tabela 1, com o valor bruto da produção total.

Por construção, o somatório dos elementos do vetor g deve ser igual ao somatório dos elementos do vetor q.
Com estas matrizes e vetores, é possível encontrar a matriz de coeficientes técnicos diretos por atividade (X = D.Bn).
Pelas notas metodológicas do IBGE sobre a matriz insumo-produto, define-se:
$Bn = Un.\langle g\rangle^{-1}$
Onde <g> é uma matriz quadrada com os valores de g na diagonal principal e zero para os outros elementos.
Sabe-se que: $q = Un.i + Fn$
Onde i é um vetor unitário.
Substituindo Un (da equação de Bn) na equação de q, tem-se:
$q = Bn.\langle g\rangle.i + Fn$
$q = Bn.g + Fn$
Sabe-se também que $D = V.\langle q\rangle^{-1}$
Multiplicando ambos os lados da equação por i e invertendo, tem-se:
$V.i = D.\langle q\rangle.i$
Como g = V.i, então:
$g = D.\langle q\rangle.i = D.q$
Portanto $D = g.q^{-1}$.
Uma vez calculado Bn e D, encontra-se X = D.Bn e conseqüentemente C:
$C_{(899x42)} = empr_{(899x42)} * X'_{(42x42)}$
Vale ressaltar que para cada regressão por setor, a variável (Demanda do setor $r)_t$ é a coluna da matriz C do respectivo setor. Além disso, a variável emprego é a projetada endogenamente no modelo.

○ Variável **(Demanda do setor r)$_t$** = f_i (1 variável)
É dada pela fórmula:

$$f_{s,z,t} \equiv \sum_{j \neq s} m_{j,s} \ln(empr_{j,z,t})$$

Onde $f_{s,z,t}$ é a oferta no setor, $m_{j,s}$ representa as unidades de insumos produzidos pelo setor s usados para produzir uma unidade do produto j (coeficiente técnico intersetorial) estimado a partir da matriz de insumo-produto (V), e $empr_{j,z,t}$ é o logaritmo natural do emprego do setor.
Assim, introduzem-se todos os setores em uma variável sintética ponderando-os pelo comércio com o setor que está sendo analisado.
Na forma matricial, considerando 42 setores, tem-se:
$empr_{(899x42)} * X_{(42x42)} = C_{(899x42)}$
Onde empr é a matriz ln(emprego) com 899 zonas e 42 setores; X é a matriz de coeficientes técnicos diretos por atividade agregada para 42 setores; e C é a matriz de oferta.
O cálculo da matriz X segue a metodologia citada na variável anterior. Vale ressaltar que a variável emprego é a projetada endogenamente no modelo e, para cada regressão por setor, a variável (Oferta do setor $r)_t$ é a coluna da matriz C do respectivo setor.

○ Variável **(Tempo de deslocamento da zona z para o 1° porto mais próximo)$_{t-1}$** = DP_1 (1 variável)
É o logaritmo natural do tempo representativo da distância da zona para o 1° porto mais próximo. Considera-se a média entre os tempos de chegada ao porto a partir da zona e de partida do porto para a zona. Estes tempos são disponibilizados pela DERSA juntamente com a matriz de tempos.

○ Variável **(Tempo de deslocamento da zona z para o 2° porto mais próximo)$_{t-1}$** = DP_2 (1 variável)

É o logaritmo natural do tempo representativo da distância da zona para o 2° porto mais próximo. Considera-se a média entre os tempos de chegada ao porto a partir da zona e de partida do porto para a zona. Estes tempos são disponibilizados pela DERSA juntamente com a matriz de tempos.

○ Variável **(Tempo de deslocamento da zona z para o 3° porto mais próximo)$_{t-1}$** = DP_3 (1 variável)

É o logaritmo natural do tempo representativo da distância da zona para o 3° porto mais próximo. Considera-se a média entre os tempos de chegada ao porto a partir da zona e de partida do porto para a zona. Estes tempos são disponibilizados pela DERSA juntamente com a matriz de tempos.

3.2- Projeção da Matriz de Tempos - W_t:

É disponibilizada pela DERSA, precisando ser transformada em uma matriz quadrada, com as linhas significando a origem e as colunas significando os destinos.

3.3- Unidades da Federação - _IUF_11...53:

São constantes no tempo.

