



UNIVERSIDADE FEDERAL DE ALAGOAS
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

ISAAC PEDRO DA SILVA

**APLICAÇÃO DE MODELOS ADITIVOS GENERALIZADOS
PARA LOCAÇÃO, ESCALA E FORMA (GAMLSS) NO
ESTUDO DA CONVERGÊNCIA DE RENDA**

Maceió – AL

2013

ISAAC PEDRO DA SILVA

**APLICAÇÃO DE MODELOS ADITIVOS GENERALIZADOS
PARA LOCAÇÃO, ESCALA E FORMA (GAMLSS) NO
ESTUDO DA CONVERGÊNCIA DE RENDA**

Dissertação apresentada ao Curso de Mestrado em Economia Aplicada da Faculdade de Economia, Administração e Contabilidade da Universidade Federal de Alagoas – UFAL, como requisito parcial para obtenção do Título de Mestre em Economia Aplicada.

Orientador: Prof. Dr. Dilson J. de Sena
Pereira

Maceió – AL

2013

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário Responsável Valter dos Santos Andrade

S586a Silva, Isaac Pedro da.

Aplicação de modelos aditivos generalizados para locação, escala e forma (GAMLSS) no estudo da convergência de renda Isaac Pedro da Silva. – 2013. 88 f. ; il., graf., tab.

Orientador: Dilson J. de Sena Pereira.

Dissertação (Mestrado em Economia Aplicada) – Universidade Federal de Alagoas. Faculdade de Economia, Administração e Contabilidade. Programa de Pós-Graduação em Economia. Maceió, 2014.

Bibliografia: f. 76-82.

Apêndice: f. 83-88.

1. Teoria do crescimento econômico. 2. Nova modelagem e estimação.
3. Hipóteses de Gauss-Markov. 4. MCRL. 5. GAM. 6. GAMLSS. I. Título.

CDU: 330.35.01

Ata da 24ª Sessão de Defesa de Dissertação de Mestrado em Economia do Programa de Pós-Graduação em Economia da Universidade Federal de Alagoas.

Aos 10 dias do mês de maio de 2013 foi instalada a 24ª sessão de Defesa de Dissertação de Mestrado em Economia Aplicada do Programa de Pós-Graduação em Economia da Universidade Federal de Alagoas, às 17h00min, no Auditório da Faculdade de Economia, Administração e Contabilidade, na FEAC, da Universidade Federal de Alagoas, a que se submeteu o mestrando **ISAAC PEDRO DA SILVA**, da linha de pesquisa Economia do Desenvolvimento apresentando o trabalho: "Aplicação de Modelos Aditivos Generalizados para Locação, Escala e Forma (GAMLSS) no estudo da convergência de renda", como requisito parcial para a obtenção do grau de Mestre em Economia, conforme o disposto no regulamento deste Programa, e tendo como Banca Examinadora já referendada pelo Colegiado do Curso, Prof. Dr. Dilson José de Sena Pereira (FEAC-UFAL), Prof. Dr. Anderson de Barros Dantas (FEAC-UFAL), Prof. Dr. Alexandre Manoel Ângelo da Silva (IPEA), sob a presidência do orientador Prof. Dr. Dilson José de Sena Pereira (FEAC-UFAL).

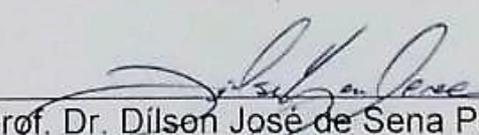
Analisando o trabalho a Banca atribui a seguinte menção:

APROVADO

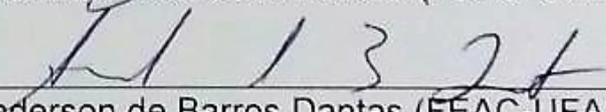
REPROVADO

OBSERVAÇÕES: Subjeto às recomendações da Banca Examinadora, a saber: reorganizar a motivação, melhorar a explicação das variáveis das variáveis, conformar os resultados e de bibliografia.

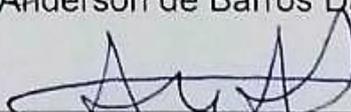
Maceió, 10 de maio de 2013.



Prof. Dr. Dilson José de Sena Pereira (FEAC-UFAL)



Prof. Dr. Anderson de Barros Dantas (FEAC-UFAL)



Prof. Dr. Alexandre Manoel Ângelo da Silva (IPEA)

Às melhores lembranças da minha querida irmã que se foi precocemente.

À vontade de potência que permanece acesa dentro de nós.

AGRADECIMENTOS

Vou buscar em outro lugar o que não me permitiram aqui. Eu mesmo vou construir meu Mestrado, onde quer que seja. Foi assim que comentei com uma amiga e comecei a pensar uma estratégia de como deixar meu Estado para disputar uma vaga no programa de pós-graduação em Economia Aplicada na UFAL, sem hesitar. É óbvio que tive dificuldades, principalmente por não conhecer ninguém em Maceió – AL. Mas, construí amizades e assim fui superando as dificuldades de entrosamento, compreensão de como as pessoas pensavam e agiam e, aos poucos, fui criando certo tipo de adaptação e tolerância respeitosa com os outros. Quem tiver a oportunidade de ler esses agradecimentos e lembrar de algum fato, vai entender ao que estou me referindo.

Como disse, as circunstâncias não me permitiram continuar um Mestrado em outro campo do conhecimento, mas parece que o destino queria que eu permanecesse estudando Economia, em outro lugar. E fui participar da seleção, em três etapas, o que permitiu testar três capacidades de um candidato ao programa. Foi então que vi a diferença em relação à minha instituição de origem, que precisaria enfrentar a prova da ANPEC, o que não acho *justo*, porque é uma seleção privada para estudar em uma universidade pública. Por isso, mesmo tendo capacidade para fazê-la, optei por um centro que buscasse do candidato o melhor que ele tem: sua capacidade de conhecimento das matérias da ciência econômica, de uma língua estrangeira e intuição para pesquisa. Portanto, meu primeiro agradecimento vai em conjunto para a Universidade Federal de Alagoas e ao Programa de Pós-Graduação em Economia da Faculdade de Economia, Administração e Contabilidade (FEAC), por ter me acolhido como aluno de pós-graduação e ter percebido em mim a possibilidade de me transformar em um profissional apto à pesquisa e ao ensino, a partir do momento que me fosse dado o *expertise*.

Acredito também que fui à UFAL somente para conhecer professores fantásticos e comprometidos com o progresso científico da instituição e cientes do seu primordial papel na construção de um desenvolvimento regional. Tenho certeza que o destino quis que eu conhecesse Professores como meu Orientador, Professor Dilson Sena, pessoa que serve como exemplo de profissional e pesquisador (obrigado pelo presente especial: um estudo dirigido em teoria do crescimento econômico. Aquele foi o pontapé para a pesquisa que se segue.). Como o Professor Reynaldo Rubem, que me presenteou com as mais calorosas discussões nas aulas de Macroeconomia (acho que hoje me inclino a pensar macroeconomia como um pós-keynesiano!). Falando de discussão acalorada, não havia coisa melhor do que discutir os problemas econômicos do Brasil nas aulas de Economia Brasileira com o

estimado Professor Fabio Guedes, que nos dava um presente em paralelo: minerar nos textos as relíquias das ideias dos maiores pensadores do problema brasileiro e discutí-los em sala. Presente, porque nos ensinou a pesquisar o que interessa, com uma simples forma de metodologia da pesquisa. Então, eram duas aulas em uma. Imagina, então, ter dois professores de Microeconomia! É preciso, de fato, várias visões de mundo para compreender o que estamos lidando. Quero sempre lembrar dos Professores Francisco Rosário, o Coordenador do curso de Mestrado, e do Professor Thierry Prates que de forma atual fez coadunar e coexistir as formas (neo)clássica e inovativa de pensar teoria econômica na ala *micro*, consolidando a compreensão, digamos, *prática e real* na economia que nos cerca.

É fato! Precisamos de ferramentas – analíticas, lógicas, matemáticas, estatísticas, econométricas, computacionais e de política – para viabilizarmos o estudo e a pesquisa em Economia, e o programa de Mestrado forneceu, em certas proporções ou, pelo menos, me instigou a buscar, mas nada como ter professores para facilitar a incursão nessas ferramentas. Esses comentários servem para lembrar das aulas de métodos quantitativos com os professores Alexandre Lima e Anderson Dantas que alertaram a nós discentes que não era porque o programa tinha uma inclinação mais heterodoxa que não era necessário conhecer como manuseá-las.

Vejam como são as coisas: tive a sorte de ter um professor de Econometria, o professor Alexandre Manoel, que me fez conhecer mais a fundo uma das ferramentas que mais me fascina e que busco usá-la e aperfeiçoá-la nas minhas pesquisas. Não posso esquecer de mencionar a professora Cecília Lustosa que, juntamente e, mais expressivamente, com o Professor Thierry Prates, definitivamente, exauriu todas as minhas dúvidas em Economia do Meio Ambiente e, além de tudo, mostraram como nós pesquisadores em economia deveríamos dar um tratamento especial a estes estudos na busca de mais ferramentas úteis para a sociedade.

Quero ser um Pesquisador e Professor como vocês. A Todos, meus sinceros e profundos agradecimentos.

Quero, também, agradecer a DEUS por ter me dado a capacidade de aprendizagem e a oportunidade de conhecer as ferramentas que o programa de código aberto **R** fornece, agradecer à existência do L^AT_EX e aos idealizadores do GAMLSS. São verdadeiros *Tesouros do Céu*.

Como a Dissertação é minha, eu menciono quem mais quiser!

Quero lembrar dos amigos no meio do caminho: ao pessoal da secretaria da Pós,

Levylna, Cléo, Paula, Alex, Becker e Mônica, esses dois últimos pessoal da secretaria da graduação em economia, Evelyni da pós em Gestão Pública e Maria da limpeza (conversava muito com ela.). Tenho mais! Danúbia, Manu, Leilane, J. Jeferson e Débora Grey, Fabrício, Bruno Pimentel, enfim, aquele abraço. Estarei sempre por aí.

Fábio Correia e Camila Ribeiro, meus amigos, sintam-se na liberdade de me visitar em Recife-PE quando quiserem e sentirem vontade. Amigo é assim mesmo. Encaixa-se no clichê do bem: *quando mais se precisa, eles estão lá.*

Quem tem amigo na praça.... Bom, na verdade quem tem um bom **networking**, não sofre (muito). Agradeço à contribuição de Bruna da Nóbrega por conhecer um amigo, que conhece outro amigo..., que teve acesso a um material indispensável para a minha pesquisa: agradeço ao colega Flavius Sodré por ter pego emprestado, na biblioteca de MEI da UFPE, o livro Generalized Additive Models, do Hastie & Tibshirani. Sem ele a parte teórica não ficaria completa e se tornariam precárias algumas explicações técnicas que somente se encontra neste material. Obrigado.

Virgínia Campos: *Cumplicidade é tudo, não dá pra fugir.*

Mãe, ainda tenho fôlego para mais. Muito mais.

Olha meu desejo traduzido na próxima página.

Só mais um beijo
Só mais um copo,
Mais um desejo que se derrama
Então, eu toco o que não vejo e acendo mais um pavio

Só mais um dia
Só mais um bloco
As odaliscas e os serafins são meus segredos

Eu tenho um mapa, só mais um passo
E eu descubro o meu tesouro
Só mais um voo
Mais um mergulho no fundo desse bar

O inferno é o céu mais sujo
Eu me lambuzo no paraíso
Só mais um gesto
Só mais um não
E o que resta no fim da festa, sem fim
Se lava com água e sabão

E, partamos, então
Pra outra estação
Outra freguesia, terra distante:
Metrópole, lua, província, ou ilha que mesmo a mais fria
Nossa poesia ainda verão

La dolce vita – a saideira
(Antônio Saraiva , Paulo Baiano)

“The very word ‘model’ implies idealization. With very few possible exceptions it would be absurd to think that a mathematical model is an exact representation of a real system.”

Cox, D. R.(2006, p.31)

RESUMO

O desenvolvimento de novos modelos e novos métodos de estimação tem sido objetivo constante da teoria e prática econométricas, a fim de atender às particularidades dos modelos em economia. Inclusive, os testes sobre os modelos da teoria do crescimento econômico trouxeram grandes avanços, instigando mudanças no paradigma da modelagem neste campo. A partir de então, suscitaram três questionamentos que podem modificar a forma de pensar a pesquisa empírica: os dados como determinantes, o modo como os modelos são concebidos e duas das hipóteses de Gauss-Markov. Isto impôs a necessidade do relaxamento em duas delas: a linearidade nos parâmetros e a normalidade do termo de erro. Muitos fenômenos estudados em crescimento econômico não têm mostrado justificativa teórica para modelagens econométricas não lineares. Com a insatisfação crescente de se trabalhar de forma limitada, pensou-se em apresentar uma generalização dos modelos lineares, em que se objetiva substituir a função linear usual das covariáveis por uma função suave não especificada, deixando que os dados mostrem a forma funcional apropriada para estudar o fenômeno, o que nos leva aos *suavizadores* em um gráfico de dispersão, que são as *splines*, e tenta mostrar uma dependência funcional sem necessitar impor uma hipótese paramétrica rígida sobre esta dependência. Com isso, abre-se a possibilidade de se libertar da linearidade nos parâmetros, ao mesmo tempo em que permite que o termo de erro siga qualquer distribuição, além da família exponencial, abandonando a hipótese de normalidade. Desta forma, propôs-se uma nova experiência de modelagem e estimação, que pode ser considerada uma melhoria técnica, a partir de uma generalização dos modelos lineares (na classe de modelos aditivos), os quais superam as rigidezes das hipóteses de Gauss-Markov e as incompletudes dos modelos aditivos generalizados (GAMs), e este modelo mais completo é conhecido por Modelo Aditivo Generalizado para Localização, Escala e Forma (GAMLSS). Para testá-lo, utilizou-se, para o período (2000 – 2009), os dados de 102 municípios do Estado de Alagoas. Assim, as estimativas informaram uma relação inversa da taxa de crescimento com seu PIB *per capita* do início do período, o que obedeceu à hipótese de convergência condicional, atingindo seu estado estacionário à taxa decrescente de 0.52. Inclusive, os sinais das estimativas foram condizentes com o que obriga a teoria econômica, destacando a contribuição ao crescimento econômico da região, revelando mais uma vez investimentos precários, porém com grande potencial alavancador. Em contrapartida, também foi revelada uma má distribuição de renda em todo o território, além de sua concentração em pontos específicos, desfavorecendo o crescimento contínuo em todas as regiões, da mesma forma que houve diminuição da força de trabalho, acompanhada pela diminuição da população adulta e jovem em idade laboral. Os resultados obtidos demonstraram que o GAMLSS forneceu ajustes superiores em

qualidade, robustez e confiabilidade em relação àqueles apresentados via Modelo Clássico de Regressão Linear (MCRL) e GAM, tendo as análises gráficas e numéricas dos resíduos indicando que a classe de modelos do tipo GAMLSS aparenta ser mais apropriada para a estimação dos parâmetros em modelos da teoria do crescimento econômico.

Palavras-chave: Teoria do crescimento econômico. Nova modelagem e estimação. Hipóteses de Gauss-Markov. MCRL. GAM. GAMLSS..

ABSTRACT

The development of new models and new estimation methods have been constant goal of econometric theory and practice in order to meet the particularities of models in economics. Including tests on models of economic growth theory brought great advances, prompting changes in the paradigm of modeling in this field. From then raised three questions that can change the way you think empirical research: Data as determinants, how models are designed and two of the Gauss-Markov assumptions. This imposes the need of relaxation in two of them: the linearity in the parameters and normality of the error term. Many phenomena studied in economic growth have shown no theoretical justification for nonlinear econometric modeling. With the growing dissatisfaction of working in a limited way, it was thought to present a generalization of linear models, where the objective is to replace the usual linear function of the covariates by a smooth function unspecified, leaving the data show that the functional form appropriate for study the phenomenon, which leads to the smoother in a scatter plot, which are the Splines, and tries to show a functional dependency without needing to impose a rigid parametric hypothesis about this dependence. With that, it opens the possibility to release the linearity in the parameters, while allowing the error term follow any distribution, beyond the exponential family, abandoning the assumption of normality. Thus, we proposed a new breakthrough modeling and estimation, which can be regarded as a technical improvement, from a generalization of linear models (in the class of additive models), which outweigh the rigidities of the Gauss-Markov assumptions and the incompleteness of generalized additive models (GAMs), and this more complete model is known as Generalized Additive Model for Lease, Scale and Shape (GAMLSS). To test it, it was used for the period (2000 – 2009), data from 102 municipalities in the state of Alagoas. Thus, the estimates reported an inverse relationship with the growth rate of its GDP *per capita* of the early period, which followed the hypothesis of conditional convergence, reaching a steady state at the rate of decreasing 0.52. Even the signs of the estimates were consistent with what economic theory requires, highlighting the contribution to economic growth in the region, revealing once more precarious investments, but with great potential leveraging. In contrast, it was also revealed a poor distribution of income across the territory beyond its concentration on specific points, discouraging the continued growth in all regions, as there was a reduction of the workforce, accompanied by the decrease of the adult population young working age. The results showed that the adjustments provided GAMLSS superior in quality, robustness and reliability compared to those presented via Classical Linear Regression Model (CLRM) and GAM, with graphical and numerical analyzes of residues indicating that the class of models of type GAMLSS appears to be more appropriate for the estimation of

parameters in models of economic growth theory.

Keywords: Economic growth theory. Modelling and estimation improvement. The Gauss-Markov assumptions. CRLM. GAM. GAMLSS.

LISTA DE FIGURAS

FIGURA 1 – Ajustamento em dois estágios e com uma <i>Spline</i>	43
FIGURA 2 – Spline Cúbica	45
FIGURA 3 – Algoritmo RS na sua forma mais simples	56
FIGURA 4 – Sintaxe da função <code>gamlss()</code> no R	58
FIGURA 5 – Perfil do desvio global para dois <i>steps</i>	63
FIGURA 6 – Perfil do desvio global para um <i>step</i>	64
FIGURA 7 – Resíduos a partir do modelo <code>gamlssDiss10</code>	65
FIGURA 8 – Resíduos a partir do modelo <code>gamlssDiss10</code> , utilizando a opção <code>ts=TRUE</code>	65
FIGURA 9 – <i>Worm plot</i> do modelo <code>gamlssDiss10</code> com valores padrão	66
FIGURA 1 – <i>Worm plot</i> da variável <code>CRIME</code>	86
FIGURA 2 – <i>Worm plot</i> da variável <code>INFRAENERGIA</code>	86
FIGURA 3 – <i>Worm plot</i> da variável <code>DENSIDEM</code>	87
FIGURA 4 – <i>Worm plot</i> da variável <code>POTMERC</code>	87
FIGURA 5 – <i>Worm plot</i> da variável <code>PIBpc2009Def</code>	88
FIGURA 6 – <i>Worm plot</i> da variável <code>PIBpc2000Def</code>	88

LISTA DE TABELAS

TABELA 1 – Distribuições do GAMLSS implementadas no R	57
TABELA 2 – Resultados das regressões para convergência condicional	69
TABELA 3 – Resultados da regressão para convergência absoluta	70

SUMÁRIO

1	INTRODUÇÃO	25
2	O LUGAR DO CRESCIMENTO NA HISTÓRIA DO PENSAMENTO ECONÔMICO	30
2.1	Metodologia clássica no cálculo da convergência de renda	32
2.2	Modos de convergência	34
3	IMERSÃO NO MÉTODO	36
3.1	O modelo linear geral	38
3.2	Modelos lineares generalizados	40
3.3	Regressão por <i>splines</i> e Modelos Aditivos Generalizados	42
4	MODELOS ADITIVOS GENERALIZADOS PARA LOCAÇÃO, ES- CALA E FORMA	50
4.1	A incompletude do GAM e a opção pelo GAMLSS	50
4.2	Modelagem e Estimação	52
4.3	Computação	55
4.3.1	<i>Algoritmo</i>	56
4.3.2	<i>Função no R para ajustamento do GAMLSS</i>	57
5	DINÂMICA DE UMA ECONOMIA	59
5.1	Um GAMLSS de crescimento econômico	59
5.1.1	<i>Convergência de renda via GAMLSS</i>	60
5.2	Resultados empíricos	67
6	CONSIDERAÇÕES FINAIS	72
	REFERÊNCIAS	76
	APÊNDICE A – ESTATÍSTICAS DAS VARIÁVEIS UTILIZADAS	83

APÊNDICE B - *GLOBAL DEVIANCE* PARA UM *STEP* 84

APÊNDICE C - *WORM PLOT* DO MODELO *GAMLSSDISS10* COM A
OPÇÃO *XVAR* PARA TODAS AS VARIÁVEIS 86

1 INTRODUÇÃO

“*Deus é sutil, mas não malicioso.*”

Albert Einstein

Em economia, a investigação sobre causalidade entre variáveis tem como ponto inicial a construção de modelos. Estes são fundamentados dentro da teoria econômica, a partir de pressupostos ou hipóteses que os sustentam.

Geralmente, a ideia inicial é encontrar uma relação de causalidade que mostre um caminho para resolução de problemas da *forma mais direta* possível. E cabe destacar que a modelagem econométrica tem um espaço privilegiado na investigação econômica. Por isso, o argumento da linearidade nos parâmetros, nessa modelagem, ainda é muito forte, sendo ela a mais comum como forma de explicar os fenômenos.

Os testes de modelos em economia deram vazão ao uso da econometria como metodologia para que se encontrasse as desejadas estimativas, a partir de um dos métodos de estimação próprio para modelos lineares, qual seja, o método dos *Mínimos Quadrados Ordinários* (MQO).

Ao mesmo tempo, foi preciso *amarrar* algumas arestas que foram surgindo naturalmente, à medida que estes modelos foram sendo testados. E, para permanecer no mundo onde a causalidade é necessária, era preciso ter sustentação em algumas hipóteses, as conhecidas *hipóteses de Gauss-Markov*.

A partir de então, os modelos lineares em econometria passaram a ser utilizados largamente, não somente em economia, como uma das principais formas de investigação de fenômenos socioeconômicos. Esta ampla aplicação fez com que esse tipo de modelagem se tornassem *geral* e seu uso maciço ao longo do tempo os fez conhecidos como *clássicos*. Por isso, muitos ainda os chamam de *modelos lineares gerais*.

À medida que o uso intensivo desses modelos lineares e seus métodos de estimação foram aplicados em importantes tomadas de decisão, falhas ocorreram na sua utilização, interpretação e nas suas implicações práticas.

Esses fatos aconteciam simplesmente devido ao uso indiscriminado e desatento por parte de alguns pesquisadores, por algumas razões: a primeira, foi a ingenuidade da crença em um modelo geral, que de fato pudesse estimar todas as respostas para todas as variáveis

desejadas pelo pesquisador. A outra, foi imaginar que eram os modelos que determinavam as condições reais. Certamente, esses pesquisadores negligenciaram as palavras de Cox (2006), citadas na epígrafe deste trabalho, no qual explica que *com muito poucas exceções possíveis, seria absurdo pensar que um modelo matemático é uma representação exata de um sistema real.*

O desenvolvimento de novos modelos e novos métodos de estimação tem sido objetivo constante da teoria e prática econométricas, a fim de atender às particularidades dos modelos em economia. Dependendo do problema enfrentado pelo pesquisador, tanto a modelagem, método de estimação ou mesmo os dados podem influenciar ou mesmo determinar os resultados. Isto é o que ocorre quando as hipóteses de *Gauss-Markov* são violadas. Por exemplo, sabe-se que os modelos clássicos lineares em econometria falham na presença de autocorrelação espacial, pois algumas dessas hipóteses e, de modo geral, todas as outras do modelo clássico de regressão linear (MCRL), não irão garantir que o estimador via MQO será o Melhor Estimador Linear Não Viesado (BLUE).

