

**SELEÇÃO DE MODELOS ECONÔMICOS
NÃO ANINHADOS:
J-TESTE E FBST**

Fernando Valvano Cerezetti

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Mestre em Ciências

Área de Concentração: Estatística
Orientador: Prof. PhD. Julio Michael Stern

São Paulo, outubro de 2007

**SELEÇÃO DE MODELOS ECONÔMICOS
NÃO ANINHADOS:
J-TESTE E FBST**

Este trabalho corresponde à redação final da dissertação devidamente corrigida e defendida por Fernando Valvano Cerezetti e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Prof. PhD. Julio Michael Stern (Orientador) – IME/USP
- Prof. PhD. Carlos Alberto de Bragança Pereira – IME/USP
- Prof. PhD. Basílio Pereira – UFRJ

DEDICATÓRIA

Dedico esse trabalho a Antonio Carlos, Maria de Lourdes, Marysnel, Renata e especialmente a Karina, quem se mostrou uma verdadeira companheira.

AGRADECIMENTOS

Agradeço especialmente ao Prof. Dr. Julio Michael Stern por todo o esforço e dedicação concedidos ao trabalho.

Agradeço aos Professores Carlos Alberto de Bragança Pereira e Basílio Pereira pelos comentários valiosos sobre o trabalho.

Agradeço ao Marcelo Laretto pelo inestimável auxílio com as técnicas de programação.

Agradeço ao Departamento de Estatística do IME/USP pelo apoio ao trabalho.

RESUMO

A comparação e seleção de modelos estatísticos desempenham um papel fundamental dentro da análise econométrica. No que se trata especificamente da avaliação de modelos não aninhados, o procedimento de teste denominado de J-Teste aparece como uma ferramenta de uso freqüente nessa literatura. De acordo com apontamentos, entre os anos de 1984 e 2004 o J-Teste foi citado em 497 artigos pertinentes. Diferentemente do J-Teste, as abordagens Bayesianas possuem um potencial de aplicabilidade ainda pouco explorado na literatura, dado que são metodologicamente coerentes com os procedimentos inferenciais da econometria. Nesse sentido, o objetivo do presente trabalho é o de avaliar a aplicabilidade do procedimento de teste Bayesiano FBST para a comparação de modelos econométricos não aninhados. Implementando-se o FBST para os mesmos dados de estudos estatísticos relevantes na Teoria Econômica, tais como Bremmer (2003) (Curva de Phillips) e Caporale e Grier (2000) (determinação da taxa de juros real), constata-se que os resultados obtidos apontam para conclusões semelhantes daquelas delineadas com a utilização do J-Teste. Além disso, ao se utilizar a noção de função poder para avaliar ambos os procedimentos de teste, observa-se que sob certas condições as chances de erro expressas pelo Erro Tipo I e Erro Tipo II se tornam relativamente próximas.

PALAVRAS-CHAVE:

- Seleção de Modelos;
- Modelos Não Aninhados;
- J-Teste;
- Abordagem Bayesiana;
- FBST.

ABSTRACT

The comparison and selection of statistical models play an important role in econometric analysis. Dealing with evaluation of non nested models, the test procedure called J-Test is a frequently used tool in the literature. Accordingly to statistics, between the years 1894 and 2004 the J-Test was cited on 497 pertinent articles. Differently from J-Test, the Bayesian theories have an unexplored applicability potential in the literature, once they are methodologically coherent with the standard procedures of inference in econometrics. In this sense, the objective of this essay is to evaluate the applicability of the Bayesian procedure FBST to comparison of non nested econometric models. Implementing the FBST to the same data of some relevant statistical studies in Economic Theory, like Bremmer (2003) (Phillips Curve) and Caporale and Grier (2000) (real interest rate determination), it can be seen that the results obtained point to the same conclusions as that attained with J-Test utilization. Besides that, when implementing the power function to evaluate both test procedures, it can be observed that under some conditions the error chances expressed by Error Type I and Error Type II become relatively close.

KEYWORDS:

- Model Selection;
- Non Nested Models;
- J-Test;
- Bayesian Theories;
- FBST.

SUMÁRIO

LISTA DE TABELAS	VI
INTRODUÇÃO	pág 01
CAPÍTULO I: TESTES DE DIAGNÓSTICOS E J-TESTE	pág 04
I.i – Primeira Idéias	
I.ii – J-Teste: A Estrutura de Especificação do Teste	
CAPÍTULO II: ABORDAGEM BAYESIANA E FBST	pág 12
II.i – Testes de Hipóteses Bayesianos	
II.ii – FBST: Formulação Metodológica	
II.iii – FBST: Seleção de Modelos	
CAPÍTULO III: SELEÇÃO DE MODELOS NÃO ANINHADOS: J-TESTE E FBST	pág 24
III.i – Curva de Phillips	
III.ii – Regime Político e Taxa de Juros	
III.iii – Avaliação dos Procedimentos de Teste	
CONSIDERAÇÕES FINAIS	pág 37
ANEXO 1	pág 40
BIBLIOGRAFIAS DE REFERÊNCIA	pág 42

LISTA DE TABELAS

Tabela III.i: Curva de Phillips – Dados Anuais 1974-2003. Variável Dependente Taxa de Inflação Corrente	pág 26
Tabela III.ii: Curva de Phillips – Dados Anuais 1974-2003. Testes Pareados para Modelos não Aninhados.....	pág 26
Tabela III.iii: Curva de Phillips – Dados Anuais 1974-2003. <i>Full Bayesian Significance Test</i> (FBST).....	pág 27
Tabela III.iv: Modelos Alternativos da Variação na Taxa de Juros Real 1961- 1986. Variável Dependente Taxa de Juros Real <i>Ex Post</i> (RRt).....	pág 29
Tabela III.v: Modelos Alternativos da Variação na Taxa de Juros Real 1961- 1986. Testes Pareados para Modelos não Aninhados.....	pág 30
Tabela III.vi: Modelos Alternativos da Variação na Taxa de Juros Real 1961- 1986. <i>Full Bayesian Significance Test</i> (FBST).....	pág 31
Tabela III.vii: Curva de Phillips – Dados Anuais 1974-2003. Chances de Erro Tipo II para um Nível de Significância de 5%.....	pág 36
Tabela III.viii: Modelos Alternativos da Variação na Taxa de Juros Real 1961- 1986. Chances de Erro Tipo II para um Nível de Significância de 5%.....	pág 36
Tabela AI.i: Modelos Alternativos da Variação na Taxa de Juros Real 1961- 1986. Chances de Erro Tipo II para um Nível de Significância de 5%.....	pág 41
Tabela AI.ii: Modelos Alternativos da Variação na Taxa de Juros Real 1961- 1986. Chances de Erro Tipo II para um Nível de Significância de 5%.....	pág 41

Introdução

Desde a publicação póstuma do trabalho de Thomas Bayes em 1773 por Richard Price, muitos cientistas e estatísticos contribuíram para o desenvolvimento da estatística Bayesiana. Contudo, Zellner (1985) argumenta que foi a partir dos estudos de autores como Jacques Drèze, Tom Rothemberg, Walter Fisher, Albert Ando, entre outros, no decorrer da década de 1960 que nasceu a moderna econometria Bayesiana. Segundo o autor, desde então pode se observar um crescente número de estudos econométricos voltados à abordagem, fruto do aumento da interação entre economistas e estatísticos.

A partir dos anos 90, o aprimoramento das técnicas computacionais permitiu que a abordagem Bayesiana amplie-se ainda mais o seu escopo de aplicação, conforme destacam Gelman *et al* (1998) e Murteira *et al* (2003). Parafraçando esses últimos autores, com a efetiva aplicação dos métodos de Monte Carlo associados às Cadeias de Markov (MCMC), técnicas computacionais bastante poderosas, houve uma transformação na investigação realizada através da estatística Bayesiana, dado que esta a permitiu atingir a maior credibilidade em diversas áreas do conhecimento científico.

Do ponto de vista pragmático, a introdução da metodologia Bayesiana na econometria está relacionada com a abundante utilização da informação *a priori*, embora em muitos casos este procedimento seja informal e até mesmo incoerente.¹ A informação inicialmente possuída pelo pesquisador é relevante para todo processo de especificação do modelo estatístico, estando presente desde a seleção das variáveis até a determinação da complexidade do modelo. Nessa linha, a abordagem Bayesiana permite aos pesquisadores a utilização da informação *a priori* formalmente, baseando-se em uma estrutura coerente e flexível, voltada para a resolução de problemas de estimação, de teste de hipóteses, predição, seleção de modelos, entre outros tipos de inferências estatísticas.

No que se trata dos aspectos de sustentação axiomáticos, Zellner (1971) destaca que o problema fundamental do progresso da ciência é o aprendizado a partir da experiência, processo no qual a lógica indutiva de inferência opera intensamente.² O elemento central na cadeia de aprendizagem é a possibilidade de os pesquisadores reverem suas crenças nas teorias por eles formuladas, com isso permitindo o seu avanço. A abordagem Bayesiana, através da lógica do Teorema de Bayes, pode ser vista como um mecanismo operacional e quantitativo para a

¹ Vide Swamy (1980) para exemplos de utilização incoerente de informação *a priori* em estudos não Bayesianos.

incorporação do efeito de novas informações nas crenças sobre as proposições teóricas. Soma-se a esse o fato de que a operacionalização do mecanismo é singular, dado ter o mesmo núcleo independente da área de implementação, Séries Temporais, Regressão ou Análise de Dados Categorizados.

Mesmo tendo o número de estudos Bayesianos aumentado significativamente nas Ciências Econômicas nos últimos anos, e axiomáticamente a abordagem se mostre coerente com a lógica indutiva de geração de conhecimento científico, estando apta a realizar os procedimentos inferenciais, dentro da modelagem econométrica essa ainda desempenha um papel tímido em relação ao seu potencial. No que se refere especificamente a avaliação da adequação do ajuste de um modelo a um conjunto de dados, constata-se uma preponderância dos procedimentos não Bayesianos quando se trata da utilização pelos economistas. Se o foco da avaliação é os modelos não aninhados, o J-Teste aparece como uma ferramenta útil aos pesquisadores envolvidos com a aplicação estatística no estudo das variáveis econômicas.

Inicialmente proposto por Davidson e MacKinnon (1981) como uma metodologia alternativa ao Teste F de restrições, na época o procedimento mais utilizado, o J-Teste tem sido utilizado como um mecanismo para se testar a plausibilidade de um modelo não aninhado frente a um ou mais modelos concorrentes. Atualmente o teste ocupa atualmente uma posição de destaque dentro das investigações econométricas. Segundo Bremmer (2003), entre 1981 e 1995 o J-Teste tinha sido citado 71 vezes em estudos pertinentes. Alguns anos mais tarde, entre 1984 e 2004, o autor revela que 497 artigos citavam o teste.

Nesse sentido, dado a consistência teórica da abordagem Bayesiana com os procedimentos inferenciais utilizados nas Ciências Econômicas e o atual *estado das artes*, o objetivo do trabalho consiste em avaliar a aplicabilidade dos mecanismos da Teoria Bayesiana para comparar e selecionar modelos estatísticos implementados sob contextos e variáveis econômicas. Especificamente, o intuito do estudo é o de analisar a performance de um procedimento de teste Bayesiano, o *Full Bayesian Significance Test* (FBST), frente ao J-Teste no que se trata da avaliação de modelos não aninhados.

O FBST foi primeiramente proposto por Pereira e Stern (1999), tendo sido aprimorado pelos próprios autores e por outros pesquisadores desde então. O FBST é metodologicamente considerado um teste “Bayesiano Completo” (*Full Bayesian*), visto que a elaboração de uma medida de evidência para a hipótese precisa depende somente da distribuição *a posteriori*. Especificamente, o teste não necessita de hipóteses adicionais, como uma *priori* positiva para a

² A lógica indutiva de inferência científica é caracterizada pelo processo no qual a partir de um conjunto de

hipótese precisa. Além disso, o FBST é coerente com o princípio da Verossimilhança Bayesiano, segundo o qual a informação obtida dos dados observados está toda representada na função de verossimilhança.