4. Calibragem

Nesta fase, compara-se a projeção feita, pelo modelo acima especificado, com a variável **(Emprego dos setores r)$_t$** calculada no modelo de população.
Para a comparação da projeção feita pelo modelo com a previsão do IBGE, trabalha-se com o caso:
Caso 6: Como a variável **(Emprego dos setores r)$_t$** é calculada por unidade da federação, que é uma agregação de zonas, e por setor, pode-se compará-la com a projeção agregada por zona agregada e setor, segundo a equação:

$$c = empr_{r,z,t} \left[\sum_{r \in R} \sum_{z \in Z} \overline{empr}_{r,z,t} \right]^{-1}$$

Onde empr $_{r,z,t}$ é o emprego previsto [variável **(Emprego dos setores r)$_t$**] para uma agregação de zonas (UF) e setor;

$\sum_{r \in R} \sum_{z \in Z} \overline{empr}_{r,z,t}$ é o somatório por setor e UF do emprego estimado pelo modelo.

A divisão do primeiro termo pelo segundo fornece um calibrador que deve ser multiplicado pelo emprego projetado, desagregado por setor e zona.

Portanto, o emprego projetado final é a projeção do modelo multiplicada por um fator de calibração, de acordo com a equação:
$EMPR_{i,z,t} = c * empr_{i,z,t}$
Onde EMPR $_{i,z,t}$ é o emprego projetado final, corrigido pelo fator de calibração; c é este fator; e empr $_{i,z,t}$ é o emprego projetado pelo modelo.

B - Modelo de projeção de população

Este modelo projeta a população para cada zona por estrato (quintil) de renda, assumindo que as interações entre regiões podem ser captadas a partir da sua relação no espaço medida pelo tempo. Além disso, considera-se que as empresas e

famílias tomam decisões de acordo com o que observam das características das regiões até o momento presente, sem abstraírem sobre o futuro.
Com base nestes pressupostos, este texto visa explicitar o modelo de projeção populacional desenvolvido para o projeto DERSA. Para tanto, contém a metodologia de construção das variáveis, as regressões e projeções do modelo e a calibragem do resultado, evitando discrepâncias com o previsto pelo IBGE.

1. Construção das Variáveis

O banco de dados principal deste estudo é o Censo Demográfico do IBGE dos anos de 1991 e 2000.
Primeiramente, restringe-se o banco para: V0302 = 1 (considera somente o chefe do domicílio) e V0345 = 1 (considera aquele que trabalhou habitualmente, ou seja, pessoa que exerceu uma ocupação remunerada, mesmo que somente durante algumas horas diárias, semanais ou mensais como assalariado, conta-própria ou empregador, e a não remunerada que trabalhou habitualmente pelo menos 15 horas semanais). Desta forma, as outras observações são desconsideradas na leitura.
Em seguida, as observações são agrupadas por código do município (CodMun), código de atividade (V4461) e quintil de renda (Quintil_RT_CD).
Contudo, trabalha-se com uma agregação de CodMun, chamada código AMC_pddt que corresponde à zona. Para tanto, há uma tabela que relaciona CodMun com AMC_pddt.

1.1- Variáveis Explicativas - Matriz X_{t-1}:

○ Variável **ln(anos de estudo)$_{t-1}$** = X_ln_anos_estudo_91 (1 variável)
Esta variável é proveniente do IPEADATA. Para construí-la, seria:
É o logaritmo natural da média da variável V3241, por zona (AMC_pddt).
V3241 = considera apenas números entre 0 e 17 (número de anos de estudo da pessoa).

○ Variável **(Emprego dos setores r)$_{t-1}$** = Empr_Ag_1...7 (7 variáveis)
Pela RAIS, agrega-se o total de pessoas empregadas nos 42 setores pddt. Em seguida, esta variável é agregada em r = 7 setores para a regressão e projeção.
Com isso, obtém-se o total de empregos formais por setor r para cada unidade da federação (UF). Para se conseguir o total de empregos, incluindo os informais, é necessário utilizar um ponderador de acordo com a equação:
$P_{I,j} = (e_{I,J})^t / (e_{I,J})^f$
Onde $P_{I,j}$ é o ponderador; $(e_{I,J})^t$ é o total de emprego dado pelo Censo/PNAD agregado por setor r e UF; e $(e_{I,J})^f$ é o emprego formal, dado pela RAIS, agregado por setor r e UF.
Este ponderador é multiplicado para todas as zonas pelo emprego agregado obtido pela RAIS, por setor r e UF.
Com isso, obtêm-se o total de emprego (formal e informal) para cada zona, por setor.

○ Variável **ln(área da zona)$_{t-1}$** = X_ln_area_91 (1 variável)
É o logaritmo natural da soma da área dos municípios que constituem cada zona (Tabela retirada do IPEADATA).

