

2 Metodologias para Projeção do Mercado da Classe Residencial

Conforme explicitado na introdução deste trabalho, é possível verificar a importância que estudos utilizando técnicas econométricas aliadas a demográficas têm para o entendimento do comportamento do mercado da classe residencial.

Tais estudos propiciam não só aquisição e aprimoramento de modelos de projeções, como também ajudam a compreender o comportamento dos consumidores residenciais, um conhecimento fundamental na definição das tarifas de energia elétrica, na formulação de políticas para a universalização do atendimento, na elaboração de políticas que incentivem a conservação da energia e no planejamento da expansão do sistema elétrica brasileiro.

No caso da classe residencial, conforme descrito por Leon *et al.* (2007) a projeção do consumo depende dos hábitos de consumo, da densidade domiciliar, da posse de eletrodomésticos, do rendimento médio domiciliar vis-à-vis do preço da energia elétrica, da distribuição de renda e do crescimento da economia. Por sua vez, a projeção do número de unidades consumidoras é fruto da projeção do número de domicílios que formam o mercado potencial da classe residencial e de um cenário para a evolução da taxa de atendimento.

Adicionalmente, as previsões de mercado baseiam-se em um conjunto de premissas básicas que incluem tanto cenários macroeconômicos quanto projeções de crescimento populacional e também do número de domicílios, sendo esta última o principal determinante para o cálculo do número de unidades consumidoras residenciais. (EPE, 2011)

Uma vez realizadas as previsões para o consumo de energia elétrica para a classe residencial (Modelagem Econométrica) e a projeção do número de domicílios que serão atendidos no decorrer dos anos 2011 a 2020 (Modelagem Demográfica), faz-se necessário uma análise conjunta destas previsões com o intuito de avaliar a consistência dos resultados obtidos.

Desta maneira, esta seção sobre as metodologias utilizadas ao longo da dissertação encontra-se dividida em três partes, a saber: a primeira compreende a modelagem econométrica de cointegração, a qual permitiu estimar as

elasticidades-preço, renda e preço de eletrodomésticos para o consumo de energia elétrica para a classe residencial, bem como projetar seu consumo futuro.

Na seqüência, é descrita a modelagem demográfica, com ênfase no método da Taxa de Chefia e no modelo de decomposição dos efeitos idade, período e coorte.

Por fim, na terceira parte é descrito o método de decomposição do índice de Divisia de média logarítmica utilizado na análise da consistência das previsões obtidas.

2.1. Modelagem Econométrica

De acordo com Andrade e Lobão (1997), objetivando especificar uma função de demanda de energia elétrica é necessário assumir duas premissas básicas, dentre elas:

- (i) assumir que, para os consumidores residenciais ligados à rede de distribuição, toda a quantidade de energia demandada é efetivamente fornecida. Ou seja, de uma forma geral ou para grande parte dos consumidores, inexistente o problema da demanda reprimida e que a oferta do serviço seja infinitamente elástica, permitindo que a quantidade consumida seja uma boa aproximação para a quantidade demandada;
- (ii) considerar que o comportamento da demanda residencial ao longo do tempo seja influenciado por três variáveis fundamentais: tarifa cobrada pelo serviço, renda familiar e preço dos aparelhos eletrodomésticos, de acordo com a seguinte função não-linear:

$$C_t = k \cdot P_t^\alpha \cdot Y_t^\beta \cdot PE_t^\delta \cdot e^{\varepsilon_t} \quad (2.1)$$

onde:

C_t : consumo residencial de energia elétrica no tempo t ;

P_t : tarifa residencial de energia elétrica no tempo t ;

Y_t : renda familiar no tempo t ;

PE_t : preço dos aparelhos eletrodomésticos no tempo t ;

e : número neperiano.

Aplicando-se a transformação logarítmica em ambos os lados da equação (2.1) a forma funcional da função demanda estimada passa ser a seguinte:

$$\ln C_t = k + \alpha \ln P_t + \beta \ln Y_t + \delta \ln PE_t + \varepsilon_t \quad (2.2)$$

$$\alpha < 0, \delta < 0 \text{ e } \beta > 0$$

Os coeficientes α , β e δ são os parâmetros a serem estimados e, conforme a teoria econômica, espera-se que o consumo reaja negativamente aos aumentos de tarifa e do preço de eletrodomésticos, positivamente, aos aumentos de renda.

Na estimação da função demanda por energia elétrica, duas abordagens econométricas se destacam: o modelo de equações simultâneas¹ e as técnicas de cointegração (SHIMIDT e LIMA, 2002).

Assim, como nos trabalhos de Andrade e Lobão (1997), Silk e Joutz (1997), Silva (2001), Schmidt e Lima (2004) e Mattos e Lima (2005), optou-se por adotar a segunda metodologia, baseada nos modelos de Johansen (1988, 1991) e Johansen e Juselius (1990). De forma bastante resumida, tal metodologia consiste em estimar por Máxima Verossimilhança, o(s) vetor(es) de cointegração, valendo-se da modelagem Vetorial Auto-Regressiva (VAR), para na sequência estimar um Mecanismo de Correção de Erros Vetorial (VECM).

Os Modelos Auto-regressivos (VAR) podem ser assim representados:

$$X_t = v + \beta_1 \cdot X_{t-1} + \xi_t \quad (2.3)$$

onde: $X_t = (X_{1t}, \dots, X_{kt})'$ é o vetor com a K de variáveis do modelo ($K \times 1$), β_i são matrizes de coeficientes a serem estimados ($K \times K$), $v = (v_1, \dots, v_k)'$ é o vetor de intercepto ($k \times 1$). Finalmente, $\xi_t = (\xi_{1t}, \dots, \xi_{kt})'$ é o termo ruído branco k -dimensional, com distribuição normal multivariada $\varepsilon \sim N(0, \Sigma)$.

Já no Modelo Correção Vetorial de Erro (VECM) parte-se da seguinte expressão:

$$\Delta X_t = v + \alpha_1 \cdot u_{t-1} + \xi_t \quad (2.4)$$

¹ Para maiores informações: Amemiya (1978), Dubin e Macfadden (1984) e Epple e McCallum (2006).

sendo: ΔX_t , o vetor das diferenças das variáveis endógenas, ν o vetor de intercepto, u_{t-1} a relação de cointegração em $t-1$, ξ_t o termo aleatório.

Cabe aqui destacar que ao diferenciar uma série econômica para torná-la estacionária, podem ocorrer perdas de informações de longo prazo entre as variáveis fazendo-se necessário a introdução do Termo de Correção de Erro, uma vez que o mesmo pode ser entendido como “*a forma a considerar as relações de equilíbrio de longo prazo estabelecida pela teoria econômica com os desvios de curto prazo*” (SILVA, 2001).

Pelo exposto, a modelagem (VAR-VECM) parece ser realmente adequada para este estudo, uma vez que considera tanto as variáveis quanto suas respectivas defasagens conciliando as tendências de curto e longo prazos das variáveis do modelo. Schmidt e Lima (2004) destacam ainda que o referido enfoque possibilita condições suficientes para se obter uma boa representação estatística dos dados, não sendo necessário introduzir variáveis *ad-hoc a priori*, como seria o caso no Modelo de Equações Simultâneas para identificação da existência de uma ou mais equações.

Para Storck (2008) a aplicação do referido modelo pode ser desenvolvida em quatro etapas, a saber:

1^a – pautados na metodologia desenvolvida por Engle e Granger (1987), na primeira etapa deve-se examinar se as séries são não estacionárias e se apresentam a mesma ordem de integração. Tal análise pode ser realizada através da aplicação do Teste de Dickey-Fuller Aumentado (ADF);

2^a – realizada a etapa anterior e identificada que as séries apresentam ordens de integração semelhantes, devem-se analisar as defasagens das mesmas objetivando especificar de maneira correta o Vetor Auto Regressivo (VAR). Além disso, outra análise, de suma importância, consiste na verificação da necessidade de se inserir ou não termos determinísticos nas séries como constante, tendências ou variáveis *dummies*;

3^a – uma vez determinada à correta especificação, na terceira etapa realiza-se os testes λ -traço e o teste λ -máximo, propostos por Johansen (1991, 1998) e Johansen e Juselius (1990), com intuito de verificar a existência de vetores de cointegração;

4^a – por fim, executadas as etapas supracitadas deve-se incluir os termos de correção de erros vetoriais, assim como especificados por Engle e Granger (1987), objetivando determinar a magnitude do desvio de curto prazo e a condução dos mesmos para o equilíbrio de longo prazo.