A fim de atingir o objetivo almejado, o presente estudo foi organizado da seguinte maneira. No capítulo I são apresentadas as técnicas tradicionais para a comparação e seleção de modelos. Uma atenção especial é dada ao J-Teste, para o qual se aprofunda nas proposições teóricas e nos métodos de implementação. No Capítulo II o foco se volta para os mecanismos Bayesianos de avaliação de modelos, sendo apresentada a abordagem tradicional do Fator de Bayes e da Teoria da Decisão, assim como a proposição alternativa do FBST.

No Capítulo III são realizados os testes práticos de implementação do FBST, confrontando os seus resultados com aplicações relevantes já existentes do J-Teste na literatura econométrica. O interesse consiste na análise de convergência dos resultados, dado tratar-se de mecanismos de seleção com princípios geradores antagônicos. Além disso nesse capítulo, utiliza-se a noção de que as chances de tomada de decisão errônea pelo pesquisador, Erro Tipo I e II, caracterizam um procedimento de teste para avaliar o J-Teste e o FBST. Por fim aspectos conclusivos do estudo são apresentados nos Comentários Finais.

fatos ou pressupostos se busca a generalização para a confecção de teorias.

Capítulo I

Testes de Diagnóstico e J-Teste

1.1 – PRIMEIRAS IDÉIAS

O avanço da visão da Teoria Econômica como uma ciência positiva levou muitos pesquisadores a ingressarem em um processo de modelagem estatística das variáveis da economia. O intuito fundamental dos autores era o de tentar descrever os fenômenos econômicos a partir de distribuições de probabilidade bem definidas estatisticamente. Assim, o referencial probabilístico e o estatístico viriam somar esforços à pesquisa econômica.

Mas quais os benefícios engendrados no desenvolvimento de modelos estatísticos para os fenômenos econômicos? De acordo com Fomby *et al* (1984), esses modelos, também denominados de modelos econométricos, teriam o controle e a predição dentre as suas funções mais relevantes de existir. No primeiro caso a modelagem permite ganho de conhecimento sobre a dimensão das variáveis relevantes para a determinação dos fenômenos econômicos. No caso da predição, a estipulação de uma estrutura de relacionamento de variáveis tende a auxiliar no julgamento de eventos futuros.

Por conseguinte, uma questão surge na seqüência é como utilizar o instrumental econométrico a fim de se obter o modelo mais próximo da realidade. Uma primeira maneira de se proceder é partir de modelos teóricos dados, tanto no que se trata da identificação de regressores como de forma funcional, e utilizar estatísticas de diagnóstico como R-quadrado, t, F e d de Durbin-Watson para avaliar sua plausibilidade. De acordo com Gujarati (2000), essa visão é conhecida na literatura como “abordagem às avessas”, ou mais propriamente Regressão Econômica Média (REM).

De um modo geral, a metodologia de REM busca analisar se no modelo inicial proposto existem erros de especificação, utilizando as estatísticas acima como forma auxílio na sua detecção. A natureza desses erros reside basicamente em quatro determinantes, a saber: i) omissão de variáveis relevantes; ii) inclusão de variáveis desnecessárias; iii) especificação de forma funcional não apropriada e; iv) erros de mensuração das variáveis analisadas. Uma vez detectados alguns desses erros de especificação, a abordagem de REM sugere mecanismos de correção.

A não observância dos erros de especificação de um modelo pode ser prejudicial para o trabalho desenvolvido pelos econométricos. Em modelos lineares, por exemplo, se uma variável relevante é omitida, os estimadores dos parâmetros da regressão se tornam viesados e inconsistentes. Além disso, a estimativa da variância do erro é errônea. Em suma, todos os

procedimentos usuais de intervalos de confiança e teste de hipóteses serão inviabilizados. No caso da inclusão de uma variável irrelevante as consequências são menos drásticas, afetando apenas a variância dos estimadores dos parâmetros, a qual se eleva.

Apesar de toda a preocupação envoltória aos erros de especificação, estes podem ser facilmente cometidos quando a teoria por detrás da construção dos modelos é inconclusiva. Além disso, novas hipóteses teóricas precisam ser testadas. Nesse caso, quando a fronteira do saber ainda não é capaz de propor um modelo final consistente, os pesquisadores se valem de instrumentos estatísticos alternativos ao REM para lhes auxiliarem na definição dos modelos. Por essas e outras razões, a abordagem de REM tem sofrido algumas críticas ao longo das últimas décadas, a partir das quais novas abordagens têm emergido.

Um aspecto característico dessas novas abordagens alternativas reside no fato de que elas sugerem que, concomitantemente ao processo de verificação da especificação do modelo, deveria existir um processo de comparação e seleção de modelos. O argumento desses autores é o de que o conhecimento possuído *a priori* sobre um fenômeno da realidade, na maioria dos casos, não é o suficiente para especificar a família de funções de distribuição de probabilidade que melhor se aproximaria do verdadeiro modelo, tal como suposto que seja pela abordagem de REM.

Dessa maneira, a proposição dos autores é o de criar um ferramental estatístico para auxiliar na identificação de um modelo para um fenômeno da realidade, mesmo quando este não pode ser especificado completamente *a priori*. Em termos ramos do conhecimento, a Especimetria é a parte da ciência que visa fornecer esses mecanismos de identificação modelos. De acordo com Leamer (1978) apud Gujarati (2000)

A especimetria descreve o processo que leva um pesquisador a escolher uma especificação de modelo em vez de outra; além disso, a especimetria tenta identificar as inferências que podem ser corretamente extraídas de um conjunto de dados quando o mecanismo gerador dos dados for ambíguo. (pág. 483)

Os testes de diagnóstico desenvolvidos pelos estatísticos não só permitem os pesquisadores avaliar a adequação de um modelo a um conjunto de dados, mas também auxiliam a Especimetria a comparar e escolher entre modelos competitivos. Basicamente a metodologia que está por detrás desses testes divide-se em dois grandes grupos, a saber, aquela que foca em modelos aninhados e uma segunda cujo foco está nos modelos não aninhados. Cada uma dessas abordagens de análise tem seus próprios objetivos, com diferentes pressupostos e procedimentos de teste.

Estatisticamente, no contexto das regressões, diz-se que um Modelo A é aninhado em um Modelo B se ele pode ser escrito em função de restrições no último modelo, ou em outros

termos, se o Modelo A está contido no Modelo B. Por exemplo, suponha dois modelos lineares expressos por

$$\text{Modelo A: } y_i = X_i\boldsymbol{\beta} + \varepsilon_i, \quad (1.1)$$

$$\text{Modelo B: } y_i = Z_i\boldsymbol{\gamma} + \tau_i, \quad (1.2)$$

nos quais y_i é a variável dependente; X_i e Z_i são vetores de variáveis exógenas; $\boldsymbol{\beta}_i$ e $\boldsymbol{\gamma}_i$ são vetores dos parâmetros a serem estimados e; ε_i e τ_i são os termos de erro. Se o vetor X_i está contido no vetor Z_i para todo i , então diz-se que o Modelo A está aninhado no Modelo B. Inversamente, se X_i e Z_i são vetores distintos de variáveis independentes os modelos são considerados não aninhados um no outro.

A avaliação e a análise comparativa de modelos não aninhados pode ser feita de várias maneiras.³ Entre os critérios chave utilizados na construção das estatísticas de diagnóstico pode-se citar o Poder de Previsão, os Critérios de Informação e o Erro Empírico. Eddy e Geisser (1979) utilizam o primeiro critério para propor um método particular para a seleção de modelos. Trabalhando com o conceito de reutilização preditiva da amostra (PSR), a idéia básica dos autores é a de escolha do modelo que maximiza o valor das densidades preditivas condicionais.

No que se refere aos critérios de informação, o princípio fundamental das estatísticas de diagnóstico é o reconhecimento da existência de divergência entre o modelo teórico proposto pelos pesquisadores e o verdadeiro modelo. Dessa maneira, as estatísticas baseadas em critérios de informação procuram construir medidas de divergência entre os modelos postulados e o real, considerando como o modelo mais adequado para estimação aquele que minimiza esta discrepância.

Especificamente, uma medida teórica de divergência bastante utilizada na formulação das estatísticas é o critério de informação de Kullback-Leibler, expresso matematicamente pela seguinte fórmula

$$I(G : F(\bullet)) = E_G \left[\log \frac{g(Y)}{f(Y/\theta)} \right], \quad (1.3)$$

no qual $g(Y)$ é a verdadeira densidade, $f(Y/\theta)$ é a densidade do modelo proposto e $E_G[\bullet]$ é a esperança com respeito a verdadeira função de distribuição de probabilidade $G(Y)$. Dentre

³ A discussão do presente trabalho ficará restrita aos critérios de comparação e seleção de modelos não aninhados. Embora alguns desses critérios também se apliquem aos modelos aninhados, o instrumental estatístico específico para os aninhados foge ao escopo do estudo, não sendo assim abordado.

os critérios mais conhecidos que utilizam Kullback-Leibler como referência, destaque para o AIC e o BIC.⁴

Por fim, o terceiro critério utilizado para se avaliar o ajuste de um modelo a um conjunto de dados analisado é a minimização do erro empírico. No entanto, como apontado por Pereira e Stern (2001)

[...] we can not select the ‘best’ model based only on the empirical error, because we would usually select a model of high complexity. [...] It is a well known fact in statistics (or learning theory), that the prediction (or generalization) power of such high dimensions models is poor. (pág 562)

Conforme destacado em Sawa (1978), não é somente o poder de predição do modelo que fica prejudicado com o aumento da sua dimensão, mas também a acurácia das estimativas dos parâmetros do verdadeiro modelo. Esses argumentos servem de suporte ao estabelecimento de um *trade-off* entre o erro empírico de um lado e a predição/acurácia do outro. Se o pesquisador deseja melhorias em um dos atributos, terá que abrir mão dos demais. Desse modo, o Princípio da Parcimônia postula um uso mais comedido dos parâmetros.

Uma maneira encontrada pelos pesquisadores para se adequar ao Princípio da Parcimônia é a utilização das funções de regularização. Essas funções são aplicadas sobre o erro empírico, penalizando a dimensão do modelo. O erro final penalizado poderia ser expresso de maneira genérica como $R_p = r(d,n) * R$, no qual R é o erro empírico calculado através dos métodos regulares e $r(d,n)$ é a função de regularização, tendo d como a dimensão do modelo e n com o número de observações da amostra.

O Coeficiente Múltiplo de Determinação Ajustado para os graus de liberdade é o exemplo mais comum das estatísticas baseadas no erro empírico que incorpora o critério da regularização. Além desse, existem outras propostas de estatísticas que se valem de funções de regularização familiares para ponderar os efeitos do super dimensionamento. Dentre as funções mais conhecidas têm-se as seguintes, nas quais $p = (d / n)$

- Erro de Predição Final de Akaike: $(1 + p)/(1 - p)$; (1.4)

- Critério Bayesiano de Schwarz: $1 + \ln(n) * p / (2 - 2p)$; (1.5)

- Modelo Seletor de Shibata: $1 + 2p$. (1.6)

Gujarati (2000) classifica os critérios de comparação e seleção de modelos não aninhados descritos acima, mais outros como a Medida de Hocking Sp e Medida de Amemiya PC , dentro de um grupo denominado por abordagem de discriminação. De acordo com o autor, a característica principal das métricas nesse grupo reside no fato de que, entre os modelos

⁴ Para mais detalhes na construção desses testes vide Sawa (1978).

concorrentes, um é selecionado se baseando apenas no valor obtido pela estatística. Todos modelos são avaliados individualmente, sendo classificados de acordo com o valor da estatística de diagnóstico trabalhada. O modelo escolhido é aquele que apresenta um melhor ajuste de acordo com o critério tratado.