Por esta vereda caminham as pesquisas nos chamados *modelos espaciais*, os quais vêm formando um corpo consistente de métodos e técnicas específicas para estudos da econometria que tem como escopo especificar, estimar, testar e prever modelos teóricos influenciados pelos efeitos espaciais, usando dados de corte transversal ou em painel. A este ramo da econometria convencionou-se chamar de Econometria Espacial.

Mas, também existem outros modelos e métodos que surgiram como resposta às falhas ou relaxamentos de algumas das hipóteses clássicas. Os livros-textos elencam as hipóteses em ordem crescente de complexidade para salvaguardar o modelo teórico econométrico. A chamada *quarta hipótese*, a que impõe sobre o termo de erro aleatório que a variância do erro condicional às variáveis explicativas seja constante, a hipótese da *homoscedasticidade*, quando relaxada ou violada, pode comprometer a forma de estimação e, com isso, os resultados das estimativas. Como forma de resposta a estas falhas, surgiram os métodos de estimação por *mínimos quadrados ponderados* (WLS), os quais se utilizam de pesos pré-definidos para proceder à estimação, na tentativa de superar as dificuldades impostas.

Além disso, podem surgir situações em que haja necessidade de nova especificação de modelos em teorias econômicas que também se utilizam da econometria para validar hipóteses: é o caso dos modelos de crescimento econômico. Esta teoria já apresenta um *core* bastante homogêneo com grandes contribuições internacionais que tem como precursores expoentes na ciência econômica moderna.

Em âmbito regional, os autores que estudam convergência têm buscado testar, por meio de vários modelos econométricos, de que forma as diversas unidades geográficas podem alcançar seu estado de equilíbrio dinâmico, em determinada economia. Cada econo-

nia estudada detém quantidades e variedades de dados disponíveis que são determinantes da sua dinâmica. Certamente, por isso, um único modelo não vai poder ser utilizado para todas as regiões ou localidades tendo, caso contrário, o risco de seus resultados serem incompatíveis com seu objetivo.

Nos diversos estudos para encontrar a convergência, foram propostos modelos econométricos que, naturalmente, atingiam alguns pressupostos do MCRL de forma singular e que, a cada passo em direção à descoberta, vinha à tona não mais o questionamento das propostas nos modelos e métodos mas, sim, sobre às principais hipóteses do modelo clássico de regressão linear. Os testes sobre os modelos da teoria do crescimento econômico trouxeram grandes avanços e instigaram mudanças no paradigma da modelagem neste campo.

A partir deste alerta, suscitaram três questionamentos importantes que podem modificar a forma de pensar a pesquisa empírica nesta linha de pensamento econômico. São três pontos básicos que devem ser questionados: os dados como determinantes; duas das hipóteses de Gauss-Markov, e o modo como os modelos são concebidos.

É importante deixar claro que variáveis que explicam, por exemplo, fenômenos do comércio internacional em regiões próximas ao litoral, quando estudadas pelo prisma de um modelo de crescimento, irão diferir das que se encontrem no interior, e os resultados podem não trazer a mesma resposta de forma significativa. Além disso, existe o risco da exclusão indevida, ausência ou mesmo negligência da inclusão de variáveis que não são importantes, o que piora ainda mais a qualidade do estudo. Ou seja, é importante alertar que as variáveis que explicam certas economias precisam ser *específicas*, pois isto interfere sobremaneira a forma como o modelo é construído.

Estamos entrando numa área de crítica quando questionamos as duas hipóteses de Gauss-Markov. Apesar de tudo, é preciso reconhecer a importância do MCRL para pesquisas em áreas como a teoria do crescimento econômico. Em contrapartida, pesquisas neste campo propuseram as maiores e mais significativas mudanças na modelagem econométrica. Os estudos mostravam que a sustentação de todas as hipóteses forçava o pesquisador a emitir justificativas demasiadas e que conflitavam com a teoria econômica.

A partir desta problemática é que foi se verificando a necessidade do relaxamento em, principalmente, duas delas: a linearidade nos parâmetros e a normalidade do termo de erro. A primeira ainda permanece usual em várias vertentes de modelagens, incluindo aspectos teóricos (mais especificamente, com grande apelo de instrumentos matemáticos e estatísticos) avançados, como Cadeias de Markov, Núcleos Estocásticos, enfim, porém mantendo-se a linearidade nos β 's. Entretanto, nada garante que a maioria das variáveis no modelo ampliado de Solow-Swan, e suas derivações, se relacionem linearmente e ainda tenha respaldo na teoria. O que se sustenta é que a clássica taxa de crescimento *per capita*

entre $t = 0$ e $t = n$ deve, sim, mostrar uma relação inversa com a taxa de crescimento *per capita* no período $t = 0$. Iremos aprofundar a discussão na Seção 5.1, do capítulo 5.

A crítica anterior sobre as duas hipóteses recaem subsequentemente na forma como os modelos lineares são concebidos. A maioria dos fenômenos estudados (se não a sua totalidade) em crescimento econômico não tem mostrado evidências de relações teóricas não-lineares que justifiquem modelagens econométricas que exijam não-linearidade nos parâmetros. Porém, isto não significa que apenas podemos trabalhar nesta classe de modelos, nem mesmo esperar que as variáveis que explicam o fenômeno pesquisado se relacione desta forma.

Esse panorama de situações-problema na econometria do crescimento econômico é a base para esta pesquisa de dissertação. As duas grandes ousadias estão, verdadeiramente, em não mais relaxar, mas *desprezar* as duas hipóteses mencionadas acima e propor tanto uma modelagem alternativa quanto seu método de estimação, o qual tem grande apelo computacional, mostrando o caráter prático do que estamos propondo.

No lugar de um modelo linear, ou alguma variante deste, para testar a convergência de renda, iremos apresentar uma generalização dos modelos lineares, qual seja, os *Modelos Aditivos* em que a ideia básica é substituir a função linear usual das covariáveis por uma função suave não especificada. A soma dessas últimas funções constitui um Modelo Aditivo, o qual é não-paramétrico, no sentido em que não é necessário impor uma forma paramétrica para as funções, porém, ao invés disso, as estima de forma iterativa através do uso de *suavizadores* em um gráfico de dispersão.

No prefácio de seu livro, Hastie & Tibshirani (1990) lembram, numa citação de Albert Einstein, que *Deus é sutil, mas não malicioso* e por isso, o modelo aditivo mantém alguma interpretabilidade advinda dos modelos lineares, quando se assume efeitos aditivos. Isto implica que os resultados das regressões a partir de modelos aditivos na econometria do crescimento econômico não afetará a maneira de como as estimativas serão interpretadas, mas haverá um ganho adicional de informação, pois cada variável aditiva mostrará seu relacionamento com a variável preditiva.

Ou seja, o que esta modelagem propõe é *deixar que os dados mostrem a forma funcional apropriada* para estudar o fenômeno. Esta é a ideia por trás dos *suavizadores* em um *gráfico de dispersão* e tenta mostrar uma *dependência funcional* sem necessitar impor uma hipótese paramétrica rígida sobre esta dependência. Assim, esta modelagem não se utiliza da linearidade nos parâmetros, ao mesmo tempo em que permite que o termo de erro siga qualquer distribuição, além da família exponencial, e não somente uma Normal, abandonando a hipótese de normalidade.

Portanto, o papel principal desta pesquisa é propor uma *nova experiência de modelagem e estimação*, a partir de uma generalização dos modelos lineares, transitando na

classe de modelos aditivos, os quais superam, além das rigidezes das hipóteses de Gauss-Markov, as incompletudes dos modelos aditivos generalizados (GAM), estabelecendo-se numa classe onde os dados mostrarão a *posição*, a *variabilidade* e o *formato* das funções suaves, e o relacionamento entre elas e a variável preditora.

Este novo modelo, com seu método de estimação, pode ser considerado uma melhoria técnica, à medida que demonstra resultados mais robustos em relação às técnicas clássicas usadas para o mesmo fim e, sobretudo, pelo seu poder preditivo¹. Este modelo mais completo é conhecido por Modelo Aditivo Generalizado para Localização, Escala e Forma (GAMLSS).

Além disso, em cinco capítulos e suas seções, mostraremos como a teoria do crescimento econômico vem se comportando ao longo da história econômica com suas contribuições importantes, tanto do ponto de vista teórico quanto da modelagem econométrica e, acima de tudo, como os resultados dos experimentos podem servir como embasamento para a elaboração de políticas públicas regionais.

¹Não nos ocuparemos, neste trabalho, do teste de previsibilidade entre os modelos apresentados, nem em comparar os Erros Quadráticos Médios (EQM).

2 O LUGAR DO CRESCIMENTO NA HISTÓRIA DO PENSAMENTO ECONÔMICO

“Toda teoria depende de hipóteses que não são totalmente verdadeiras. É isso que a faz teoria. A arte de bem teorizar é fazer as inevitáveis hipóteses simplificadoras de tal maneira que os resultados finais não sejam muito sensíveis.”

Robert Solow (1956), p. 65.

Já no século XVIII, conhecidos economistas clássicos com Adam Smith (1776) e Thomas Malthus (1798) e, em seguida, David Ricardo (1817), haviam fornecido os ingredientes das muitas ideias básicas que aparecem nas modernas teorias do crescimento econômico. No início do século XX, estudiosos como Frank Ramsey (1928), Allyn Young (1928), Joseph Schumpeter (1934) e Frank Knight (1944) também contribuíram sobremaneira para o avanço deste campo.

Essas ideias incluem a abordagem básica do comportamento competitivo e equilíbrio dinâmico, o papel dos retornos decrescentes e sua relação com a acumulação de capital físico e humano, a interação entre a renda *per capita* e a taxa de crescimento da população, os efeitos do progresso tecnológico na forma de aumento da especialização produtiva e as descobertas de novos produtos e métodos de produção, além do papel do poder de monopólio como um incentivo ao avanço tecnológico (BARRO & SALA-I-MARTIN, 2004).

Cronologicamente, porém, o ponto de partida da moderna teoria do crescimento é o artigo clássico de Ramsey (1928), trabalho este que se situava a várias décadas à frente do seu tempo. O tratamento de Ramsey para otimização das famílias ao longo do tempo vai muito além da sua aplicação na teoria do crescimento.

Entre o trabalho de Ramsey e a década de 1950, Harrod (1939) e Domar (1946) tentaram integrar a análise Keynesiana com elementos do crescimento econômico, utilizando funções de produção com pouca substitutibilidade entre insumos para argumentar que o sistema capitalista é inerentemente instável.

As duas mais importantes contribuições após este período são atribuídas a Solow (1956) e Swan (1956). O principal aspecto do conhecido modelo de Solow-Swan é a forma neoclássica da função de produção, especificação que assume retornos constantes de escala, retornos decrescentes para cada insumo e elasticidade positiva de substituição entre os

insumos, que é combinada a uma regra de taxa de poupança constante para fornecer um modelo extremamente simples de equilíbrio geral de uma economia, além da ausência de aumentos na tecnologia, em que é previsível que o crescimento possa eventualmente cessar.

Barro & Sala-I-Martin (2004) explicam que Cass (1965) e Koopmans (1965) trouxeram a análise de Ramsey da otimização do consumidor de volta aos modelos neoclássicos e, com isso, forneceram a determinação endógena da taxa de poupança e esta extensão permitiu uma melhor dinâmica de transição, porém tendeu a preservar a hipótese de convergência condicional. Essa endogeneidade da taxa de poupança também não elimina a dependência da taxa de crescimento *per capita* de longo prazo sobre o progresso tecnológico exógeno.

Os trabalho de Cass (1965) e Koopmans (1965) completaram o modelo básico de crescimento neoclássico e depois disso, a teoria do crescimento se tornou excessivamente técnica e, paulatinamente, perdeu o contato com as aplicações empíricas. Em contrapartida, os economistas que estudavam desenvolvimento ainda mantiveram uma perspectiva aplicada e continuaram a utilizar os modelos que não eram sofisticados, porém empiricamente úteis.

Imagina-se que, devido a esta ausência de relevância empírica, a teoria do crescimento tivesse, efetivamente, morrido como uma atividade de pesquisa, durante os anos da década de 1970, período anterior a revolução das expectativas racionais e do choque do petróleo.

Porém, após meados da década de 1980, a pesquisa em crescimento econômico experimentou uma súbita expansão, a partir dos trabalhos de Romer (1986) e Lucas (1988), cuja motivação era a observação de que os determinantes do crescimento econômico de longo prazo são questões cruciais, muito mais importantes do que a mecânica do *business cycles* ou dos efeitos contracíclicos das políticas monetária e fiscal.

A onda inicial da nova agenda de pesquisa foi fundamentada a partir dos trabalhos de Arrow (1962), Sheshinski (1967) e Uzawa (1965) e, de fato, não introduziu a teoria da mudança tecnológica. A incorporação das teorias de P&D e competição imperfeita na abordagem do crescimento começou com Romer (1987, 1990) e incluiu também as contribuições de Aghion & Howitt (1992) e Grossman & Helpman (1991, caps 3 e 4) (BARRO & SALA-I-MARTIN, 2004, p.19).

Este programa de pesquisa permaneceu ativo durante toda a década dos anos 1990 e tem se aplicado, por exemplo, à compreensão dos efeitos de escala no processo de crescimento com Jones (1990), e também analisando se o progresso tecnológico seria *capital augmenting* ou *labor augmenting* (idem, 2004, p. 20).

Portanto, Barro & Sala-I-Martin (2004) concluem que a mais clara distinção entre aquela teoria do crescimento dos anos 1960 e esta nova agenda dos anos 1990 é que as pesquisas recentes prestam uma atenção especial para as implicações empíricas e à relação entre a teoria e os dados. Contudo, muita dessa perspectiva aplicada envolve aplicações das hipóteses empíricas a partir das velhas teorias, notadamente sobre a previsão da convergência condicional do modelo de crescimento neoclássico.

2.1 Metodologia clássica no cálculo da convergência de renda

Com o intuito de fornecer uma avaliação quantitativa da velocidade de convergência, é proposta uma extensão do modelo de Solow-Swan. Segundo Barro & Sala-I-Martin (2004), o processo se inicia com uma versão log-linearizada do sistema dinâmico¹ das equações de transição (ou de movimento) no modelo de Ramsey, quando expandidas na posição de estado estacionário. Resolvendo as equações do modelo, os resultados podem ser escritos por

$$\log[\hat{y}(t)] = e^{-\beta t} \cdot \log[\hat{y}(0)] + (1 - e^{-\beta t}) \cdot \log(\hat{y}^*) \quad (2.1)$$

onde $\beta > 0$ e $\forall t \geq 0$, $\log[\hat{y}(t)]$ se define como a proporção média dos valores iniciais e de estado estacionário, respectivamente $\log[\hat{y}(0)]$ e $\log(\hat{y}^*)$, tendo a proporção do valor inicial decaindo exponencialmente à taxa β (velocidade de convergência), a qual pode ser definida como a mudança na taxa de crescimento do produto *per capita* à medida que este muda. Esta *velocidade* depende dos parâmetros tecnologia e preferência (BARRO & SALA-I-MARTIN, 2004, p.111).

Esta última equação implica que a taxa média de crescimento do produto *per capita*, num determinado intervalo a partir de $t = 0$ para qualquer tempo futuro $T \geq 0$, é dada por

$$(1/T) \cdot \log[y(T)/y(0)] = x + T^{-1} \cdot (1 - e^{-\beta T}) \cdot \log[\hat{y}^*/\hat{y}(0)] \quad (2.2)$$

Mantendo-se fixos a taxa de crescimento de estado estacionário, x , a velocidade de convergência, β , e o intervalo médio, T , a equação (2.2) informa que a taxa média de crescimento do produto *per capita* depende negativamente da taxa $\hat{y}(0)$ para \hat{y}^* e, com isso, da mesma forma que no modelo de Solow-Swan, o efeito da posição inicial, $\hat{y}(0)$, é condicionado à posição de estado estacionário, \hat{y}^* , querendo dizer, com isso, que o modelo de Ramsey prevê tanto a convergência condicional quanto a absoluta (*Idem*, 2004, p.112).

Assim, para efeito de estimação da β -convergência, pode-se seguir a especificação

¹Para \hat{k} e \hat{c} , níveis efetivos das variáveis consumo e estoque de capital, ambas *per capita*.

do modelo econométrico abaixo

$$\frac{1}{T} \cdot \log\left(\frac{y_{it}}{y_{i0}}\right) = \alpha + \frac{(1 - e^{-\beta t})}{T} \cdot \log(y_{i0}) + \epsilon_{i0,t} \quad (2.3)$$

tendo y_{i0} e y_{it} como o produto *per capita* da economia i no período $t = 0$ e $t = T$, respectivamente. Esta especificação segue as implicações do modelo de Solow-Swan, tal que a taxa de crescimento do PIB *per capita* é decrescente com o tamanho do produto, ou seja, a taxa de convergência da economia é menor à medida que ela se aproxima do estado estacionário.

Em outras palavras, a taxa de crescimento média cai quando T aumenta, ao passo que quando $T \rightarrow \infty$, a taxa de crescimento do estado estacionário, α (equivalente a x , no modelo de Ramsey), domina a média e, assim, o coeficiente $(1 - e^{-\beta t})/T$ se aproxima de zero e $T^{-1} \cdot \log[y_{it}/y_{i0}] \rightarrow \alpha$.

A partir da constatação da possibilidade de calcular um segundo modo de convergência pelo modelo de Ramsey, como visto acima, surgiram as Novas Teorias do Crescimento Econômico (NTCE)², as quais abriram novas possibilidades para analisar o processo de convergência de renda e mais notadamente pelo fato de negar a existência de uma convergência absoluta, dando lugar à convergência condicional.

Em relação a esta situação, Grolli, Oliveira & Jacinto (2006) explicam que, neste caso, cada economia teria seus próprios parâmetros, implicando que elas apresentariam um estado estacionário próprio. Ou seja, a existência desta convergência condicional significaria que, possuindo parâmetros em comum, as economias tenderiam a crescer mais rapidamente quanto maior fosse o *gap* do seu estado inicial ao estado estacionário. Tendo esta consideração, e como não mais se considera a convergência absoluta dada pela equação (2.1), é possível construir uma nova especificação econométrica obtendo um novo modelo da seguinte forma

$$\frac{1}{T} \cdot \log\left(\frac{y_{it}}{y_{i0}}\right) = \alpha_0 + \alpha_1 \cdot \log(y_{i0}) + \Theta' \mathbf{X}_{i0} + \epsilon_{i0,t} \quad (2.4)$$

em que $\alpha_1 = T^{-1} \log(1 - e^{-\beta t})$, Θ é um vetor de parâmetros e \mathbf{X}_{i0} é uma matriz das variáveis de controle, a qual mantém constante o estado estacionário das economias.

Este fato abre a possibilidade da adição de outras variáveis explicativas específicas das economias ao modelo econométrico e irá diferenciar os estados estacionários, permitindo apenas a existência da convergência condicional.

Grolli, Oliveira & Jacinto (2006) advertem para a inclusão de variáveis explicativas

²Destacando-se os modelos AK de crescimento endógeno (Frankel-Romer) e o modelo de dois setores com crescimento endógeno (Lucas e Uzawa).

que não tenham fundamento econômico teórico para justificar o modelo. Inclusive, sobre este fato, Levine & Renelt (1992) identificam como resultados *espúrios* a inclusão ou exclusão dessas variáveis de controle alterando significativamente os sinais obtidos e a direção do efeito sobre a taxa de crescimento do produto.

Frequentemente, quando se utiliza base de dados com a desagregação municipal, acontecem problemas de endogeneidade na estimação por mínimos quadrados ordinários (MQO). Dentre outros, os principais são: a pouca homogeneidade entre as unidades geográficas, que pode causar problemas de heteroscedasticidade e a presença de *outliers*, afetando a eficiência e consistência dos estimadores. Com a estimação da velocidade por MQO não é possível testar a convergência do tipo clube, já que a estimação está à média.

Além disso, como identificado por Quah (1993), a relação negativa entre $T^{-1} \cdot \log[y_{it}/y_{i0}]$ e $\log(y_{i0})$ pode não implicar convergência, caracterizando a *Falácia de Galton*³.

2.2 Modos de convergência

No que se refere à literatura empírica, uma parte significativa dos esforços busca validar uma das três versões da hipótese de convergência, com o objetivo de legitimar algum modelo teórico em detrimento de outro. Seguiremos a estratégia proposta por Johnson & Takeyama (2003) para testar a hipótese de convergência, tendo como base a especificação usual de um modelo descrito pela equação a seguir:

$$g_{y_{i,t}} = \frac{1}{T} \cdot \log\left(\frac{y_{it}}{y_{i0}}\right) = \beta_0 + \beta_1 \cdot \log(y_{i0}) + \beta_2 X_{i0} + \epsilon_{i0} \quad (2.5)$$

Toda decisão se foca no comportamento das estimativas dos parâmetros deste modelo. As hipóteses de convergência absoluta, condicional e clubes de convergência, todas originárias de Solow (1956), podem ser testadas de acordo com as variações da equação (2.3). Por exemplo, quando β_0 e β_1 forem iguais para todos os municípios, mas com $\beta_2 = 0$ e se $\beta_1 < 0$, tem-se o indicativo da ocorrência de *convergência absoluta*. Com isso, segue a definição abaixo:

Definição 1 (Convergência absoluta) *As rendas per capita de duas economias quaisquer tenderão para o mesmo nível no longo prazo (estado estacionário) independentemente de suas condições iniciais.*

³A Falácia de Galton tem um significado histórico particular no caso referente aos estudos originais de Francis Galton, primo de Charles Darwin, devido a uma inferência incorreta que é feita a partir de suas observações. Galton avaliou a altura dos pais contra a altura dos filhos. O resultado de sua investigação é que filhos de pais altos tendem a ser altos, mas menores que seus pais, e que filhos de pais baixos tendem a ser também baixos, mas maiores que seus pais e, desta forma, a altura da população tenderia a convergir para a média. A falácia ocorre porque a seleção de observações tomando seus extremos levaria naturalmente a essa tendência (COSTA, 2009, p.9).

Retirando-se a restrição de β_2 , sendo $\beta_1 < 0$ e considerando que as variáveis de controle se mantenham estatisticamente significativas, há a ocorrência de *convergência condicional*, cuja definição mais formal segue:

Definição 2 (convergência condicional) *Duas economias com características estruturais comuns (mesmas preferências, tecnologias, taxas de crescimento populacional, políticas públicas etc.) terão o mesmo nível de renda per capita no longo prazo, independentemente de suas condições iniciais.*

Sustenta-se, na literatura, a premissa de heterogeneidade espacial quando no estudo da convergência se incluem aspectos espaciais. Por isso, Ferreira e Silva & Oliveira Junior (2010) comentam que as economias somente transitam para uma condição de equilíbrio se dispuserem de características estruturais comuns e de mesma condição inicial gerando, desta forma, múltiplos equilíbrios de longo prazo. Este pressuposto faz culminar o conceito de *Clube de convergência*, de acordo com a definição abaixo.

Definição 3 (clube de convergência) *As rendas per capita de duas economias quaisquer tenderão para um mesmo nível de longo prazo somente se estas partilharem das mesmas características estruturais e, ao mesmo tempo, possuírem condições iniciais semelhantes.*

Para os propósitos desta dissertação, dar-se-á foco no estudo das duas primeiras hipóteses ou modos de convergência, definidos acima.

3 IMERSÃO NO MÉTODO

“[...] a insistência de se testar as teorias econômicas, frequentemente encontradas em escritos deste gênero, atesta a fé na existência de fatores estocásticos. Se as relações teóricas fossem, de fato, determinísticas, não surgiria a questão do teste estatístico.”