Realizada esta descrição sucinta das etapas a serem seguidas, as mesmas serão apresentadas de uma maneira mais abrangente.

2.1.1. Testes de Raízes Unitárias

Para a maior parte das técnicas de cointegração as propriedades das séries temporais (processos estocásticos) precisam ser investigadas. Em particular, precisa-se determinar se as variáveis são estacionárias em níveis e, portanto, integradas de ordem zero, $I(0)$ ou são não estacionária e, portanto, têm raiz unitária, necessitando de diferenciação para alcançar tal estacionariedade sendo, então, integradas de ordem d , $I(d)$ onde d representa o número de vezes que a série temporal necessita ser diferenciada para tornar-se estacionária. Contudo, é importante mencionar que ao se remover a tendência, elementos de longo prazo entre as variáveis podem ser eliminados. Isto é necessário porque, segundo Gujarati e Porter (2011), “*se uma série temporal é não estacionária, podemos estudar seu comportamento apenas no período de tempo em consideração*”; isto é, o conjunto de dados da série temporal será específico, impossibilitando generalizá-lo para outros períodos. Em contrapartida, caso a série temporal seja estacionária, sua média, variância e autocovariâncias (em várias defasagens) serão invariantes no tempo.

Harris (1995) salienta que, a utilização de modelos de regressão envolvendo séries temporais não estacionárias pode conduzir a um fenômeno denominado Regressão Espúria ou sem sentido. Por exemplo, se Y_t e X_t são duas séries temporais não estacionárias do tipo passeio aleatório (*random walk*), a regressão de uma variável na outra proporciona resultados inconsistentes, em que os testes de significância convencionais apontam para relações entre as variáveis que, na verdade, inexistem.

Neste contexto, o estudo de cointegração torna-se importante na análise de séries econômicas não estacionárias, pois, assim como supracitado, a presença de

raiz unitária na série temporal ocasiona resultados viesados, invalidando os pressupostos da estatística clássica de que a média e a variância são constantes ao longo do tempo, e, com isto, distorce o relacionamento entre duas, ou mais, variáveis.

Com isso, torna-se claro que devido aos riscos que se têm ao trabalhar com séries erroneamente identificadas, é fundamental estudar previamente as propriedades univariadas das séries individuais antes de se iniciar efetivamente análises mais complexas como, por exemplo, análises multivariadas. Em outras palavras, antes de fazer inferência com base em modelos de regressão, faz-se necessário efetuar testes de raízes unitárias de todas as séries envolvidas.

Com intuito de testar se uma série temporal é integrada ou não, é necessário realizar os chamados testes de raízes unitárias, sendo os mais usuais os testes Dickey-Fuller (DF) e Dickey-Fuller Aumentado (ADF), que serão detalhados a seguir.

2.1.1.1.

Testes Dickey-Fuller (DF) e Dickey-Fuller Aumentado (ADF)

Considere um modelo auto-regressivo de primeira ordem ou $AR(1)$, tal como:

$$Y_t = \rho Y_{t-1} + u_t \quad -1 \leq \rho \leq 1 \quad (2.5)$$

em que u_t é um termos de erro Ruído Branco.

Nota-se que, se $\rho = 1$, isto é, no caso de raiz unitária, a equação (2.5) torna-se um modelo de passeio aleatório sem deslocamento, ou seja, um processo estocástico não estacionário. Subtraindo Y_{t-1} em ambos os lados da expressão (2.5) obtêm-se a seguinte expressão matemática:

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\ &= (\rho - 1)Y_{t-1} + u_t \end{aligned} \quad (2.6)$$

que pode também ser escrita como

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (2.7)$$

sendo $\delta = (\rho - 1)$ e Δ o primeiro operador da diferença.

Uma vez estimada a equação (2.6), deve-se analisar a hipótese nula de que há uma raiz unitária ($H_0 : \delta = 0$) contra a hipótese alternativa, de que a série temporal é estacionária, possivelmente, em torno de uma tendência determinística ($H_1 : \delta < 0$). Para realização do referido teste é importante saber que, sob a hipótese nula $\delta = 0$ (ou seja, $\rho = 1$), o valor t do coeficiente estimado de Y_{t-1} não segue uma distribuição t , mesmo em grandes amostras, isto é, não possui uma distribuição assintótica normal.

Dickey e Fuller (1979) a fim de solucionar o problema demonstraram que sob a hipótese nula, o valor estimado t do coeficiente Y_{t-1} na equação (2.7) segue uma estatística τ (tau)². Entretanto, de forma curiosa, caso a hipótese de que $\delta = 0$ seja rejeitada (caso em que a série temporal é estacionária), pode-se utilizar o teste $t - Student$.

Também é possível elaborar o teste partindo-se de um processo estocástico com constante:

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t \quad (2.8)$$

Ou ainda, considerar além da constante, uma tendência linear determinística:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t \quad (2.9)$$

Nesses casos caber ressaltar que o procedimento do teste é o mesmo para todas as equações (2.7), (2.8) ou (2.9); isto é, testa-se a hipótese nula de que δ seja estatisticamente igual a zero. Contudo, os valores críticos tabelados para cada equação são diferentes.

Caso o verdadeiro processo gerador dos dados históricos seja expresso por um modelo auto-regressivo de ordem p ; ou seja, $AR(p)$, estes testes não podem

² DICKEY, D. A.; FULLER, W.A. "Distribution of the estimators for autoregressive times series with a unit root." *Journal of the American Statistical Association*, v. 74, p. 427-431. Ver também, FULLER, W. A. *Introduction to statistical time series*. Nova York: John Wiley & Sons, 1976.

ser aplicados, uma vez que os resíduos da regressão obtidos a partir das equações (2.7), (2.8) e (2.9) apresentarão autocorrelação serial, o que tornará o teste não eficiente. A fim de resolver o problema, a solução proposta pelos próprios autores do teste foi adicionar ao lado direito de cada uma das equações, defasagens da primeira diferença da variável Y_t . Dessa maneira, as equações do teste de Dickey-Fuller aumentado, comumente chamado de ADF (*Augmented Dickey-Fuller*), poderão ser reescritas como:

$$\Delta Y_t = \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + u_t \quad (2.10)$$

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + u_t \quad (2.11)$$

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + u_t \quad (2.12)$$

em que t é o tempo ou a variável de tendência e o subscrito i indica o número de defasagens utilizadas. O número de defasagens a ser incluída deverá ser o necessário para corrigir a autocorrelação dos resíduos das equações (2.10), (2.11) e (2.12), de modo a obter uma estimativa não viesada de δ , coeficiente defasado de Y_{t-1} . Em linhas gerais, para se determinar o número ideal de defasagens utiliza-se algum critério de informação como, por exemplo, o Critério de Akaike³ (AIC) ou de Bayesiano de Schwarz⁴ (SBC).

No que tange, aos valores tabelados da estatística t , assim como no teste de Dickey-Fuller (DF), também não podem ser utilizados como valores críticos para

³ Conforme Akaike (1974), seu critério de informação é dado por $AIC = -2 \log L(\hat{\theta}) + 2(p)$, em que o viés é dado assintoticamente por p , onde p é o número de parâmetros a serem estimados no modelo. Pelo critério de Akaike, o melhor modelo é o que possui menor valor de AIC.