O fato de um modelo ser analisado individualmente, não utilizando a informação contida nos demais modelos concorrentes para avaliar o seu grau de ajuste, representa um inconveniente das técnicas baseadas no processo de discriminação. Esse aspecto abre espaço para a utilização de outras técnicas de comparação de modelos não aninhados que buscam selecionar os modelos considerando-os conjuntamente. Dentre os métodos tradicionais pertencentes a essa classe de testes estão o Teste F e o J-Teste.

A utilização do Teste F para a seleção de modelos não aninhados tem sido uma longa prática na econometria, datando mesmo de antes do desenvolvimento do J-Teste. A principal idéia do teste é a de construir um modelo híbrido expandido com as variáveis de todos os modelos considerados. Na seqüência testa-se restrições sobre os coeficientes das variáveis dos demais modelos frente a um considerado como referência. O resultado do teste indicará se a restrição é plausível, apontando para o modelo com melhor ajuste aos dados.⁵ O segundo teste tradicional apontado pela literatura é o J-Teste, o qual será estudado em maiores detalhes na próxima seção.

LII – J-TESTE: A ESTRUTURA DE ESPECIFICAÇÃO DO TESTE

O J-Teste foi inicialmente proposto por Davidson e Mackinnon (1981) como uma simplificação do trabalho anterior de Pesaran e Deaton (1978). Esse último estudo sugeriu uma nova metodologia de teste de validade de modelos de regressão, possivelmente multivariado e não lineares, para os quais a hipótese alternativa não está aninhada. De acordo com Davidson e Mackinnon (1981) esse novo método era uma exceção importante aos procedimentos padrões de teste de restrições em modelos expandidos, os quais eram a prática predominante na econometria naquele momento.

Diferentemente do período em que o teste foi proposto, recentemente tem sido evidenciado na literatura econométrica que a aceitação do J-Teste aumentou durante o transcorrer dos anos. De acordo com dados apontados em Bremmer (2003), entre 1981 e 1995 o J-Teste tinha sido citado 71 vezes em estudos pertinentes. Alguns anos mais tarde, entre 1984 e 2004, o autor revela que 497 artigos citavam o teste, valor aproximadamente sete vezes maior que o do primeiro período.

⁵ Para um exemplo de utilização do Teste F direcionado à seleção de modelo não aninhados vide Bremmer (2003).

Com intuito de apresentar as técnicas de análise do J-Teste, considere uma situação na qual a teoria econômica é inconclusiva a respeito da correta forma funcional estatística de um fenômeno econômico, como é o caso da Teoria do Investimento, na qual se discute a sua determinação financeira, ou na Teoria do Consumo, na qual renda permanente e a renda corrente são duas abordagens competitivas na explanação dos gastos do consumidor. Os exemplos de situações similares são abundantes, destacando-se os acima devido a sua proeminência.

Inicialmente suponha que existam duas formas distintas de se modelar um fenômeno, embora as idéias aqui expostas possam ser generalizadas para cenários envolvendo mais modelos. Nesse caso, os pesquisadores são confrontados com a necessidade de testar a validade do modelo que a priori avaliam ser mais plausível. Considere que esse possa ser matematicamente expresso por

$$H0: y_i = f_i(X_i, \beta) + \varepsilon_{0i}, \quad (1.7)$$

no qual y_i é a i -ésima observação da variável dependente; X_i é o i -ésimo vetor de valores para as variáveis exógenas; β é o vetor de parâmetros de tamanho k considerado e; ε_{0i} é o termo de erro suposto ser independente e normalmente distribuído com média zero e variância σ_0^2 para todo i , $\varepsilon_{0i} \sim N(0, \sigma_0^2)$, $i = 1, 2, \dots, n$.

Suponha que o modelo alternativo seja dado por

$$H1: y_i = g_i(Z_i, \gamma) + \varepsilon_{1i}, \quad (1.8)$$

no qual Z_i é o i -ésimo vetor de observações para as variáveis exógenas, tendo pelo menos um campo distinto daqueles contidos no vetor X_i ; γ é o vetor de parâmetros de tamanho l a ser estimado e; ε_{0i} é o termo de erro suposto ser independente e normalmente distribuído com média zero e variância σ_1^2 para todo i , $\varepsilon_{1i} \sim N(0, \sigma_1^2)$, $i = 1, 2, \dots, n$, se $H1$ for o modelo verdadeiro.

Na formulação do J-Teste, Davidson e Mackinnon (1981) utilizaram uma versão modificada do que Fomby *et al* (1984) definiram como “Aninhamento Mecânico”, no qual a idéia básica é a construção de um modelo expandido contendo as variáveis de ambos modelos concorrentes. Em vez de introduzirem no modelo geral as variáveis de $H1$, os autores empregaram a variável dependente predita pelo modelo $H1$. Em termos algébricos, o modelo expandido pode ser expresso como

$$y_i = (1 - \alpha)f_i(X_i, \beta) + \alpha \hat{g}_i + \varepsilon_i, \quad (1.9)$$

no qual $\hat{g}_i = g_i(Z_i, \hat{\gamma})$ e $\hat{\gamma}$ é o estimador de máxima verossimilhança (EMV) de γ .

A suposição de que $H0$ não está aninhado em $H1$ e que $H1$ não está aninhado em $H0$ implica que a validade de um modelo é automaticamente a não veracidade do outro. Em outras palavras, ao rejeitar a hipótese de que $H1$ é a melhor forma funcional de se ajustar os dados analisados está-se considerando $H0$ como o modelo mais plausível. Nesse sentido, para se avaliar a veracidade de $H0$ basta se testar a não significância estatística de $H1$, através do teste da hipótese $\alpha = 0$.

Considerando o fato de que assintoticamente \hat{g}_i será independente de ε_i e alguns outros pressupostos apontados, Davidson e Mackinnon (1981) propõem um Lema que afirma que, se $H0$ é o verdadeiro modelo, a estatística t para o estimador de máxima verossimilhança de α gerada através do modelo de regressão expandido será normalmente distribuída, com média zero e variância um, $\hat{\alpha} \sim N(0,1)$. Dessa maneira, para se testar $\alpha = 0$ pode-se implementar simplesmente um teste t assintótico convencional, ou equivalentemente um teste de razão de verossimilhança.

Se $H0$ é o verdadeiro modelo, então se tenderá a aceitar a hipótese de que $\alpha = 0$. Contrariamente, se $H1$ é a melhor forma funcional para se descrever o fenômeno analisado, as estimativas de α irão convergir assintoticamente para um. Embora o primeiro resultado permita o pesquisador corroborar a validade de $H0$, a obtenção de estimativas para α próximas a um não implica na aceitação do modelo $H1$ como verdadeiro.

O aspecto destacado acima deriva do fato de que a estatística t inerente ao teste é condicional na veracidade do modelo $H0$, e não na validade de $H1$. Uma estatística de teste que seja pertinente para a avaliação da plausibilidade de $H0$ não será coerente para o teste de veracidade de $H1$. Os pesquisadores também interessados em estudar a validade do modelo $H1$ devem implementar o teste novamente, invertendo os papéis das hipóteses consideradas inicialmente.

Embora a idéia de se utilizar o J-Teste para avaliar o ajuste de vários modelos ao mesmo tempo, ou em outras palavras, selecionar a forma funcional mais apropriada para descrever os dados, seja bastante tentadora, Davidson e Mackinnon (1981) enfatizaram que o procedimento de teste foi realmente desenvolvido para análise de especificação do modelo. Os autores argumentaram que se o objetivo é escolher um modelo dentre uma série de concorrentes, os pesquisadores deveriam focar seus esforços em outro tipo de critério de informação.

É interessante notar que as precauções alinhavadas pelos autores parecem ter sido esquecidas na literatura pertinente. Muitos dos textos discutidos até o momento, assim como alguns que ainda serão tratados, tendem a implementarem o J-Teste como forma de seleção de

modelos. O pressuposto de uma crença maior sobre a validade de um modelo frente aos demais argumentada acima, e conseqüentemente o teste das hipóteses alternativas contra este, não tem sido prática nos textos estudados. O procedimento mais comumente evidenciado é uma seqüência de testes em pares, nos quais a cada repetição um novo modelo é tratado como referência.

Um aspecto negativo decorrente do uso do J-Teste como mecanismo de seleção de modelo, tal como evidenciado na literatura, é que resultados inconclusivos podem aparecer. Especificamente, existe a possibilidade de que todos os modelos sob análise sejam rejeitados, isto é, não sejam considerados válidos para descrever os dados. Por outro lado, também há chances de se avaliar tanto o modelo H_0 quanto os alternativos como plausíveis. Em ambos os casos o J-Teste não é capaz de apontar para uma conclusão decisiva.

Além desse aspecto negativo apontado acima, existe outra peculiaridade do J-Teste que exige atenção e consideração especial por parte dos pesquisadores, a saber, o fato de o teste ser definido para amostras grandes de dados. Conforme argumentado acima, a estatística de teste tem distribuição apropriada somente quando tratada assintoticamente. Em conseqüência, a redução do tamanho da amostra acaba por levar a rejeição mais freqüente do verdadeiro modelo. Por esses aspectos, e pela necessidade de se ampliar o conhecimento sobre procedimentos de comparação e seleção de modelos não aninhados, no próximo capítulo estudar-se-á outras abordagens de teste.

Capítulo II

Abordagem Bayesiana e FBST

II.1 – TESTES DE HIPÓTESES BAYESIANOS

O paradigma Bayesiano de análise estatística tem na distribuição *a posteriori* do parâmetro avaliado o seu objeto chave de investigação. Para os pesquisadores bayesianos essa distribuição incorpora toda a informação disponível sobre o parâmetro, a qual advém da combinação da informação inicial sobre o mesmo, a distribuição *a priori*, juntamente com a informação contida nos dados, materializada na verossimilhança.

Nesse paradigma as avaliações sobre o parâmetro são realizadas através de cálculos sobre a distribuição *a posteriori*, sendo esta o único estimador no processo inferencial. No entanto, a informação contida nesta distribuição pode ser sumarizada graficamente ou por medidas quantitativas de localização, dispersão e forma funcional. Especificamente, pode-se trabalhar com medidas de estimação pontual, de estimação por intervalo, teste de hipóteses ou medidas de predição.

No que se refere aos testes de hipóteses, a avaliação das hipóteses é realizada simplesmente através do cálculo de suas respectivas probabilidades, isto é, $\Pr(H_0)$, $H_0 : \theta \in \Theta_0$, e $\Pr(H_1)$, $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$, no qual Θ é o espaço paramétrico. A escolha por uma das hipóteses em confronto está baseada na construção de uma grandeza relativa, a qual é representada pelo Fator de Bayes. A medida visa quantificar o peso relativo da evidência contida nos dados a favor de uma das hipóteses avaliadas, sendo expressa matematicamente por

$$B(x) = \frac{\Pr(H_0/x) / \Pr(H_1/x)}{\Pr(H_0) / \Pr(H_1)}, \quad (2.1)$$

no qual $\Pr(H_i/x)$ e $\Pr(H_i)$, $i=1$ e 2 , representam respectivamente a probabilidade *a posteriori* e *a priori* das hipóteses consideradas.

O processo de escolha de uma das hipóteses em confronto a partir do Fator de Bayes é complementado com a implementação dos fatores analíticos da Teoria da Decisão, tais como a função de perda natural, a função crítica do teste e o cálculo dos riscos *a posteriori* da decisão. Para um dado vetor de observações da variável aleatória X , a melhor escolha dentre as hipóteses em questão é aquela que minimiza o risco *a posteriori*.