Jam Kmenta (1988, p.244)

Por que não mais um modelo linear? Esta é a questão chave, já que todas as pesquisas feitas para se testar convergência de renda ainda se inspiram em variantes do MCRL, para buscar alternativas que possam responder de forma satisfatória o porquê de as economias se encontrarem em diferentes equilíbrios, em situações específicas e cujo período somente agrava suas distâncias, em relação às economias mais avançadas.

Existem várias razões para isto. Entretanto, iremos nos debruçar apenas sobre os aspectos técnicos, teóricos e empíricos, sobre a modelagem e seus resultados, o que se adere à esta pesquisa de dissertação.

O fato mais importante é que, como também nos referimos acima, não existe um modelo único que possa, com seus resultados, servir de base para responder a problemática de todas as regiões que precisam ser investigadas. Este é o primeiro ponto. Outro é a desconfiança no modo como se embaraçou a teoria e a prática do teste econométrico.

Como bem observa Kmenta (1988), em linhas gerais, o fato de os economistas teóricos apresentarem as relações de forma determinística, não por acreditarem em ausência de acaso, mas devido às influências sistemáticas terem mais peso na teorização, não implica que a estrutura rígida, mas lógica, da teoria precise impor aos testes as mesmas rigidezes e determinar, desta forma, o modo como a modelagem econométrica deva se comportar, se adequando, ao seu máximo, à mesma estrutura de um modelo teórico.

É a partir desta mistura, ou má-interpretação de conceitos, que se abriu caminho para equívocos nos resultados dos testes econométricos. É preciso também compreender o que de fato a teoria econômica nos ensina e o que podemos (e devemos) contribuir com ela.

Mesmo que a teoria econômica defina que exista uma relação direta entre uma variável e as outras que a explicam, nem sempre a resposta caminha, em termos econométricos, da mesma forma, o que se traduz em construções analíticas na classe de modelos

lineares. A experiência vem comprovando que esta classe é imensa, mas limitada à poucas respostas teóricas e que, por sorte, tem respaldo econométrico. Mas, a mesma experiência prática nos leva a crer na limitação da classe de modelos lineares para aplicações da teoria econômica.

Os fatos concretos da experiência da modelagem, em que se incluem os conhecidos *fatos estilizados* apresentados em Jones (2001), mostram ao mesmo tempo os avanços no campo de estudo da teoria do crescimento, como também sua consolidação como campo teórico, ao passo que se apresentam técnicas que vem pondo em xeque a usabilidade da modelagem linear.

Como se pode observar, o desdobramento da metodologia da convergência de renda na seção 2.1, o avanço das pesquisas no tempo se faz utilizar de vários artifícios e estruturas novas, mas o tradicionalismo ainda persiste em utilizar ou misturar estruturas poderosas em uma classe limitada, como a dos modelos lineares.

Houve quem inovasse incluindo estruturas espaciais no modelo, o que apontou para mais avanços nas pesquisas e na importância de se considerar esses aspectos quando se tratar de estudos regionais. Essa grande contribuição fez também suscitar desconfianças no modelo teórico utilizado, tanto quanto fez lembrar da imprescindibilidade da inclusão de variáveis inerentes à estas localidades.

Como bem exposto no final do capítulo 2, Barro & Sala-I-Martin (2004), ao explicar a diferença da abordagem dos anos 1960 para a nova agenda de pesquisa pós-1990, a qual esta pesquisa segue, é que a implicação empírica é que dá destaque a esta nova vertente, sempre caminhando para inovações quanto à modelagem e métodos no qual envolva a retroalimentação advinda desta prática empírica.

Por isso, mesmo que a teoria econômica seja constituída por modelos determinísticos, não necessariamente a modelagem econométrica precisa ser especificada de acordo com um modelo linear, pois as barreiras que impediam a prática empírica já foram superadas, graças ao avanço das técnicas computacionais e pela sofisticação dos modelos, permitindo também a inclusão de variáveis importantes que de outra forma não seria possível para se chegar às estimativas.

Aprendemos que a ciência se faz quando nos apoiamos em ombros de gigantes e, por isso, apresentamos uma pequena evolução de como caminhar com testes econométricos, utilizando a sofisticação de novos modelos e métodos, sem esquecer que o MCRL foi o princípio desta mudança de conduta, para propor um avanço na consistência e robustez dos resultados.

3.1 O modelo linear geral

Uma das maiores preocupações da teoria econômica, sem dúvida, é a relação entre variáveis, ao passo que a econometria tem o papel de testar as proposições da teoria que se incorporam nessas relações e, com isso, estimar os parâmetros nela envolvidos.

Essas relações são a representação sistemática dos fenômenos traduzidas em *relações funcionais* e formalizadas em equações. Por exemplo, uma destas equações pode ser caracterizada como $y = \beta + \zeta x$. Com isso, percebe-se que a forma da equação caracterizadora é que vai nominar a relação funcional, como foi o exemplo de uma equação linear que descreve uma relação linear, e assim por diante.

Existe uma classificação dicotômica para essas relações funcionais, quais sejam: elas podem ser *determinísticas*, quando cada elemento do domínio se associa com um e apenas um elemento da imagem. Isto é, a relação entre as variáveis X e Y caracterizada como $Y = f(x)$ é uma relação determinística se para cada valor de X somente acontece um valor correspondente de Y .

Inclusive, pode haver a situação em que a relação seja determinística mesmo quando ambas as variáveis forem estocásticas. Mas, se ambas as variáveis forem estocásticas quando a relação for determinística, a distribuição condicional de Y dado X será degenerada. Por outro lado, diz-se que a relação entre X e Y é *estocástica* se para cada valor de X houver uma distribuição de probabilidade total dos valores de Y (KMENTA, 1988, p.241). Com isso, para cada valor dado de X , a variável Y , pode assumir um valor específico, ou cair dentro de um intervalo, com certa probabilidade.

Na regressão paramétrica, as distribuições condicionais da relação entre as variáveis preditoras e a variável resposta são modeladas por meio de famílias de distribuições de probabilidades paramétricas, cuja forma da função distribucional é conhecida. Os valores dos seus parâmetros são desconhecidos, tendo também a forma funcional f conhecida e que pode ser escrita por:

$$y_i = f(x_i, \beta_1, \dots, \beta_k) + \epsilon_i, \quad \forall i = 1, 2, \dots, n. \quad (3.1)$$

Interessante notar que no exercício da teoria econômica, as relações são apresentadas de forma determinística, de maneira geral. Kmenta (1988) comenta que isto não ocorre pelo fato de os economistas acreditarem na ausência completa de acaso, mas pelo fato de considerarem as perturbações estocásticas menos importantes que as influências sistemáticas, pois a introdução de perturbações estocásticas, nas relações econômicas, complicaria sobremaneira a tarefa do teórico.

Entretanto, é importante expor que, em econometria, as relações que são tratadas

detém, exclusivamente, a forma estocástica. Assim, na sua versão mais simples, uma relação estocástica entre duas variáveis X e Y , conhecida como *modelo simples de regressão linear*, pode ser descrita formalmente do seguinte modo:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (3.2)$$

com Y sendo a *variável dependente ou preditora*, X *variável independente*, ϵ a perturbação estocástica, e β_0 e β_1 são os parâmetros do modelo, desconhecidos, sendo o i referente à i -ésima observação.

Kmenta (1988) acrescenta que a natureza estocástica do modelo de regressão implica que para que cada valor de X haja uma distribuição de probabilidade total dos valores de Y , significando que não pode haver previsibilidade exata para esta última variável, pois a incerteza relativa a Y surge por causa da presença da perturbação estocástica, ϵ que, sendo causal, provoca causalidade naquela variável.

Geralmente, a distribuição de probabilidade de Y e suas características são determinadas pelos valores de X e pela distribuição de probabilidade de ϵ , e a forma matemática da relação funcional precisa ser pressuposta, sendo os parâmetros estimados de acordo com os valores de X e Y .

Dadas as características funcionais, é preciso deixar claro que a especificação plena do modelo de regressão inclui, não somente a forma da equação de regressão mas, sobretudo, a especificação da distribuição do termo de perturbação aleatório e as *proposições* as quais indicam como os valores da variável independente são determinados (KMENTA, 1988, p. 245).

Essas proposições são os *pressupostos básicos*, também conhecidos como as *Hipóteses de Gauss-Markov*, para o Modelo Clássico de Regressão Linear (MCRL). Como bem explica Hayashi & Fox (2000), as hipóteses de Gauss-Markov, em ordem crescente e de complexidade, são:

1. **Linearidade** nos parâmetros, ou seja $y_i = \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \epsilon_i$, onde os β 's são parâmetros a ser estimados e ϵ_i é o termo de erro, como mencionado acima.
2. **Exogeneidade estrita**, ou seja, $\mathbb{E}(\epsilon_i | \mathbf{X}) = 0, \forall i = 1, 2, \dots, n.$, o que significa que o valor esperado do termo de erro é condicional a *todos* os regressores para *todas* as observações.
3. **Não-Multicolinearidade**, em que o posto da matriz de dados $n \times K$, \mathbf{X} , é K , com probabilidade 1.
4. **Variância de erro esférico**, o que implica:

- *Homoscedasticidade* ou $\mathbb{E}(\epsilon_i^2 | \mathbf{X}) = \sigma^2 > 0, \forall i = 1, 2, \dots, n.,$ e $\sigma^2 < \infty.$
- *Erros não correlacionados* ou *Não-auto-regressão*, implicando que $\epsilon_i \perp \epsilon_j$ ou $\mathbb{E}(\epsilon_i \epsilon_j | \mathbf{X}) = 0, \forall i, j = 1, 2, \dots, n; i \neq j.$

5. **Normalidade do termo de erro**, ou seja, a distribuição de ϵ , condicional a \mathbf{X} , segue uma *Normal* conjunta, notada por $\epsilon | \mathbf{X} \stackrel{iid}{\sim} \mathbb{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n).$

O MCRL tem um método próprio para estimar os parâmetros, os *mínimos quadrados ordinários*, que implica a minimizar o quadrado do termo de erro estocástico, afim de ajustar a reta característica de uma função linear.

Esse método de estimação pode bem ser acompanhado em Hayashi & Fox (2000), Kmenta (1988), Wooldridge (2006), Stock & Watson (2004), Gujarati (2000; 2012), Greene (2012), Baltagi (2011), entre vários que dão ênfase à teoria econômica.

Outros, porém, tendem mais para a teoria da regressão, como Souza (1988), Bingham & Fry (2010), Rencher & Schaalje (2007), somente para citar alguns, e os que utilizam aplicação com o software livre **R**, dentre eles Faraway (2005), Hay-Jahans (2012) e Wright & London (2009).

3.2 Modelos lineares generalizados

Dada sua abrangência, Nelder & Wedderburn (1972) *Apud* Wood (2006) definem os Modelos Lineares Generalizados como aqueles que permitem outras considerações sobre as distribuições, além da Normal, inclusive permitindo um grau de não-linearidade na estrutura do modelo. Com isso, um GLM apresenta a seguinte estrutura

$$g(\mu_i) = \mathbf{X}_i \beta \tag{3.3}$$

onde $\mu_i \equiv \mathbb{E}(Y_i)$, $g(\cdot)$ é uma função-link de suavização monotônica, \mathbf{X}_i é a i -ésima linha da matriz do modelo, \mathbf{X} e β é um vetor de parâmetros desconhecido.

Adicionalmente, um GLM geralmente faz hipóteses sobre a distribuição em que os Y_i são independentes e que

$$Y_i \sim \text{alguma distribuição na família exponencial}$$

Esta família inclui muitas distribuições que são úteis para a modelagem prática, como a Poisson, Binomial, Gama e a Normal. Para uma referência completa sobre GLM, a consulta em McCullagh & Nelder (1989), Casela & Berger (2011) e Dobson (2001) permite uma conhecimento aprofundado.

Devido ao fato de os Modelos Lineares Generalizados serem especificados em termos de um preditor linear, $\mathbf{X}\beta$, muitas das ideias e conceitos gerais da modelagem linear se incluem, com pequenas modificações, para a modelagem linear generalizada.

As formulações são praticamente as mesmas para modelos lineares, exceto pelo fato da função *link* e a distribuição precisarem ser escolhidos. Certamente, se a função identidade é escolhida como *link*, juntamente com a distribuição Normal, então os modelos lineares ordinários se reduzirão a um caso especial.

Esta generalização é conseguida com algum custo: agora, o ajustamento do modelo precisa ser feito *iterativamente* e os resultados das distribuições, usadas para fins de inferência, são agora aproximações e justificadas pelos resultados advindos do limite da teoria assintótica, ao invés de resultados exatos.

Teoricamente, a estimação e inferência no GLM é baseada na teoria da estimação pela *máxima verossimilhança*, embora este tipo de maximização acaba por requerer uma abordagem iterativa pelos mínimos quadrados.

Para estimar um GLM, é preciso estabelecer um vetor de tamanho n de variáveis resposta independente, Y , onde $\mu \equiv \mathbb{E}(Y)$, via $g(\mu_i) = \mathbf{X}_i\beta$ e $Y_i \sim f_{\theta_i}(y_i)$, no qual $f_{\theta_i}(y_i)$ indica uma distribuição na família exponencial, com um parâmetro *canônico* θ_i , que é determinado por μ_i e, conseqüentemente, por β . Dado um vetor y com uma observação de Y , a estimação por máxima verossimilhança de β é possível.

Considerando Y_i mutuamente independente, a verossimilhança de β é

$$\mathcal{L}(\beta) = \prod_{i=1}^n f_{\theta_i}(y_i) \quad (3.4)$$

e, com isso, a log-verossimilhança de β é

$$l(\beta) = \sum_{i=1}^n \log[f_{\theta_i}(y_i)] \quad (3.5)$$

$$= \sum_{i=1}^n [y_i\theta_i - b_i(\theta_i)]/a_i(\phi) + c_i(\phi, y_i) \quad (3.6)$$

onde, $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções arbitrárias, ϕ é o parâmetro de escala e θ é conhecido como parâmetro *canônico* da distribuição (WOOD, 2006).

Na equação (3.3), a dependência do lado direito em β acontece através da dependência indireta via θ_i . Inclusive, os parâmetros a, b e c variam de acordo com i , o que possibilitaria diferentes denominadores binomiais, n_i , para cada observação de resposta binomial ou diferentes variâncias de resposta para a Normal (*id.*, p.63).

O parâmetro ϕ , por exemplo, é assumido para todo i . Em situações práticas, é suficiente considerar somente casos onde se encontre $a_i(\phi) = \phi/w_i$, onde w_i é uma

constante, geralmente estipulada em 1, no caso em que

$$l(\beta) = \sum_{i=1}^n w_i [y_i \theta_i - b_i(\theta_i)] / \phi + c_i(\phi, y_i) \quad (3.7)$$

A maximização se viabiliza pela diferenciação parcial de l com respeito a cada elemento de β , igualando o resultado a zero e resolvendo para o próprio β . Assim

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} \cdot \sum_{i=1}^n w_i \left(y_i \cdot \frac{\partial \phi_i}{\partial \beta_j} - b'_i(\theta_i) \cdot \frac{\partial \phi_i}{\partial \beta_j} \right) \quad (3.8)$$

E, com alguma álgebra, chegamos a

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} \cdot \sum_{i=1}^n \frac{[y_i - b'_i(\theta_i)]}{b''_i(\theta_i)/w_i} \cdot \frac{\partial u_i}{\partial \beta_j} \quad (3.9)$$

Agora, levando em consideração $\mathbb{E}(Y) = b'(\theta)$ e $Var(Y) = b''(\theta) \cdot \phi/w$ para substituir na última equação, implica que para resolver β segue

$$\sum_{i=1}^n \frac{(y_i - \mu_i)}{Var(\mu_i)} \cdot \frac{\partial \mu_i}{\partial \beta_j} = 0 \quad \forall j \quad (3.10)$$

Neste caso, o objetivo dos mínimos quadrados seria

$$\mathcal{S} = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{Var(\mu_i)} \quad (3.11)$$

com μ_i dependente não-linearmente de β , porém os pesos $Var(\mu_i)$ são tratados como fixos. Assim, para resolver (3.11) é preciso utilizar métodos iterativos, os quais podem ser vistos em Wood (2006) e Nelder & Wedderburn (1972).

3.3 Regressão por *splines* e Modelos Aditivos Generalizados

Os trabalhos de Hastie & Tibshirani (1990) e, depois, Wood (2006), estenderam a ideia do modelo linear na classe de modelos aditivos. No modelo linear, cada variável \mathbf{X} é multiplicada por um escalar, o valor β . Isto é o que faz o modelo ser linear (linearidade nos parâmetros). Entretanto isto restringe outras relações entre qualquer variável \mathbf{X} e \mathbf{Y} , condicionada a todas as outras.

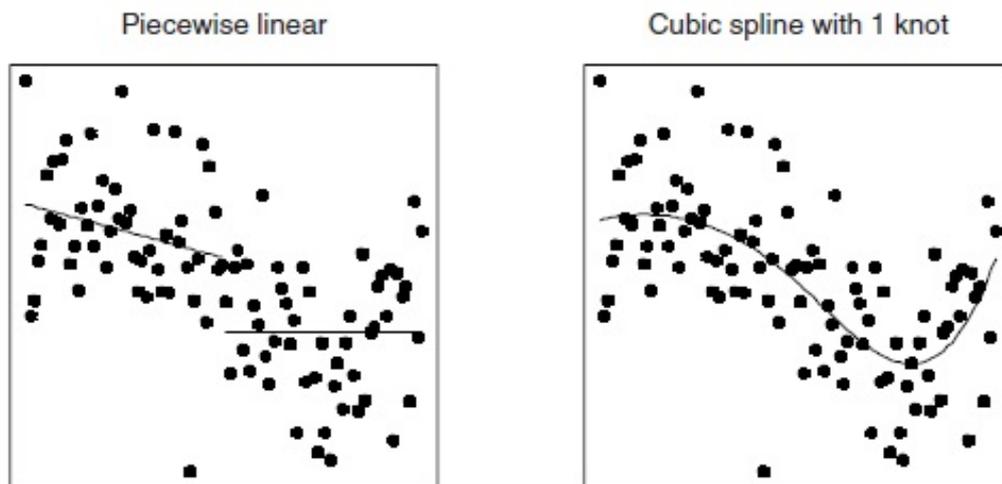
No caso dos modelos aditivos, os valores dos β 's são substituídos por uma função regular simples, em termos de graus de liberdade. Dessa forma, o modelo linear ganha interpretação aditiva e pode ser escrito como:

$$y_i = \beta_0 + f_1(X_{1i}) + \dots + f_k(X_{ki}) + \epsilon_i \quad (3.12)$$

As funções $f_k(X_{ki})$ são habitualmente assumidas como *splines*, com pequenos números de *nós*. Funções mais complexas podem ser usadas, porém isto pode causar ao modelo *sobre ajustes* nos dados observados e, portanto, não generalizar bem para novos conjuntos.

O gráfico resultante típico, como a Figura 1, abaixo, mostra as funções e os resultados numéricos mostram os ajustes dos componentes lineares e não lineares. A escolha das funções, que frequentemente cai para o tipo e complexidade de *splines*, é crítica. Portanto, para entender a lógica dos modelos aditivos, é preciso primeiro entender como o ajuste se dá por meio das *splines*.

Figura 1 – Ajustamento em dois estágios e com uma *Spline*



O painel esquerdo mostra regressões lineares separadas para a primeira e a segunda metade dos dados. O painel do lado direito mostra uma *Spline* cúbica com um único nó na mediana, que é o local padrão para um único nó.

Fonte: London & Wright (2009)

Embora existam diferentes tipos de *splines* utilizadas em estatística, enquanto todas são matematicamente complexas, para o nosso propósito utilizaremos as duas mais simples, que permitem ao mesmo tempo flexibilidade e simplicidade de compreensão, chamadas **B-spline** e **C-spline**.

O propósito da regressão por *splines* é desenhar uma curva em um gráfico de dispersão, do tipo polinomial com diferentes graus: $y_i = \beta_0$ (polinomial de grau zero); $y_i = \beta_0 + \beta_1 x_i$ (grau = 1), etc. O número de β 's adicionais estimado é uma medida de complexidade do modelo e é chamado de *número de graus de liberdade* para aquela variável.

Wright & London (2009) esclarecem que um dos maiores problemas com regressões polinomiais é que elas são bastante *inflexíveis* e raramente ajustam bem todo o conjunto

de dados. Pode-se aumentar a complexidade de uma polinomial ajustando seus graus, porém isso traz dois problemas associados. Primeiramente, pode haver pouca relação econômica que poderia ser prevista no caso de esta função se tornar mais complexa e os polinômios tendem a atirar para cima e para baixo nas extremidades e, portanto, ser muito influenciados por pontos nessas extremidades (*outliers*). A solução para isto é juntar (somar) pedaços de polinômios. Esta é a base das *splines*.

As *splines* não somente conseguem fazer com que as curvas se encontrem¹, nos chamados *nós*, mas o fazem *da forma mais suave possível*. Inclusive, um aspecto bastante inteligente é que cada curva adicional requer a adição de apenas um grau (*df*), tal que aumentando o número de nós não aumenta muito a complexidade do modelo em termos de graus de liberdade.

Os métodos de suavização² ou *smoothing methods* são frequentemente dependentes dos dados ao invés da teoria, o que significa que com grandes quantidades de dados de alta qualidade se pode construir uma ferramenta exploratória muito útil, porém com pequenas amostras (o que é típico em Economia), é preciso usá-los cuidadosamente.

A representação dessas funções é melhor introduzida considerando um modelo contendo uma *função suave*³ de uma variável.

$$y_i = f(x_i) + \epsilon_i \quad (3.13)$$

onde y_i é a variável resposta, x_i a covariável, $f(\cdot)$ uma função *suave* e $\epsilon_i \stackrel{iid}{\sim} \mathbb{N}(0, \sigma^2)$ uma variável aleatória. E, neste caso, a *função* é que será *estimada*.

Para estimar $f(\cdot)$, é requerido que ela seja representada de uma forma que (3.13) se torne um modelo linear. Isto pode ser feito pela escolha de uma *base*, definindo o espaço de funções, nas quais $f(\cdot)$ (ou uma boa aproximação dela) é um elemento. Escolhendo essa base é o equivalente a escolher algumas funções básicas que serão tratadas como completamente conhecidas: Se uma função $b_i(x)$ for a tal i -ésima função básica, então $f(\cdot)$ é assumida ser representada por:

$$f(x) = \sum_{i=1}^q b_i(x)\beta_i \quad (3.14)$$

¹No caso em que uma parte da curva estimada em uma das metades dos dados e, a outra metade, numa segunda parte.

²Existem, além das *splines*, várias funções de suavização (*smoothing methods*). Uma forma muito popular é conhecida como *Kernel methods*. Estes calculam uma linha para cada ponto ao longo do eixo das abscissas \mathbf{x} e os juntam. Essa linha é baseada nos dados dentro de uma distância do ponto até o eixo \mathbf{x} . A suavização da curva fica dependente do quão estreita ou larga esta região (ou *kernel*) é. Quanto mais ampla a região, mais suavizada será a curva. Esses métodos são frequentemente chamados de *métodos ponderados localmente* (*locally weighted methods*), porque somente valores próximos do ponto ao eixo \mathbf{x} são usados para calcular a linha. Às vezes, as funções usadas irão ponderar igualmente todos os valores nas regiões e, às vezes, seus pesos serão dependentes do quão distante estão do eixo \mathbf{x} .