○ Variável **(Pop nos outros estratos)$_{t-1}$** = X_Pop91_0...5 (sem considerar o estrato i) (5 variáveis)

É a agregação do número de domicílios por quintil de renda (estrato), por zona. Na matriz X_{t-1} com as variáveis explicativas, não há a população do estrato i considerado na variável dependente da regressão.

○ Variável **ln(renda per capita domiciliar)$_{t-1}$** = X_ln_renda_pc_91 (1 variável)
Esta variável é proveniente do IPEADATA. Para construí-la, seria:
É o logaritmo natural da média, por zona, de (renda total do domicilio dividida pelo número de pessoas moradoras no domicílio).
A renda total do domicílio é o somatório da renda total (V3561) de todos os membros do domicílio e o número de pessoas moradoras no domicílio abrange pessoas em domicílios particulares exceto domésticas, pensionistas e com renda missing. Com isso, precisa-se considerar V0201 = 1 (situação do domicílio = particular permanente), V0202 = 1 a 6 (não considera cômodos e domicílios improvisados ou domicílios coletivos) e V0302 = 1 a 13, e 20 (condição da pessoa no domicílio, desconsiderando pensionista, empregado(a) doméstico(a) e parente de empregado(a) doméstico(a)).
Além disso, o valor está representado em moeda de julho de 1991, precisando ser convertido para julho de 2000. Para isto, multiplica-se o valor por (IPCA do período) / 2750000.

○ Variável **ln(Salário Formal Total)$_{t-1}$** = Sal_Tot (1 variável)
É o logaritmo natural de (massa salarial formal dividida por emprego formal), por zona.
Inicialmente, a massa salarial e o emprego estão abertos em 42 setores com base na RAIS.
O salário formal médio de todos os setores em t-1 é calculado através de:

$$Sal\_Tot_{t-1} = \frac{\sum_{1}^{43} massa\_salarial\_formal_{t-1}}{\sum_{1}^{43} emprego\_formal_{t-1}}$$

○ Variável **Dummy se município no litoral** = X_costeira (1 variável)
Há uma tabela (fonte IBGE) que indica se o município está no litoral ou não. Na agregação por zona, X_costeira = 1 se pelo menos um dos municípios, que constituem a zona, está no litoral, caso contrário X_costeira = 0.

○ Variável **ln(Salário Formal dos setores r)$_{t-1}$** = Sal_Ag_1...7 (7 variáveis)
É o logaritmo natural da massa salarial dividida pelo emprego, por setor e zona.
Inicialmente, a massa salarial e o emprego estão abertos em 42 setores com base na RAIS, sendo agregados em 7 setores para a regressão e projeção.
Com isso, chega-se ao salário formal de cada setor em t-1, que é calculado através de:

$$Sal\_Ag_{t-1} = \frac{massa\_salarial\_formal_{t-1}}{emprego\_formal_{t-1}}$$

○ Variável **(% pessoas em domicílio com banheiro e água encanada)$_{t-1}$** = X_struc_91 (1 variável)
Esta variável é proveniente do IPEADATA. Para construí-la, seria:
Para a população total, precisa-se considerar V0201 = 1 (situação do domicílio = particular permanente), V0202 = 1 a 6 (tipo de domicílio, não considera cômodos e domicílios improvisados ou domicílios coletivos) e V0302 = 1 a 13, e 20 (condição da

pessoa no domicílio, desconsiderando pensionista, empregado(a) doméstico(a) e parente de empregado(a) doméstico(a)).
Para o número de pessoas que moram em domicílio com banheiro e água encanada, considera-se V0206 = 1 a 3 (tem instalação sanitária do tipo rede geral, fossa séptica ligada à rede pluvial ou fossa séptica sem escoadouro), V0207 = 1 (uso da instalação sanitária somente pelo domicílio), V0213 > 0 (há pelo menos um banheiro no domicílio) e V0205 = 1 a 3 (água encanada = rede geral com canalização interna, poço ou nascente com canalização interna ou outra forma com canalização interna).

Com isto, tem-se todas as variáveis explicativas que constituem a matriz X_{t-1}. Para a construção desta matriz, basta posicionar cada variável ao lado da outra, criando a matriz com o número de zonas como linhas e as 25 variáveis supracitadas como colunas.

1.2- Matriz de Tempos - W_{t-1}:

É preciso transformar a informação disponibilizada pela DERSA (3 colunas: origem, destino e tempo ponderado) em uma matriz quadrada, com as linhas significando a origem e as colunas significando os destinos.