⁴ O Critério de Informação Bayesiano (BIC), proposto por Schwarz (1978) pode ser representado da seguinte forma: $BIC = -2 \log f(x_n | \theta) + p \log n$, em que $f(x_n | \theta)$ é o modelo escolhido, p é o número de parâmetros a serem estimados e n é o número de observações da amostra. De maneira similar ao Critério de Akaike, deve-se escolher para a determinação do número de defasagens do respectivo modelo ARMA o menor valor obtido pelo BIC.

a decisão do testes, uma vez que a presença da variável dependente defasada do lado direito das equações referidas implicaria em termos aleatórios autocorrelacionados viesando, desta maneira, a estimativa tanto da própria variância quanto da variância do coeficiente de interesse estimado. O fato, por sua vez, comprometeria a estatística t calculada.

Objetivando solucionar o problema supracitado, novos valores críticos foram obtidos quando os erros do modelo, sob a hipótese nula (H_0), se mostram independentes e identicamente distribuídos (*i.i.d*). Os mesmos podem ser conferidos em MacKinnon (1991).

2.1.2. Análise de cointegração

Como exposto na subseção anterior, uma variável é integrada de primeira ordem se após a aplicação de uma única diferenciação obtém-se uma série estacionária. Contudo, quando duas variáveis são integradas de ordem um ($I(1)$), mas a combinação linear entre elas é estacionária, isto é, apesar de serem ambas integradas de ordem um e a sua combinação for integrada de ordem zero, elas serão denominadas como Cointegradas. A principal implicação da cointegração entre duas variáveis reside no fato de passar a existir um equilíbrio de longo prazo entre elas.

Uma vez analisada a condição de estacionariedade das séries, bem como identificadas suas respectivas ordens de integração, as propriedades da cointegração das mesmas também necessitam ser estudadas. Para isso, a segunda etapa tem por objetivo construir um vetor auto-regressivo (VAR), sendo necessário determinar a(s) ordem(ns) da(s) defasagem(ns) das variáveis endógenas, já que o referido procedimento tem como base a hipótese de que, colocando-se um número suficiente de defasagens, é possível se adquirir uma estrutura de resíduos bem comportados, ou seja, estacionários.

Para Pesaran e Smith (1999) a identificação adequada do número de vetores depende sobremaneira de algumas decisões tomadas por parte do analista. Por exemplo, as diversas combinações possíveis entre o número de variáveis adotadas no modelo VAR, o número de defasagens (k) destas variáveis inseridas ao modelo, assim como, a inclusão ou não de termos determinísticos, tais como: constante, tendência linear, *dummies* sazonais, *dummies* de intervenção, ou

qualquer outro tipo de regressor que seja considerado fixo e não estocástico, possibilitam o encontro de resultados bastante distintos.

Margarido (2004) corroborando a assertiva dos autores supracitados destaca que, assim como no caso univariado, o número de defasagens a ser introduzida, a fim de se adquirir uma estrutura de ruído branco para o resíduo, deve ser obtido através dos critérios de informação, uma vez que cada regressor adicional no modelo VAR ocasiona uma estagnação da soma dos resíduos ou, em muitos casos, até mesmo diminuição das mesmas. Dessa forma, buscando uma solução de compromisso entre a redução dos erros e o aumento do número de regressores deve-se associar uma penalidade a esse aumento, isto é, quando a penalidade decorrente da incorporação de um regressor adicional for menor que a diminuição da soma de resíduos, deve-se adicioná-lo ao modelo. Caso contrário, o regressor não deve ser incorporado.

Cabe ressaltar que os critérios de informação são generalizações diretas dos modelos uniequacionais, sendo os mais utilizados na literatura os AIC (*Akaike Information Criterion*), o SBC (*Schwarz Bayesian Criterion*), ou então, o HQ (*Hannan-Quinn*).

A fim de obter um melhor entendimento do teste de Johansen (1988), considere o modelo VAR de dimensão (k), com erros gaussianos:

$$X_t = \Gamma_1 X_{t-1} + \dots + \Gamma_k X_{t-k} + \phi D_t + u_t \quad (2.13)$$

onde: $u_t \sim IN(0, \Sigma)$, X_t é um vetor ($n \times 1$), cada elemento Γ_i é uma matriz de parâmetros de ordem ($n \times n$), D_t representa os termos determinísticos e ϕ seus parâmetros.

Caso as variáveis não apresentem a mesma ordem de integração ou, mesmo apresentando, não se verifique relações de cointegração entre as mesmas, o modelo VAR, na forma padrão, acima apresentado é apropriado. Caso contrário, se pelo menos dois vetores forem cointegrados, torna-se necessário introduzir no sistema VAR o vetor de correção dos erros (ΠX_{t-1}).

Desenvolvidos por Engle e Granger (1987), os modelos VECM têm como finalidade a inserção de ajustes de curto prazo, devido à presença de cointegração. Dessa forma, um modelo VECM pode ser representado da seguinte forma:

$$\Delta X_t = \Pi_1 X_{t-k} + \Gamma \Delta X_{t-1} + \dots + \Gamma_{k-1} \Delta_{t-(k-1)} + \phi D_t + u_t \quad (2.14)$$

em (2.14), X_t é um vetor ($n \times 1$) que deve ser testado por cointegração, e $\Delta X_t = X_t - X_{t-1}$; D e ϕ são o termo determinístico que contém variáveis exógenas ao sistema e seu parâmetro, respectivamente; Γ representa a matriz dos coeficientes que medem a velocidade de ajustamento do modelo no curto prazo e, por fim, Π é a matriz contém os coeficientes que medem o ajustamento de longo prazo.

De acordo com Duarte, Lima e Alves (2007) em relação ao vetor de correção de erros (ΠX_{t-1}), a matriz Π pode ser escrita como $\Pi = \alpha \beta'$, então quando o posto é igual a um ($r = 1$) é possível obter:

$$\Pi X_{t-1} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} (\beta_1 \dots \beta_n) \begin{pmatrix} X_{t-1} \\ \vdots \\ X_{nt-n} \end{pmatrix} \quad (2.15)$$

Normalizando pelo coeficiente de uma das variáveis de X_t , por exemplo, temos:

$$\begin{aligned} &= \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} (\beta_1 X_{1t-1} + \dots + \beta_n X_{nt-1}) \beta_1 / \beta_1 \\ &= \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \beta_1 \left(X_{1t-1} + \dots + \frac{\beta_n}{\beta_1} X_{nt-1} \right) \\ &= \begin{pmatrix} \alpha_1 \cdot \beta_1 \\ \vdots \\ \alpha_n \cdot \beta_1 \end{pmatrix} \left(X_{1t-1} + \dots + \frac{\beta_n}{\beta_1} X_{nt-1} \right) \\ &= \begin{pmatrix} \tilde{\alpha}_1 \\ \vdots \\ \tilde{\alpha}_n \end{pmatrix} (VC) \end{aligned} \quad (2.16)$$

em (2.16), VC representa o vetor de cointegração do sistema e $(\tilde{\alpha}_1 \cdots \tilde{\alpha}_n)$ são os coeficientes que medem o ajustamento de longo prazo das variáveis do sistema. Realizando a substituição da equação (2.16) em (2.14) obtém-se:

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma \Delta X_{t-i} + \tilde{\alpha}(VC) + \phi D + u_t + \mu \quad (2.17)$$

onde μ representa o intercepto do sistema.

Conforme Enders (1996 e 2004) o número de vetores de cointegração pode ser determinado a partir do posto (r) da matriz Π . Isto é, caso o posto seja igual a zero ($r = 0$), a matriz é nula e a equação de número (2.17) será um modelo VAR em primeira diferença. Caso Π tenha posto completo, isto é, se ($r = n$) as variáveis de X_t serão estacionárias, não cabendo qualquer análise de cointegração. Por fim, nas situações em que ($1 \leq r < n$), as relações de longo prazo entre as variáveis são determinadas por r vetores de cointegração. Nesse caso, ΠX_{t-1} é o termo de correção de erros, responsável por relacionar a dinâmica de curto prazo com equilíbrio de longo prazo.