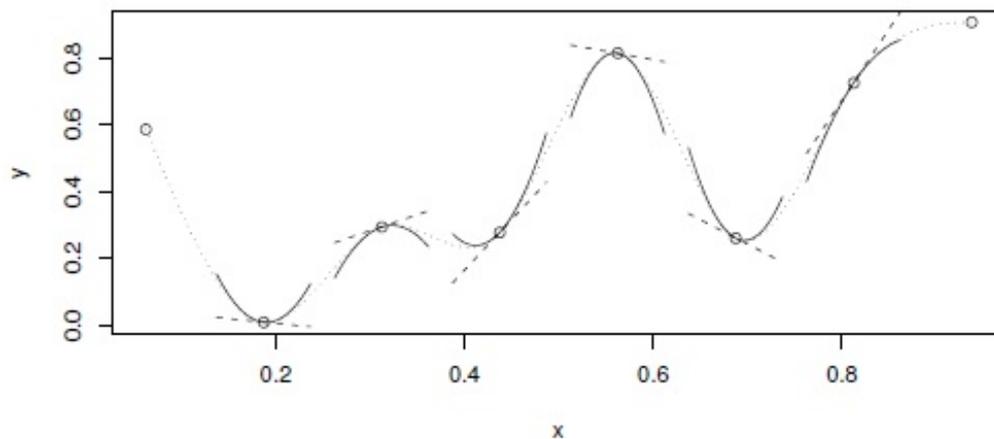
³Consideram-se todas as funções contínuas.

para alguns valores de parâmetros desconhecidos β_i . A substituição de (3.14) em (3.13), claramente resulta em um modelo linear.

Uma possibilidade óbvia para a escolha do grau de suavização é tentar fazer uso de métodos de teste de hipóteses para selecionar um q por *backwards selection* ou *seleção reversa*. Entretanto, Wright & London (2009) alertam que esta abordagem é problemática, dado que o modelo baseado em $k - 1$ observações, com nós uniformemente espaçados, não irão, geralmente, estar aninhados dentro de um modelo baseado em k com nós igualmente espaçados.

Por outro lado, é possível iniciar com uma malha fina de nós e, simplesmente, ir soltando sequencialmente, como uma parte da *seleção reversa*. Porém o espaçamento desigual de nós pode, ele mesmo, levar a uma má *performance* do modelo. Além disso, para os modelos de regressão por *splines*, o ajuste tende a depender fortemente da localização escolhida para os nós (WRIGHT & LONDON, 2009).

Figura 2 – Spline Cúbica



Os pontos em que elas são unidas (\circ) (e os dois pontos finais) são conhecidos como os nós da *splines*. Cada seção tem coeficientes diferentes, mas nos nós elas irão corresponder a seções de sua vizinha e às duas primeiras derivadas. As linhas retas tracejadas mostram os gradientes da *spline* nos nós e as linhas curvas contínuas são quadráticas, que correspondem às primeira e segunda derivadas nos nós: estes ilustram a continuidade da primeira e segunda derivadas através dos nós. Esta *spline* tem zero como segunda derivada nos nós finais, o que a caracteriza como uma *spline* natural.

Fonte: Wood (2006)

Uma alternativa para controlar a suavização pela alteração da dimensão da base é manter esta dimensão fixa com um tamanho um pouco maior do que se acredita razo-

avelmente ser necessário. Mas isso para controlar a suavização do modelo pela adição de uma *dobratura* como penalidade⁴ para o objetivo do ajuste por mínimos quadrados.

Por exemplo, ao invés de ajustar o modelo pela minimização de

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 \quad (3.15)$$

poderia ser ajustado pela minimização de

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \cdot \int_0^1 [f''(x)]^2 dx \quad (3.16)$$

onde a integral da segunda derivada penaliza os modelos que são muito *contorcidos*. O *trade-off* entre ajustamento de modelo e sua suavização é controlado pelo parâmetro de suavização, λ .

Quando $\lambda \rightarrow \infty$, conduz a uma estimação linear para $f(\cdot)$, enquanto $\lambda = 0$ resulta em uma simples estimação de regressão por MQO. Quando $\lambda = 1$, isso vai conduzir a uma estimação não penalizada por *splines*.

Devido ao fato de $f(\cdot)$ ser linear nos parâmetros β_i , a penalidade pode sempre ser escrita como uma forma quadrática em β

$$\int_0^1 [f''(x)]^2 dx = \beta' \mathbf{S} \beta \quad (3.17)$$

onde \mathbf{S} é uma matriz de coeficientes conhecida. Assim, o problema de ajustamento da regressão por *splines* penalizada se reduz a minimizar

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \cdot \beta' \mathbf{S} \beta \quad (3.18)$$

com relação a β . O problema de o grau de suavização para o modelo é agora um problema de estimação do parâmetro de suavização λ . Mas, antes de abordar o estimador de λ , considera-se a estimação de β , dado λ .

Agora, suponha que duas variáveis explanatórias x e z estejam disponíveis para a variável resposta y e que uma estrutura aditiva simples sejam apropriada, como

$$y_i = f_1(x_i) + f_2(z_i) + \epsilon_i \quad (3.19)$$

As f_i são funções suaves e $\epsilon_i \stackrel{iid}{\sim} \mathbb{N}(0, \sigma^2)$ uma variável aleatória. Este é um típico modelo aditivo.

Existem dois pontos a ser percebidos, nesse modelo. Primeiramente, a hipótese

⁴É preciso deixar claro que a soma do produto do parâmetro de suavização e a integral de Lebesgue, que é uma noção de integral numa classe muito mais ampla de funções, não implica que o modelo seja restrito, porque esta restrição se dirige, especialmente, a evitar dobraduras excessivas no caminho percorrido pela *Spline* no ajustamento.

de efeitos aditivos é bem forte: $f_1(x) + f_2(z)$ é um caso especial bem restrito da função suave geral de duas variáveis $f(x, z)$. Além disso, o fato de agora o modelo conter mais de uma função introduz um problema de identificabilidade: f_1 e f_2 cada uma apenas pode ser estimável para dentro de uma constante aditiva. Para visualizar isto, bastaria que qualquer constante fosse adicionada a f_1 e subtraída de f_2 , sem modificar a predição do modelo. Portanto, restrição de identificabilidade deve ser imposta ao modelo antes do ajuste.

No entanto, desde que questões de identificabilidade sejam tratadas, o modelo aditivo pode ser representado usando regressões *splines* penalizadas, estimadas pelos mínimos quadrados penalizados e os graus de suavização estimado via *cross validation*, da mesma forma que no modelo simples univariado, dando vazão à possibilidade de sua generalização.

Conceitualmente, os modelos aditivos generalizados seguem a partir dos modelos aditivos, da mesma forma que os modelos lineares generalizados seguem a partir dos modelos lineares. Isto é, o preditor linear, agora, prevê alguma função suave monotônica conhecida do valor esperado da variável resposta, e esta precisa seguir qualquer distribuição na *família exponencial* ou simplesmente ter uma relação média-variância, permitindo o uso da abordagem quasi-verossimilhança.

Um modelo aditivo generalizado (HASTIE & TIBISHIRANI, 1990) é um modelo linear generalizado com um preditor linear envolvendo a soma de funções suaves de covariáveis. Em geral, o modelo apresenta a estrutura:

$$g(\mu_i) = \mathbf{X}_i^* \Theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}; x_{4i}) + \dots \quad (3.20)$$

onde $\mu_i \equiv \mathbb{E}(Y_i)$ e $Y_i \sim$ alguma distribuição na *família exponencial*. Y_i é a variável resposta; \mathbf{X}_i^* é uma matriz-linha para qualquer componente paramétrico no modelo; Θ é o vetor de parâmetros correspondente, e os f_i s são funções suaves das covariáveis, x_k .

O modelo permite uma especificação bastante flexível da dependência da variável resposta sobre as covariáveis e a especificação do modelo deve ser feita em termos de *funções suaves*, ao invés de relações paramétricas detalhadas, sendo possível evitar modelos carregados e difíceis de manejar.

Com isso, algumas limitações impostas pelas hipóteses dos modelos tradicionais podem ser superadas e promover melhores ajustes e respostas das estimativas para os problemas propostos. Uma delas, a que o erro deve seguir ou convergir para uma distribuição normal é substituída por uma situação mais ampla, pois agora basta que a distribuição pertença à família exponencial, e a Normal⁵ se inclui.

⁵Segundo Casella & Berger (2011), uma família de fdp ou fps é chamada de *exponencial*, caso seja

Agora, dado um modelo centrado matricialmente, para cada termo de suavização, a equação (3.20), pode ser escrita como

$$g(\mu_i) = \mathbf{X}_i \boldsymbol{\beta} \quad (3.21)$$

onde $\mathbf{X} = [\mathbf{X}^* : \mathbf{X}_1 : \mathbf{X}_1 : \dots]$ e $\boldsymbol{\beta}' = [\boldsymbol{\theta}', \boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots]$. Dessa forma, pode-se verificar que (3.21) recai para um GLM, e se pode escrever sua função de verossimilhança, $l(\boldsymbol{\beta})$. Inclusive, se os q_i forem razoavelmente amplos para que tenhamos uma possibilidade razoável de representar apropriadamente as funções f_j 's desconhecidas, e $\boldsymbol{\beta}$ for estimado pela maximização da verossimilhança ordinária, então provavelmente há boas chances de acontecer substanciais *sobre ajustes*. Por essa razão, Wood (2006) recomenda que os GAMs devam ser usualmente estimados pela maximização da *verossimilhança penalizada*, onde essas penalidades são idealizadas para sobrepor as estimativas com *dobraturas* excessivas dos termos de f_j .

Wood (2006) ainda comenta que as penalidades mais convenientes de se trabalhar são aquelas nas quais a função de dobratura assume uma forma quadrática nos coeficientes.

O modelo da equação (3.21), por exemplo, pode ser ajustado pela minimização de

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda \cdot \int_0^1 [f''(x)]^2 dx \quad (3.22)$$

onde a integral da segunda derivada penaliza os modelos que são muito *contorcidos* ou acentuada *dobratura*. O *trade-off* entre ajustamento de modelo e sua suavização é controlado pelo parâmetro de suavização, λ . Devido ao fato de $f(\cdot)$ ser linear nos parâmetros β_i , a penalidade pode sempre ser escrita como uma forma quadrática em $\boldsymbol{\beta}$

$$\int_0^1 [f''(x)]^2 dx = \boldsymbol{\beta}' \mathbf{S} \boldsymbol{\beta} \quad (3.23)$$

expressa como

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right)$$

Seja, agora, $f(x|\mu, \sigma^2)$ a família $\mathbb{N}(\mu, \sigma^2)$ de fdps, onde $\boldsymbol{\theta} = (\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma > 0$. Então,

$$\begin{aligned} f(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{-\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right) \end{aligned}$$

Definindo $h(x) = 1, \forall x$; $c(\boldsymbol{\theta}) = c(\mu, \sigma) = (\sqrt{2\pi}\sigma)^{-1} \exp\left(-\frac{-\mu^2}{2\sigma^2}\right)$; $w_1(\mu, \sigma) = \frac{1}{\sigma^2}$ $\sigma > 0$; $w_2(\mu, \sigma) = \frac{\mu}{\sigma^2}$ $\sigma > 0$; $t_1(x) = -x^2/2$ e $t_2(x) = x$, então,

$$f(x|\mu, \sigma^2) = h(x)c(\mu, \sigma) \exp[w_1(\mu, \sigma)t_1(x) + w_2(\mu, \sigma)t_2(x)]$$

c.q.d.

onde \mathbf{S} é uma matriz de coeficientes conhecida. Assim, o problema de ajustamento da regressão por *splines* penalizada se reduz a minimizar

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \cdot \beta' \mathbf{S} \beta \quad (3.24)$$

com relação a β , dado λ .

A reparametrização centralizada poderia converter esta penalidade para a forma $\beta'_j \bar{\mathbf{S}}_j \beta_j$, onde $\bar{\mathbf{S}}_j = \mathbf{Z}' \bar{\mathbf{S}}_j \mathbf{Z}$. Porém, para facilitar a notação, é conveniente reescrever esta penalidade em termos do vetor de coeficiente β completo, de forma que ele venha a se tornar $\beta' \mathbf{S} \beta$, onde \mathbf{S} é apenas $\bar{\mathbf{S}}_j$ inflada de zeros, de forma que $\beta' \mathbf{S} \beta \equiv \mathbf{Z}' \bar{\mathbf{S}}_j \mathbf{Z}$ (WOOD, 2006).

Com isso, a expressão formal para minimizar (3.24) é a estimação por mínimos quadrados penalizados de β_i , descrita abaixo:

$$\hat{\beta} = (\mathbf{X}' \mathbf{X} + \lambda \mathbf{S})^{-1} \mathbf{X}' \mathbf{y} \quad (3.25)$$

Portanto, dada a medida de *dobratura* para cada função, é possível definir a verossimilhança penalizada por

$$l_p(\beta) = l(\beta) - \frac{1}{2} \cdot \sum_j \lambda_j \beta' \mathbf{S}_j \beta \quad (3.26)$$

Onde o λ_j são parâmetros de suavização, controlando o *trade-off* entre ajuste e suavização do modelo. Dados valores para o λ_j , então l_p é maximizado para encontrar $\hat{\beta}$, porém os λ_j precisam ser estimados por métodos computacionais iterativos.

O método de estimação dos parâmetros dos GAMs é conhecido como IRLS⁶ ou *mínimos quadrados penalizados iterativamente reponderados*, descrito pela equação (3.26).

⁶Do original *Iteratively re-weighted least squares*.

4 MODELOS ADITIVOS GENERALIZADOS PARA LOCAÇÃO, ESCALA E FORMA

“Man can learn nothing unless he proceeds from the known to the unknown.”

Claude Bernard

De maneira geral, os Modelos Aditivos Generalizados para Locação, Escala e Forma (GAMLSS) são uma classe de modelos estatísticos gerais para variáveis de resposta univariada. O modelo assume observações independentes das variáveis de resposta y , dados os parâmetros, das variáveis explicativas e dos vetores de efeitos fixos.

São também conhecidos como modelos de regressão do tipo semi-paramétricos. Eles são *paramétricos* no sentido de que eles requerem a hipótese de uma distribuição paramétrica para a variável resposta e *semi* porque a modelagem dos parâmetros da distribuição, como função das variáveis explicativas, pode envolver o uso de funções suaves não-paramétricas (STASINOPOULOS & RIGBY, 2007).

A distribuição da variável resposta, no GAMLSS, pode ser selecionada a partir de uma família de distribuição muito ampla, incluindo aquelas com grande assimetria e curtose, em distribuições discretas e contínuas.

A parte sistemática do modelo é expandida para permitir não somente a média (ou o *parâmetro de locação*), mas também outros parâmetros da distribuição de y , como, por exemplo, funções paramétricas e/ou aditivas não-paramétricas (que são as funções suaves) das variáveis explicativas e/ou termos de efeito aleatório.

O GAMLSS foi introduzido por Rigby & Stasinopoulos (2001, 2005) e Akantzi-liotou, Rigby & Stasinopoulos (2002) como uma forma de superar algumas limitações associadas com os populares Modelos Lineares Generalizados, os GLMs, e os Modelos Aditivos Generalizados, GAMs¹.

4.1 A incompletude do GAM e a opção pelo GAMLSS

A quantidade de dados disponíveis e as análises estatísticas contemporâneas tem crescido vertiginosamente ao longo dos anos e requerido cada vez mais ajustes de modelos mais complexos e potencialmente mais realistas.

¹Ver Nelder & Wedderburn (1972) e Hastie & Tibshirani (1990), respectivamente.

Neste contexto, a abordagem das técnicas de modelagem da regressão univariada, os Modelos Lineares Generalizados, GLM, e os Modelos Aditivos Generalizados, GAM, ocupam um espaço proeminente (NELDER & WEDDERBURN, 1972; HASTIE & TIBSHIRANI, 1990). Ambos assumem uma distribuição pertencente à família exponencial para a variável de resposta, y , no qual a média, μ de y é modelada como função das variáveis explicativas e de sua variância, dada por $Var(y) = \phi \cdot \nu(\mu)$, dependendo de um parâmetro constante de dispersão ϕ e da média μ , através da função de variância $\nu(\mu)$.

Além disso, para uma distribuição na família exponencial, ambas assimetria e curtose de y são, em geral, função de μ e de ϕ . Nessa abordagem, os modelos GLM e GAM, suas variâncias, curtoses e assimetrias não são modeladas explicitamente em termos das variáveis explicativas, mas implicitamente pela sua dependência de μ .

Outra classe importante de modelos, os modelos mistos (com efeitos aleatórios), os quais fornecem uma abordagem bastante ampla para modelar dados dependentes, particularmente associados com esquemas espaciais, hierárquicos e amostragem longitudinais, assumem normalidade para a distribuição condicional de y dados os efeitos fixos e, com isso, não podem modelar a assimetria e curtose explicitamente.

O modelo linear misto generalizado, GLMM, combina o GLM e o modelo misto linear, porém introduz um termo de efeito aleatório no termo preditor linear para a média do GLM. O modelo misto se utiliza de procedimento Bayesianos para fazer o ajuste utilizando algoritmos EM e métodos MCMC², descritos em McCulloch (1997) e Zeger & Karim (1991).

Fahrmeir & Lang (2001) se utilizaram do mesmo procedimento para o GAM, usando inferência Bayesiana, que se tornou GAMM. Fahrmeir & Tutz (2001) discutiram procedimentos de estimação alternativos para o GLMM e o GAMM, porém mesmo estes dois últimos modelos sendo mais flexíveis que suas versões originais, também assumem uma família exponencial condicional para y e, raramente, permitem modelagem de outros parâmetros exceto a média (locação) da distribuição da variável resposta, como função das variáveis explicativas.

Seu ajustamento frequentemente depende de MCMC ou verossimilhanças integradas³, tornando sua operacionalização intensiva computacionalmente e altamente consumidora de tempo, mostrando ineficiência, pelo menos até o momento contemporâneo, para grandes bancos de dados onde a seleção de modelos requer a investigação de muitos modelos alternativos.

²Monte Carlo com Cadeias de Markov, classe de algoritmos para a amostragem de distribuições de probabilidade com base na construção de uma cadeia de Markov, que tem a distribuição desejada como a sua distribuição de equilíbrio. O estado da cadeia após um grande número de passos é então utilizado como uma amostra da distribuição desejada.

³Distribuições marginais; isto é, quadraturas gaussianas.

Vários procedimentos de aproximação para ajustar o GLMM e o GAMM têm sido propostos (BRESLOW & CLAITON, 1993; BRESLOW & LIN, 1995; LEE & NELDER, 1996, 2001a, b), entretanto uma abordagem alternativa é utilizar máxima verossimilhança não paramétrica baseada em *mistura finita* (AITKIN, 1999).

Rigby & Stanisopolous (2005) propuseram uma classe de modelos de regressão univariado, que é o modelo base utilizado nesta dissertação, o GAMLSS, em que o pressuposto da família exponencial requerido pelo GAM é relaxado⁴.

Dentro desta nova abordagem, o modelo expande sua parte sistemática para permitir *todos* os parâmetros da distribuição condicional de y para ser modelada como funções paramétricas e aditivas não paramétricas (suaves), ou outras possibilidades, das variáveis explicativas, como também termos de efeitos aleatórios. Porém, o mais importante é que o GAMLSS faz o ajuste por meio de dois procedimentos algorítmicos diferentes. O primeiro, conhecido como RS, baseia-se no algoritmo que foi utilizado para ajustar média e a dispersão em modelos aditivos de Rigby & Stanisopolous (1996a), enquanto o segundo algoritmo, CG, é baseado em Cole & Green (1992).

4.2 Modelagem e Estimação

Considerando os p parâmetros $\Theta' = (\theta_1, \theta_2, \dots, \theta_p)$ de uma função de densidade de probabilidade $f(y|\Theta)$ de uma população, para este estudo, eles são estimados a partir da construção de Modelos Aditivos.

Especificamente, o modelo assume que, $\forall i = 1, 2, \dots, n$, as observações y_i são independentes e condicionais a Θ^i , com função de densidade de probabilidade $f(y_i|\Theta^i)$, onde $\Theta^i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})$ é um vetor de parâmetros p relacionado às variáveis explicativas e aos efeitos aleatórios⁵.

Seja $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ um vetor de observações da variável resposta. Inclusive, para $k = 1, 2, \dots, p$, seja $g_k(\cdot)$ uma função-link monotônica conhecida, relacionando θ_k às variáveis explicativas e aos efeitos aleatórios através de um modelo aditivo dado por

$$g_k(\theta_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \gamma_{jk} \quad (4.1)$$

onde $\boldsymbol{\theta}_k$ e $\boldsymbol{\eta}_k$ são vetores de tamanho n , isto é, $\boldsymbol{\theta}'_k = (\theta_{1k}, \dots, \theta_{nk})$, $\boldsymbol{\beta}'_k = (\beta_{1k}, \dots, \beta_{j'_k k})$

⁴É importante esclarecer que tanto o GAM, com também o GLM, não foram esquecidos, porque existem aplicações em que não é necessária a utilização de um modelo mais geral. Inclusive, o próprio GAMLSS detecta, nas distribuições, a existência dos parâmetros a serem utilizados na estimação e, com isso, pode haver o caso em que modelos menos complexos sejam usados, não necessitando da utilização do GAMLSS, e passe para uma classe anterior, como o GAM ou GLMM, por exemplo.

⁵Se os valores das covariáveis são estocásticos ou observações de y_i dependendo de seus valores passados, então $f(y_i|\Theta^i)$ é concebido como condicional a esses valores.

é um vetor de parâmetros de tamanho J'_k , \mathbf{X}_{jk} é uma matriz de variáveis conhecida de ordem $n \times J'_k$, \mathbf{Z}_{jk} é uma matriz de variáveis fixa conhecida $n \times q_{jk}$, onde contém os termos aditivos e γ_{jk} é um vetor de variáveis aleatórias de tamanho q_{jk} . Este modelo (4.1) é conhecido como GAMLSS.

O vetor γ_{jk} , para $j = 1, 2, \dots, j_k$ poderia ser combinado em um único vetor, γ_k , com uma única matriz \mathbf{Z}_{jk} ; entretanto, a formulação (4.1) aqui é preferida, porque facilita o processo computacional que utiliza o algoritmo *Backfitting* e permite combinações de diferentes tipos de termos aditivos de efeitos fixos a ser incorporado facilmente ao modelo.

Se, para $k = 1, 2, \dots, p$, $j_k = 0$, então o modelo (4.1) se reduz a um modelo paramétrico completo, dado por

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \beta_k \quad (4.2)$$

Se $\mathbf{Z}_{jk} = \mathbf{I}_n$, onde \mathbf{I}_n é uma matriz identidade $n \times n$ e $\gamma_{jk} = \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$, para todas as combinações de j e k , no modelo (4.1), fornece

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \beta_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}) \quad (4.3)$$

onde \mathbf{x}_{jk} , para todo $j = 1, 2, \dots, j_k$ e $k = 1, 2, \dots, p$ sendo vetores de tamanho n . A função h_{jk} é uma função desconhecida das variáveis explicativas \mathbf{X}_{jk} e $\mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ é um vetor no qual avalia a função h_{jk} em \mathbf{x}_{jk} .

Os vetores explanatórios \mathbf{x}_{jk} são assumidos conhecidos. O modelo da equação (4.3) é chamado GAMLSS *semiparamétrico*. Este modelo é um caso especial do modelo (4.1). Caso $\mathbf{Z}_{jk} = \mathbf{I}_n$ e $\gamma_{jk} = \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ para combinações específicas de j e k no modelo (4.1), então, o modelo resultante contém termos paramétricos, não-paramétricos e de efeitos aleatórios.

Os dois principais parâmetros populacionais, θ_1 e θ_2 , no modelo (4.1), são normalmente caracterizados como parâmetros de *locação* e *escala*, denotados por μ e σ , em que os parâmetros restantes, se existirem, são caracterizados como *parâmetros de forma*, embora o modelo possa ser aplicado mais geralmente para os parâmetros de qualquer distribuição da população.