1.3- Unidades da Federação - _IUF_11...53:

O modelo possui as 27 dummies de unidade da federação (UF), buscando captar o efeito da UF, que é fixo no tempo, sobre a variável dependente. Na programação, para cada observação, calculam-se as 27 dummies de unidade da federação, da seguinte forma:
Se V1101=11 então _IUF_11=1, caso contrário, _IUF_11=0;
Se V1101=12 então _IUF_12=1, caso contrário, _IUF_12=0;
Se V1101=13 então _IUF_13=1, caso contrário, _IUF_13=0;
Se V1101=14 então _IUF_14=1, caso contrário, _IUF_14=0;
Se V1101=15 então _IUF_15=1, caso contrário, _IUF_15=0;
Se V1101=16 então _IUF_16=1, caso contrário, _IUF_16=0;
Se V1101=17 então _IUF_17=1, caso contrário, _IUF_17=0;
Se V1101=21 então _IUF_21=1, caso contrário, _IUF_21=0;
Se V1101=22 então _IUF_22=1, caso contrário, _IUF_22=0;
Se V1101=23 então _IUF_23=1, caso contrário, _IUF_23=0;
Se V1101=24 então _IUF_24=1, caso contrário, _IUF_24=0;
Se V1101=25 então _IUF_25=1, caso contrário, _IUF_25=0;
Se V1101=26 então _IUF_26=1, caso contrário, _IUF_26=0;
Se V1101=27 então _IUF_27=1, caso contrário, _IUF_27=0;
Se V1101=28 então _IUF_28=1, caso contrário, _IUF_28=0;
Se V1101=29 então _IUF_29=1, caso contrário, _IUF_29=0;
Se V1101=31 então _IUF_31=1, caso contrário, _IUF_31=0;
Se V1101=32 então _IUF_32=1, caso contrário, _IUF_32=0;
Se V1101=33 então _IUF_33=1, caso contrário, _IUF_33=0;
Se V1101=35 então _IUF_35=1, caso contrário, _IUF_35=0;
Se V1101=41 então _IUF_41=1, caso contrário, _IUF_41=0;
Se V1101=42 então _IUF_42=1, caso contrário, _IUF_42=0;
Se V1101=43 então _IUF_43=1, caso contrário, _IUF_43=0;
Se V1101=50 então _IUF_50=1, caso contrário, _IUF_50=0;
Se V1101=51 então _IUF_51=1, caso contrário, _IUF_51=0;
Se V1101=52 então _IUF_52=1, caso contrário, _IUF_52=0;

Se V1101=53 então _IUF_53=1, caso contrário, _IUF_53=0.

1.4- População Inicial = X_Pop_91_0...5 (só o estrato i):

É a agregação do número de domicílios do Censo de 1991 por quintil de renda, por zona. Considera somente a população do estrato em questão na regressão.

1.5- População Final = X_Pop_00_0...5 (só o estrato i):

É a agregação do número de domicílios do Censo de 2000 por quintil de renda, por zona. Considera somente a população do estrato em questão na regressão.

2. Regressões

Para a obtenção do modelo de projeção populacional, inicialmente é necessário utilizar um modelo de regressão com variáveis instrumentais em 2 estágios.
Vale destacar que tanto para a regressão quanto para a projeção da população, há a agregação em 7 setores. Com relação às zonas, para a regressão utiliza-se 788 AMC_pddt, devido a esta ser a máxima abertura possível a partir do Censo de 1991, e para a projeção 899 AMC_pddt, uma vez que as informações utilizadas são de 2000 em diante. Este modelo é estimado para cada um dos 6 estratos (i=0,1,2,...5), conforme a seguir.

○ 1° Estágio:

$$W_{t-1}\ln(p_{i,t-1}) = \gamma_0 + \gamma_1 W_{t-1}(X_{t-1}) + \gamma_2 \ln(p_{i,t-1}) + \gamma_3 X_{t-1} + \gamma_4 \_IUF\_$$

○ 2° Estágio:

$$\ln(p_{i,t}) = \beta_0 + \beta_1 \ln(p_{i,t-1}) + \beta_2 X_{t-1} + \beta_3 \_IUF\_ + \beta_4 [W_{t-1}\widehat{\ln(p_{z,i,t-1})}]$$

Exemplo: Se i=1, então tem-se:

1° Estágio:

- Variável Dependente: W_{t-1}.População Inicial = $W_{t-1}.\ln(p_{1,t-1})$ = W * X_Pop_91_1
- Variáveis Independentes:

W_{t-1}.Variáveis Matriz X = $W_{t-1}.\ln(X_{t-1})$:
W * X_ln_anos_estudo_91; W * X_ln_area_91; W * X_ln_renda_pc_91; W * X_costeira; W * X_struc_91; W * X_Empr_Ag_1; W * X_Empr_Ag_2; W * X_Empr_Ag_3; W * X_Empr_Ag_4; W * X_Empr_Ag_5; W * X_Empr_Ag_6; W * X_Empr_Ag_7; W * Sal_Tot; W * X_Sal_Ag_1; W * X_Sal_Ag_2; W * X_Sal_Ag_3; W * X_Sal_Ag_4; W * X_Sal_Ag_5; W * X_Sal_Ag_6; W * X_Sal_Ag_7; W * X_Pop_91_0; W * X_Pop_91_2; W * X_Pop_91_3; W * X_Pop_91_4; W * X_Pop_91_5.
População Inicial: $\ln(p_{1,t-1})$ = X_Pop_91_1
Variáveis Matriz X = X_{t-1}:
X_ln_anos_estudo_91; X_ln_area_91; X_ln_renda_pc_91; X_costeira; X_struc_91; X_Empr_Ag_1; X_Empr_Ag_2; X_Empr_Ag_3; X_Empr_Ag_4; X_Empr_Ag_5; X_Empr_Ag_6; X_Empr_Ag_7; Sal_Tot; X_Sal_Ag_1; X_Sal_Ag_2; X_Sal_Ag_3; X_Sal_Ag_4; X_Sal_Ag_5; X_Sal_Ag_6; X_Sal_Ag_7; X_Pop_91_0; X_Pop_91_2; X_Pop_91_3; X_Pop_91_4; X_Pop_91_5.
Unidades da Federação (sem dummy SP): _IUF_11; _IUF_12; _IUF_13; _IUF_14; _IUF_15; _IUF_16; _IUF_17; _IUF_21; _IUF_22; _IUF_23; _IUF_24; _IUF_25;

_IUF_26; _IUF_27; _IUF_28; _IUF_29; _IUF_31; _IUF_32; _IUF_33; _IUF_41; _IUF_42; _IUF_43; _IUF_50; _IUF_51; _IUF_52; _IUF_53.

2° Estágio:

- Variável Dependente: População Final = $\ln(p_{1,t})$ = X_Pop_00_1
- Variáveis Independentes:

População Inicial = $\ln(p_{1,t-1})$: X_Pop_91_1
Variáveis Matriz X = X_{t-1}:
X_ln_anos_estudo_91; X_ln_area_91; X_ln_renda_pc_91; X_costeira; X_struc_91; X_Empr_Ag_1; X_Empr_Ag_2; X_Empr_Ag_3; X_Empr_Ag_4; X_Empr_Ag_5; X_Empr_Ag_6; X_Empr_Ag_7; Sal_Tot; X_Sal_Ag_1; X_Sal_Ag_2; X_Sal_Ag_3; X_Sal_Ag_4; X_Sal_Ag_5; X_Sal_Ag_6; X_Sal_Ag_7; X_Pop_91_0; X_Pop_91_2; X_Pop_91_3; X_Pop_91_4; X_Pop_91_5.
Unidades da Federação (sem dummy SP): _IUF_11; _IUF_12; _IUF_13; _IUF_14; _IUF_15; _IUF_16; _IUF_17; _IUF_21; _IUF_22; _IUF_23; _IUF_24; _IUF_25; _IUF_26; _IUF_27; _IUF_28; _IUF_29; _IUF_31; _IUF_32; _IUF_33; _IUF_41; _IUF_42; _IUF_43; _IUF_50; _IUF_51; _IUF_52; _IUF_53.
Variável dependente prevista no 1° Estágio: $\{W_{t-1}.\ln(p_{1,t-1})\}$.

3. Projeções

O IBGE divulga a previsão de população para 2005 por município, para 2010, 2015 e 2020 por zona agregada (unidade da federação) e para 2005, 2010, 2015, 2020 e 2025 do Brasil. Com o modelo especificado acima, projeta-se o n° de domicílios por estrato de renda e por zona. Em seguida, compara-se a projeção feita com o resultado da divisão da previsão de população do IBGE pela relação n° de pessoas por domicílio do censo de 2000, para que seja possível comparar os resultados no nível domicílio.
Para cada ano, calcula-se:

$$\ln(p_{i,t+1}) = \hat{\beta}_0 + \hat{\beta}_1 \ln(p_{i,t}) + \hat{\beta}_2 X_t + \hat{\beta}_3\_IUF\_ + \hat{\beta}_4 [W_t \widehat{\ln}(p_{z,i,t})]$$