Segundo Mattos e Lima (2005), a matriz Π é estimada por máxima verossimilhança e o (s) vetor (es) de cointegração, valendo-se da modelagem de um vetor auto-regressivo (VAR). Num segundo momento, com base na estimativa do vetor de cointegração, estima-se um Mecanismo de Correção de Erros (MCE), com vetor de cointegração.

A determinação do número de vetores de cointegração consiste na solução de um problema de autovalor em que o posto da matriz Π é igual ao número de autovalores não nulos.

Partindo deste princípio, Johansen e Juselius (1990) desenvolveram dois testes estatísticos a fim de avaliar a significância do posto da matriz Π . O primeiro, conhecido como Teste do Traço e usualmente indicado por λ_{trace} , objetiva testar a hipótese nula de que o número de vetores de cointegração é menor ou igual a r . Ao passo que o segundo, denominado de Teste do Máximo Autovalor (λ_{max}), tem por finalidade testar a hipótese nula de r vetores de cointegração contra a alternativa de $r+1$. Vale ressaltar que para implementação

desses testes, os autovalores são ordenados de forma decrescente e os testes sequenciais são realizados.

As estatísticas dos testes assim podem ser representadas:

$$\text{Estatística do Traço: } \lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i) \quad (2.18)$$

$$\text{Estatística do máximo autovalor: } \lambda_{\text{max}}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (2.19)$$

onde: λ_i são valores estimados das raízes características (autovalores) da matriz Π , T é o número de observações e n o número de variáveis adotadas.

Por fim, é importante mencionar que os resultados dos testes supracitados podem ser conflitantes, uma vez que o teste de máximo autovalor possui hipóteses alternativas mais rigorosas se comparado com a hipótese do teste do traço. Dessa forma, prefere-se o primeiro para determinação do número de vetores cointegrantes (ENDERS, 1996).

2.2. Modelagem Demográfica

Uma projeção de mercado da classe residencial deve indicar a tendência esperada para o crescimento do consumo de energia elétrica bem como apontar a tendência futura para o crescimento do número de unidades consumidoras.

O conjunto de unidades consumidoras da classe residencial é formado pelos domicílios que contam com o fornecimento regular de energia elétrica por parte de alguma concessionária de distribuição. No caso do Brasil, uma pequena parcela dos domicílios ainda não é atendida, assim a relação entre o número de unidades consumidoras residenciais (NCR) e o número de domicílio (DOM) é estabelecida pela seguinte relação, onde TA denota a taxa de atendimento, ou seja, a parcela dos domicílios que contam com o fornecimento de energia elétrica:

$$NCR = DOM \times TA \quad (2.20)$$

Para um mercado universalizado a taxa de atendimento vale 1 (100%), significando que o total de unidades consumidoras residenciais é o próprio número de domicílios.

Conforme ilustrado em Leon *et al.* (2006), os censos demográficos classificam os domicílios segundo as condições de ocupação. Por exemplo, na classificação por espécie, os domicílios podem ser particulares e coletivos. Os domicílios particulares podem ser permanentes (casa, apartamento ou cômodo destinado à moradia) ou improvisados (imóvel não destinado exclusivamente à moradia, lojas, indústrias, salas comerciais, barracas, carroças, grutas, etc).

Por sua vez, os domicílios permanentes subdividem-se em ocupados e não ocupados, sendo que estes últimos são categorizados em vagos, fechados e de uso ocasional. Os domicílios vagos são aqueles que não tinham moradores na data de referência, enquanto os domicílios fechados são aqueles que na data de referência estavam ocupados, porém seus moradores se encontravam ausentes durante o período de coleta do censo. No Censo 2010 os domicílios fechados foram classificados como sendo domicílios ocupados sem entrevista. Por fim, os domicílios ocasionalmente ocupados incluem as residências utilizadas para descanso de fins de semana, férias e outros fins.

Leon & Pessanha (2005) estabelecem uma correspondência entre a classificação dos domicílios nos censos demográficos e as classes de consumo de energia elétrica definidas na Resolução ANEEL nº 456/2000. A partir do desenho conceitual da classificação dos domicílios nos censos, Leon *et al.* (2006) concluem que os domicílios coletivos, os particulares improvisados e os permanentes vagos não são unidades consumidoras de energia elétrica da classe residencial, conforme as razões a seguir:

- O domicílio coletivo não tem o seu consumo de energia elétrica classificado na classe residencial, mas é contabilizado nas classes comercial, industrial ou outras classes;
- O domicílio particular improvisado pode ter seu consumo de energia elétrica classificado nas classes comercial e industrial ou pode ter uma ligação clandestina e, neste caso, é computado nas perdas;

- O domicílio vago é aquele que não têm moradores na data de referência do Censo e mesmo que tenha medidor de consumo de energia elétrica instalado não é uma unidade consumidora de eletricidade, pois não há um responsável em firmar um contrato de fornecimento com uma concessionária de distribuição. Um bom exemplo de domicílio vago é um imóvel para aluguel, onde o inquilino é o responsável por contratar o fornecimento de energia elétrica e encerrá-lo após sua desocupação.

Logo, os domicílios ocupados, fechados (ocupados, mas sem entrevista) e de uso ocasional formam o conjunto de clientes potenciais da classe residencial (Leon *et al.*, 2006). Este conjunto de domicílios totalizava 91,0% dos domicílios particulares recenseados no ano 2010.

Na projeção do número de domicílios potencialmente consumidores da classe residencial os domicílios ocupados devem ser tratados de forma diferente dos domicílios de uso ocasional, pois não há uma contagem da população residente nesta última categoria (Leon *et al.*, 2006). Assim, é necessário inicialmente projetar o número de domicílios particulares, permanentes ocupados (com os domicílios fechados inclusive) a partir da projeção da população residente disponibilizada pelo IBGE. Na seqüência são elaboradas as projeções anuais do número de domicílios ocasionalmente ocupados com base nas participações destes no total de domicílios.

2.2.1. Projeção do número de domicílios ocupados

Para a projeção do número de domicílios ocupados (com os fechados inclusive) propõem-se o emprego do método da taxa de chefia (Kono, 1987, Rodriguez *et al.*, 1991, Givisiez *et al.*, 2005, Pessanha *et al.*, 2007) o qual assume o número de domicílios em uma população como sendo igual ao número de chefes de domicílios na mesma população. De forma bastante resumida, um chefe de domicílio é qualquer pessoa acima de 10 anos de idade responsável pelo domicílio. Este pressuposto permite classificar os domicílios ocupados segundo as idades e os sexos dos seus respectivos chefes. Assim, seja ${}_n H_x^s(t)$ o total de domicílios chefiados por indivíduos do sexo s e com idade entre x e $x+n$ anos completos no ano t .

O total de domicílios no ano t , $H(t)$, pode ser expresso como a soma dos totais de domicílios ao longo de todos os grupos etários nas populações masculina e feminina:

$$H(t) = \sum_{idade} \sum_{sexo} {}_n H_x^s(t) \quad (2.21)$$

A taxa de chefia é expressa como a razão entre o número de chefes de domicílio e o número de habitantes. Esta taxa deve ser calculada por sexo e por faixa etária, conforme indicado a seguir:

$${}_n T_x^s(t) = \frac{{}_n H_x^s(t)}{{}_n P_x^s(t)} \quad (2.22)$$

onde

${}_n T_x^s(t)$ é a Taxa de chefia do sexo s , no grupo etário de x a $x+n$ anos completos, no ano t ;

${}_n H_x^s(t)$ é o número de domicílios chefiados por indivíduos do sexo s e com idade entre x e $x+n$ anos completos no ano t ;

${}_n P_x^s(t)$ é a população do sexo s , na faixa etária de x a $x+n$ anos completos no ano t .