Para muitas famílias de distribuição da população existem, no máximo, dois parâ-

metros de forma $\boldsymbol{\nu}(= \theta_3)$ e $\boldsymbol{\tau}(= \theta_4)$ e fornece o modelo completo

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1}\gamma_{j1} \quad (4.4)$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2}\gamma_{j2} \quad (4.5)$$

$$g_3(\boldsymbol{\nu}) = \boldsymbol{\eta}_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3}\gamma_{j3} \quad (4.6)$$

$$g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \mathbf{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4}\gamma_{j4} \quad (4.7)$$

onde $\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}$ e $\boldsymbol{\eta}_k$ são vetores de tamanho n , $\boldsymbol{\beta}'_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{J'_k k})$ é um vetor paramétrico de tamanho J'_k , \mathbf{X}_k é uma matriz projetada de tamanho $n \times J'_k$, \mathbf{Z}_{jk} é uma matriz projetada dos termos aditivos de tamanho $n \times q_{jk}$ e γ_{jk} é uma variável aleatória dimensional na qual é assumida como seguindo uma distribuição $\gamma_{jk} \sim \mathbb{N}_{q_{jk}}(\mathbf{0}, \mathbf{G}_{jk}^{-1})$, onde \mathbf{G}_{jk}^{-1} é a inversa generalizada de uma matriz simétrica $q_{jk} \times q_{jk}$, $\mathbf{G}_{jk} = \mathbf{G}_{jk}(\lambda_{jk})$, no qual pode depender de um vetor de hiperparâmetros λ_{jk} , e quando \mathbf{G}_{jk} for singular, então γ_{jk} se interpreta como sendo uma função de densidade imprópria proporcional a $\exp\left(-\frac{1}{2}\gamma'_{jk}\mathbf{G}_{jk}\gamma_{jk}\right)$ ⁶.

O GAMLSS da equação (4.1) é mais geral que o GLM, GAM, GLMM ou GAMM no qual a distribuição da variável dependente não é limitada à família exponencial e *todos* os parâmetros (não apenas a média) são modelados em termos de ambos efeitos fixos e aleatórios.

Na abordagem do GAMLSS, é imprescindível para a forma como os componentes aditivos são ajustados o algoritmo *backfitting* e o fato de que as penalidades quadráticas da verossimilhança resulta da consideração de um efeito aleatório normalmente distribuído no preditor linear. A estimação resultante usa matrizes reduzidas (suavizadas) que fazem parte do algoritmo (RIGBY & STASINOPOULOS, 2005).

A hipótese de dependência entre diferentes vetores γ_{jk} de efeitos aleatórios é essencial no GAMLSS. Entretanto, se para um valor particular para k , dois ou mais vetores de efeitos aleatórios não sejam independentes, eles podem ser combinados em um único vetor e sua matriz de correspondência, \mathbf{Z}_{jk} , é única, para que a condição de independência seja satisfeita.

Disto resulta que, para um valor de λ_{jk} fixo, os β_k s e os γ_{jk} s são estimados no

⁶Às vezes, é bem melhor se referir à \mathbf{G}_{jk} ao invés de $\mathbf{G}_{jk}(\lambda_{jk})$ por simplicidade notarial, embora a dependência de \mathbf{G}_{jk} , do hiperparâmetro λ_{jk} , permaneça.

GAMLSS pela maximização de uma função de verossimilhança penalizada, l_p , dada por

$$l_p = l - \frac{1}{2} \cdot \sum_{k=1}^p \sum_{j=1}^{j_k} \lambda_{jk} \gamma'_{jk} \mathbf{G}_{jk} \gamma_{jk} \quad (4.8)$$

onde $l = \sum_{i=1}^n \log f(y_i | \boldsymbol{\theta}^i)$ é uma função log-verossimilhança dos dados, dado $\boldsymbol{\theta}^i$, $\forall i = 1, 2, \dots, n$. Isso equivale à maximização da verossimilhança hierárquica ou estendida, definida por

$$l_h = l_p - \frac{1}{2} \cdot \sum_{k=1}^p \sum_{j=1}^{j_k} \left\{ \log |\mathbf{G}_{jk}| - q_{jk} \cdot \log(2\pi) \right\} \quad (4.9)$$

os hiperparâmetros $\boldsymbol{\lambda}$ podem ser fixados ou estimados. O l_p é alcançado utilizando o algoritmo CG, que pode ser visto no Apêndice B, de Rigby & Stasinopolous (2005).

4.3 Computação

O modelo da equação (4.3) pode ser ajustado pela minimização de

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \cdot \int_0^1 [f''(x)]^2 dx \quad (4.10)$$

onde a integral da segunda derivada penaliza os modelos que são muito *contorcidos*. O *trade-off* entre ajustamento de modelo e sua suavização é controlado pelo parâmetro de suavização, λ . Devido ao fato de $f(\cdot)$ ser linear nos parâmetros β_i , a penalidade pode sempre ser escrita como uma forma quadrática em β

$$\int_0^1 [f''(x)]^2 dx = \beta' \mathbf{S} \beta \quad (4.11)$$

onde \mathbf{S} é uma matriz de coeficientes conhecida. Assim, o problema de ajustamento da regressão *spline* penalizada se reduz a minimizar

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \cdot \beta' \mathbf{S} \beta \quad (4.12)$$

com relação a β , dado λ . A expressão formal para minimizar (4.12) é a estimação por mínimos quadrados penalizados de β_i , descrita abaixo:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{S})^{-1} \mathbf{X}'\mathbf{y} \quad (4.13)$$

A expressão (4.13) não é a única utilizada para computação. Para motivos práticos,

$$\left\| \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{X} \\ \sqrt{\lambda}\mathbf{B} \end{bmatrix} \beta \right\|^2 = \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \cdot \beta' \mathbf{S} \beta \quad (4.14)$$

onde \mathbf{B} é qualquer raiz quadrada da matriz \mathbf{S} tal que $\mathbf{B}'\mathbf{B} = \mathbf{S}$ e pode ser obtida facilmente pela decomposição espectral ou decomposição de Choleski; e, uma vez obtida, o problema de mínimos quadrados aumentados pode ser resolvido utilizando métodos ortogonais. A rotina usual Newton-Raphson é usada para maximizar log-verossimilhança nesses modelos e pode facilmente ser convertido em uma forma P-IRLS (mínimos quadrados penalizados iterativamente reponderados).

4.3.1 Algoritmo

São dois os algoritmos usados por Rigby & Stasinopoulos (2005) para maximizar a verossimilhança penalizada, como mostrado na equação (4.14). O primeiro deles, o algoritmo CG, é uma generalização do algoritmo de Cole & Green (1992), e usa o valor esperado (ou aproximado) da primeira e segunda derivadas cruzadas da função de verossimilhança com respeito ao parâmetro θ .

Entretanto, para as muitas funções de densidade de probabilidade da população $f(y|\theta)$, os parâmetros θ são informações ortogonais⁷, ou seja, são modelos de locação e escala, e pertencem à família de dispersão, ou aproximação dela (RIGBY & STASINOPOULOS, 2005).

Figura 3 – Algoritmo RS na sua forma mais simples

A simple version of the RS algorithm

- **Start:** Initialize fitted values $\theta_k^{(1,1)}$ for $k = 1, 2, 3, 4$ for distributional parameter vectors of length n , μ , σ , ν and τ respectively. Evaluate the initial linear predictors $\eta_k^{(1,1)} = g_k \left[\theta_k^{(1,1)} \right]$, for $k = 1, 2, 3, 4$.
- **START OUTER CYCLE** $r = 1, 2, \dots$ **UNTIL CONVERGENCE.**
FOR $k = 1, 2, 3, 4$
 - **START INNER CYCLE** $i = 1, 2, \dots$ **UNTIL CONVERGENCE .**
 - * Evaluate the current $\mathbf{u}_k^{(r,i)}$, $\mathbf{W}_{kk}^{(r,i)}$ and $\mathbf{z}_k^{(r,i)}$
 - * Regress the current $\mathbf{z}_k^{(r,i)}$ against design matrix \mathbf{X}_k using the iterative weights $\mathbf{W}_{kk}^{(r,i)}$ to obtain the updated parameter estimates $\beta_k^{(r,i)}$.
 - **END INNER CYCLE** on convergence of $\beta_k^{(r,\cdot)}$ and set $\beta_k^{(r+1,1)} = \beta_k^{(r,\cdot)}$, $\eta_k^{(r+1,1)} = \eta_k^{(r,\cdot)}$ and $\theta_k^{(r+1,1)} = \theta_k^{(r,\cdot)}$, otherwise update i and continue inner cycle.
- UPDATE** value of k

- **END OUTER CYCLE:** if the change in the (penalized) likelihood is sufficiently small, otherwise update r and continue outer cycle.

Fonte: Rigby & Akantziliotou (2009)

⁷Dado que os valores esperados das derivadas cruzadas da função de verossimilhança são zero.

Neste caso, o algoritmo RS na sua forma mais simples, o qual é uma generalização do algoritmo que foi utilizado por Rigby & Stasinopoulos (1996a, b) para ajustar a média e a dispersão em modelos aditivos (mas que não usam a derivação cruzada), é mais adequado. O parâmetro θ tem informação ortogonal completa somente para distribuições binomiais negativas, gama, Gaussiana inversa, logística e distribuições normais, como se pode vê-las na Tabela 1.

Tabela 1 – Distribuições do GAMLSS implementadas no R

Nº parâmetros	Distribuições
Discreto com um parâmetro	Poisson(PO), Positive.Poisson(PP), Geometric(GO), Logarithmic (LG), Yule (YU), Binomial (BI),
Contínuo com um parâmetro	Exponential (EX), Pareto (PA)
Discreto com dois parâmetros	Negative.Binomial.type.I (NB), Negative.Binomial.type.II (BN), Poisson.Inverse.Gaussian (PI), Beta.Binomial (BB)
Contínuo com dois parâmetros	Normal (NO), Gamma (GA), Inverse.Gaussian (IG), Gumbel (GU), Reverse.Gumbel (RG) Logistic (LO), Log.Logistic (LL), Weibull (WE), Box.Cox (BC)
Discreta com três parâmetros	Sichel (SI)
Contínuo com três parâmetros	Cole.Green (i.e. Box-Cox Normal) (CG) Generalized.Gamma.Family (GG) Exponential.Power.Family (EP) <i>t</i> .Family (TF), Generalized.Extreme.Family (GE)
Contínuo com quatro parâmetros	Generalized. <i>t</i> .Family (i.e. Box-Cox <i>t</i>) (GT)

Fonte: Adaptado de Akanztiliotou, Rigby & Stasinopoulos (2002)

4.3.2 Função no R para ajustamento do GAMLSS

A função de densidade de probabilidade da população, $f(y|\theta)$, é deliberadamente deixada geral, sem a forma da distribuição condicional explícita para a variável resposta y . A única restrição que tem a implementação do R para o GAMLSS, para especificar a distribuição de y , é que a função $f(y|\theta)$ e, opcionalmente o valor esperado da primeira e segunda derivadas cruzadas, com respeito a cada parâmetro θ , seja computável (Stasinopoulos *et al.*, 2004). As derivadas explícitas são preferíveis, mas as derivadas numéricas podem ser utilizadas, resultando no aumento da eficiência computacional.

Para algumas das distribuições que são apresentadas na Tabela 1, mais de uma parametrização foi implementada. Rigby & Stasinopoulos (2005) utilizam a notação $y \sim \mathcal{D}\{g_1(\theta_1) = t_1, g_2(\theta_2) = t_2, \dots, g_p(\theta_p) = t_p\}$ para identificar unicamente um GAMLSS, onde \mathcal{D} é a distribuição da variável resposta, $(\theta_1, \theta_2, \dots, \theta_p)$ sendo os parâmetros de \mathcal{D} , (g_1, g_2, \dots, g_p) são funções-*link* e (t_1, t_2, \dots, t_p) são a forma dos modelos para os termos exploratórios e/ou efeitos aleatórios nos preditores $(\eta_1, \eta_2, \dots, \eta_p)$, respectivamente.

Como exemplo, $y \sim TF\{\mu = cs(x, 3), \log(\sigma) = x, \log(\nu) = 1\}$, mostra um modelo onde a variável resposta, y , tem um distribuição t , com o parâmetro de locação μ , usando uma função-*link*, como uma *spline* cúbica com três graus de liberdade efetivos em x , no topo do termo linear x , isto é, $cs(x, 3)$; o parâmetro de escala, σ , usando um modelo log-linear em x , e uma distribuição t tendo ν graus de liberdade, usando um modelo com constante denotado por 1, porém em escala log (RIGBY & STASINOPOULOS, 2005).

Figura 4 – Sintaxe da função `gamlss()` no R

```
gamlss(formula = formula(data), sigma.formula = ~1,
       nu.formula = ~1, tau.formula = ~1, family = NO(),
       data = sys.parent(), weights = NULL,
       contrasts = NULL, method = RS(), start.from = NULL,
       mu.start = NULL, sigma.start = NULL,
       nu.start = NULL, tau.start = NULL,
       mu.fix = FALSE, sigma.fix = FALSE, nu.fix = FALSE,
       tau.fix = FALSE, control = gamlss.control(...),
       i.control = glim.control(...), ...)
```

Fonte: Stasinopoulos, Rigby & Akantziliotou (2012)

Assim, no R, a função que será utilizada para a classe dos modelos aditivos generalizados para locação, escala e forma é `gamlss()`, conforme se pode verificar a sintaxe na Figura (4) e que é utilizada para ajustar um GAMLSS, cujos argumentos mais importantes são, primeiro, **formula** o qual especifica o modelo para o parâmetro de locação, μ , **sigma.formula**, **nu.formula** e **tau.formula**, como argumentos opcionais que especificam os modelos para os parâmetros σ , ν e τ , apropriados. Outro argumento especial também é **family** que identifica a distribuição, de acordo com o mostrado na Tabela 1. A função **summary** é usada após a iteração para resumir os resultados provenientes do ajustamento do GAMLSS.

Pode-se perceber claramente que a Tabela 1 fornece uma seleção ampla de distribuições que podem ser escolhidas, porém, adicionalmente, é possível fazer a escolha própria da distribuição, caso o pesquisador tenha informação antecipada, de acordo com a experiência, ou mesmo que haja respaldo na teoria econômica para tal. Para mudar este argumento, basta que se modifique uma das famílias de funções existentes.

5 DINÂMICA DE UMA ECONOMIA

$F = m \cdot \vec{a}$, um corpo sob dinâmica.

$\dot{\hat{k}}/\hat{k} = s \cdot A \cdot (\hat{k})^{-(1-\alpha)} - (x + n + \delta)$, o capital sob dinâmica de transição, numa economia.

Os conceitos de convergência de renda, definidos na seção 2.2, têm sido utilizados para explicar como as diferentes economias tendem a atingir seu equilíbrio de longo prazo ou equalizar seu nível de desenvolvimento econômico. Ou seja, o modo como as regiões mais atrasadas tendem, de acordo com certas circunstâncias, a crescer mais rápido que as ricas, a fim de diminuir o hiato entre elas. A este fenômeno de superação dos considerados mais pobres é denominado convergência, e é desta dinâmica que estaremos tratando.

Isto é, conforme as variáveis específicas selecionadas para compor o modelo, as estimativas irão comprovar, ou não, como e o quanto cada região estudada precisaria de intervenção externa para alcançar seu *steady state* e em que velocidade.

Especialmente, iremos utilizar uma das economias regionais mais peculiares do Nordeste do Brasil, a do Estado de Alagoas – AL, para estudar o fenômeno da convergência de renda. E, de posse dos resultados, analisaremos como os efeitos de cada uma das variáveis impacta no desenvolvimento econômico, inclusive dando sugestões de como o Estado, como agente externo, pode diminuir o *gap* em relação a outras economias mais avançadas regionalmente, a partir de estímulos e atuação direta e indireta nas causas que impactam às variáveis causadoras de diminuição do desenvolvimento.

5.1 Um GAMLSS de crescimento econômico

Vamos reconsiderar a especificação econométrica descrita na equação (2.4). Ela está de acordo com o modelo teórico de Solow-Swan, apresentada em Barro & Sala-I-Martin (2004).

$$\frac{1}{T} \cdot \log\left(\frac{y_{it}}{y_{i0}}\right) = \alpha_0 + \alpha_1 \cdot \log(y_{i0}) + \Theta' \mathbf{X}_{i0} + \epsilon_{i0,t} \quad (5.1)$$

Como discutido acima, entretanto, a natureza deste modelo se fragiliza, caso seja tratada na classe de modelos lineares, porque a matriz de variáveis de controle, \mathbf{X}_{i0} , a qual mantém constante o estado estacionário, é composta de variáveis que não detém

causalidade direta entre si, e entre elas e a variável preditora. Esta covarição dentro da matriz é imprescindível para que a informação seja extraída.

Seria bem forçoso imaginar variáveis como escolaridade média e tamanho do governo, por exemplo, como poderia pressupor um modelo linear de crescimento econômico, apresentar correlação, covariação e causar, de forma direta, a variável preditora. Ao invés disso, é bem razoável que cada variável *adicione* informação e, em seu conjunto, promova *causação*¹ à variável dependente.

O modelo descrito pela equação (5.1), pode ser adaptado a um GAMLSS, modelo que permite associar tanto partes lineares e/ou aditivas, na mesma estrutura. A matriz \mathbf{X}_{i0} pode ser desmembrada para dar lugar à soma de vetores-coluna de variáveis

$$\frac{1}{T} \cdot \log\left(\frac{y_{it}}{y_{i0}}\right) = \alpha_0 + \alpha_1 \cdot \log(y_{i0}) + \mathbf{X}_{1t} + \cdots + \mathbf{X}_{nt} + \epsilon_{i,t} \quad (5.2)$$

Após a consideração da aditividade, a equação 5.1 se reduz à 5.2, recaindo à forma

$$\frac{1}{T} \cdot \log\left(\frac{y_{it}}{y_{i0}}\right) = \underbrace{\alpha_0 + \alpha_1 \cdot \log(y_{i0})}_{\text{parte linear}} + \underbrace{f(\mathbf{X}_{1t}) + \cdots + f(\mathbf{X}_{nt})}_{\text{parte aditiva}} + \epsilon_{i,t} \quad (5.3)$$

E, por fim, ao modelo abaixo

$$g_{i;0,T} = \alpha_0 + \alpha_1 \cdot \log(y_{i0}) + \sum_{k=1}^n f_k(\cdot) + \epsilon_{i,t} \quad (5.4)$$

onde $g_{i;0,T} = T^{-1} \cdot \log(y_{it}/y_{i0})$ e $f_k(\cdot)$ são as variáveis covariadas explicativas do crescimento econômico da região em análise.

5.1.1 Convergência de renda via GAMLSS

O teste vai utilizar, para o período que compreende de 2000 a 2009, os dados de 102 municípios do Estado de Alagoas. A maioria dos dados, como densidade demográfica, indicador de infraestrutura, a *proxy* para o tamanho do governo, a produção industrial municipal (VA) e os produtos internos brutos² dos municípios, foi fornecida por banco de dados oficiais e seguem de acordo com as mesmas especificações descritas em Oliveira & Monasterio (2011).

A variável escolaridade média foi fornecida pelas séries históricas do IBGE, enquanto que a taxa de homicídio foi extraída do banco de dados do IPEAdata; a primeira faz parte da seleção clássica sugerida nos trabalhos sobre convergência e a segunda por ter a peculiaridade de influenciar outras como densidade demográfica e potencial de mercado.

¹Não aplicaremos, neste texto, teste de causalidade de Granger.

²Devidamente deflacionados pelo deflator implícito do PIB fornecido pelo IPEAdata, para os anos de 2000 e 2009.

A partir da contribuição de Harris (1954) à economia regional, foi fornecida a variável potencial de mercado, cuja demanda potencial é representada pelo produto interno bruto *per capita* calculado para cada município, de acordo com a metodologia a seguir:

$$\text{POTMERC}_{in} = \begin{cases} \sum_n \frac{\text{PIB per capita}_{i0}}{d_i^2} & , n \leq (i - 1) \\ 1 & , \text{caso contrário} \end{cases} \quad (5.5)$$

no qual d_i é a distância do município i à capital do Estado. Incluindo todas as variáveis logaritmizadas no modelo de convergência condicional na equação (5.4), este apresentará a especificação:

$$g_{i;0,T} = \alpha_0 + \alpha_1 \cdot \log(y_{i0}) + \sum_{k=1}^7 f_k(\cdot) + \epsilon_{i,t} \quad (5.6)$$

sendo f_1 : escolaridade média, f_2 : taxa de criminalidade, f_3 : produção industrial municipal; f_4 : densidade demográfica, f_5 : potencial de mercado; f_6 : participação do governo e f_7 : indicador de infraestrutura.

O modelo *prático* que será lançado no console do **R** deverá seguir a especificação:

```
gamlssDiss10 = gamlss(g ~ cs(PIBpc2000Def, df= 1) + cs(ESCMEDIA,
df= 3) + cs(CRIME, df= 3) + cs(POTMERC, df= 3) + cs(INFRAENERGIA,
df= 3) + cs(DENSIDEM, df= 5), family = NO(), data=gamlss4)
```

Este modelo segue o que supõe a equação (5.6), pois a indicação $df = 1$ exclui a possibilidade do uso de uma *spline* cúbica e vai testar a relação linear existente entre a taxa de crescimento, g , e o PIB *per capita* do período inicial, tudo de acordo com o que prescreve as hipóteses que permeiam o modelo de Solow-Swan.

Foi feita uma cuidadosa análise estatística de dispersão em cada uma das variáveis e se percebeu dispersão homogênea em todas, permitindo supor, com certo grau de cautela, que a elas se deveria impor $df = 3$. Foi reconhecido, na análise gráfica, que a variável DENSIDEM tem uma peculiaridade (geográfica) teórica para a economia em estudo e, portanto, sugeriu-se que se deveria impor $df = 5$, já que o território apresentou densidade esparsa, com uma parte concentrada no litoral, na capital do estado e adjacências.

O comando para chamar a função GAMLSS que ajusta o modelo anterior, o seu carregamento e o lançamento do modelo descrito acima, no console do **R**, segue conforme abaixo:

```
> library(gamlss)
Carregando pacotes exigidos: splines
Carregando pacotes exigidos: gamlss.dist
Carregando pacotes exigidos: MASS
Carregando pacotes exigidos: gamlss.data
```

```

Carregando pacotes exigidos: nlme
*****      GAMLSS Version 4.2-0      *****
For more on GAMLSS look at http://www.gamlss.org/
Type gamlssNews() to see new features/changes/bug fixes.

> gamlssDiss10 = gamlss(g ~ cs(PIBpc2000Def, df= 1) + cs(ESCMEDIA, df= 3)
+ cs(CRIME, df= 3) + cs(POTMERC, df= 3) + cs(INFRAENERGIA, df= 3) +
  cs(DENSIDEM, df= 5),family = NO(), data=gamlss4)

GAMLSS-RS iteration 1: Global Deviance = 175.9561
GAMLSS-RS iteration 2: Global Deviance = 175.9626
GAMLSS-RS iteration 3: Global Deviance = 175.9799
GAMLSS-RS iteration 4: Global Deviance = 175.9957
GAMLSS-RS iteration 5: Global Deviance = 176.0072
GAMLSS-RS iteration 6: Global Deviance = 176.0149
GAMLSS-RS iteration 7: Global Deviance = 176.0197
GAMLSS-RS iteration 8: Global Deviance = 176.0227
GAMLSS-RS iteration 9: Global Deviance = 176.0245
GAMLSS-RS iteration 10: Global Deviance = 176.0256
GAMLSS-RS iteration 11: Global Deviance = 176.0262

```

Com uma competência impressionante, em termos de desempenho computacional, o GAMLSS atingiu a convergência com apenas 11 iterações, com um *Global Deviance*³ = 176.0262.