Se t=2000, então a variável considerada é a observada. Caso contrário, se t ≥ 2005, então a variável é a projetada.
No caso da população, a variável foi anteriormente projetada, pela equação acima. Assim, para se projetar a população em t+1, utiliza-se como variável explicativa a população em t, que foi projetada utilizando as variáveis explicativas em t-1, ou seja, a variável de população é endógena.
Pela equação acima, tem-se:
$\ln(p_{i,t})$ é a população no estrato i e no tempo anterior (t);
X_t é a matriz de variáveis explicativas em t;
IUF representa as dummies de unidade da federação;
$W_t.\widehat{\ln}(p_{z,i,t})$ é a previsão do produto da matriz de tempos W_t com a população no estrato i e no tempo anterior (t), calculado a partir do 1° estágio.
Dois pontos precisam ser frisados na equação em questão. Primeiramente, com relação a esta previsão do termo $W_t.\ln(p_{z,i,t})$, utiliza-se os coeficientes estimados no 1° estágio, com as variáveis independentes em t. Ou seja, no 1° estágio a equação é:

$$W_t \widehat{\ln}(p_{z,i,t}) = \hat{\gamma}_0 + \hat{\gamma}_1 W_t(X_t) + \hat{\gamma}_2 \ln(p_{i,t}) + \hat{\gamma}_3 X_t + \hat{\gamma}_4\_IUF\_$$

O segundo ponto relevante na projeção da população é a projeção das variáveis independentes de ambas as equações acima.

3.1- Projeção das Variáveis Explicativas - Matriz X_t:

○ Variável **ln(anos de estudo)$_t$** = 3,1.[ln(anos de estudo)$_{t-1}$]$^{3/2}$

○ Variável **(Emprego dos setores r)$_t$** = (Emprego por Setor)$_{t-1}$. (1 + taxa de crescimento de cada setor) (7 variáveis)

Inicialmente, esta variável está aberta em 42 setores com base na RAIS.

A taxa de crescimento de cada setor representa a elasticidade-produto em relação ao emprego de acordo com a equação:

$$\ln(w_{ijt}) = \beta_{ij} + \beta_{it} + \lambda_i \ln(Y_{ijt}) + \gamma_i \ln(L_{ijt}) + \varepsilon_{ijt}$$

Através de uma regressão em painel com as informações do Censo 1991, PNAD 1996 e Censo 2000.

A elasticidade-produto em relação ao emprego é $\varepsilon^Y = -\lambda_i/\gamma_i$ e a elasticidade-salário em relação ao emprego é $\varepsilon^w = 1/\gamma_i$.

Por teoria, a elasticidade ε^Y = (variação percentual do produto) / (variação percentual do emprego), então a taxa de crescimento do emprego por setor = (variação percentual do emprego) = (variação percentual do produto) / ε^Y.

Após o cálculo do emprego dos setores no tempo t, agrega-se os 42 setores da RAIS em 7 setores para a projeção da população.

○ Variável **ln(área da zona)$_t$** = ln(área da zona)$_{t-1}$

É constante no tempo.

○ Variável **(Pop nos outros estratos i)$_t$**

É a projeção do número de domicílios por quintil de renda e por zona para o período anterior. Vale lembrar que se t=2000, então esta variável é a observada com base no Censo 2000. Caso contrário, se t ≥ 2005, então esta variável é a projetada.

○ Variável **ln(renda per capita domiciliar)$_t$** = ln { (renda per capita domiciliar)$_{t-1}$. (salário total)$_t$ / [(salário total)$_{t-1}$. (1 + taxa de crescimento da população da UF)] }

Onde salário total é calculado a partir da massa salarial e emprego, em t.

Para tanto, calcula-se a massa salarial da seguinte maneira:

(massa salarial)$_t$ = (massa salarial)$_{t-1}$. (1 + taxa de crescimento do salário), com taxa de crescimento do salário = 0,144552314 (fonte FIESP).

Inicialmente, a massa salarial e o emprego estão abertos em 42 setores com base na RAIS.

Com isso, chega-se ao salário formal de cada setor em t que é calculado através de:

$$Sal\_Tot_t = \frac{\sum_1^{43} massa\_salarial\_formal_t}{\sum_1^{43} emprego\_formal_t}$$

Para o cálculo da taxa de crescimento da população da UF, é utilizada a relação entre as populações em t e t-1, por UF, com base nas previsões populacionais do IBGE.

○ Variável **(Salário Formal Total)$_t$** = Sal_Tot (1 variável)

É o logaritmo natural de (massa salarial dividida por emprego), por zona, de acordo com o que foi explicitado na variável anterior.

○ Variável **Dummy se município no litoral** = X_costeira (1 variável)

É constante no tempo.