Os procedimentos inferenciais dos testes de hipóteses bayesianos acima descritos podem ser facilmente generalizados para serem utilizados na comparação e seleção de modelos. Em termos similares, o próprio teste de hipóteses paramétrico expresso anteriormente pode ser encarado como uma avaliação de dois grupos de modelos distintos, os quais se matizam devido à diferença nos parâmetros indexadores das famílias de distribuições amostrais. Com o intuito de ilustrar a aplicação suponha a existência de n modelos concorrentes, $M = \{M_i, i = 1, 2, \dots, n\}$, originando as seguintes distribuições preditivas *a priori*

$$P(x/M_i) = \int_{\Theta} f_i(x/\theta_i)h(\theta_i)d\theta_i, \quad i = 1, 2, \dots, n, \quad (2.2)$$

nas quais $f_i(x/\theta_i)$ e $h(\theta_i)$ são respectivamente a densidade de x e a *priori* de θ sob a ótica do modelo M_i . Com base nessas e na distribuição *a priori* para os modelos,

$P(M_i)$, $i = 1, 2, \dots, n$, é possível se definir a densidade preditiva global de x , expressa por

$$P(x) = \sum_{i=1}^n P(M_i)P(x/M_i). \quad (2.3)$$

As distribuições acima expressas servem de base para a configuração da Fórmula de Bayes, na qual a probabilidade dos modelos é atualizada a partir da informação contida nos dados. Em termos matemáticos tem-se

$$P(M_i/x) = \frac{P(M_i)P(x/M_i)}{P(x)}, \quad i = 1, 2, \dots, n. \quad (2.4)$$

Do mesmo modo que os procedimentos ordinários de teste de hipóteses bayesianos, pode-se utilizar as distribuições *a priori* e *a posteriori* dos modelos para se construir o Fator de Bayes, permitindo desta forma criar a medida de grandeza relativa necessária para comparação entre eles. Em termos concretos, no caso de dois modelos distintos, o Fator de Bayes é dado por

$$B_{ij}(x) = \frac{P(M_i/x)/P(M_j/x)}{P(M_i)/P(M_j)}, \quad (2.5)$$

sendo a escolha de um deles determinada a partir da incorporação do cenário de decisão e seus respectivos atributos.

Até o momento considerou-se que as hipóteses avaliadas possuíam probabilidade *a priori* positiva, tanto no caso mais geral, $H_0 : \theta \in \Theta_0$ e $H_1 : \theta \in \Theta_1$, quanto no caso da seleção de modelos, $H_0 : M = M_i$ e $H_1 : M \neq M_i$. Contudo, se se estiver utilizando uma *priori* contínua, a definição de uma hipótese simples do tipo $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ pode inviabilizar os procedimentos inferenciais abordados. Esse aspecto decorre do fato de que a Fórmula de Bayes não permite atualizar uma probabilidade nula *a priori*.

Uma primeira maneira de se contornar o problema reside na incorporação de uma distribuição *a priori* mista, na qual uma massa de probabilidade pontual é dada a H_0 , e o restante da probabilidade distribuída de forma contínua nos demais valores de θ . Esse procedimento é fundamentado na noção de que a própria hipótese $H_0 : \theta = \theta_0$ enseja informação relevante sobre o parâmetro. Especificamente, definida como tal, a hipótese atribui um grau de importância maior a θ_0 do que aos demais valores de θ , justificando assim a massa concentrada de probabilidade.

Uma maneira distinta de se pensar sobre o problema das hipóteses categóricas é tratar $H_0 : \theta = \theta_0$ como irrealista. Na verdade, ao se formular uma hipótese com tal não se está considerando somente o θ_0 , mas toda uma gama de valores de θ próximos a este. Em outros termos, tanto θ_0 quanto os seus vizinhos mais próximos possuem crenças *a priorísticas* relativamente homogêneas. Nesse sentido, a hipótese $H_0 : \theta = \theta_0$ deve no fundo ser interpretada como $H_0 : \theta = V_\varepsilon(\theta_0)$, no qual $V_\varepsilon(\cdot)$ representa uma bola centrada em θ_0 com raio ε .⁶

Um terceiro modo de se avaliar as hipóteses simples, dado distribuições *a priori* contínuas para os parâmetro, é construir uma região de credibilidade γ para θ , aceitando a hipótese nula se θ_0 estiver contido na região. Ao se proceder dessa maneira está se alterando a forma de analisar as hipóteses, de um cálculo inicial do peso relativo dos dados a favor de H_0 , medido pelo Fator de Bayes, para uma mensuração da plausibilidade de θ_0 frente aos demais valores do parâmetro.

A construção da região de credibilidade γ para θ pode estar alicerçada em um conhecimento *a priori* sobre o parâmetro, assim como no desejo de um determinado nível de confiança para o intervalo. A fim de se mitigar o inconveniente de se possuir nesse intervalo valores de θ que tenham menor credibilidade do que os valores não contidos no mesmo, os pesquisadores têm se valido dos intervalos HPD (*Highest Posterior Density*), nos quais para um nível dado de credibilidade do intervalo, os valores de θ apresentam as maiores densidades.

Assim como a construção de intervalos de credibilidade para θ , a plausibilidade de θ_0 também pode ser analisada calculando-se a probabilidade dos valores de θ com densidade maior ou igual à densidade de θ_0 , expresso da seguinte maneira:

$$R(x) = \{\theta : h(\theta/x) \geq h(\theta_0/x)\}. \quad (2.6)$$

⁶ Para maiores detalhes vide Casella e Berger (1987).

A probabilidade de que $\theta \in R(x)$ pode ser encarada como uma medida da não pertinência da hipótese $H_0 : \theta = \theta_0$, no sentido de que se está for muito grande, elevado também é o volume de valores de θ mais plausíveis do que θ_0 .

De acordo com Murteira *et al* (2003), este último método tem sido utilizado no caso de estudos multiparamétricos, nos quais os procedimentos de teste de hipóteses ordinários não são pertinentes, e o cálculo das probabilidades pode ser bastante trabalhoso. O *Full Bayesian Significance Test* (FBST) busca avançar com estas técnicas no caso de vários parâmetros e hipóteses precisas, $\dim \Theta_0 < \dim \Theta$. Devido a sua relevância para o presente trabalho, o teste será estudado em detalhes na próxima seção.

II.II – FBST: FORMULAÇÃO METODOLÓGICA

Nos parágrafos anteriores procurou-se evidenciar que o tratamento e a mensuração do suporte estatístico de hipóteses precisas coloca uma série de dificuldades metodológicas para a abordagem Bayesiana tradicional, a qual procura responder às questões relacionadas com os testes de hipóteses através do Fator de Bayes e da utilização de procedimentos conceituais da Teoria da Decisão. A introdução de *prioris* com massa de probabilidade positiva na hipótese precisa foi uma das soluções encontradas por essa abordagem para contornar o problema.⁷

A mudança no modo de se avaliar a hipótese nula, partindo do cálculo de uma grandeza relativa baseada nas probabilidades das hipóteses em direção à mensuração da plausibilidade da hipótese nula em relação aos demais valores do parâmetro, tem se apresentado como uma alternativa de sucesso no tratamento das hipóteses precisas. Dentre as abordagens Bayesianas, distintas daquelas voltadas ao Fator de Bayes e à Teoria da Decisão, está o FBST. O teste foi inicialmente proposto por Pereira e Stern (1999) como uma forma de se avaliar hipóteses precisas através do cálculo de medidas de evidência.

Embora o FBST tenha alterado a maneira de como se extrair informação dos dados sobre a hipótese a ser testada, cabe observar que o teste não se mostra incompatível com a Teoria da Decisão, a qual segundo Stern (2004) se apresenta como alicerce da abordagem Bayesiana tradicional. Tanto o FBST quanto a Teoria da Decisão provisionam um sistema coerente para utilização dos conceitos da Probabilidade. Além disso, Madruga *et al* (2001), ao definir uma função perda para as decisões feitas a partir da implementação do FBST, garante a coerência do teste com a teoria.

De acordo com Stern (2004), o FBST também apresenta atributos positivos em relação aos procedimentos tradicionais Bayesianos no que se refere à metodologia estatística e a consistência lógica. Esses serão introduzidos ao longo da seção, à medida que se for

apresentando os conceitos do teste. Do ponto de vista epistemológico e ontológico, o autor argumenta que o teste está alinhado com a perspectiva construtivista da geração do conhecimento científico, visto que trata de modo coerente as hipóteses precisas, experimentalmente e teoricamente.

O FBST é metodologicamente considerado um teste “Bayesiano Completo” (*Full Bayesian*), visto que a elaboração de uma medida de evidência para a hipótese precisa depende somente da distribuição *a posteriori*. Especificamente, o teste não necessita de hipóteses adicionais, como uma *priori* positiva para a hipótese precisa. Além disso, o FBST é coerente com o princípio da Verossimilhança Bayesiano, segundo o qual a informação obtida dos dados observados está toda representada na função de verossimilhança.

A caracterização do FBST fica facilitada se pensa-se no teste sendo composto por duas partes, a primeira definida pela determinação de uma região tangente a hipótese nula, denominada *Highest Posterior Density Set* (HPDS), e a segunda pelo cálculo da credibilidade do conjunto HPDS. Embora as definições propostas nessas duas partes do teste tratem de procedimentos gerais, merece ser destacado que o teste foi concebido para se avaliar hipóteses precisas, nas quais a dimensão da hipótese nula é inferior à dimensão do espaço paramétrico.

Com o intuito de apresentar metodologicamente o teste, considere inicialmente a hipótese nula precisa, $H_0 : \theta \in \Theta_0$, no qual $\Theta_0 \subset \Theta \subseteq R^n$ e Θ é o espaço paramétrico. Seja θ^* o argumento máximo da *posteriori* $h(\theta/x)$ sob a hipótese nula, $\theta^* = \arg \max_{\theta \in \Theta_0} h(\theta/x)$. Assim, pode-se definir o intervalo HPDS como

$$HPDS = \{\theta \in \Theta / h(\theta/x) > h(\theta^*/x)\}. \quad (2.7)$$

O intervalo contém todos os valores do parâmetro θ para os quais a densidade *a posteriori* assume valores superiores ao maior ponto dentro do intervalo definido pela hipótese nula.

Para se calcular a credibilidade do conjunto HPDS integra-se a densidade *a posteriori* neste intervalo, sendo matematicamente expressa por

$$k = \int_{\theta \in HPDS} h(\theta/x) d\theta. \quad (2.8)$$

O complemento da probabilidade do conjunto HPDS é a medida de evidência da hipótese nula proposta por Pereira e Stern (1999),

$$Ev(H_0) = 1 - k. \quad (2.9)$$

Como uma medida de probabilidade do HPDS, a evidência a favor da hipótese nula varia no intervalo $[0, 1]$, sendo a hipótese sob avaliação tanto mais plausível quanto mais o valor da

⁷ A Estatística frequentista clássica também se depara com dificuldades metodológicas ao tratar das hipóteses precisas ou categóricas.

evidência se aproximar da unidade. A lógica inerente à medida de evidência considerada é a de que se a credibilidade do HPDS é relativamente elevada, então o conjunto de valores dos parâmetros pertencentes a Θ_0 ocupa uma região de baixa probabilidade na *posteriori*. Inversamente, quanto menor for a probabilidade do HPDS, mais plausível se torna a hipótese nula, visto que os valores dos parâmetros sob esta possuem densidades relativas elevadas.

A implementação dos cálculos para a obtenção da evidência a favor da hipótese nula é realizada em duas etapas, a primeira de otimização numérica e a segunda de integração não analítica, tais como integração numérica ou aproximação de Monte Carlo. Essas etapas descrevem respectivamente os procedimentos para consecução das partes do teste acima expostas. Através do processo de otimização se obtém o argumento θ^* que maximiza a *posteriori* sob H_0 , permitindo a definição do intervalo HPDS. No caso da integração, esta serve para a consecução da credibilidade do conjunto HPDS, e conseqüentemente da $Ev(H_0)$.⁸

Assim como nos demais procedimentos inferenciais Bayesianos, a resolução de integrais desempenha um papel importante no FBST. No entanto, em muitos casos o cálculo analítico das integrais para a obtenção da credibilidade do conjunto HPDS é problemática. Nesses casos, entre as alternativas disponíveis estão a integração numérica e os métodos de Monte Carlo. Em cenários multidimensionais mais complexos, até mesmo a resolução numérica se torna inviável, restando apenas as aproximações obtidas via simulação estocástica através dos procedimentos de Monte Carlo.⁹

No caso dos procedimentos de Monte Carlo com amostragem de importância, técnica aplicada quando se tem dificuldade em simular-se diretamente da distribuição *a posteriori* de interesse, Zacks e Stern (2003) expõem mecanismos para reduzir a variabilidade do estimador, tornando o procedimento de aproximação de integrais mais eficiente. Dessa forma, nos casos em que esse instrumental é necessário, a determinação da medida de evidência a favor da hipótese nula se torna mais precisa.