Da equação (2.22) tem-se que o número de domicílios chefiados por indivíduos do sexo s na faixa etária de x a $x+n$ anos completos no ano t é o produto da taxa de chefia pela respectiva população, conforme apresentado:

$${}_n H_x^s(t) = {}_n T_x^s(t) \cdot {}_n P_x^s(t) \quad (2.23)$$

Substituindo a equação (2.23) na equação (2.21) obtém-se a equação de projeção do total de domicílios como resultado da soma dos produtos entre as projeções das taxas de chefia e das populações em cada sexo e em cada faixa etária:

$$H(t) = \sum_{idade} \sum_{sexo} {}_n T_x^s(t) \cdot {}_n P_x^s(t) \quad (2.24)$$

Objetivando projetar o total de domicílios através das equações acima são necessárias as projeções da população e da taxa de chefia por grupo etário e por sexo.

As projeções populacionais por sexo e por faixa etária, ${}_n P_x^s(t)$, são fornecidas pelo Instituto Brasileiro de Pesquisa e Estatística (IBGE) na forma de pirâmides etárias, conforme ilustrado na figura 5 para os anos de 2010 e 2020.

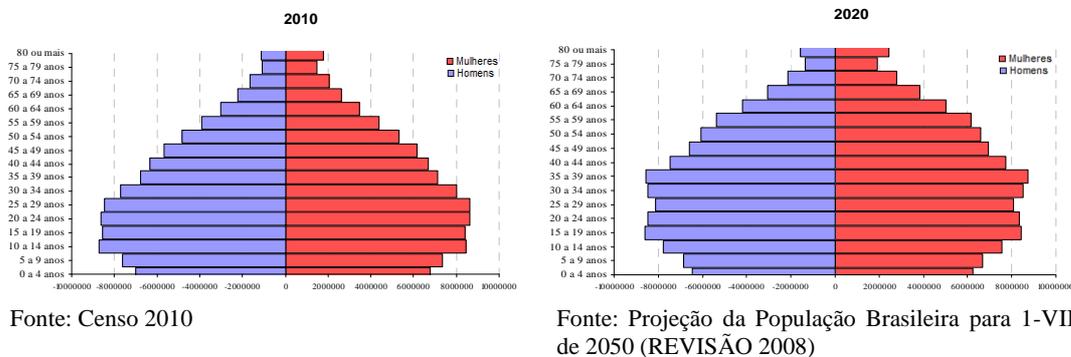


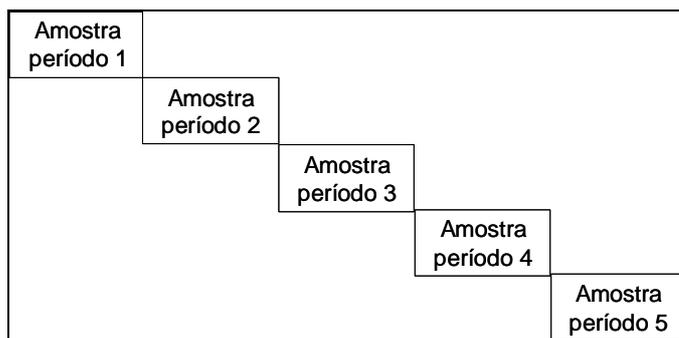
Figura 1: Pirâmides etárias da população brasileira em 2010 e 2020 (IBGE)

Por sua vez, a projeção da taxa de chefia por sexo e faixa etária (${}_n T_x^s(t)$) deve ser elaborada pelo próprio analista e usualmente a sua projeção fundamenta-se na aplicação de uma taxa constante aplicada na projeção populacional. Segundo Giviziez *et al.* (2005), tal procedimento pode representar uma importante fonte de erro na projeção do número de domicílios ocupados.

Desta forma, objetivando solucionar as deficiências da hipótese da taxa de chefia constante, Rios-Neto *et al.* (2005) e Giviziez *et al.* (2005) propõem o emprego do modelo idade, período e coorte – IPC (Deaton, 1997, Glenn, 2005) na projeção das taxas de chefia (${}_n T_x^s(t)$).

A coleção dos censos ao longo de algumas décadas fornece os dados necessários para ajustar um modelo IPC e com ele projetar as taxas de chefia nas populações masculina e feminina, em cada uma das faixas etárias acima de 10 anos de idade.

Os censos realizados pelo IBGE têm dois questionários: o principal e o da amostra. O questionário principal tem por finalidade fazer o censo de um seletor conjunto de variáveis, já o questionário da amostra, mais amplo que o principal, é respondido apenas pelos indivíduos selecionados em uma amostra da população. Neste caso, os dados oriundos da pesquisa por amostragem formam uma seqüência de cortes transversais repetidos (*repeated cross-section*), um conjunto de dados ao longo do tempo, mas sem uma sobreposição entre os casos (indivíduos da população), pois a amostra é diferente a cada período, conforme ilustrado a seguir na figura 6:



Fonte: Adaptado de Menard (2002)

Figura 2: Seqüência de cortes transversais

Como os mesmos indivíduos não são medidos nas várias edições dos censos, os dados de cortes transversais repetidos permitem identificar tendências na evolução do comportamento, tanto transversalmente quanto longitudinalmente, porém apenas em níveis mais agregados como, por exemplo, grupos étnicos, classes sociais e coortes de nascimento.

Uma coorte é definida como um grupo de pessoas dentro de uma população que experimentaram o mesmo evento significativo em suas vidas em um dado período de tempo (GLENN, 2007). Por exemplo, uma coorte de nascimento abrange um grupo de pessoas que nasceram em um período de tempo particular, por exemplo, a mesma década.

As tendências na evolução do comportamento podem refletir mudanças transversais e/ou longitudinais, por exemplo, podem expressar uma mudança na composição etária da população ou uma mudança do comportamento ao longo do tempo.

Transversalmente, as mudanças expressam as diferenças entre indivíduos de idades diferentes observados em um mesmo momento (MENARD, 2002), por exemplo, diferenças entre os sujeitos com 40 anos de idade e os sujeitos com 50 anos de idade em 2010. Neste caso, as diferenças estão confundidas, pois podem ser devido às diferenças entre as distintas coortes de nascimento (efeito coorte) ou devido aos grupos etários em um particular instante de tempo (efeito idade).

Por sua vez, longitudinalmente as mudanças expressam as diferenças entre os mesmos indivíduos ao longo do tempo (MENARD, 2002), por exemplo, diferenças entre indivíduos com 40 anos de idade em 2000 e estes mesmos indivíduos quando eles têm 50 anos de idade em 2010. Neste caso, as mudanças também estão confundidas, pois as diferenças podem ser interpretadas como o

desenvolvimento comportamental dentro de uma mesma coorte de nascimento (efeito idade) ou entre os grupos etários ao longo do tempo (efeito período).

Portanto, para esclarecer a natureza das tendências é preciso examinar os efeitos idade, período e coorte.

A idade fornece uma explicação para a evolução do comportamento e o efeito idade reflete as conseqüências do processo de envelhecimento dos indivíduos (GLEEN, 2007).

O período fornece uma explicação de natureza histórica, estando associado a eventos que ajudam a moldar o comportamento. Assim, o efeito período se relaciona com as mudanças da sociedade ao longo do tempo e que afetam simultaneamente toda a população independentemente da faixa etária (GLEEN, 2007).

Por fim, o efeito coorte está associado a mudança de comportamento compartilhada pelos indivíduos que nasceram em uma mesma década (GLEEN, 2007). O efeito coorte de nascimento explica o comportamento em termos da idade de um indivíduo em um particular ano ou período, em outras palavras, ele é uma *proxy* da interação dos efeitos idade e período.

Para ajustar um modelo IPC, as estatísticas de população e taxas de chefia registradas nos censos devem ser organizadas em uma tabela de contingência na qual as linhas representam as faixas etárias, as colunas os períodos e as diagonais as coortes de nascimento, conforme ilustrado na tabela 1. Cada coorte representa um período de nascimento, por exemplo, a coorte 1 (C1) abrange os anos de 1991 até 2000, enquanto a coorte 11 (C11) começa em 1891 e termina em 1900, conforme ilustrado na tabela 2.