Vamos proceder ao uso de uma função, *prof.dev*, que fornece um gráfico com o perfil do desvio de qualquer um dos parâmetros de distribuição, μ , σ , ν e τ , de uma família de ajustamento e que é útil para checar a confiabilidade do modelo, nos quais um (ou mais) parâmetros nas distribuições sejam constantes. As Figuras 7 e 8 mostram o comportamento dos perfis dos desvios com um e dois *steps*, respectivamente.

```

> pdDissgamlss = prof.dev(gamlssDiss10, "mu", min=2, max=10, step=2)

*****
mu.start=( 2 )
GAMLSS-RS iteration 1: Global Deviance = 406.7957
GAMLSS-RS iteration 2: Global Deviance = 406.7957
*****
mu.start=( 4 )
GAMLSS-RS iteration 1: Global Deviance = 307.9639
GAMLSS-RS iteration 2: Global Deviance = 307.9639

```

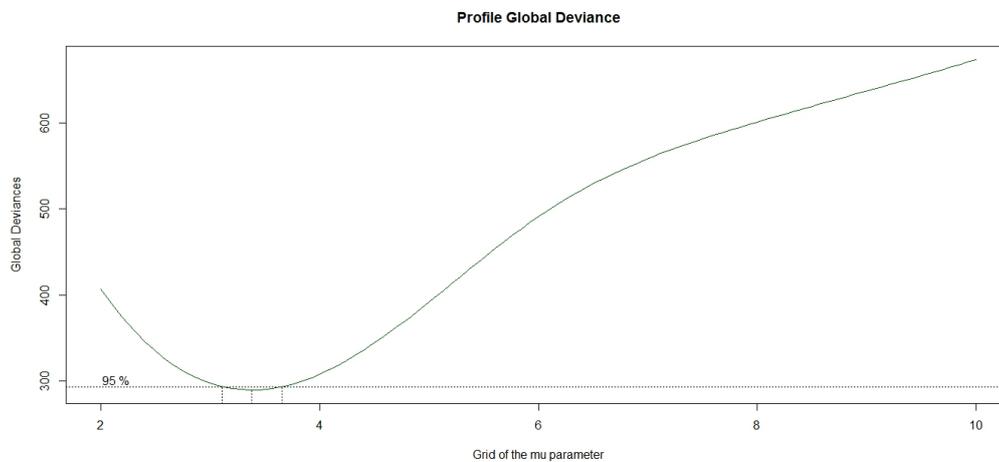
³O desvio (ou *Deviance*), também conhecido como a *estatística da razão de verossimilhança*, para um modelo ajustado $\hat{\mu}$ é definido por $D(\mathbf{y}; \hat{\mu}) = 2\{l(\boldsymbol{\mu}_{max}; \mathbf{y}) - l(\hat{\boldsymbol{\mu}}; \mathbf{y})\}$, com $\boldsymbol{\mu}_{max}$ sendo o valor do parâmetro que maximiza $l(\boldsymbol{\mu}; \mathbf{y})$ sob todos os valores de $\boldsymbol{\mu}$, situação conhecida como *modelo saturado*. O *desvio* toma o papel da soma dos quadrados dos resíduos para modelos generalizados, e pode ser utilizado para avaliar o quão bom é o ajuste do modelo, e também para comparação entre eles (HASTIE & TIBSHIRANI, 1990).

```

*****
mu.start=( 6 )
GAMLSS-RS iteration 1: Global Deviance = 491.3643
GAMLSS-RS iteration 2: Global Deviance = 491.3643
*****
mu.start=( 8 )
GAMLSS-RS iteration 1: Global Deviance = 601.3472
GAMLSS-RS iteration 2: Global Deviance = 601.3472
*****
mu.start=( 10 )
GAMLSS-RS iteration 1: Global Deviance = 673.8572
GAMLSS-RS iteration 2: Global Deviance = 673.8572
*****
*****
The Maximum Likelihood estimator is 3.374638
with a Global Deviance equal to 289.4795
A 95 % Confidence interval is: ( 3.107626 , 3.652258 )
*****

```

Figura 5 – Perfil do desvio global para dois *steps*



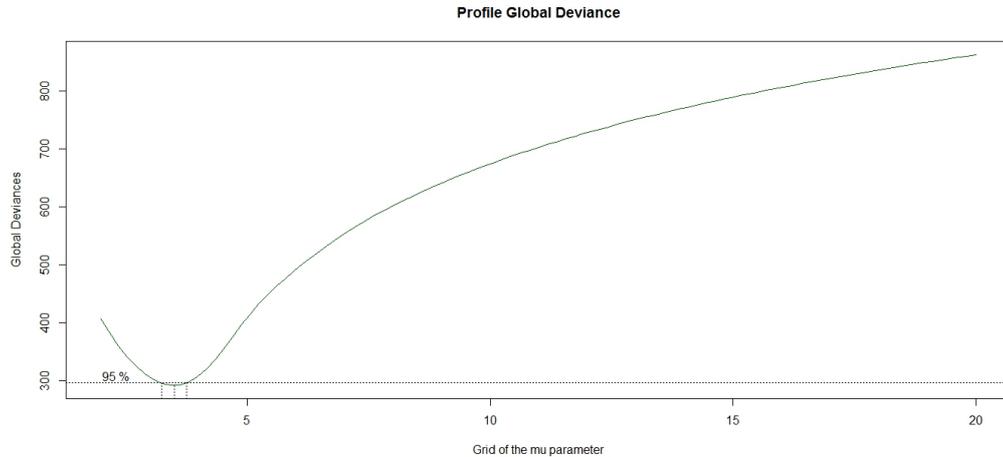
Fonte: elaboração do autor.

Com o pacote `gamlss`, do **R**, é possível utilizar mais duas funções, a `plot()` e `wp()`, que também servem como ferramentas de diagnóstico de modelos. Neste caso, elas se utilizam dos resíduos do modelo ajustado pelo `gamlss()`, a partir de *quantis* normalizados aleatoriamente.

A função `plot()`⁴, pode assim ser chamada já que seu primeiro argumento é um objeto ajustado via `gamlss`, produz um mosaico com quatro gráficos para checar os quantis residuais normalizados aleatoriamente, sendo esta randomização somente procedente para variáveis-resposta discretas.

⁴Cujo nome real é `plot.gamlss`.

Figura 6 – Perfil do desvio global para um *step*



Fonte: elaboração do autor.

Percebe-se, com os resultados apresentados na chamada da função, que os resíduos do modelo mostram um bom comportamento. Isto se conclui à medida que a média se mostra praticamente nula (zero), sua variância é aproximadamente um, porém se apresentou um coeficiente de assimetria muito pequeno e um índice de momento de curtose superior a quatro, mostrando concentração em torno da média e pouca variação, além de a distribuição apresentar um *pico*, o que neste caso específico vai demonstrar o quão *pesadas* são as caudas da distribuição, afastando uma convergência em distribuição para uma Normal.

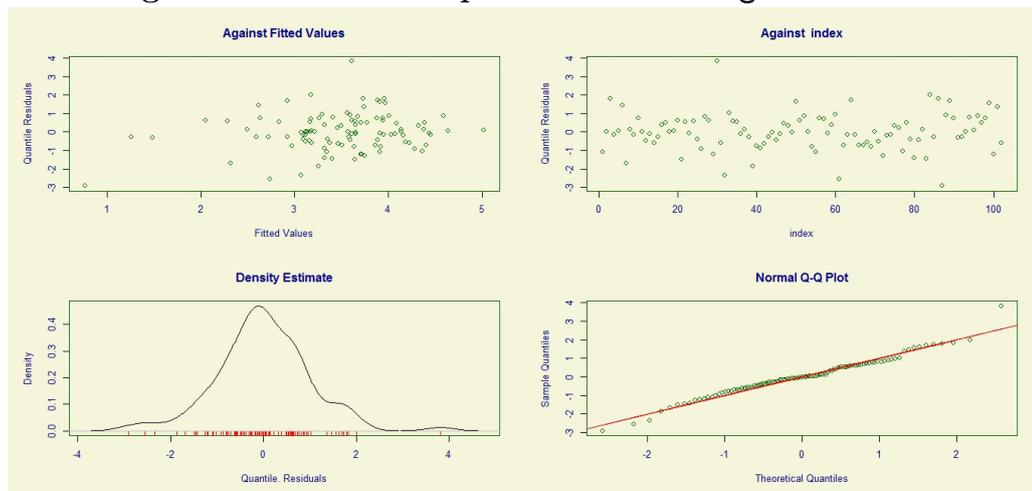
```
> plot(gamlssDiss10)
> plot(gamlssDiss10, ts=TRUE)
> wp(gamlssDiss10)
```

```
*****
Summary of the Quantile Residuals
      mean = 6.468648e-12
      variance = 1.009901
      coef. of skewness = 0.1885691
      coef. of kurtosis = 4.677698
Filliben correlation coefficient = 0.9841803
*****
```

A função `wp()`, *worm plot* dos resíduos, foi primeiramente introduzida por Buuren *et al.* (2001) e permite identificar regiões (intervalos) das variáveis explicativas nas quais o modelo não ajusta adequadamente aos dados, chamado de *violação do modelo*.

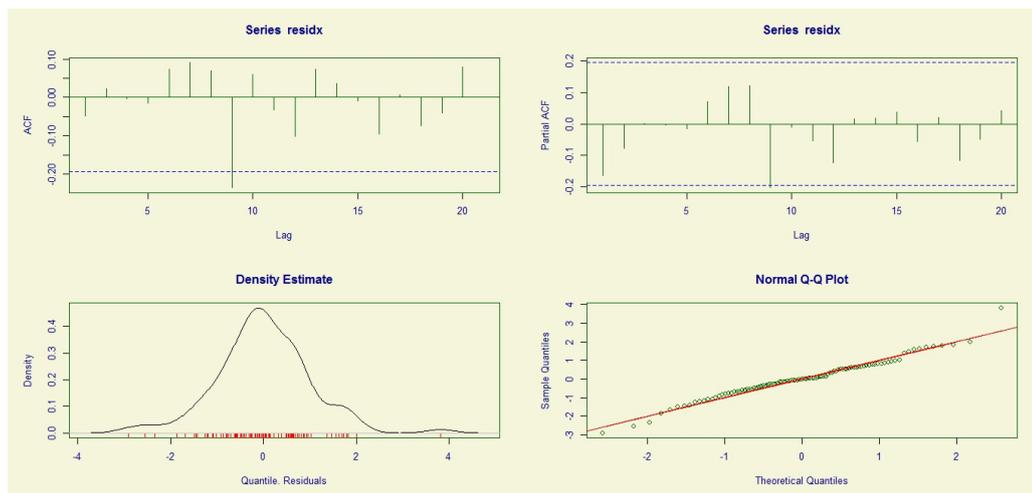
Esta função também é uma ferramenta de diagnóstico para se fazer a verificação para diferentes alcances da variável exploratória. Na Figura 7, dado que não foi especi-

Figura 7 – Resíduos a partir do modelo `gamLssDiss10`



Fonte: elaboração do autor.

Figura 8 – Resíduos a partir do modelo `gamLssDiss10`, utilizando a opção `ts=TRUE`



Fonte: elaboração do autor.

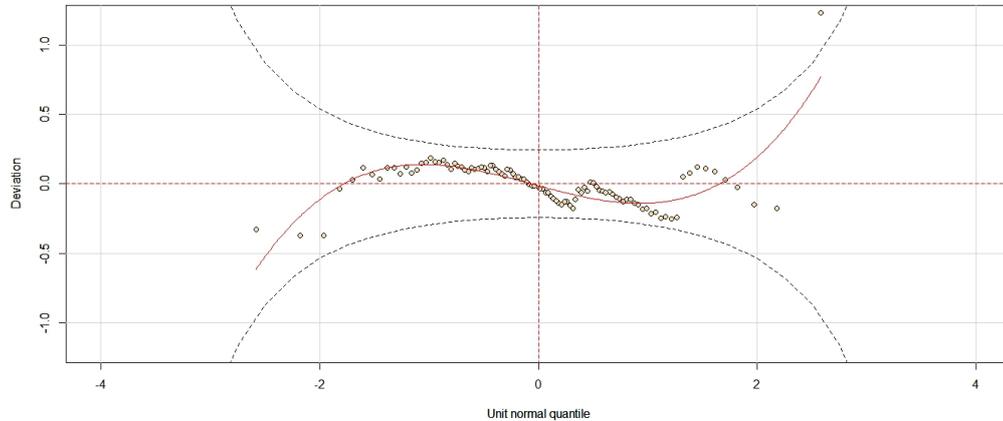
ficado nenhum argumento na chamada da função `wp()` somente um único *worm plot* de todos os resíduos foi plotado.

Observa-se que a quase totalidade dos resíduos caíram na região de aceitação, que são as duas curvas elípticas, excetuando-se um, localizada no primeiro quadrante e que se trata de um *outlier*, provavelmente indicando alguma característica concernente à capital do Estado, mas nada que prejudique o bom ajustamento do modelo. A curva que passa pela nuvem de dispersão é uma polinomial cúbica ajustada.

A função `summary()` resume o ajuste do objeto `gamLss`, mostrando as estimativas da equação (5.6), traduzido pelo modelo *prático* inserido no console do **R**, fornecendo os resultados para a hipótese da convergência condicional.

```
> summary(gamLssDiss10)
```

Figura 9 – *Worm plot* do modelo `gamlssDiss10` com valores padrão



Fonte: elaboração do autor.

The following object(s) are masked from 'gamlss4':

g, CRIME, DENSIDEM, ESCMEDIA, INDPARTVA, INFRAENERGIA, PARTGOV,
PIBpc2000Def, PIBpc2009Def, POTMERC

Family: c("NO", "Normal")

Call: `gamlss(formula = g ~ cs(PIBpc2000Def, df=1) + cs(ESCMEDIA, df=3) +
cs(CRIME, df=3) + cs(POTMERC, df=3) + cs(INFRAENERGIA, df=3) +
cs(DENSIDEM, df=5), family = NO(), data = gamlss4)`

Fitting method: RS()

Mu link function: identity

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.300667	0.70848	10.3047	4.413e-16
cs(PIBpc2000Def, df = 1)	-0.683592	0.08939	-7.6475	5.193e-11
cs(ESCMEDIA, df = 3)	1.048358	0.30619	3.4239	9.979e-04
cs(CRIME, df = 3)	-0.008329	0.03998	-0.2083	8.355e-01
cs(POTMERC, df = 3)	-0.254971	0.04671	-5.4584	5.800e-07
cs(INFRAENERGIA, df = 3)	0.015606	0.02989	0.5221	6.031e-01
cs(DENSIDEM, df = 5)	-0.052950	0.08405	-0.6300	5.306e-01

Sigma link function: log

Sigma Coefficients:

Estimate	Std. Error	t value	Pr(> t)
-5.561e-01	5.748e-02	-9.674e+00	6.892e-15

No. of observations in the fit: 102

Degrees of Freedom for the fit: 26.00143

Residual Deg. of Freedom: 75.99857

```

                                at cycle: 11
Global Deviance:                176.0262
                                AIC:      228.0291
                                SBC:      296.2821
*****
Mensagens de aviso perdidas:
In vcov.gamlss(object, "all") :
  additive terms exists in the mu formula standard errors for
the linear terms maybe are not appropriate
-----

```

Em relação aos resultados exibidos pela saída do `gamlss`, o primeiro aspecto que deve ser observado é a ausência de duas variáveis, a `PARTGOV` e a `INDPARTVA`. Inicialmente, o primeiro teste não requer que se especifique qualquer família de distribuição, deixando os dados mostrarem por si a qual delas se aproxima, dando o número de iterações parcial, próximo da convergência, e com isso sugerindo a família de distribuição.

Neste mesmo momento, o modelo faz o teste de quais variáveis, de fato, irão adicionar informação para a estimação. Essas duas variáveis foram desconsideradas aos menores níveis de AIC e BIC. Após este teste, ao se sugerir uma família ampla, de acordo com as mostradas na Tabela (1), as iterações continuam e chega-se à convergência, completando-se o ciclo da estimação.

O segundo aspecto vai para o fato de os *sinais* se comportarem de acordo com a teoria econômica do crescimento. Vê-se, claramente, que o sinal da variável `PIBpc2000Def` obedeceu o que espera a hipótese de convergência com a relação inversa entre aquela variável e g , a taxa de crescimento no período estudado.

As variáveis aditivas também se comportaram como esperado: a `ESCMEDIA` com sinal positivo indicando sua contribuição positiva ao crescimento e desenvolvimento econômico de uma região, a `CRIME`, como esperado, implicando negativamente o crescimento econômico do período estudado, o `POTMERC` exibindo como o PIB *per capita* ainda está disperso no território, porém seguindo em direção aos municípios mais importantes, em detrimento dos demais; a `INFRAENERGIA` sinalizando positivamente caso mais incentivos fossem dado para aumentar a produtividade nas atividades aceleradoras de crescimento econômico e, por fim, a `DENSIDEM` com sinal negativo podendo indicar queda contínua da população por m^2 , pois a área territorial permaneceu constante no período estudado.

5.2 Resultados empíricos

A Tabela 2 vai mostrar os resultados das estimações de três modelos: o GAMLSS, o modelo aditivo generalizado (GAM) e do modelo linear (MCRL), cujos métodos de estimação são feitos via P-IRLS, IRLS e OLS, respectivamente.

Destaca-se, nesta síntese, a distante diferença entre as velocidades de convergência fornecida pelo modelo GAMLSS, em comparação ao modelo linear, mas com a proximidade relativa ao GAM. Esta proximidade entre o GAMLSS e o GAM era esperada, já que os dois se utilizam de *splines* para proceder aos ajustes.

Com as variáveis aditivas $f(\text{PARTGOV})$ e $f(\text{INDPARTVA})$ não foi possível alcançar a convergência em 30 iterações e, por isso, o modelo sugeriu a exclusão dessas duas variáveis, sem prejuízo ao ajuste.

Outra questão a ser observada é a diferença no resultado das estimativas. Como explicado no texto, o GAM promove o melhor ajustamento possível e da forma mais suave, obtendo com isso maior *absorção* da informação dos dados, o que pode ser visto na primeira coluna da Tabela 2. Porém, como neste modelo é necessário impor previamente uma distribuição na família exponencial, isto vai restringir o caminho trilhado pela *splines*, o que comprometeu, em certos casos, o resultado da estimativa, tanto quanto seu sinal, da mesma forma que ocorreram equívocos com os resultados do modelo linear.

É preciso atenção redobrada quando se trata de levar em consideração uma estimativa por um modelo errado, principalmente do ponto de vista prático, porque os resultados podem nos levar a conclusões equivocadas e, o pior, a tomar decisões levando por base parâmetros vindos de resultados errados. Por exemplo, a estimativa de $f_1(\text{ESCMEDIA}) = -0.520$, no GAM, sugerindo que o indicador de educação teria uma relação inversa com a taxa de crescimento, o que teoricamente seria um contrassenso, já que se reconhece que este indicador se comporta exatamente ao contrário do que mostra o resultado. Por tanto, dentre as incompletudes desses modelos, este fato mostrado é um dos inconvenientes.

Ainda com relação à Tabela 2, é interessante notar o \bar{R}^2 no GAMs é menor que no modelo linear. Uma análise desavisada poderia chegar a conclusão que o primeiro modelo não ajustou tão bem quanto se pensava, porém, a explicação no modelo aditivo se faz de forma conjunta e, principalmente, em função do *desvio explicado*, que conta com 98.6%. Isto implica dizer que em relação à *banda de ajuste*, no caminho trilhado pela *Spline* no gráfico, apenas 1.4% não contribui para explicação do fenômeno. Assim, este é o fato que atesta a confiabilidade dos resultados. Já para o GAMLSS, a explicação do ajuste se dá de maneira conjunta, a partir dos resultados da Tabela 3 e dos gráficos das Figuras 7, 8 e 9, mostrados anteriormente.

Tabela 2 – Resultados das regressões para convergência condicional

Variável	Estimativas (GAM)	Variável	Estimativas (MCRL)	Variável	Estimativas (GAMLSS)
Constante	0.75890 ($< 2e-16$)***	Constante	0.877109 ($< 2e-16$)***	Constante	7.30067 (4.413e-16)
α_1	-0.63915 (1.35e-11)***	α_1	-0.09711 ($< 2e-16$)***	α_1	-0.68359 (5.193e-11)
β_{IRLS}	0.20003	β_{OLS}	0.06786	β_{P-IRLS}	0.52093
f_1 (ESCMEDIA)	-0.52000 (0.22137)	ESCMEDIA	0.00861 (0.01220)*	f (ESCMEDIA)	1.04836 (9.979e-04)
f_2 (CRIME)	0.15046 (0.00206)**	CRIME	0.001159 (0.38280)	f (CRIME)	-0.00833 (8.355e-01)
f_6 (POTMERC)	0.01522 (0.39735)	POTMERC	0.00381 (0.03370)*	f (POTMERC)	-0.25497 (5.800e-07)
f_3 (INFRAENERGIA)	0.16755 (0.03624)*	INFRAENERGIA	0.00168 (0.21050)	f (INFRAENERGIA)	0.01560 (6.031e-01)
f_4 (PARTGOV)	-2.77640 (0.79049)	PARTGOV	-0.02590 (1.78e-07)***	f (PARTGOV)	— ***
f_5 (DENSIDEM)	0.03319 (0.12374)	DENSIDEM	0.00159 (0.55930)	f (DENSIDEM)	-0.05295 (5.306e-01)
f_2 (INDPARTVA)	0.12374 (0.77703)	INDPARTVA	0.01769 (1.25e-05)***	f (INDPARTVA)	— ***
\bar{R}^2	0.77703	\bar{R}^2	0.94820	Df do ajuste	26.00143
Deviance explained	98.6%	F(8,93)	232	Residual df	75.99857
GCV score	0.00033	R^2 -Múltiplo	0.95230	Círculos de iteração	11
Scale est.	0.00018	RSE, 93df	0.01962	Global Deviance	176.0262
n	102	n	102	AIC	228.0291
		n	102	BIC	296.2821
		n	102	n	102

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

“” 1; p-valores entre parênteses; ‘***’

Fonte: elaboração do autor

Tabela 3 – Resultados da regressão para convergência absoluta

Variável	Estimativa GAM/MCRL	Variável	Estimativa GAMLSS
Constante	0.79428 ($< 2e-16$)***	Constante	5.5972 (1.244e-08)
α_1	-0.84706 ($< 2e-16$)***	α_1	-0.2814 (2.050e-02)
$\frac{\beta_{IRLS}}{\beta_{OLS}}$	0.22482	β_{P-IRLS}	0.24795
$\overline{R^2}$	0.86320	Df do ajuste	4.00022
F(1,100)	638.4	Residual df	97.99978
R^2 -Múltiplo	0.86460	Ciclos de iteração	2
RSE, 100df	0.03188	Global Deviance	229.6703
Deviance explained	86.5%	AIC	237.6708
GCV score	0.00103	BIC	248.1712
Scale est.	0.00101	—	—
n	102	n	102

Signif. codes: 0 “***” 0.001 “**” 0.01 “*” 0.05 “.” 0.1 “ ” 1

p -valores entre parênteses.

Fonte: elaboração do autor

A Tabela 3 exibe outra síntese de resultados da estimação, porém desta vez com o objetivo de comparar a convergência absoluta utilizando a modelagem GAMLSS, o GAM, como também o modelo clássico de regressão linear. Neste caso, é importante perceber que o modelo linear e o GAM apresentaram resultados em comum, pois a escolha do modelo aditivo seguiu a teoria econômica no que se refere a possibilidade de haver relação inversa entre a taxa de crescimento do produto e o PIB *per capita* do período inicial. Enquanto no modelo linear segue a especificação do modelo econométrico da equação (5.1), isso se traduz no GAM como uma *Spline* com um grau de liberdade, sem nós (`degree = 1`, `df=1`).

Mesmo apresentando a Tabela 3, não se pode fazer uso desses resultados, pois não é razoável acreditar que o crescimento econômico dos municípios seja explicado somente pelo seu PIB inicial ou mesmo pelo seu estoque de capital. A Tabela 2 mostra resultados mais fidedignos dado que fez uso de variáveis próprias para explicar suas relações com a variável dependente.