Além das melhorias nos procedimentos de implementação do FBST, tal como a apresentada por Zacks e Stern (2003), após a divulgação do teste outros estudos buscaram elevar o potencial de aplicabilidade. Madruga *et al* (2003) apresentaram uma distinta formulação do FBST, a qual permite tornar a medida de evidência invariante para transformações de coordenadas gerais do espaço paramétrico. Seja $\omega = \phi(\theta)$ uma

⁸ De acordo com Pereira e Stern (1999) e (2001), algoritmos computacionais eficientes estão disponíveis para otimização local e global, assim como para integração numérica, podendo ser implementadas em ambientes de programação amigáveis ao usuário.

⁹ Murteira *et al* (2003) discutem as técnicas de implementação do algoritmo de Monte Carlo ordinário e por amostragem de importância. Gelman *et al* (1998) avançam com essas técnicas tratando de Monte Carlo com Cadeias de Markov.

reparametrização própria, na qual ϕ é uma função mensurável e integrável. Os autores demonstram que sob certas condições a medida de evidência da $H_0 : \theta \in \Theta_0$ é a mesma da $H_0 : \omega \in \Omega_0$, na qual $\Omega_0 = \phi(\Theta_0)$.

Lauretto *et al* (2003) avançam na proposição de propriedades para a medida de evidência gerada através do FBST quando abordam o erro total decorrente da aceitação ou rejeição da hipótese nula incorretamente. Contudo, diferentemente da medida de evidência inicialmente proposta por Pereira e Stern (1999), os autores trabalham diretamente com a credibilidade do intervalo de valores do parâmetro com densidade superior ao máximo obtido sob a hipótese nula. Nesse sentido, estão medindo a evidência contrária a hipótese em consideração, e não a seu favor como até então se vinha tratando.

No caso de se procurar estimar as chances de rejeição de H_0 quando está for verdadeira, Erro Tipo I, pode-se optar pela determinação de um nível crítico para a evidência contrária a hipótese nula a partir do qual se decide pela sua não aceitação. A definição desse nível é feita através do cálculo de quantis. Utilizando os valores de θ^* como parâmetros da função de verossimilhança $L(x/\theta)$, simulam-se novos conjuntos de valores para a variável X, ou seja, $x_{ij}, i = 1, 2, \dots, n$ e $j = 1, 2, \dots, m$, sendo n o tamanho de cada amostra e m o número de amostras simuladas. Para cada nova amostra de $x_{\bullet j}^*$ calcula-se a medida de evidência $\eta_j^* = Evc_{\theta^*}(H_0)_j$. Com todas as medidas determinadas, estabelece-se o quantil $\lambda = q_{\alpha; n}(\theta^*)$, no qual $(1 - \alpha)$ é o nível de credibilidade desejado.

A estimativa do poder do teste, e conseqüentemente das chances de aceitação indevida de H_0 , Erro Tipo II, pode ser feita de modo semelhante a da determinação do nível de credibilidade. Inicialmente simulam-se conjuntos de valores $\hat{x}_{ij}, i = 1, 2, \dots, n$ e $j = 1, 2, \dots, m$, a partir da verossimilhança parametrizada por $\hat{\theta}$, o valor do parâmetro que maximiza a *posteriori* sem restrições. Para cada conjunto $\hat{x}_{\bullet j}$, determina-se a evidência contrária a hipótese testada, $\hat{\eta}_j = Evc_{\hat{\theta}}(H_0)_j$. A estimativa do poder do teste, $(1 - \beta)$, é dada pela fração das medidas de evidência $\hat{\eta}_j$ que são superiores ao nível λ . Nesse caso, β é uma estimativa da chance de se incorrer no Erro Tipo II.

Diferentemente de Lauretto *et al* (2003) que utilizam a simulação como forma de se avaliar as chances de erro ao se optar pela rejeição ou aceitação da hipótese nula, Stern (2006) desenvolve uma formulação analítica para determinar intervalos de credibilidade para a

medida de evidência contrária a H_0 . De acordo com o autor, para um conjunto inicial de valores de X crescente, $x_l, l \rightarrow \infty$, somada a condições de regularidades apropriadas e a um valor verdadeiro do parâmetro dado por θ^0 , a distribuição cumulativa da evidência contrária, $\bar{V}(c) = \Pr(Evc(H_0) \leq c)$, tem as seguintes características:

- Se H_0 é falsa, $\theta^0 \notin H_0$, então $Evc(H_0)$ converge em probabilidade para 1, e

$$\bar{V}(0 \leq c < 1) \rightarrow 0 ;$$

- Se H_0 é verdadeira, $\theta^0 \in H_0$, então $\bar{V}(c)$ converge em distribuição para

$$QQ(t, h, c) = Q(t - h, Q^{-1}(t, c)), \text{ na qual} \quad (2.10)$$

$$Q(k, x) = \frac{\Gamma(k/2, x/2)}{\Gamma(k/2, \infty)}, \quad \Gamma(k, x) = \int_0^x y^{k-1} e^{-y} dy, \quad (2.11) \quad (2.12)$$

$t = \dim(\Theta)$, $h = \dim(H_0)$ e $Q(k, x)$ é a distribuição cumulativa Qui-quadrado com k graus de liberdade.

II.III – FBST: SELEÇÃO DE MODELOS

O argumento de que os mecanismos de teste de hipóteses Bayesianos podem ser facilmente estendidos para o processo de seleção de modelos, conforme descrito no início do capítulo com o Fator de Bayes, também é válido no caso do FBST. Pereira e Stern (2001) expandem o escopo da aplicabilidade do teste, utilizando-o para comparar e selecionar modelos concorrentes de regressão linear múltipla.

Um aspecto positivo da utilização do FBST na seleção de modelos é a sua capacidade de penalizar a parametrização excessiva dos modelos. De acordo com Pereira e Stern (2001), sob condições gerais, o critério de avaliação do FBST possui um mecanismo de regularização intrínseco. Esse mecanismo de penalização está alicerçado no tratamento de distribuições *a posteriori* completas, ou seja, na permanência na dimensão original do vetor de parâmetros. Diferentemente do sugerido por Basu (1988), os parâmetros perturbadores não precisam ser eliminados dos cálculos estatísticos para a implementação do FBST.¹⁰

Pereira e Stern (2001) tratam no artigo da comparação e seleção de modelos aninhados, isto é, caso em que a partir de restrições no modelo mais geral obtém-se um modelo mais reduzido, denominado de aninhado. A fim de se ilustrar a técnica do FBST para esses tipos de modelos, suponha inicialmente que se tem um Modelo Geral dado por

¹⁰ Denomina-se parâmetro perturbador, ou *nuisance parameter* no inglês, o parâmetro que não faz parte do interesse do pesquisador para uma dada análise estatística.

$$y_i = X_i \boldsymbol{\beta} + \varepsilon_i, \quad (2.13)$$

no qual y_i é a variável dependente; X_i é um vetor de variáveis exógenas; $\boldsymbol{\beta}$ é vetor dos parâmetros a serem estimados de tamanho k ; ε_i é o termo de erro, suposto ser independente e normalmente distribuído com média zero e variância σ_0^2 para todo i , $\varepsilon_{oi} \sim N(0, \sigma_0^2)$, $i = 1, 2, \dots, n$.

Imagine que se deseja comparar o Modelo Geral com um Modelo Aninhado, entre os quais a única diferença reside no vetor de parâmetros. Seja $\boldsymbol{\beta}'$ esse vetor para o Modelo Aninhado. Especificamente, considere que $\beta_k = 0$ no vetor $\boldsymbol{\beta}'$.

No caso de se possuir uma informação inicial vaga ou difusa sobre os parâmetros em questão, $\boldsymbol{\beta}$ e σ , pode-se atribuir uma distribuição *a priori* não informativa, do tipo

$$h(\boldsymbol{\beta}, \sigma) \propto \frac{1}{\sigma}, \quad -\infty < \beta_j < \infty, \quad j = 1, 2, \dots, k \quad \text{e} \quad 0 < \sigma < \infty. \quad (2.14)$$

Considerando-se essa distribuição com a função de verossimilhança do Modelo Geral, e aplicando a fórmula de Bayes, obtém-se a seguinte densidade *a posteriori*, conforme descrito em Zellner (1971),

$$h(\boldsymbol{\beta}, \sigma / y, x) \propto \frac{1}{\sigma^{n+1}} \exp \left\{ -\frac{1}{2\sigma^2} \left[v s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right] \right\}, \quad (2.15)$$

na qual $v = n - k$

$$\hat{\boldsymbol{\beta}} = (X' X)^{-1} X Y, \quad (2.16)$$

$$s^2 = \frac{(Y - X \hat{\boldsymbol{\beta}})' (Y - X \hat{\boldsymbol{\beta}})}{v}. \quad (2.17)$$

A *posteriori* poder ser fatorada em uma parte com distribuição normal multivariada em $\boldsymbol{\beta}$ dado σ , com média $\hat{\boldsymbol{\beta}}$ e matriz de covariância $(X' X)^{-1} \sigma^2$, e uma distribuição marginal gama inversa para σ , na qual os parâmetros são dados por v e s^2 .

A implementação do FBST para a comparação e seleção de modelos é feita de maneira semelhante à aplicação do teste para avaliação de hipóteses precisas, descrito na seção anterior. A primeira etapa consiste na maximização da distribuição *a posteriori* sob a hipótese nula $H_0: \beta_k = 0$, resultando no vetor de parâmetros ótimos $\boldsymbol{\theta}^* = [\beta_0^*, \beta_1^*, \dots, \beta_{k-1}^*, \sigma^*]$ e concomitantemente na determinação do intervalo HPDS.

A etapa subsequente é dada pela integração da *posteriori* completa no intervalo HPDS, obtendo-se a credibilidade do conjunto de valores do parâmetro, ou seja, a evidência contrária à hipótese nula. Em casos multidimensionais como o apresentado, a aplicação do algoritmo de Monte Carlo com amostragem por importância é uma boa opção para se poder estimar a integral em questão.

A medida de evidência contrária a H_0 , gerada pela implementação do FBST, é um indicador da falta de plausibilidade do Modelo Aninhado. Valores elevados, isto é, próximos à unidade sugerem que o modelo reduzido não se ajusta bem ao conjunto de dados analisados, evidenciando que a variável X_k agrega informação para a modelagem de Y . Em termos da densidade, uma medida elevada para a evidência representa o fato de que a hipótese nula permeia uma região de baixa densidade. O inverso das constatações é válido quando a evidência contrária se aproxima de zero.

As técnicas de implementação do FBST para avaliação e seleção de modelos aninhados servem de base para a análise comparativa de modelos não aninhados, procedimentos estes que são o foco do presente trabalho. Conforme apontado anteriormente, modelos não aninhados são aqueles em que um modelo não é subconjunto do outro, não podendo obter um através da imposição de restrições no outro. Seguindo a linha da ilustração apresentada até o momento, um exemplo de modelo não aninhado pode ser expresso segundo

$$y_i = Z_i \gamma + \tau_i, \quad (2.18)$$

no qual y_i é a variável dependente; Z_i é um vetor de variáveis exógenas; γ é vetor dos parâmetros a serem estimados de tamanho l ; τ_i é o termo de erro, suposto ser independente e normalmente distribuído com média zero e variância σ_1^2 para todo i , $\tau_i \sim N(0, \sigma_1^2)$, $i = 1, 2, \dots, n$. Esse modelo não está aninhado no Modelo Geral dado que a matriz de valores $Z_{n \times l}$ não é um subconjunto de $X_{n \times k}$.