Tabela 1: Idades, Período e Coortes

Idade (Anos)	Período				
	1970	1980	1990	2000	2010
10 a 19	Coorte 5	Coorte 4	Coorte 3	Coorte 2	Coorte 1
20 a 29	Coorte 6	Coorte 5	Coorte 4	Coorte 3	Coorte 2
30 a 39	Coorte 7	Coorte 6	Coorte 5	Coorte 4	Coorte 3
40 a 49	Coorte 8	Coorte 7	Coorte 6	Coorte 5	Coorte 4
50 a 59	Coorte 9	Coorte 8	Coorte 7	Coorte 6	Coorte 5
60 a 69	Coorte 10	Coorte 9	Coorte 8	Coorte 7	Coorte 6
70 +	Coorte 11	Coorte 10	Coorte 9	Coorte 8	Coorte 7

Fonte: Elaboração Própria do Autor

Tabela 2: Coorte de nascimento

Coortes	Períodos de nascimento	Coortes	Períodos de nascimento
C1	1991 a 2000	C7	1931 a 1940
C2	1981 a 1990	C8	1921 a 1930
C3	1971 a 1980	C9	1911 a 1920
C4	1961 a 1970	C10	1901 a 1910
C5	1951 a 1960	C11	1891 a 1900
C6	1941 a 1950		

Fonte: Elaboração Própria do Autor

O contingente populacional e o número de chefes de domicílio em cada célula são informados nos censos, logo é possível preencher cada célula da tabela de contingência com a respectiva estimativa para a taxa de chefia. Considerando as estatísticas disponíveis nos censo de 1970, 1980, 1991, 2000 e 2010 as tabelas de contingências contendo as taxas de chefia das populações masculina e feminina são apresentadas a seguir nas tabelas 3 e 4.

Tabela 3: Taxa de chefia da população masculina

Idade (Anos)	Período				
	1970	1980	1990	2000	2010
10 a 19	1,06%	1,50%	1,45%	1,39%	2,12%
20 a 29	44,15%	48,76%	41,83%	37,44%	29,56%
30 a 39	85,87%	87,54%	80,84%	73,99%	57,82%
40 a 49	91,75%	93,13%	90,44%	84,01%	68,07%
50 a 59	93,77%	94,61%	92,45%	87,87%	72,94%
60 a 69	92,47%	92,91%	91,76%	88,30%	76,45%
70 +	79,52%	85,83%	84,65%	81,92%	72,63%

Fonte: Elaboração Própria do Autor

Tabela 4: Taxa de chefia da população feminina

Idade (Anos)	Período				
	1970	1980	1990	2000	2010
10 a 19	0,33%	0,68%	0,31%	0,53%	2,53%
20 a 29	3,40%	5,63%	4,31%	7,30%	16,98%
30 a 39	7,29%	9,39%	10,74%	16,13%	29,27%
40 a 49	13,25%	15,13%	17,71%	24,60%	36,12%
50 a 59	20,58%	22,19%	25,28%	31,61%	42,12%
60 a 69	26,82%	30,38%	33,30%	39,97%	48,78%
70 +	27,58%	34,69%	38,08%	44,73%	51,62%

Fonte: Elaboração Própria do Autor

A análise das estatísticas das tabelas 3 e 4 pode revelar alguma estrutura latente que permita inferir sobre a dinâmica de evolução da taxa de chefia ao longo do tempo (períodos) e ao longo das idades e das coortes, ou seja, permite avaliar os efeitos idade, período e coorte.

A identificação de estruturas em tabelas de contingência é facilitada pelo ajuste de um modelo linear generalizado (Agresti, 2002), uma extensão dos tradicionais modelos de regressão linear em que o valor esperado da variável

dependente relaciona-se com as variáveis independentes por meio de uma função não linear denominada por Função de Ligação, sendo que a distribuição de probabilidade da variável resposta pertence à família exponencial (Normal, Poisson, Binomial e Gama). Neste caso, a variável dependente é a taxa de chefia e as variáveis explicativas são as faixas etárias, os períodos e as coortes.

Porém, a estimação de tal modelo de regressão não é uma tarefa fácil, pois os efeitos idade, período e coorte não são separáveis. Os efeitos são linearmente dependentes, pois conhecidos a idade de um indivíduo e o ano do Censo em que ele foi entrevistado, pode-se conhecer a sua coorte de nascimento, ou seja, os três efeitos estão relacionados pela seguinte combinação: idade = período – coorte.

Assim, a estimação dos efeitos se dá mediante a imposição de algumas restrições de identificação. O método mais utilizado para estimação dos efeitos idade, período e coorte foi proposto por Mason *et al.* em 1973 apud GLEEN (2005). Segundo este método as faixas etárias (idade), os períodos e as coortes são variáveis categóricas codificadas em um conjunto de variáveis *dummy*. As *dummy* entram como variáveis independentes em um modelo de regressão em que a variável explicada é a taxa de chefia ou qualquer função desta. Visando evitar a multicolinearidade perfeita (GUJARATI e POTER, 2011) deve-se omitir uma *dummy* de cada variável categórica, ou seja, uma da idade, outra do período e uma da coorte. Contudo, em função da combinação linear dos efeitos, uma *dummy* adicional de alguma das variáveis categóricas (idade, período ou coorte) ainda deve ser omitida a critério do analista. Em resumo, quatro variáveis *dummy* devem ser omitidas de forma a permitir a identificação do modelo. A omissão de uma variável adicional introduz o pressuposto de que os efeitos das duas categorias em uma mesma variável categórica são iguais.

Em linhas gerais, o modelo IPC decompõe a taxa de chefia em três efeitos distintos: idade, período e coorte de nascimento, conforme indicado a seguir.

$$\text{Taxa de Chefia} = \text{Efeito Idade} + \text{Efeito Período} + \text{Efeito Coorte} \quad (2.25)$$

Na família dos modelos lineares generalizados, a regressão logística e a regressão de Poisson são os mais indicados neste caso, pois a variável dependente expressa a participação dos chefes de domicílios na população total.

Seja $T_{ijk} = H_{ijk}/P_{ijk}$ a taxa de chefia da idade i , no período j e na coorte k , a equação de regressão que relaciona o logit da taxa de chefia com os efeitos tem a seguinte especificação no caso do modelo de regressão logística:

$$\text{Ln}\left(\frac{T_{ijk}}{1-T_{ijk}}\right) = \mu + \sum_{\text{idades}} \alpha_i \cdot D_i + \sum_{\text{períodos}} \beta_j \cdot D_j + \sum_{\text{coortes}} \lambda_k \cdot D_k \quad (2.26)$$

onde

$D_i = 1$ - se a variável dependente é referente a i -ésima faixa etária, caso contrário

$D_i = 0, i=1, \dots, 6;$

$D_j = 1$ - se a variável dependente é referente ao j -ésimo período, caso contrário

$D_j = 0, j=2, \dots, 5;$

$D_k = 1$ - se a variável dependente é referente a k -ésima coorte, caso contrário

$D_k = 0, k=1, \dots, 9;$

μ - é o efeito médio;

α_i - é o efeito da idade $i, i=1, \dots, 6;$

β_j - é o efeito do período $j, j=2, \dots, 5;$

λ_k - é o efeito da coorte $k, k=1, \dots, 9.$

No caso da regressão de Poisson tem-se a seguinte especificação de relação entre o total de domicílios, a população e os efeitos, onde as variáveis *dummies* (D) e os coeficientes α_i, β_j e λ_k são os mesmos definidos acima:

$$\text{Ln}(H_{ijk}) = \mu + \sum_{\text{idades}} \alpha_i \cdot D_i + \sum_{\text{períodos}} \beta_j \cdot D_j + \sum_{\text{coortes}} \lambda_k \cdot D_k + \text{Ln}(P_{ijk}) \quad (2.27)$$

Para evitar a multicolinearidade entre as variáveis idade, período e coorte e possibilitar a estimação dos modelos 2.26 e 2.27, Deaton (1997) sugere a retirada de uma variável *dummy* da idade, outra da coorte e duas variáveis *dummies* do período. No entanto, como o foco do trabalho consiste na projeção do número de domicílios, a alternativa mais interessante consiste na retirada de apenas uma variável *dummy* do período. Assim, as variáveis explicativas omitidas do modelo referem-se ao grupo etário com idades iguais ou superiores a 70 anos, no ano de 1970, bem como às duas coortes mais velhas (coortes 10 e 11), cujos efeitos são considerados idênticos.