Percebe-se, ainda na Tabela 3, que a β -convergência absoluta na primeira coluna mostrou uma velocidade de $\beta_{IRLS} = \beta_{OLS} = 0.22482$, enquanto que $\beta_{P-IRLS} = 0.24795$, tendo $\alpha_{1(GAM/MCRL)} = -0.84706 < -0.2814 = \alpha_{1(GAMLSS)}$, comprovando que mesmo em se tratando de uma teste *bruto* de convergência, a abordagem GAMLSS ainda promove maior velocidade, em comparação à métodos menos sofisticados. O mesmo aconteceu

com os resultados da velocidade de convergência condicional da Tabela 2, quando $\beta_{OLS} < \beta_{IRLS} < \beta_{P-IRLS} = 0.52093$, mesmo quando $\alpha_{1(OLS)} < \alpha_{1(GAMLSS)} < \alpha_{1(GAM)}$.

6 CONSIDERAÇÕES FINAIS

“Eu sei, meu caro Watson, que você compartilha minha paixão por tudo o que é bizarro e o que rompe com as convenções e a rotina monótona da vida diária.”

Sherlock Holmes (*A liga da cabeça vermelha*)

O principal objetivo desta pesquisa foi propor uma *nova experiência de modelagem e estimação*, a partir da utilização de uma *nova* modelagem econométrica, ferramenta esta que transita na classe de modelos aditivos, conhecido por Modelos Aditivos Generalizados para Localização, Escala e Forma (GAMLSS), cuja intenção é eliminar as incompletudes dos Modelos Aditivos Generalizados (GAM) e, em consequência, dos Modelos Lineares Generalizados (GLM), superando, assim, as rigidezes de duas das principais hipóteses de Gauss-Markov, quais sejam, a linearidade nos parâmetros e a normalidade do termo de erro no Modelo Clássico de Regressão Linear.

Este GAMLSS, com seu método próprio de estimação, o P-IRLS, desempenhou papel fundamental e forneceu as estimativas de forma iterativa, a partir de algoritmos específicos, o que garantiu grande competência em desempenho computacional, apresentando resultados fidedignos e confiáveis.

Toda rotina de programação foi possível devido a pacotes fornecidos pelos idealizadores do modelo, e que se utiliza de um programa estatístico de código aberto, conhecido como **R**, para *rodar* uma rotina que fornece, além dos resultados analíticos, gráficos precisos que ajudam no diagnóstico do modelo proposto e a visualizar vários resultados possíveis.

A ideia também era mostrar que para certas modelagens econométricas, devido à sua especificidade, a linear não mais se adequava. Isto serviu de base para a proposta de um modelo que pudesse, ao mesmo tempo, captar a maior quantidade de informações advindas dos dados, deixando que eles falassem por si, e também pudesse aderir à teoria econômica do crescimento, obedecendo aos preceitos das hipóteses básicas dos testes de convergência de renda, porém dando respostas mais robustas e mais precisas.

Os dados utilizados foram selecionados de acordo com a especificidade da economia em estudo, inclusive, empregando as sugestões da Nova Geografia Econômica e das Novas Teorias do Crescimento Econômico fazendo, assim, uso de variáveis de alto potencial explicativo do crescimento econômico, e também utilizando sugestões clássicas da economia regional.

A história do pensamento econômico já assinalava a necessidade de novas propostas de modelagem nesta área, tendo as primeiras manifestações acontecidas na final do século XX, porém ainda enveredando pelo *mundo* dos modelos lineares. Esta nova proposta feita neste trabalho de dissertação apareceu como alternativa ou complemento aos modelos lineares no mesmo período histórico, mas até os dias atuais, pouca ou nenhuma aplicação em testes na teoria econômica ainda foram apresentados, porém em outras áreas de pesquisa recentemente, principalmente no começo deste século XXI, o GAMLSS tem sido amplamente utilizado com sucesso.

Foi argumentado no texto que este tipo de modelagem pode conviver, e às vezes melhor se ajusta, com muitas situações em que a teoria econômica ainda não experimentou se utilizar, devido a barreiras impostas por certas hipóteses. Também foi deixado claro que o GAMLSS não elimina a utilização de modelos lineares em economia, mas é um complemento ou uma alternativa para situações em que não seja necessário levar em conta grandes quantidades de restrições nos modelos teóricos que serão testados empiricamente.

Para apresentar o GAMLSS foi preciso trilhar o caminho da *evolução* das classes de modelo em econometria, desde os modelos lineares e sua generalização, até a apresentação da ideia de *splines*, que é a base da modelagem aditiva, mostrada suas inconveniências e incompletudes, mas também a sofisticação dentro do GAM e GLM, com seus níveis de poder de predição e aplicação.

O modelo aditivo generalizado para locação, escala e forma foi apresentado no Capítulo 4 e foram mostradas todas as nuances do seu método específico de estimação, juntamente com seu ferramental analítico e computacional para se chegar às estimativas por métodos iterativos, como também a apresentação do programa de código aberto, o **R**, as rotinas nele implementadas, a partir de pacotes feitos pelos idealizadores do GAMLSS, seus algoritmos e funções usadas para o ajustamento.

No último capítulo foi proposto e discutido um GAMLSS de crescimento econômico a partir da especificação clássica de um modelo Solow-Swan. Além disso, foi apresentado um *modelo prático* de como inserir no *console* do **R** as linhas de programação que permitem *rodar* o modelo. Foi explicado como *chamar* os pacotes necessários, as funções que forneceram resultados e gráficos que permitiram análises e diagnósticos, além da discussão acerca dos resultados econométricos, os testes dos sinais e a aderência às hipóteses testadas.

A partir do resumo do ajuste pode-se inferir sobre algumas questões: o modelo proposto forneceu estimativas consistentes em relação ao crescimento econômico da economia estudada, revelando fatos, em relação às variáveis estudadas, para potenciais ajustes na condução de políticas de Estado que, quando feitos, poderão romper com a desigualdade de renda entre os municípios, abrindo caminho para o desenvolvimento particular de cada

localidade e, em consequência, da unidade regional por completo.

As estimativas são fornecidas em escala logarítmica, de acordo com a construção do modelo em *log-log*, de forma a facilitar a interpretação. Assim, quando a saída da estimativa informa que $PIBpc2000Def = -0.6835$, além de ter obedecido à hipótese de convergência condicional, para se chegar ao estado estacionário, esta economia o fará à taxa decrescente de 0.52, meio por cento, aproximadamente. Ou seja, existiram algumas situações na economia da região que dificultou o caminho percorrido pelos fatores básicos de produção, no período estudado, para que a economia alcance uma situação de desenvolvimento comum em todas as localidades.

Essas situações de dificuldade, em parte, podem ser atribuídas às variáveis inseridas no modelo e que forneceram estimativas com sinais negativos, enquanto que as outras com sinal positivo, exceto aquela da parte linear do modelo, informam que contribuíram positivamente e àquela taxa. As outras situações de dificuldade não puderam ser contempladas no modelo e fazem parte das variáveis exógenas de origem política, cultural e da estrutura do Estado e que precisaria de outro estudo para complementar a análise, mas que foge do escopo deste estudo, no momento.

É inegável que há uma relação positiva entre taxa de crescimento e algum indicador para o nível de educação. Tal hipótese foi comprovada quando a estimativa $ESCMEDIA = 1.048$, destacando a contribuição em, aproximadamente, 1.05% da escolaridade média ao crescimento econômico da região, e a estimativa $INFRAENERGIA = 0.015$ como os investimentos ainda eram precários, porém tinham grande potencial alavancador de crescimento. Em contrapartida, $POTMERC = -0.254$ implica que, no período estudado, havia má distribuição de renda em todo o território e concentração em pontos específicos, desfavorecendo o crescimento contínuo. Isto se acentua quando observamos $DENSIDEM = -0.053$ que pode refletir diminuição da força de trabalho, pessoas em idade produtiva, muito provavelmente pelo que exibe $CRIME = -0.008$, pois se sabe que no período em análise a maioria dos estados do Nordeste tem aumentado a taxa de homicídios de pessoas em idade laboral, e Alagoas tem sido fortemente afetada.

Mesmo com muitas dificuldades de atacar estes problemas, é possível uma ação integrada entre os municípios e o governo do Estado, com a ajuda da União, para viabilizar projetos de Estado em âmbito municipal, estabelecendo metas administrativas no que cabe como responsabilidade a cada ente estatal, no período de vigência de cada governo, dentro de uma ou mais legislaturas. Sabendo-se que dentro de dez anos, ou seja, duas legislaturas e meia, estas macrovariáveis atingem seu *estado estacionário* à taxa de 0.52, é possível idealizar uma engenharia de recursos para aplicação no fortalecimento da educação básica, pois é um dos ingredientes que mais fornecem retornos em produtividade, no médio e longo prazos. Está é a obrigação primordial dos Municípios.

Mesmo cabendo ao Estado a educação secundária, instrução técnica e formação superior, não é necessário demonstrar a viabilidade política e social de manter equilibrada estas três esferas dentro de uma rede que somem recursos próprios e da União para uma aplicação satisfatória e equilibrada durante o tempo de *maturação* do crescimento em um período idêntico ao estudado.

Vai caber ao Estado fornecer meios para que a população não seja estimulada a se deslocar de seu local de origem em busca de oportunidades em outros estados da federação ou mesmo em cidades maiores dentro do estado, como a capital, porque isto provoca mais desequilíbrio na variável que explica a perda de mercado, por falta de potenciais consumidores. Além disso, mais negligência no fornecimento da infraestrutura adequada para atração, fixação e consolidação de novos empreendimentos reais de alto valor agregativo de riqueza através da produção, e não somente de serviços e do setor público, fará estagnar a economia dos municípios mais frágeis.

Como visto no texto, duas variáveis escolhidas em princípio pela relevância teórica e que foram *descartadas* do modelo preocupam sobremaneira por terem mostrado insignificância *prática* na análise. O conjunto **INFRAENERGIA + POTMERC + DENSIDEM** de retornos enfraquecidos pode levar a crer que, de fato, não teria como haver alguma participação da indústria no crescimento do PIB, já que a quantidade relativa de indústrias instaladas no estado é muito pequena em relação a Estados com as economias uma pouco mais avançadas. Além disso, como o Estado é pouco desenvolvido, também o é na estrutura administrativa do Estado, com altos entraves burocráticos e gestões pouco produtivas que fazem com que a participação do governo também prejudique uma provável possibilidade de tirar a economia do trajeto atual.

Todas estas *sugestões* e comentários podem talvez não ser bem-vindas, mas é possível instigar a reflexão, mesmo porque a ferramenta mostrou a viabilidade técnica, socioeconômica e política de uma incursão séria nas variáveis clássicas de política pública para que o projeto do partido político, personalizado nos gestores municipais e estadual, venha a continuar no comando do poder executivo.

Salientamos, mais uma vez, que este projeto de pesquisa apenas inicia sua exploração com aplicações no campo da teoria econômica na sua completude e, mais precisamente, na teoria do crescimento e desenvolvimento agora de forma experimental e aplicada. Inclusive, nada impede de esta modelagem se coadunar com as tradicionais, porque seu diferencial é exatamente sua flexibilidade. É bem possível combinar o GAMLSS com aspectos espaciais e testes para detecção de clubes de convergência. Outro diferencial está no seu método de estimação que se torna eficiente pelo uso maciço de métodos computacionais com rotinas numéricas poderosas, o que promete se tornar mais uma ferramenta indispensável na pesquisa empírica em economia.

REFERÊNCIAS

- AITKIN, M. **A general maximum likelihood analysis of variance components in generalized linear models.** *Biometrics*, 55: 117–128. 1999.
- AGHION, Philippe. HOWITT, Peter. **A Model of Growth Through Creative Destruction.** *Econometrica*, 60, March, 323–351, 1992.
- AKANTZILIOTOU, K. RIGBY, R. A. and STASINOPOULOS, D. M. **The R implementation of Generalized Additive Models for Location, Scale and Shape.** *In*: Stasinopoulos, M. and Touloumi, G. (eds.), *Statistical modelling in Society: Proceedings of the 17th International Workshop on statistical modelling*, pp. 75–83. Chania, Greece. 2002.
- ANSELIN, L. **Spatial Econometrics: Methods and Models.** Dordrecht: Kluwer Academic, 284p. 1988.
- ANSELIN, L.; BERA, A. **Spatial dependence in linear regression models with an introduction to spatial econometrics.** *In*: ULLAH, A.; GILES, D. (Eds.). *Handbook of applied economic statistics*. New York: Marcel Dekker, 1998.
- ARROW, Kenneth J. **The Economic Implications of Learning by Doing.** *Review of Economic Studies*, 29, June, 155–173, 1962.
- AZZONI, C. R. **Crescimento econômico e convergência das rendas regionais: o caso brasileiro a luz da nova teoria do crescimento.** *Anais da ANPEC. Florianópolis.* v. 1, p. 185–205, 1994.
- BALTAGI, Badi H. **Econometrics**, 5ª edição. Springer-Verlag Berlin Heidelberg, 2011
- BARRO, Robert, SALA-I-MARTIN, Xavier. **Economic Growth.** 2.ed. Cambridge, Mass.: MIT, 2004. 608p.
- BATISTA DA SILVA, M. V.; SILVEIRA NETO, R. M. **Dinâmica da concentração industrial no Brasil entre 1994 e 2004: uma análise a partir de economias de aglomeração e nova geografia econômica.** *Economia Aplicada*, v. 13, n. 2, p. 299–331, abr./jun. 2007.
- BINGHAM, N.H. FRY, John M. **Regression – Linear Models in Statistics.** (Springer Undergraduate Mathematics). Springer-Verlag London Limited, 2010.
- BORGES, A. F. H. **O debate sobre a convergência de rendas per capita.** 1995. Disponível em: < http://www.cepe.ecn.br/conv_ne.pdf >.
- BRESLOW, N. E. CLAYTON, D. G. **Approximate inference in generalized linear mixed models.** *J. Am. Statist. Ass.*, 88, 9–25. 1993.
- BRESLOW, N. E. Lin, X. **Bias correction in generalized linear mixed models with a single component of dispersion.** *Biometrika*, 82, 81–91. 1995.

- BUUREN, S. Van. Fredriks, M. **Worm plot: a simple diagnostic device for modeling growth reference curves**. *Statistics in Medicine*, 20: 1259–1277. 2001.
- CASELA, George. BERGER, Roger L. **Inferência Estatística** – Tradução da segunda edição Americana. São Paulo: Cengage Learning, 2011.
- CASS, David. **Optimum Growth in an Aggregative Model of Capital Accumulation**. *Review of Economic Studies*, 32, July, 233–240, 1965.
- COLE, T. J. GREEN, P. J. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statist. Med.*, 11: 1305–1319. 1992.
- COSTA, Leticia Magalhães da. **Análise do processo de convergência de renda nos estados brasileiros: 1970–2005**. 2009. Dissertação (Mestrado). Escola de Pós-Graduação em Economia (EPGE), Fundação Getúlio Vargas, Rio de Janeiro, Rio de Janeiro-RJ, 2009. 48 f.
- COX, D.R. **Principles of Statistical Inference**. Cambridge University Press, 2006.
- DOMAR, Evsey D. **Capital Expansion, Rate of Growth, and Employment**. *Econometrica*, 14, April, 137–147, 1946.
- DOBSON, A. J. **An Introduction to Generalized Linear Models**, Second Edition. Chapman and Hall/CRC (November 2001), London, 2001.
- ELLERY JUNIOR, R. FERREIRA, P. C. **Convergência entre a renda per-capita dos estados brasileiros**. *Revista de Econometria*, v. 16, n. 1, p. 83–103, 1996.
- FERREIRA, A. H. B. DINIZ, C. C. **Convergência entre as rendas per capita estaduais no Brasil**. *Revista de Economia Política*, v. 15, n. 4, out./dez. 1995.
- FAHRMEIR, L. TUTZ, G. **Multivariate Statistical Modelling based on Generalized Linear Models**, 2nd ed. New York: Springer, 2001.
- FAHRMEIR, L. LANG, S. **Bayesian inference for generalized additive mixed models based on Markov random field priors**. *Appl. Statist.*, 50, 201–220. 2001.
- FRIEDMAN, M. **Do old fallacies ever die?** *Journal of Economic Literature*, v. 30, n. 4, p. 2.129–2.132, dez. 1992.
- FARAWAY, Julian James. **Linear models with R**. Chapman & Hall/CRC, 2005.
- FARAWAY, Julian James. **Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models**. Chapman & Hall/CRC, 2006.
- GREENE, William H. **Econometric analysis**, 7th ed. Pearson Education, Inc, 2012.
- GROLI, Priscila Albina. OLIVEIRA, Cristiano Aguiar de. JACINTO, Paulo Andrade. **Crescimento econômico e convergência com a utilização de regressões quantílicas: um estudo para os municípios do Rio Grande do Sul (1970–2001)**. In: ENCONTRO NACIONAL DE ECONOMIA - ANPEC, 34., 2006, Salvador, Bahia, 2006. Anais... Salvador: ANPEC, 2006.

- GROSSMAN, Gene M. HELPMAN, Elhanan. **Innovation and Growth in the Global Economy**. Cambridge, MA: MIT Press. 1991.
- GUJARATI, Damodar N. **Econometria Básica**, 3ª Edição. Makron: São Paulo, 2000.
- GUJARATI, Damodar N. **Econometria Básica**, 5ª Edição. McGraw Hill – Artmed: São Paulo, 2012.
- HARRIS, C.D. **The market as a factor in the localization of industry in the United States..** 1954.
- HARROD, Roy F. **An Essay in Dynamic Theory**. *Economic Journal*, 49, June, 14–33, 1939.
- HAYASHI, Fumio. FOX, Robert. GUINNESS, Ivor Grattan. **Econometrics**, 1st Edition. Princeton University. 2000.
- HAY-JAHANS, Christopher. **An R Companion to Linear Statistical Models**. CRC Press Taylor & Francis Group. 2012.
- HASTIE, T. TIBSHIRANI, R. J. **Generalized Additive Models**. Chapman & Hall, 1990.
- HASTIE, T. TIBSHIRANI, R. J. **Varying-coefficient models** (with discussion). *J. R. Statist. Soc. B*, 55, 757–796. 1993.
- HASTIE, T. TIBSHIRANI, R. J. **Bayesian backfitting**. *Statist. Sci.*, 15, 213–223. 2000.
- HASTIE, T. TIBSHIRANI, R. J. FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference and Prediction**. New York: Springer, 2001.
- HASTIE, T. **gam: Generalized Additive Models**. R package version 0.98.
- JOHNSON, P. TAKEYAMA, L. **Convergence among the US States: absolute, conditional or club?** Poughkeepsie: Department of Economics Vassar College, 2003. (Vassar College Economics Working Paper Series, 50).
- JONES, C. I. **R&D-Based Models of Economic Growth**. *Journal of Political Economy*, 103, August, 759–784., 1995.
- JONES, C. I. **On the evolution of world income distribution**. *Journal of Economic Perspectives*, v. 11, n. 3, p. 19–36, 1997.
- JONES, C. I. **Growth: With or Without Scale Effects**. *American Economic Review*, 89, May, 139–144, 1999.
- JONES, C. I. **Was an Industrial Revolution Inevitable? Economic Growth over the Very Long Run**. *Advances in Economics*, 1(2), Article 1. 2001.
- KNIGHT, Frank H. **Diminishing Returns from Investment**. *Journal of Political Economy*, 52, March, 26–47. 1944.
- KMENTA, Jan. **Elementos de Econometria: teoria estatística básica**, volume 1, Editora Atlas S. A.: São Paulo, 1994.

- KMENTA, Jan. **Elementos de Econometria: teoria econométrica básica**, volume 2, Editora Atlas S. A.: São Paulo, 1994.
- KOOPMANS, Tjalling C. **Stationary Ordinal Utility and Impatience**. *Econometrica*, 28, April, 287–309, 1960.
- KOOPMANS, Tjalling C. **On the Concept of Optimal Economic Growth**. *In: The Econometric Approach to Development Planning*. Amsterdam: North Holland, 1965.
- LEVINE, R. RENELT, D. **A Sensitivity Analysis of Cross-country Growth Regressions**. *American Economic Review*, v.82, p.942–963, 1992.
- LEE, Y. NELDER, J. A. **Hierarchical generalized linear models**, (with discussion). *J. R. Statist. Soc. B.*, 58: 619–678, 1996.
- LEE, Y. NELDER, J. A. **Two ways of modelling overdispersion in non-normal data**. *Appl. Statist.*, 49, 591–598, 2000.
- LEE, Y. NELDER, J. A. **Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions**. *Biometrika*, 88, 987–1006, 2001a.
- LEE, Y. NELDER, J. A. **Modelling and analysing correlated non-normal data**. *Statist. Modelling*, 1, 3–16. 2001b.
- LEE, Y. NELDER, J. A. **Double Hierarchical generalized linear models**, (with discussion). *Appl. Statist.*, 55: 139–185. 2006.
- LUCAS, ROBERT E. **On the Mechanics of Economic Development**. *Journal of Monetary Economics*, v.22, p.3–42, 1988.
- McCULLAGH, P. NELDER, J. A. **Generalized Linear Models**, 2nd ed. Chapman and Hall, London. 1989.
- MALTHUS, Thomas R. **An Essay on the Principle of Population**. London: W. Pickering (1986), 1798.
- NELDER, J. A. WEDDERBURN, R. W. M. **Generalized linear models**. *J. R. Statist. Soc. A.*, 135: 370–384. 1972.
- OLIVEIRA, C. **Externalidades espaciais e o crescimento econômico das cidades do estado do ceará**. *In: ENCONTRO REGIONAL DE ECONOMIA DO NORDESTE*, 2005, Fortaleza, CE. Anais. Fortaleza: Banco do Nordeste.
- OLIVEIRA, Carlos Wagner de Albuquerque. MONASTERIO, Leonardo (Orgs). **Dinâmica regional e convergência de renda: uma análise para os municípios brasileiros selecionados no período 2002–2007**. Brasília: Ipea, 2011. 270 p.
- PORTO JÚNIOR, S. S. **Dinâmica de crescimento e convergência de renda per capita no Brasil: o caso da região sul**. Porto Alegre, 2000.
- PEREIRA, Dilson. J. de Sena. **Um estudo para municípios alagoanos no período recente** *In* MONASTERIO, L. M. OLIVEIRA, C. W. A. (Orgs.) **Dinâmica regional e convergência de renda: Uma análise para os municípios brasileiros selecionados no período 2002-2007**. Brasília: IPEA, 2011. v. 1. 210 p.