Considere que se deseja comparar o Modelo Geral com o Modelo Não Aninhado através da utilização do FBST, avaliando qual dos modelos melhor se ajusta a um determinado conjunto de dados. Nesse caso, o primeiro procedimento a ser executado é a definição de um modelo conjunto para os dados, denominado daqui a diante de Modelo Expandido. Esse modelo será formado pela junção de todas as variáveis explicativas singulares do Modelo Geral, com as variáveis singulares do Modelo Não Aninhado, mais as variáveis comuns em ambos os modelos. Estatisticamente esse processo é denominado de aninhamento mecânico. Algebricamente, para um cenário no qual a única variável em comum entre os modelos é a constante, tem-se

$$y_i = \omega_0 + X_i^{[c]} \omega_x + Z_i^{[c]} \omega_z + \mu_i, \quad (2.19)$$

no qual $Z_i^{[c]}$ é o vetor de variáveis independentes presentes somente no Modelo Não Aninhado, com exceção da constante; $X_i^{[c]}$ é o vetor de variáveis explicativas do Modelo Geral sem a constante; ω_x e ω_z são vetores de parâmetros; e $\mu_i \sim N(0, \sigma_E^2)$, para $i = 1, 2, \dots, n$.

Com a utilização de uma distribuição *a priori* não informativa nos parâmetros $\omega = [\omega_0, \omega_x, \omega_z]$ e σ_E do tipo apresentada anteriormente, a densidade *a posteriori* do Modelo Expandido tem a mesma forma da distribuição do Modelo Geral, alterando apenas o número de parâmetros a serem estimados. Em outros termos, a densidade do Modelo Expandido pode ser fatorada em uma parte com distribuição normal multivariada em ω , condicionada ao valor de σ_E , e uma distribuição marginal gama invertida para σ_E .

A implementação do FBST para a obtenção das evidências contrárias a cada um dos modelos considerados é realizada nos mesmos moldes daquela exposta para a comparação de modelos aninhados. Primeiramente se maximiza *a posteriori* sob a hipótese nula, determinando os parâmetros ótimos e o intervalo HPDS. Na seqüência integra-se essa densidade no intervalo descrito, resultando no cálculo da credibilidade do HPDS e conseqüentemente na medida de evidência desejada.

Se o intuito é avaliar a plausibilidade do Modelo Geral, pode-se utilizar o FBST para testar se a hipótese de que todas as variáveis singulares a este modelo não agregam informação ao Modelo Expandido, ou seja, $H_0 : \omega_x = \mathbf{0}$. Através da implementação do FBST calcula-se a evidência que os dados observados dão a essa hipótese, apontando para adequação ou não do modelo a estes. Identicamente se procede no caso da avaliação do Modelo Não Aninhado, testando-se $H_0 : \omega_z = \mathbf{0}$.

No caso da avaliação da plausibilidade do Modelo Geral através da hipótese $H_0 : \omega_x = \mathbf{0}$, a obtenção de uma evidência contrária à hipótese próxima a zero é um indicativo de que o modelo pode estar mal especificado, dado que a área correspondente a nulidade dos parâmetros singulares ao Modelo Geral permeia uma região de alta densidade na *posteriori*. Nesse sentido, a obtenção de um resultado como o acima descrito sugere que os dados provenientes do Modelo Geral agregam pouco poder explicativo frente às variáveis do Modelo Não Aninhado.

Por fim cabe destacar o fato de que a implausibilidade do Modelo Geral não remete a adequação do Modelo Não Aninhado. Caso se deseje avaliar o Modelo Não Aninhado, é necessário calcular a evidência da hipótese $H_0 : \boldsymbol{\omega}_z = \mathbf{0}$.

Capítulo III

Seleção de Modelos Não Aninhados: J-Teste e FBST

Nos capítulos anteriores foram apresentadas sucintamente diversas técnicas de validação, comparação e seleção de modelos. Destaque especial foi dado aos mecanismos dedicados à avaliação de modelos não aninhados. Especificamente, uma abordagem mais detalhada incidiu sobre o J-Teste e sobre o FBST. Utilizando estes testes pode-se analisar, em um conjunto de modelos não aninhados concorrentes, qual é a forma funcional mais adequada para se ajustar os dados considerados.

Ao longo das últimas décadas a importância desses mecanismos de avaliação comparativa de modelos tem aumentado. No que se trata do J-Teste, o cenário aparenta ser ainda mais promissor. No intervalo de vinte anos, 1984 a 2004, foram constatadas 497 citações do teste na literatura pertinente. As Ciências Econômicas aparecem como uma área do conhecimento na qual a aplicação do J-Teste tem se tornado evidente, dado a profusão dos debates teóricos que culminam na geração de modelos concorrentes não aninhados.

Com o intuito de comparar dois mecanismos de seleção de modelos cujos princípios geradores são antagônicos, o foco do capítulo foi a implementação do J-Teste e do FBST sobre áreas do conhecimento econômico nas quais o debate têm sido intenso. A proposição dessas aplicações é a de analisar a convergência de resultados no que se trata de identificar o modelo econométrico que melhor se ajusta a um conjunto de dados.

III.1 – CURVA DE PHILLIPS

Os Economistas têm discutido ao longo dos últimos 40 anos algumas das idéias decorrentes da proposição da Curva de Phillips. Durante todos esses anos muito foi escrito sobre o tema, concordância e oposição de argumentos foram observadas entre as distintas linhas de pensamento que estudam o assunto, e até mesmo um Prêmio Nobel foi conquistado a partir do desenvolvimento das conjecturas feitas por Phillips nos anos 50.¹¹

Embora existam argumentos sobre uma postulação anterior da teoria, a Curva de Phillips, tal como pensada contemporaneamente, foi inicialmente proposta por Phillips (1958). Em termos gerais a principal idéia do trabalho foi a conjectura de uma relação antagônica entre desemprego e taxa de variação dos salários, posteriormente tratada como inflação. A modificação em sentidos opostos das duas variáveis aconteceria devido ao fato de que em períodos de início de inflação as receitas das empresas cresceriam mais rapidamente do que os custos, com uma situação inversa no caso da deflação.

¹¹ Na década de 70 Milton Friedman foi agraciado com o Prêmio Nobel escrevendo a respeito do tema.

Bremmer (2003) usa a teoria inerente à Curva de Phillips para apresentar uma proposta de utilização do J-Teste para seleção de modelos. O autor especifica dois diferentes, mas não teoricamente implausíveis, modelos empíricos para a curva. Esses são não aninhados, possuindo formas funcionais distintas.

A primeira especificação, Modelo A, pode ser expresso segundo

$$\pi_t = \beta_0 + \beta_1\pi_{t-1} + \beta_2U_t + \varepsilon_t, \quad (3.1)$$

no qual π_t é inflação corrente no período t ; π_{t-1} é a inflação no período anterior, $t-1$; U_t é o desemprego corrente; e ε_t é um erro aleatório suposto ser normalmente distribuído com média zero e variância dada por σ_ε^2 .

O Segundo modelo configura uma variação sutil do Modelo A, sendo que a única modificação se refere ao modo como U_t entra na equação. Enquanto no Modelo A o desemprego é linearmente relacionado com a inflação corrente, no Modelo B é o seu inverso que faz a variável dependente se modificar. Em termos matemáticos o Modelo B é escrito da seguinte maneira

$$\pi_t = \gamma_0 + \gamma_1\pi_{t-1} + \gamma_2\frac{1}{U_t} + \mu_t, \quad (3.2)$$

no qual as variáveis têm os mesmos significados do modelo anterior, e μ_t é também um erro aleatório suposto ser normalmente distribuído com média zero e variância dada por σ_μ^2 .

Em ambos modelos a relação entre inflação corrente e desemprego corrente é oposta, mas no Modelo A o coeficiente β_2 é suposto negativo, enquanto que no Modelo B o parâmetro γ_2 é esperado ser positivo. Os dois outros parâmetros, β_1 and γ_1 são supostamente positivos, uma vez que visam representar a inércia das séries de inflação de um período de tempo para o outro.

Utilizando os dados para a economia dos EUA no período de 1974 a 2003, os resultados da estimação de mínimos quadrados das funções de regressão de ambos modelos são apresentados na Tabela III.i. Os valores das Estatísticas F mostram que ambas as formas funcionais são plausíveis para a modelagem dos dados. Além disso o poder de explicação é adequado, R^2 é 0,75 para o Modelo A e 0,71 para o Modelo B. No que se trata dos coeficientes das variáveis, pode ser observado que todos eles possuem os sinais esperados, sendo significantes ao nível de 10% ou menos. Por fim, os métodos de regularização como Critério de Schwarz e AIC são minimizados para o Modelo A, embora estes estejam muito próximos aos seus respectivos no Modelo B.

Tabela III.i
Curva de Phillips – Dados Anuais : 1974-2003
Variável Dependente: Taxa de Inflação Corrente (π_t)

Variáveis	Modelo A	Modelo B
Constante	4.91* (3.27)	-3.75 (-1.64)
Inflação Anterior (π_{t-1})	1.06* (8.47)	0.98* (7.55)
Taxa de Desemprego Corente (U_t)	-0.83* (3.00)	
Inverso da Taxa de Desemprego Corrente ($\frac{1}{U_t}$)		22.76*** (2.05)
R Quadrado	0.75	0.71
Estatística F	40.15♣	33.01♣
Critério da Informação de Akaike (AIC)	3.96	4.10
Critério de Schwarz	4.10	4.24

Fonte: Bremmer (2003). Os caracteres *, **, *** indicam que o coeficiente estimado é estatisticamente diferente de zero a um nível de significância de 1%, 5% e 10% respectivamente, utilizando teste *t* bi-caudal. ♣ indica que a estatística F é significativa ao nível de 1%.

Com a plausibilidade de ambos modelos, um questionamento relevante é a identificação de qual forma funcional melhor se ajusta aos dados. Bremmer (2003) utiliza o J-Teste para avaliar o poder de explanação do Modelo A e do Modelo B, para então apontar para a especificação da Curva de Phillips mais adequada. A teoria e a interpretação dos resultados do J-Teste foram apresentadas anteriormente, estando presente na Tabela III.ii as resultantes para este teste.

Tabela III.ii
Curva de Phillips – Dados Anuais: 1974-2003
Testes de Hipóteses Pareados para Modelos não Aninhados

	Modelo A	Modelo B
Modelo A	—	-2.58*** (-1.98)
Modelo B	2.78* (2.92)	—

Fonte: Bremmer (2003). Os valores apresentados são os coeficientes dos valores preditos pelo modelo da coluna inseridos dentro do modelo especificado na linha. Os caracteres *, **, *** indicam que o coeficiente estimado é estatisticamente diferente de zero a um nível de significância de 1%, 5% e 10% respectivamente, utilizando teste *t* bi-caudal. Entre parênteses está expresso o valor da estatística *t*.

As entradas da tabela mostram que o Modelo A tende a ter um ajuste superior ao modelo B. Quando os valores preditos de π_t pelo Modelo A são inseridos no Modelo B como regressores, a hipótese de não adição de poder explicativo, ou de um $\alpha = 0$ conforme descrito na teoria apresentada anteriormente, pode ser rejeitada a um nível de 1%. Inversamente,

quando os valores preditos de π_t pelo Modelo B são introduzidos no Modelo A, ao nível de 1% não se pode rejeitar a hipótese de nulidade do coeficiente. A constatação implica que a forma funcional do Modelo A contém informação importante na determinação da inflação corrente

Uma segunda maneira de se realizar a avaliação de modelos não aninhados é através da utilização do FBST, conforme exposto anteriormente. A fim de se comparar os resultados de dois mecanismos teoricamente distintos, implementou-se esse teste sobre os mesmos modelos e dados trabalhos em Bremmer (2003), e cujos valores resultantes são apresentados na Tabela III.iii.