Uma vez estimado o modelo, o logito da taxa de chefia na idade i (faixa etária), dos respectivos efeitos α_i , β_j e λ_k , podem ser determinados pela soma dos respectivos efeitos adicionada do intercepto μ , conforme demonstrado na tabela 5.

Tabela 5: Composição dos efeitos na formação do logito da taxa de chefia

Idade (Anos)	Período					
	1970	1980	1990	2000	2010	2020
10 a 19	$\mu + \alpha_1 + \lambda_5$	$\mu + \alpha_1 + \beta_2 + \lambda_4$	$\mu + \alpha_1 + \beta_3 + \lambda_3$	$\mu + \alpha_1 + \beta_4 + \lambda_2$	$\mu + \alpha_1 + \beta_5 + \lambda_1$	$\mu + \alpha_1 + \beta_6 + \lambda_0$
20 a 29	$\mu + \alpha_2 + \lambda_6$	$\mu + \alpha_2 + \beta_2 + \lambda_5$	$\mu + \alpha_2 + \beta_3 + \lambda_4$	$\mu + \alpha_2 + \beta_4 + \lambda_3$	$\mu + \alpha_2 + \beta_5 + \lambda_2$	$\mu + \alpha_2 + \beta_6 + \lambda_1$
30 a 39	$\mu + \alpha_3 + \lambda_7$	$\mu + \alpha_3 + \beta_2 + \lambda_6$	$\mu + \alpha_3 + \beta_3 + \lambda_5$	$\mu + \alpha_3 + \beta_4 + \lambda_4$	$\mu + \alpha_3 + \beta_5 + \lambda_3$	$\mu + \alpha_3 + \beta_6 + \lambda_2$
40 a 49	$\mu + \alpha_4 + \lambda_8$	$\mu + \alpha_4 + \beta_2 + \lambda_7$	$\mu + \alpha_4 + \beta_3 + \lambda_6$	$\mu + \alpha_4 + \beta_4 + \lambda_5$	$\mu + \alpha_4 + \beta_5 + \lambda_4$	$\mu + \alpha_4 + \beta_6 + \lambda_3$
50 a 59	$\mu + \alpha_5 + \lambda_9$	$\mu + \alpha_5 + \beta_2 + \lambda_8$	$\mu + \alpha_5 + \beta_3 + \lambda_7$	$\mu + \alpha_5 + \beta_4 + \lambda_6$	$\mu + \alpha_5 + \beta_5 + \lambda_5$	$\mu + \alpha_5 + \beta_6 + \lambda_4$
60 a 69	$\mu + \alpha_6$	$\mu + \alpha_6 + \beta_2 + \lambda_9$	$\mu + \alpha_6 + \beta_3 + \lambda_8$	$\mu + \alpha_6 + \beta_4 + \lambda_7$	$\mu + \alpha_6 + \beta_5 + \lambda_6$	$\mu + \alpha_6 + \beta_6 + \lambda_5$
70 +	μ	$\mu + \beta_2$	$\mu + \beta_3 + \lambda_9$	$\mu + \beta_4 + \lambda_8$	$\mu + \beta_5 + \lambda_7$	$\mu + \beta_6 + \lambda_6$

Fonte: Elaboração Própria do Autor

A partir da análise da tabela 5, verifica-se também que tanto o efeito de período em 2020 (β_6) quanto o efeito coorte de nascimento (λ_0) não são projetados através do processo de estimação da equação (2.26). Tais projeções podem ser obtidas pela extrapolação das tendências dos efeitos estimados (ALBUQUERQUE, 2009).

Com todos os efeitos estimados e projetados para 2020, a projeção da taxa de chefia para o grupo de idade i , no ano t ($t = 2020$) na coorte k , pode ser determinada a partir das funções a seguir:

$$\text{Modelo de regressão logística: } \hat{T}_{ijk} = \frac{\hat{H}_{ijk}}{P_{ijk}} = \frac{\exp(\mu + \alpha_i + \beta_j + \gamma_k)}{1 + \exp(\mu + \alpha_i + \beta_j + \gamma_k)} \quad (2.28a)$$

$$\text{Modelo de Poisson: } \hat{T}_{ijk} = \frac{\hat{H}_{ijk}}{P_{ijk}} = \exp(\mu + \alpha_i + \beta_j + \gamma_k) \quad (2.28b)$$

Por fim, a previsão total de domicílios ocupados para o ano de 2020 é obtida pela seguinte equação:

$$H(2020) = \sum_{idade \times sexo} \sum_s T_x^s(2020) \cdot {}_n P_x^s(2020) \quad (2.29)$$

onde ${}_n P_x^s(2020)$ é a projeção do total de indivíduos do sexo s e com idade entre x e $x+n$ anos completos no ano 2020 fornecida pelo IBGE e ${}_n T_x^s(t)$ é a respectiva taxa de chefia projetada com o auxílio do modelo IPC.

Conforme as Nações Unidas (UNITED NATIONS, 1973) o método baseado em Taxas de Chefia é o mais adequado para se projetar demanda por habitação, uma vez que reflete tanto mudanças na estrutura etária da população quanto na razão de sexos existente na mesma.

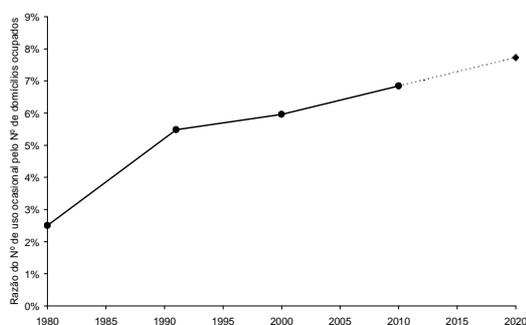
Assim, utilizando-se de uma projeção confiável da pirâmide etária, é possível obter um cenário acerca do total de domicílios ocupados compatível com o processo de transição demográfica da população brasileira, uma vez que o método da Taxa de Chefia reflete as mudanças populacionais esperadas da estrutura etária e razão do sexo e incorporam, de maneira indireta, as tendências passadas e futuras das componentes demográficas: fecundidade, mortalidade e migração.

2.2.2.

Projeção do número de domicílios de uso ocasional

Conforme observado por Leon e Pessanha (2005), os contingentes populacionais nos domicílios de uso ocasional não são enumerados, portanto, o quantitativo destes domicílios não se relaciona com a população recenseada, ao contrário do que ocorre nos domicílios ocupados. Neste caso, o método de projeção utilizado baseia-se na projeção tendencial da razão entre o total de domicílios de uso ocasional e o número de domicílios ocupados.

Conforme indicado na figura 7 a razão entre estas duas categorias de domicílios apresentam uma tendência de crescimento desde o censo de 1980.



Fonte: Elaboração Própria do Autor com base nos dados dos Censos 1980, 2000 e 2010.

Figura 3: Razão do número de domicílios de uso ocasional pelo número de domicílios ocupados

Com isso, admitindo-se a tendência linear deste indicador após o ano de 2010 projeta-se um valor de aproximadamente 7,73% para a razão entre o número de domicílios de uso ocasional e o número de domicílios ocupados. Desta forma, tomando-se como base o total dos domicílios ocupados projetados através do método da taxa de chefia, é possível obter uma projeção para o total de domicílios de uso ocasional, por meio da seguinte fórmula:

$$N^{\circ} \text{ de domicílio de uso ocasional} = 0,0773 \times N^{\circ} \text{ de domicílio ocupados} \quad (2.30)$$

Assim, a projeção do número total de clientes potencialmente atendidos na classe residencial para o ano de 2020 é a soma das projeções do número de domicílios ocupados com o total de domicílios de uso ocasional.