- QUAH, D. T. **Galton's fallacy and tests of convergence hypothesis**. *Scandinavian Journal of Economics*, v. 95, n. 4, p. 427-443, 1993a.
- QUAH D. T. **Regional convergence clusters across Europe**. *European Economic Review*, v. 40, p. 951-958, 1996.
- QUAH, D. T. **Empirics for growth and distribution: stratification, polarization and convergence clubs**. *Journal of Economic Growth*, v. 2, n. 1, p. 27-59, Mar. 1997.
- REY, S. J., MONTOURY, B. D. **U.S. Regional income convergence: a spatial econometric perspective**. *Regional Studies*, v.33, n.2, p.143-156, Apr.1999.
- REY, S. J. **Spatial empirics for economic growth and convergence**. *Geographical Analysis*, v. 33, n. 3, p. 195-214, 2001.
- RAMSEY, Frank. **A Mathematical Theory of Saving**. *Economic Journal*, 38, December, 543-559, 1928.
- RENCHER, Alvin C. SCHAALJE, G. Bruce. **Linear models in statistics**, second edition. John Wiley & Sons, Inc., 2008.
- RICARDO, David. **On the Principles of Political Economy and Taxation**. Cambridge: Cambridge University Press (1951), 1817.
- ROMER, PAUL. **Increasing returns and long-run growth**. *Journal of Political Economy*, vol.94, n.5, p. 1002-1037, 1986.
- RIGBY, R. A. STASINOPOULOS, D. M. **Robust fitting of an additive model for variance heterogeneity**. *In: Dutter, R. and Grossmann, W. (eds.), COMPSTAT: Proceedings in Computational Statistics*, pp. 263-268. Physica, Heidelberg. (1994).
- RIGBY, R. A. STASINOPOULOS, D. M. **A semi-parametric additive model for variance heterogeneity**. *Statist. Comput.*, 6: 57-65, 1996a.
- RIGBY, R. A. STASINOPOULOS, D. M. **Mean and dispersion additive models**. *In: Hardle, W. and Schimek, M. G. (eds.), Statistical Theory and Computational Aspects of Smoothing*, pp. 215-230. Physica, Heidelberg, 1996b.
- RIGBY, R. A. STASINOPOULOS, D. M. **Construction of reference centiles using mean and dispersion additive models**. *Statistician*, 49: 41-50, 2000.
- RIGBY, R. A. STASINOPOULOS, D. M. **The GAMLSS project: a exible approach to statistical modelling**. *In: Klein, B. and Korsholm, L. (eds.), New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling*, pp. 249-255. Odense, Denmark. 2001.
- RIGBY, R. A. STASINOPOULOS, D. M. **Smooth centile curves for skew and kurtotic data modelled using the Box-Cox Power Exponential distribution**. *Statistics in Medicine*, 23: 3053-3076, 2004.
- RIGBY, R. A. STASINOPOULOS, D. M. **Generalized additive models for location, scale and shape**, (with discussion). *Appl. Statist.*, 54: 507-554, 2005.

- RIGBY, R. A. and STASINOPOULOS, D. M. **Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis.** *Statistical Modelling*, 6: 209–229, 2006.
- RIGBY, R. A. STASINOPOULOS, D. M. **A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution.** *Computational Statistics and Data analysis*, 2008.
- SCHUMPETER, Joseph A. **The Theory of Economic Development.** Cambridge, MA: Harvard University Press. 1934.
- SHESHINSKI, Eytan. **Optimal Accumulation with Learning by Doing.** *In* Karl Shell, ed., *Essays on the Theory of Optimal Economic Growth*, 31–52. Cambridge, MA: MIT Press, 1967.
- SMITH, Adam. **An Inquiry into the Nature and Causes of the Wealth of Nations.** New York: Random House (1937), 1776.
- STOCK, James H. WATSON, Mark W. **Econometria** (trad.) São Paulo: Addison Wesley, 2004.
- SOLOW, Robert M. **A Contribution to the Theory of Economic Growth.** *Quarterly Journal of Economics*, 70, February, 65–94, 1956.
- SOLOW, Robert M. **Technical Change and the Aggregate Production Function.** *Review of Economics and Statistics*, 39, August, 312–320, 1957.
- SOLOW, Robert M. **Investment and Technical Change.** *In* Kenneth J. Arrow et al., eds., *Mathematical Methods in the Social Sciences*. Palo Alto, CA: Stanford University Press, 1969.
- SOUZA, Geraldo da Silva e. **Introdução aos modelos de regressão lineares e não-lineares.** Brasília-DF: Empbrapa-SP/Embrapa-SEA, 1998.
- STASINOPOULOS, D. Mikis. RIGBY, Robert A. FAHRMEIR, L. **Modelling rental guide data using mean and dispersion additive models.** *Statistician*, 49: 479–493. 2000.
- STASINOPOULOS, D. Mikis. RIGBY, Robert A. **Generalized additive models for location, scale and shape.** *Appl. Statist.* 54, Part 3, pp. 507 - 554, 2005.
- STASINOPOULOS, D. Mikis. RIGBY, Robert A. **Contribution to the discussion of the paper by Lee and Nelder, Double hierarchical generalized linear models.** *Appl. Statist.*, 55: 171–172, 2006.
- STASINOPOULOS, D. Mikis. RIGBY, Robert A. AKANTZILIOTOU, C. **Instructions on how to use the GAMLSS package in R**, Second Edition. Technical Report 01-08, STORM Research Centre, London Metropolitan University, London. 2007.
- STASINOPOULOS, D. Mikis. RIGBY, Robert A. **Generalized Additive Models for Location Scale and Shape (GAMLSS) in R.** *Journal of Statistical Software*, December 2007, Volume 23, Issue 7. Disponível em <http://www.jstatsoft.org>
- STASINOPOULOS, D. Mikis. RIGBY, Robert A. AKANTZILIOTOU, C. **Instructions on how to use the gamlss package in R**, Second Edition. February 25, 2012.

SWAN, Trevor W. **Economic Growth and Capital Accumulation**. *Economic Record*, 32, November, 334–361. 1956.

UZAWA, Hirofumi. **Optimal Technical Change in an Aggregative Model of Economic Growth**. *International Economic Review*, 6, January, 18–31, 1965.

WOOD, Simon. **Generalized additive models: an introduction with R**. *Texts in Statistical Science*, Chapman & Hall/CRC, 2006.

WOOLDRIDGE, Jeffrey M. **Introdução à econometria: uma abordagem moderna**, 1ª ed. São Paulo-SP: Thomson Learning, 2007.

WRIGHT, Daniel B. LONDON, Kamala. **Modern Regression Techniques Using R: A Practical Guide for Students and Researchers**. SAGE Publications Ltd, 2009.

YOUNG, Allyn. **Increasing Returns and Economic Progress**. *Economic Journal*, 38, December, 527–542, 1928.

APÊNDICE A – ESTATÍSTICAS DAS VARIÁVEIS UTILIZADAS

Estatísticas descritivas						
Variável	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
PIBpc2000Def	0.000	7.300	7.468	7.488	7.750	8.991
PIBpc2009Def	2.840	3.447	4.104	4.606	4.773	13.965
POTMERC	-4.169	-3.012	-2.059	-1.888	-0.978	2.312
DENSIDEM	0.000	3.804	3.804	4.217	4.541	7.354
PARTGOV	0.000	8.465	9.105	9.033	9.541	12.950
INDPARTVA	0.000	6.919	7.667	7.903	8.504	13.350
CRIME	0.000	0.000	2.943	2.168	3.702	4.576
ESCMEDIA	0.000	2.118	2.543	2.669	2.978	6.560
INFRAENERGIA	0.000	2.724	4.600	4.788	7.101	11.689

Fonte: elaborado pelo autor.

APÊNDICE B - *GLOBAL DEVIANCE* PARA UM *STEP*

```
> pdDisssgamlss = prof.dev(gamlssDiss10, "mu", min=2, max=20, step=1)
```

```
*****
mu.start=( 2 )
GAMLSS-RS iteration 1: Global Deviance = 406.7957
GAMLSS-RS iteration 2: Global Deviance = 406.7957
*****
mu.start=( 3 )
GAMLSS-RS iteration 1: Global Deviance = 306.2621
GAMLSS-RS iteration 2: Global Deviance = 306.2621
*****
mu.start=( 4 )
GAMLSS-RS iteration 1: Global Deviance = 307.9639
GAMLSS-RS iteration 2: Global Deviance = 307.9639
*****
mu.start=( 5 )
GAMLSS-RS iteration 1: Global Deviance = 408.6993
GAMLSS-RS iteration 2: Global Deviance = 408.6993
*****
mu.start=( 6 )
GAMLSS-RS iteration 1: Global Deviance = 491.3643
GAMLSS-RS iteration 2: Global Deviance = 491.3643
*****
mu.start=( 7 )
GAMLSS-RS iteration 1: Global Deviance = 553.0957
GAMLSS-RS iteration 2: Global Deviance = 553.0957
*****
mu.start=( 8 )
GAMLSS-RS iteration 1: Global Deviance = 601.3472
GAMLSS-RS iteration 2: Global Deviance = 601.3472
*****
mu.start=( 9 )
GAMLSS-RS iteration 1: Global Deviance = 640.7062
GAMLSS-RS iteration 2: Global Deviance = 640.7062
*****
mu.start=( 10 )
GAMLSS-RS iteration 1: Global Deviance = 673.8572
GAMLSS-RS iteration 2: Global Deviance = 673.8572
*****
mu.start=( 11 )
GAMLSS-RS iteration 1: Global Deviance = 702.4573
```

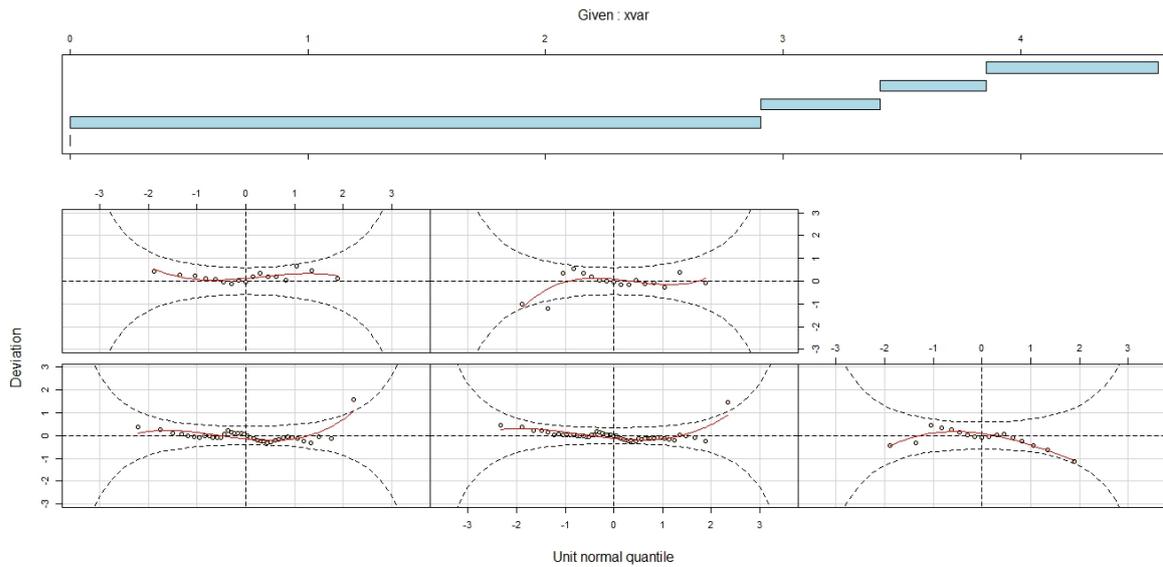
```

GAMLSS-RS iteration 2: Global Deviance = 702.4573
*****
mu.start=( 12 )
GAMLSS-RS iteration 1: Global Deviance = 727.5887
GAMLSS-RS iteration 2: Global Deviance = 727.5887
*****
mu.start=( 13 )
GAMLSS-RS iteration 1: Global Deviance = 749.9933
GAMLSS-RS iteration 2: Global Deviance = 749.9933
*****
mu.start=( 14 )
GAMLSS-RS iteration 1: Global Deviance = 770.1999
GAMLSS-RS iteration 2: Global Deviance = 770.1999
*****
mu.start=( 15 )
GAMLSS-RS iteration 1: Global Deviance = 788.5984
GAMLSS-RS iteration 2: Global Deviance = 788.5984
*****
mu.start=( 16 )
GAMLSS-RS iteration 1: Global Deviance = 805.484
GAMLSS-RS iteration 2: Global Deviance = 805.484
*****
mu.start=( 17 )
GAMLSS-RS iteration 1: Global Deviance = 821.0853
GAMLSS-RS iteration 2: Global Deviance = 821.0853
*****
mu.start=( 18 )
GAMLSS-RS iteration 1: Global Deviance = 835.5831
GAMLSS-RS iteration 2: Global Deviance = 835.5831
*****
mu.start=( 19 )
GAMLSS-RS iteration 1: Global Deviance = 849.1226
GAMLSS-RS iteration 2: Global Deviance = 849.1226
*****
mu.start=( 20 )
GAMLSS-RS iteration 1: Global Deviance = 861.8223
GAMLSS-RS iteration 2: Global Deviance = 861.8223
*****
*****
The Maximum Likelihood estimator is 3.501693
with a Global Deviance equal to 292.4989
A 95 % Confidence interval is: ( 3.240982 , 3.753872 )
*****

```

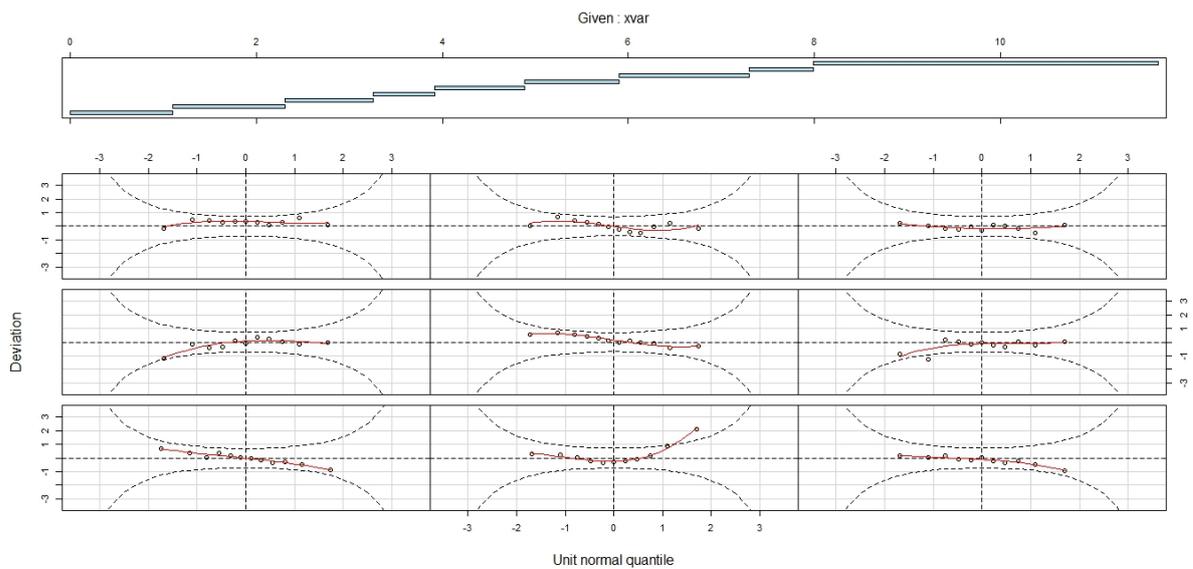
APÊNDICE C – *WORM PLOT* DO MODELO GAMLSSDISS10 COM A
OPÇÃO *XVAR* PARA TODAS AS VARIÁVEIS

Figura 1 – Worm plot da variável CRIME



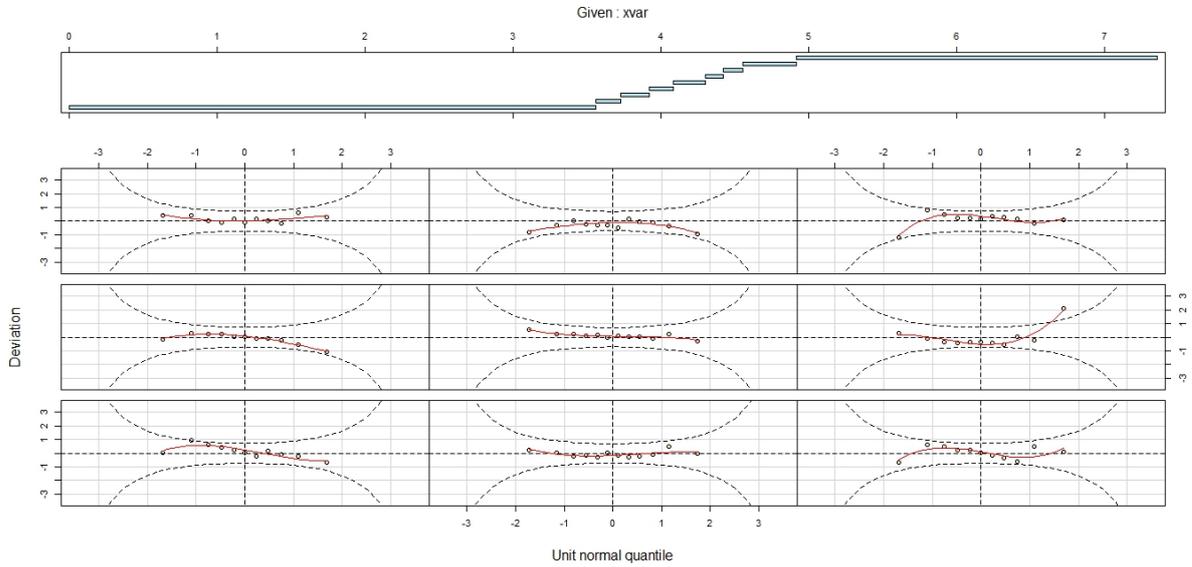
Fonte: elaborado pelo autor.

Figura 2 – *Worm plot* da variável INFRAENERGIA



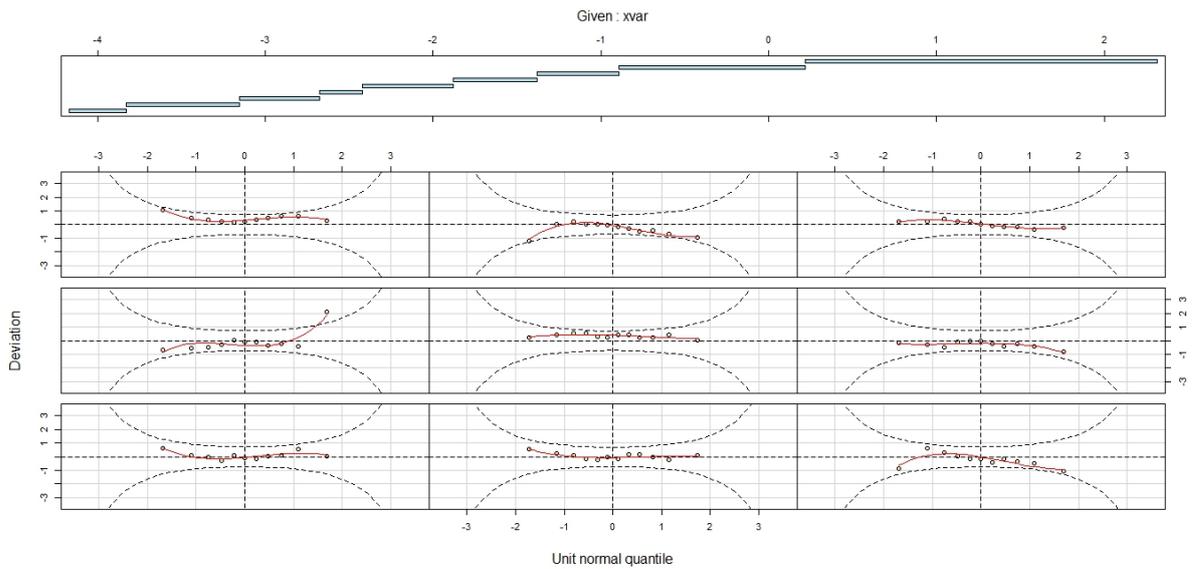
Fonte: elaborado pelo autor.

Figura 3 – Worm plot da variável DENSIDEM



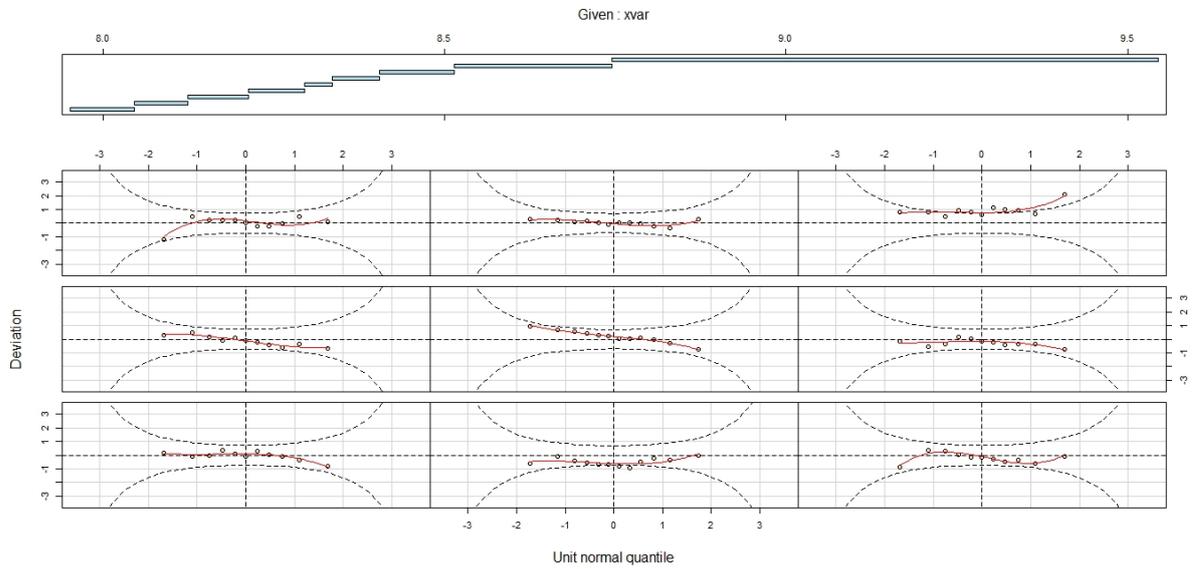
Fonte: elaborado pelo autor.

Figura 4 – Worm plot da variável POTMERC



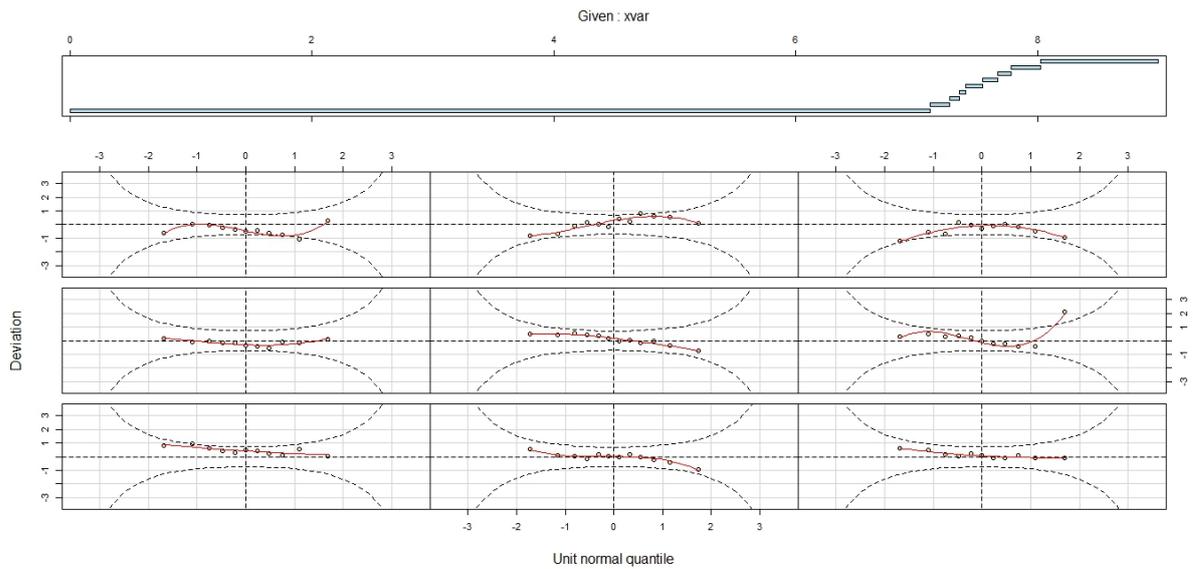
Fonte: elaborado pelo autor.

Figura 5 – *Worm plot* da variável PIBpc2009Def



Fonte: elaborado pelo autor.

Figura 6 – *Worm plot* da variável PIBpc2000Def



Fonte: elaborado pelo autor.