No caso específico do problema considerado, o teste da significância estatística do Modelo A calculada a partir do FBST é realizado através da avaliação da hipótese nula de que o coeficiente do desemprego, U_t , seja igual a zero no modelo expandido. Do mesmo modo, a análise da plausibilidade do Modelo B é feita através do teste da hipótese de que o coeficiente do inverso do desemprego, $1/U_t$, seja igual a zero. Essa maneira de avaliação do ajuste dos modelos é pertinente somente devido ao fato de que a inflação do período anterior está presente em ambas as formas funcionais.

Os resultados da Tabela III.iii expressam um cenário no qual o Modelo A aparenta se adequar melhor aos dados trabalhados do que o Modelo B. A evidência contrária à hipótese de que o Modelo A seja inadequado, ou de que o coeficiente do desemprego U_t seja zero, está bastante próxima do valor 1, indicando a pouca plausibilidade da mesma. Inversamente, no caso do Modelo B, a evidência contrária à hipótese de nulidade deste é baixa, o que aponta para a aceitação de que o inverso do desemprego, $1/U_t$, tenha pouca influência sobre a inflação corrente. Dessa maneira, observa-se que tanto o J-Teste quanto o FBST apontam para um melhor ajuste do Modelo A sobre os dados.

Tabela III.iii

Curva de Phillips – Dados Anuais : 1974-2003
Full Bayesian Significance Test (FBST)

	Modelo A	Modelo B
Evidência Contrária a H0 (FBST)	0,989	0,014

Nota: A Evidência Contrária a H0 (FBST) de nulidade do modelo da coluna, a qual mostra a área da densidade *a posteriori* acima do máximo valor da densidade sob a Hipótese Nula.

III.II – REGIME POLÍTICO E TAXA DE JUROS

Um outro debate importante na Teoria Econômica está voltado para o entendimento do comportamento das taxas de juros reais. A questão básica levantada está relacionada com a

constância ou variabilidade dessas taxas ao longo do tempo. No caso de taxas passíveis de alteração, se torna relevante determinar quais as variáveis que tendem a afetar os juros. As mudanças políticas são apontadas pela literatura pertinente como uma importante fonte de variação, embora o modo como se dá a relação entre esta e as taxas de juros reais seja ainda motivo de não concordância entre autores.

Para os autores que trabalham com modelos Macroeconômicos, as mudanças nos regimes políticos tendem a ser consideradas como um fator relevante para o entendimento do comportamento das taxas de juros reais, tal como acontece nos modelos Keynesianos e Monetaristas com as alterações na política monetária. No sentido contrário estão os modelos financeiros, como as teorias de precificação de ativos, os quais desqualificam quaisquer efeitos dos fatores políticos sobre as taxas reais. Nesses modelos, as taxas de juros reais são afetadas apenas por fatores econômicos reais, e a política não é um deles.

Caporale e Grier (2000) são pesquisadores que procuram responder às questões apontadas acima a respeito da relação entre taxas de juros reais e mudanças políticas. No decorrer dos anos entre 1961 e 1986 nos EUA os autores identificam quatro administrações presidenciais (Kennedy-Johnson, Nixon-Ford, Carter, Reagan) e quatro diferentes lideranças no Banco Central Americano, FED, (Martin, Burns, Miller, Volcker). Sobre cada uma dessas esferas da administração pública a relação entre taxa de juros e alterações políticas é modelada no estudo dos pesquisadores.

O primeiro modelo investigado por Caporale e Grier (2000) é o Modelo da Administração Presidencial. Esse procura capturar a idéia de que modificações nos governantes da nação afetam a determinação das taxas de juros. Especificamente, o modelo por ser expresso através de

$$RR_t = \beta_0 + \beta_1(Reagan) + \beta_2(Carter) + \beta_3(NixonFord) + \varepsilon_t, \quad (3.3)$$

no qual RR_t é a taxa de juros real pós-fixada no tempo t ; *Reagan*, *Carter* e *NixonFord* são variáveis dummy que representam os distintos períodos de administração presidencial; e ε_t é o termo de erro com média zero e variância fixa.

O segundo modelo considerado segue a mesma intuição de modelagem da relação entre modificações políticas e taxas de juros reais, sendo as primeiras frutos de alterações na liderança administrativa do FED. O presidente do Banco Central Americano tem o poder de alterar a política monetária, aspecto que pode impactar nas taxas reais segundo os autores. Algebricamente o modelo pode ser escrito como

$$RR_t = \gamma_0 + \gamma_1(Burns) + \gamma_2(Miller) + \gamma_3(Volcker) + \mu_t, \quad (3.4)$$

no qual novamente RR_t é a taxa de juros real pós-fixada no tempo t ; Burns, Miller e Volcker são variáveis dummy que representam os diferentes períodos de lideranças do FED; e μ_t é o termo de erro com média zero e variância fixa.

Um último modelo trabalhado pelos autores está relacionado com a estimação estatística pura do número e localização dos pontos de quebra da série temporal analisada. Aplicando o método de otimização global desenvolvido por Bai e Perron (1998) nos dados trimestrais do período de 1961.1 até 1986.3, Caporale e Grier (2000) encontraram três quebras significativas situadas em 1967.1, 1972.4 e 1980.2.

Tabela III.iv

Modelos Alternativos da Variação na Taxa de Juros Real: 1961-1986

Variável Dependente : Taxa de Juros Real Ex Post (RR_t)

Variáveis	Presidential	FED	Estatístico Puro
Constante	1.59 (8.02)	1.52 (7.55)	1.82 (10.76)
NixonFord	-1.89 (3.26)		
Carter	-2.81 (5.37)		
Reagan	3.66 (10.56)		
Burns		-2.08 (3.84)	
Miller		-3.33 (9.07)	
Volcker		2.71 (3.05)	
1967.1-1986.3			-0.92 (4.67)
1972.4-1986.3			-2.46 (7.17)
1980.2-1986.3			6.27 (10.27)
R^2 Ajustado	0.630	0.476	0.632

Notas: Os números entre parênteses são os valores absolutos das estatísticas t .
Fonte: Caporale and Grier (2000).

Cada um dos pontos de quebra da série obtido via processo de otimização é considerado como um regressor das taxas de juros, mecanismo através do qual os autores configuram o terceiro modelo de análise. Esse modelo de estimação puramente estatística é tratado como uma referência para os outros dois modelos. Matematicamente, pode-se enunciar o modelo como

$$RR_t = \alpha_0 + \alpha_1(1967.1) + \alpha_2(1972.4) + \alpha_3(1980.2) + \delta_t, \quad (3.5)$$

no qual RR_t é definido como anteriormente; 1961.1, 1972.4 e 1980.2 são variáveis dummy que expressam os pontos de quebra da série; e δ_t é o termo de erro com média zero e variância fixa.

Os resultados provenientes da estimação de mínimos quadrados dos modelos de regressão apresentados acima para o período entre 1961.1 e 1986.3 são expressos na Tabela III.iv. Inicialmente, pode-se visualizar que os três modelos possuem valores relevantes para o R^2 Ajustado. Além disso, observa-se valores elevados da estatística t para os coeficientes de todos os modelos, os quais são significantes ao nível de 1%.

Considerando a significância estatística dos coeficientes dos três modelos e o poder de explicação pertinente do R^2 Ajustado, um questionamento natural que surge se refere à comparação do ajuste dos modelos. Uma forma que os autores propuseram para lidar com essa indagação é a utilização do J-Teste, o qual serve como mecanismo para confrontar os distintos modelos de alterações políticas, e também para avaliar a performance de cada um destes frente ao modelo puramente estatístico desenvolvido por Bai e Perron (1998).

Na tabela III.v se apresentam os resultados para os testes pareados entre os três modelos. Avaliando-se os valores, observa-se que o Modelo da Administração Presidencial se adéqua melhor aos dados do que o modelo de mudanças no FED. Quando os valores preditos para as taxas de juros reais, calculados a partir do Modelo do FED, são adicionados como regressores na equação presidencial, o coeficiente estimado para esta variável não é considerado estatisticamente significativo a um nível inferior a 10%. Contrariamente, visualiza-se na tabela que os valores preditos dos juros pelo Modelo Presidencial aumentam o poder de explicação da equação da gestão do FED.

Tabela III.v
Modelos Alternativos da Variação na Taxa de Juros Real: 1961-1986
Testes de Hipóteses Pareados para Modelos não Aninhados

	Estatístico Puro	Presidencial	FED
Estatístico Puro	—	3.34 (0.001)	-0.47 (0.64)
Presidencial	3.29 (0.001)	—	1.62 (0.11)
FED	12.52 (0.0001)	5.09 (0.0001)	—

Notas: Os valores apresentados são os resultados dos testes t para os coeficientes dos valores preditos pelo modelo da coluna inseridos dentro do modelo especificado na linha. As entradas em parênteses são os valores de significância do teste.
Fonte: Caporale and Grier (2000).

A conclusão parcial de um ajuste aos dados mais adequado obtida anteriormente se reforça quando o J-Teste é utilizado para comparar os modelos de mudanças políticas com o de estimação puramente estatística. Observa-se na Tabela III.v que o modelo de alterações no

comando o FED se mostra com pior ajuste do que o estatístico puro. No caso da avaliação do Modelo Presidencial frente ao de Estimação Estatística, os resultados são inconclusivos, haja vista que ambos os modelos incrementam o poder de explicação dos dados um do outro.

Seguindo o mesmo caminho analítico aplicado no caso do estudo da Curva de Phillips, o teste FBST é implementado como uma forma alternativa de avaliação e comparação do ajuste de modelos não aninhados. A idéia é a de novamente estar analisando como duas técnicas distintas de seleção de modelos se comportam na investigação da modelagem de um mesmo banco de dados.

Tal como o J-Teste, o FBST é implementado através de testes pareados. Para cada par a ser testado um modelo expandido é formado, sendo composto do intercepto mais as variáveis regressoras de ambos modelos a serem avaliados. Para se testar a hipótese de que o Modelo Presidencial se adéqua melhor aos dados do que o Modelo do FED, por exemplo, duas conjecturas são feitas. Primeiramente testa-se que todos os coeficientes das variáveis originárias do Modelo do FED são nulos. Posteriormente um segundo teste é realizado, no qual se avalia a nulidade dos coeficientes do Modelo Presidencial.

Tabela III.vi

Modelos Alternativos da Variação na Taxa de Juros Real:1961-1986
Full Bayesian Significance Test (FBST)

Modelos	Estatístico Puro	Presidencial	FED
Estatístico Puro	—	0.965	0
Presidencial	0.003	—	0.052
FED	1	1	—

Notas: A conjunção do modelo da coluna com o modelo da linha forma o modelo expandido. As entradas nas células mostram a Evidência Contrária a H0 (FBST) de nulidade do modelo da coluna, a qual representa a área da densidade a posteriori acima do máximo valor da densidade sob a Hipótese Nula.

Conclusões referentes a um melhor ajuste do Modelo Presidencial somente podem ser levadas a cabo se for possível rejeitar a falta de significância dos seus parâmetros no modelo expandido, assim como se aceitar a nulidade dos coeficientes provenientes do Modelo do FED. Os valores apresentados na Tabela III.v demonstram que esta é a situação encontrada no uso do FBST. A evidência contrária a H0 de que o Modelo do FED não se adéqua aos dados é de apenas 0,05, enquanto que a evidência contrária à nulidade do Modelo Presidencial é de 1,00.

A constatação de um ajuste mais apropriado do Modelo Presidencial é reforçada no momento em que comparações com o modelo estatístico puro são realizadas. O teste pareado do Modelo Presidencial e do puro estatístico através do FBST mostra que os coeficientes do último têm uma maior probabilidade de serem nulos do que os respectivos coeficientes da

forma funcional presidencial. No entanto, um resultado oposto a esse último é encontrado no caso do teste entre o modelo puro e o Modelo do FED. A evidência contrária à nulidade dos coeficientes provenientes da forma funcional do FED é nula, indicando que esta seja uma hipótese plausível.