2.3. Método Índice de Divisia de Média Logarítmica I (LMDI I)

O consumo de energia elétrica na classe residencial em um ano t (E_t) pode ser expresso pela seguinte identidade:

$$E_t = \frac{E_t}{NC_t} \times \frac{NC_t}{ND_t} \times ND_t \quad (2.31)$$

onde

NC_t é o número de unidades consumidoras na classe residencial;

ND_t é o número de domicílios ocupados e de uso ocasional;

$\frac{E_t}{NC_t}$ é o consumo médio por unidade consumidora (CPC_t);

$\frac{NC_t}{ND_t}$ é a taxa de atendimento (TA_t).

Na identidade acima, o consumo é definido pelo produto de três fatores: o consumo médio, a taxa de atendimento e o número de domicílios.

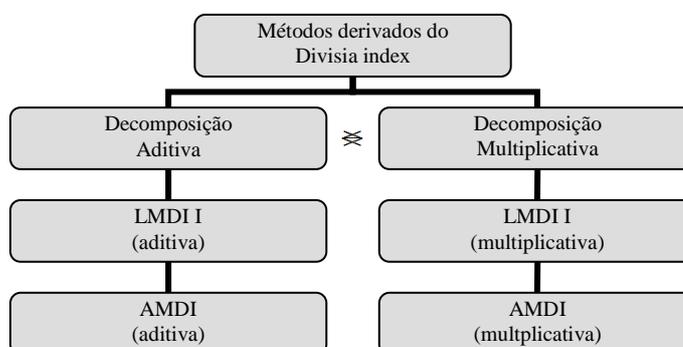
Objetivando um maior entendimento da dinâmica do consumo da classe residencial é interessante quantificar qual a contribuição de cada um destes fatores

para a variação anual do consumo, dada pela primeira diferença $\Delta E_t = E_t - E_{t-1}$. Assim, a formulação geral do método de decomposição por Índice de Divisia (Ang, 2004, Leon & Pessanha, 2005, Achão 2009) é dada por:

$$\Delta E_t = E_t - E_{t-1} = \Delta_{E|CPC} + \Delta_{E|TA} + \Delta_{E|ND} \quad (2.32)$$

Onde $\Delta_{E|CPC}$, $\Delta_{E|TA}$ e $\Delta_{E|ND}$ são, respectivamente, as contribuições das variações no consumo médio por consumidor, na taxa de atendimento e no número de domicílios para a variação total do consumo.

Ang (2004) classifica o conjunto de métodos relacionados ao índice de Divisia segundo os critérios de fundamentação teórica e aplicabilidade, conforme ilustrado na figura 8.



Nota: LMDI I - *Logarithmic Mean Divisia Index*

AMD I - *Arithmetic Mean Divisia Index*

Fonte: Elaboração própria a partir de Ang (2004).

Figura 4: Métodos derivados do índice de Divisia úteis na análise de decomposição do uso da energia

Ang, Liu e Chung. (2002) *apud* Ang (2004) definiram quatro testes ligados à teoria de números índices a fim de determinar o método mais adequado:

- reversibilidade dos fatores;
- reversibilidade no tempo;
- proporcionalidade; e
- testes de agregação.

Ang (2004) ressalta que dentre os quatro testes supracitados, o mais importante corresponde à reversibilidade dos fatores, pois segundo o autor, os métodos que passam neste teste podem ser considerados altamente recomendados.

Adiciona-se a isto, o fato de os métodos poderem apresentar uma decomposição aditiva ou multiplicativa, possibilitando uma escolha arbitrária. Segundo o autor, em alguns casos, a decomposição aditiva pode ser preferível à decomposição multiplicativa, ou vice-versa, conforme a facilidade e compreensão dos resultados apresentados.

Conforme observado por Achão (2009), o método do índice de Divisia de média aritmética (AMDI) utiliza-se de uma função com pesos obtidos a partir de médias aritméticas, ao passo que no método de Divisia de média logarítmica (LMDI I) os pesos são oriundos de médias logarítmicas. Como consequência, o método AMDI apresenta fórmula mais simples do que àquela correspondente ao método LMDI I.

Ang (2004) *apud* Achão (2009) salienta que o método AMDI pode ser empregado no lugar do LMDI I em muitas situações, pois frequentemente apresenta resultados similares para a decomposição daqueles obtidos a partir da aplicação do método LMDI I.

Contudo, a decomposição pelo AMDI, tanto na forma multiplicativa quanto na forma aditiva, apresenta duas importantes limitações (Achão, 2009): o método se mostra inadequado no teste de reversibilidade dos fatores e o método não pode ser aplicado quando há valores nulos no conjunto de dados analisados.

O método se mostra inadequado no teste de reversibilidade dos fatores, uma vez que não fornece uma decomposição perfeita, isto é, apresenta termos residuais expressivos, sobretudo, em duas situações, a saber: (i) quando a decomposição anual se dá de maneira encadeada sobre um período relativamente extenso, no qual o termo residual vai se acumulando ao longo do tempo, e; (ii) quando, apesar da decomposição não ocorrer de maneira encadeada, os dois anos analisados estão separados por um significativo período de tempo, fazendo com que as variações nos dados tornem-se relevantes.

Com relação a segunda limitação do método AMDI Achão (2009) destaca que, ao contrário do método AMDI, o referido problema pode ser solucionado no método LMDI I, a medida que o mesmo permite substituir os “zeros” por pequenos valores ou por limites analíticos, conforme estabelecido por Ang, Zhang e Choi (1998) e corroborado por Wood e Lenzen (2006).

No que tange ao método índice de Divisia de média logarítmica (LMDI I), Achão (2009) conclui explanando que o mesmo é indicado para uso geral, tanto

em sua forma aditiva (Ang e Choi, 1997) quanto em sua forma multiplicativa (Ang e Liu, 2001). Portanto, ambas as formas atendem o critério de reversibilidade dos fatores, proporcionando perfeita decomposição a partir da qual não há termo residual.

Assim, como no estudo realizado por Leon & Pessanha (2005) e Achão (2009), optou-se pela adoção do método índice de Divisia de média logarítmica a fim de analisar a evolução das variações no consumo de energia elétrica no setor residencial brasileiro. Os fatores preponderantes para esta decisão consistem no fato do método LMDI I possuir a propriedade de efetuar uma decomposição simétrica e perfeita da variação da variável endógena, isto é, uma decomposição independente da escolha de uma data de referência e sem termo residual, além disso, permite uma desagregação em períodos com durações distintas, possibilitando análises para eventos históricos específicos, por exemplo, o caso do racionamento no uso da energia elétrica vivenciado nos anos 2001 e 2002.

Pelo método LMDI I as contribuições das variações do consumo médio por consumidor, da taxa de atendimento e do número de domicílios para a variação total do consumo na classe residencial são calculadas pelas seguintes fórmulas:

$$\Delta_{E|CPC} = \frac{E_t - E_{t-1}}{\ln(E_t) - \ln(E_{t-1})} \cdot \ln\left(\frac{CPC_t}{CPC_{t-1}}\right) \quad (2.33)$$

$$\Delta_{E|TA} = \frac{E_t - E_{t-1}}{\ln(E_t) - \ln(E_{t-1})} \cdot \ln\left(\frac{TA_t}{TA_{t-1}}\right) \quad (2.34)$$

$$\Delta_{E|ND} = \frac{E_t - E_{t-1}}{\ln(E_t) - \ln(E_{t-1})} \cdot \ln\left(\frac{ND_t}{ND_{t-1}}\right) \quad (2.35)$$

Substituindo as expressões dos efeitos na variação do consumo tem-se que

$$\Delta E_t = E_t - E_{t-1} = \frac{E_t - E_{t-1}}{\ln(E_t) - \ln(E_{t-1})} \cdot \left\{ \ln\left(\frac{CPC_t}{CPC_{t-1}}\right) + \ln\left(\frac{TA_t}{TA_{t-1}}\right) + \ln\left(\frac{ND_t}{ND_{t-1}}\right) \right\} \quad (2.36)$$

ou seja, a variação do consumo é expressa pela soma das taxas de crescimento de cada um dos efeitos.