○ Variável **(Salário Formal dos setores r)$_t$** = Sal_Ag_1...7 (7 variáveis)

É o logaritmo natural de (massa salarial dividida por emprego), por setor e zona. Com isso, chega-se ao salário total em t que é calculado através de:

$$Sal\_Ag_t = \frac{massa\_salarial\_formal_t}{emprego\_formal_t}$$

Variável **(% pessoas em domicílio com banheiro e água encanada)$_t$** = 1,44 . ln [1 + (% pessoas em domicílio com banheiro e água encanada)$_{t-1}$].

3.2- Projeção da Matriz de Tempos - W_t:

É disponibilizada pela DERSA, precisando ser transformada em uma matriz quadrada, com as linhas significando a origem e as colunas significando os destinos.

3.3- Unidades da Federação - _IUF_11...53:

São constantes no tempo.

4. Calibragem

Nesta fase, inicialmente, é necessário converter a previsão de população do IBGE para n° de domicílios. Para tanto, divide-se pelo n° de pessoas por domicílio, resultando no n° de domicílios previsto.
Para a comparação da projeção feita pelo modelo com a previsão do IBGE, trabalha-se com três casos:

→ **Caso 1**: Como há a informação divulgada por município para a população de 2005, compara-se a projeção agregada de todos os estratos, por AMC_pddt, com esta informação também agregada por AMC_pddt, de acordo com a equação:

$$c = p_{z,t}\left[\sum_i \overline{p}_{i,z,t}\right]^{-1}$$

Onde $p_{z,t}$ é o n° de domicílios a partir da população prevista pelo IBGE agregado por zona (AMC_pddt);

$\sum_i \overline{p}_{i,z,t}$ é o somatório de todos os estrados do n° de domicílio projetado pelo modelo, agregado por zona.

A divisão do primeiro termo pelo segundo fornece um calibrador que deve ser multiplicado pelo n° de domicílios projetado, desagregado por estrato e zona.

→ **Caso 2**: Para os anos de 2010, 2015 e 2020, o IBGE divulga a previsão populacional por unidade da federação. Podendo-se compara-la com a projeção agregada de todos os estratos, por zona agregada (unidade da federação), segundo a equação:

$$c = p_{Z,t}\left[\sum_i \sum_{z\in Z} \overline{p}_{i,z,t}\right]^{-1}$$

Onde $p_{Z,t}$ é o n° de domicílios a partir da população prevista pelo IBGE por zona agregada (UF);

$\sum_i \sum_{z\in Z} \overline{p}_{i,z,t}$ é o somatório de todos os estratos do n° de domicílio projetado pelo modelo, agregado por UF.

A divisão do primeiro termo pelo segundo fornece um calibrador que deve ser multiplicado pelo n° de domicílios projetado, desagregado por estrato e UF.

→ **Caso 3**: Para o último ano de projeção, há apenas a informação da população total do Brasil prevista para 2025. Com isso, compara-se a projeção agregada de todos os estratos e códigos AMC_pddt, com esta informação, através da equação:

$$c = p_t \left[\sum_i \sum_z \overline{p}_{i,z,t} \right]^{-1}$$

Onde p_t é o n° de domicílios a partir da população prevista pelo IBGE para o Brasil;

$\sum_i \sum_z \overline{p}_{i,z,t}$ é o somatório de todos os estrados e códigos AMC_pddt do n° de domicílios projetado pelo modelo, totalizado para o Brasil.
A divisão do primeiro termo pelo segundo fornece um calibrador que deve ser multiplicado por todas as projeções de n° de domicílio, desagregada por estrato e zona.

Portanto, o n° de domicílios projetado final é a projeção do modelo multiplicada por um fator de calibração, de acordo com a equação:
$P_{i,z,t} = c * p_{i,z,t}$
Onde $P_{i,z,t}$ é o n° de domicílios projetado final, corrigido pelo fator de calibração; c é este fator; e $p_{i,z,t}$ é o n° de domicílios projetado pelo modelo.

1. Variáveis de entrada

1.1 – Descrição das variáveis:

- Inicial: Matriz ZxR de População inicial na zona z no extrato/setor r

- comeca_c: coluna da matriz X na qual os parâmetros de Consumidor (C) começam

- Correlação_r_gr: Matriz RxL (onde L é 1 + o número de diferentes agregações de r)

- Correlação z_gz: Matriz ZxM (onde M é 1 + o número de diferentes agregações de z)

- Crescimento Natural: Matriz (R+1)xR com os seguintes parâmetros para cada extrato/setor r:
  - ○ alpha: fator fixo de crescimento natural;
  - ○ betas: taxa natural de crescimento da população em relação ao ano anterior (r betas).

- Gama: Matriz (K+1)xR com os parâmetros das K variáveis explicativas
  ATENÇÃO: o último parâmetro de Gama é o de WY (uma linha apenas)

- Lambda: Matriz (2xR+2xK+1)xR com os parâmetros dos Z instrumentos (1º estágio) cada extrato/setor r.
  Ordem: constante, Pop, W*Pop, X, WX

- Matriz c_e_f: matriz de coeficientes técnicos da Insumo Produto. A diagonal desta matriz deve estar zerada.

- Nível_Calibragem: Matriz (Tx2) indicando:
  - ○ coluna 1: Nível de Calibragem a ser usado (1 a 5);
  - ○ coluna 2: Apenas para níveis 2 e 4, qual a coluna da matriz de correlação correspondente a ser usada.

- X: Matriz ZxKxT das K variáveis explicativas por zona z no instante t (não inclui a constante nem W*X, nem os extratos de pop)

- Y_Calibragem: struct com os seguintes itens correspodentes ao nível de calibragem (use P como o max(t) que necessita esse nivel de calibragem):
  .c1 = Matriz Z x (P<=T);
  .c2 = Matriz GZ x (P<=T) (GZ é o número de agregações de Z, número máximo no caso de diferentes agregações);
  .c3 = Matriz R x (P<=T);
  .c4 = Matriz GR x (P<=T) (GR é o número de agregações de R, número máximo no caso de diferentes agregações);
  .c5 = Vetor (P<=T) com elementos;
  .c6 = Matriz(GZ x R) Ex: UF por Setor.

- W: Matriz ZxZxT de correlação espacial i,j no instante t

1.2 – Tópicos especiais:

○ todas as variáveis de entrada devem estar somente em nível ou somente em logaritmo natural;
○ a ordenação de setores e zonas pddt devem ser as mesmas para todas as variáveis de entrada;
○ a matriz_c_e_f é opcional para a projeção de População.

Sobre a formatação dos arquivos correspondentes às variáveis de entrada:

- Inicial:

- comeca_c:

- Correlação_r_gr:

- Correlação z_gz:

- Crescimento Natural:

- Gama:

- Lambda:

- Matriz c_e_f:

- Nível_Calibragem:

- X:

- Y_Calibragem:

- W:

2. Saídas

2.1 – Descrição dos arquivos de saída:

- Saídas gráficas: 2 mapas de projeção de emprego ou população por zona pddt e por quintil de emprego ou população.

- Arquivos de texto:
  - ○ y_hat: projeção de emprego ou população para 5 e 10 anos por zona pddt e por setor de atividades;
  - ○ fatores_calibragem: armazena fatores de calibragem resultantes do modelo, utilizados para checagem de consistência dos resultados.

2.2 – Procedimentos para exportação das saídas:

Clique em ...

3. Descrição da rotina em Matlab®

OBJETIVO: Gerar a Projeção de População em z, r, t utilizando as equações de BIDERMAN, Modelo de decisão de localização de empresas e domicílios por zona de tráfego para o Brasil.

Opcionalmente salva os fatores de calibragem gerados internamente.

ATENÇÃO:
1) Os dados das diferentes Matrizes devem obedecer ao mesmo critério de ordenação

2) Para as Variáveis:
- X
- W

t vai de t=1 até t=T, onde Var(t) indica dados em t para a previsão em t+1

3) Para as Variáveis:
- Nivel_Calibragem
- Y_Calibragem

t vai de t=1 até t=T, onde Var(t) indica previsão para t+1 a partir dos dados em t

Exemplo:
- Projeçoes para 3 anos t:
- Yhat(t=t) usará dados de X(t=t), W(t=t), Nivel_Calibragem(t=t), Y_Calibragem(t=t)
- Com isso, todas essas matrizes, na sua dimensão T irão de t=1 até t=3, sendo Var(t=3) os dados necessários para a última previsão de Yhat(t=3)

Usos:

(1) [Yhat Fatores_Calibragem] = projecoes(LN, Inicial, Crescimento_Natural, X, gama, lambda, W, Nivel_Calibragem, Y_Calibragem, Correlacao_z_gz (opcional), Correlacao_r_gr (opcional))

(2) Yhat = projecoes(LN, Inicial, Crescimento_Natural, X, gama, lambda, W, Nivel_Calibragem, Y_Calibragem, Correlacao_z_gz (opcional), Correlacao_r_gr (opcional))

onde: LN = 0 se Y, Y_Calibragem e Inicial estão em Nível
1 se Y, Y_Calibragem e Inicial estão em Ln