III.III – AVALIAÇÃO DOS PROCEDIMENTOS DE TESTE

Conforme discutido no Capítulo II, quando se realiza um teste de hipóteses com o intuito de escolher uma das alternativas estipuladas existem dois tipos de erro que o pesquisador pode incorrer. Em um primeiro caso, denominado de Erro do Tipo I, o pesquisador rejeita a hipótese nula quando na realidade esta é verdadeira. Uma segunda possibilidade de erro decorrente da decisão do pesquisador acontece quando o mesmo aceita a hipótese nula, contudo é a hipótese alternativa que é verdadeira. Esse erro é tratado como Erro do Tipo II.

Em linhas gerais seria interessante desenvolver um procedimento de teste que reduzi-se ao máximo esses dois tipos de erro, ou seja, que no limite os torna-se nulos. No entanto, de acordo com DeGroot e Schervish (2002) a possibilidade de erro existe virtualmente em todo problema de teste de hipóteses, sendo improvável reduzi-los a zero. Nesse sentido, uma vez determinado os dois tipos de erro para os procedimentos de teste a serem adotados, cabe ao pesquisador escolher o procedimento que minimize o valor total do erro a ser incorrido.

O exposto acima remete ao pressuposto conhecido entre os estatísticos de que todo procedimento de teste pode ser caracterizado pelas chances de aceitação e rejeição da hipótese nula, ou em termos pragmáticos, pela determinação dos erros do Tipo I e II. A fim de auxiliar na caracterização dos testes, a teoria estatística se utiliza do conceito da função poder. Essa função específica, para cada valor do parâmetro a ser tratado, a probabilidade de que um dado procedimento de teste rejeite a hipótese nula.

Em termos matemáticos, suponha $H_0: \theta \in \Theta_0$ como a hipótese nula e $H_1: \theta \in \Theta_1$ como hipótese alternativa, na qual Θ_0 e Θ_1 é uma partição disjunta do espaço paramétrico Θ . Seja C a região crítica, isto é, o conjunto de valores de variável aleatória X analisada para os quais se rejeita H_0 . Nesses termos, a função que satisfaz a seguinte definição é denominada de função poder:

$$\pi(\theta/\delta) = \Pr(X \in C/\theta) \quad \text{para todo } \theta \in \Theta, \quad (3.6)$$

na qual δ é um dado procedimento de teste.

Utilizando-se da definição da função poder, o cenário desejado, apontado acima, de se minimizar a possibilidade de erro nos testes de hipóteses poderia ser expresso em termos de baixos valores de $\pi(\theta/\delta)$ para $\theta \in \Theta_0$, probabilidade do Erro do Tipo I, e elevados valores de $\pi(\theta/\delta)$ quando $\theta \in \Theta_1$, o complementar da probabilidade do Erro Tipo II. No entanto,

novamente esse não é um objetivo de um todo factível, dado que ao se tentar reduzir o valor de $\pi(\theta/\delta)$ para $\theta \in \Theta_0$ incorre-se também na baixa do valor da função poder na hipótese alternativa. Nesse sentido, torna-se central para um procedimento de teste achar um equilíbrio entre os valores de $\pi(\theta/\delta)$ na hipótese nula e na alternativa.

Nos capítulos anteriores argumentou-se que os procedimentos inferenciais dos testes de hipóteses podem ser facilmente generalizados para a comparação e seleção de modelos estatísticos distintos. Dessa maneira, torna-se pertinente aplicar o conceito da função poder para se avaliar os mecanismos de seleção de modelos descritos até o momento, ou seja, o FBST e o J-Teste.

Especificamente, no procedimento do J-Teste, a avaliação da plausibilidade de um modelo é feita testando-se as hipóteses $H_0: \alpha = 0$ e $H_1: \alpha \neq 0$, nas quais α representa o parâmetro dos valores preditos da variável Y inseridos no modelo de referência. Fixando-se um nível de significância de 5%, as chances do Erro Tipo I, pode-se construir a função poder para os demais valores do parâmetro contidos no espaço amostral, $\alpha \neq 0$, e se analisar o comportamento do Erro Tipo II a medida em que o valor de α se altera.

Gráfico III.i
Função Poder:
J-Teste para Avaliar Modelo A

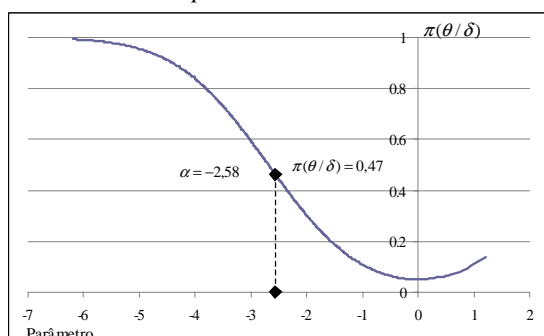
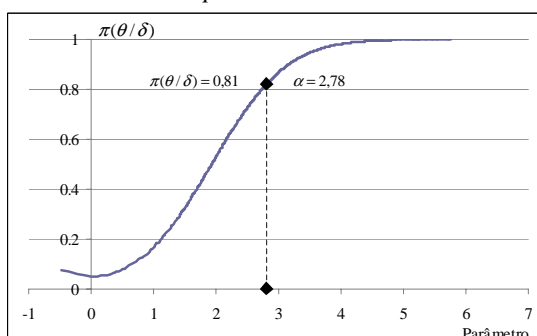


Gráfico III.ii
Função Poder:
J-Teste para Avaliar Modelo B



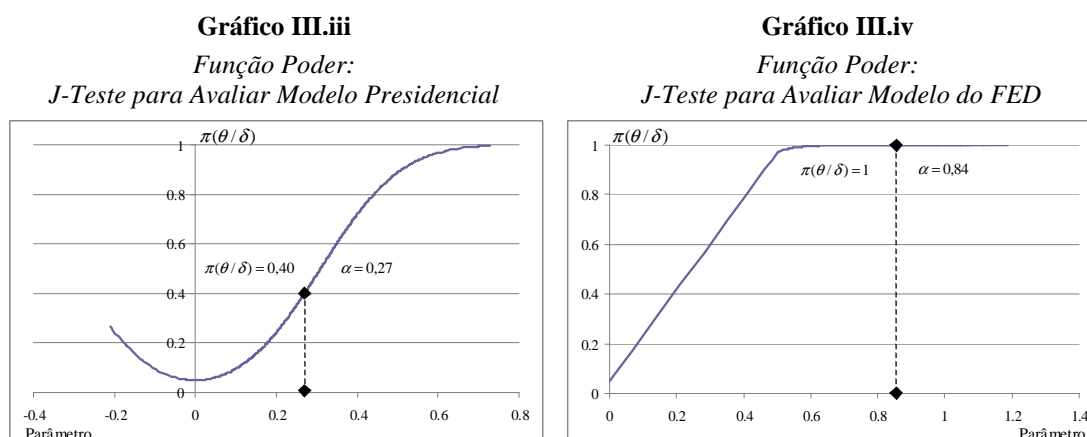
No caso da comparação dos modelos que buscam tratar da relação entre inflação e desemprego, Curva de Phillips, o teste da plausibilidade do Modelo A produz a função poder expressa no Gráfico III.i. Conforme pode ser observado, a medida em que o valor do parâmetro α se distancia de zero, a probabilidade do Erro Tipo II diminui paulatinamente, atingindo níveis inferiores a 1% quando $|\alpha| > 5,93$. Considerando-se o parâmetro no valor do estimador de mínimos quadrados (EMQ), $\hat{\alpha} = -2,58$, a função poder é 47%.

Do mesmo modo, pode-se calcular a função poder para o teste que avalia a pertinência do Modelo B, ou seja, no qual é o inverso da taxa de desemprego que se associa com a inflação corrente. O Gráfico III.ii apresenta a referida função para o teste tomando o nível de

significância como 5%. Nessa, constata-se que a probabilidade do Erro Tipo II também decresce com o distanciamento do parâmetro do valor zero. O erro torna-se inferior a 1% para $|\alpha| > 3,17$. Além disso, no valor do EMQ, $\hat{\alpha} = 2,78$, a função assume 81%.

No que se trata do problema de comparação e seleção dos modelos para a taxa de juros reais, pode-se utilizar os mesmos conceitos da função poder para avaliar o procedimento do J-Teste. O Gráfico III.iii expressa essa função para o teste pareado entre o Modelo da Administração Presidencial e o Modelo da Liderança Administrativa do FED, no qual o primeiro é modelo de referência a ser testado.¹² Observa-se que a função cresce rapidamente quando o valor do parâmetro se distancia de zero, tornando-se a probabilidade de Erro Tipo II inferior a 1% para $|\alpha| > 0,68$. No estimador de mínimos quadrados, $\hat{\alpha} = 0,27$, a função atinge 40%.

Invertendo-se os papéis dos dois modelos considerados, pode-se determinar a função poder para o teste de um melhor ajuste do Modelo da Administração do FED aos dados. A função para esse teste é demonstrada no Gráfico III.iv, a partir do qual se constata que o valor das chances de ocorrência do Erro Tipo II ficam menores que 1% para $|\alpha| > 0,56$. No que se trata da avaliação da função no EMQ, $\hat{\alpha} = 0,84$, o valor da função é 100%.



Assim como acontece com os procedimentos do J-Teste, a utilização do FBST para a comparação e subsequente seleção de modelos também está sujeita à ocorrência dos erros Tipo I e II. Em linhas gerais, para ambos os testes a origem dos erros reside no fato de que as observações da variável em análise possuídas pelo pesquisador representam apenas uma das suas inúmeras realizações possíveis. Especificamente no caso do FBST, a constatação acima

¹² Utilizou esses dois modelos em específico apenas como um exemplo da análise do J-Teste no problema em questão. As funções poder para os testes entre os demais modelos são apresentadas no Anexo I.

implica que para cada diferente elemento do espaço amostral ter-se-ia uma nova função de verossimilhança, alterando o valor da evidência. Dessa maneira, no FBST a avaliação dos erros é uma forma de quantificar o impacto de outras funções de verossimilhança sobre a evidência.

Diferentemente do J-Teste, para análise do qual é possível a construção da função poder objetiva, no caso do FBST o cálculo das chances dos erros Tipo I e II é feita através da simulação estocástica, conforme apresentado no capítulo 2. Basicamente, ambos os erros são obtidos com a determinação de novas evidências contrárias a hipótese nula a partir da simulação de conjuntos de amostras para a variável em questão. Inicialmente simulam-se novos dados para a variável através da função verossimilhança parametrizada pelo estimador de máxima verossimilhança (EMV) calculado na hipótese nula. O próximo passo é repetir a simulação, substituindo o valor dos parâmetros pelo EMV obtido no espaço paramétrico completo. Após determinar-se a evidência para cada novo conjunto amostral, estipula-se a região crítica, avaliando as chances da evidência contrária a hipótese nula pertencer a esta.

Simulando-se 1000 novos conjuntos de valores para a inflação corrente baseados no EMV obtido na hipótese nula e 1000 para o EMV calculado sobre o espaço paramétrico, pode-se determinar as chances dos erros para o teste que analisa a hipótese de não significância do Modelo A no tocante à Curva de Phillips. Para um nível de significância de 5%, a probabilidade estimada de uma medida de evidência estar localizada na região crítica, dado que os parâmetros pertencem à hipótese alternativa, é de 75%. Em outros termos, as chances do Erro Tipo II são de 25%. Inversamente, quando o FBST é implementado para avaliar o ajuste do Modelo B aos dados, isto é, a nulidade de seus parâmetros no modelo expandido, sob as mesmas condições acima expressas, as chances do Erro Tipo II sobem para 51%.

No caso dos modelos que analisam a relação entre taxa de juros reais e cenário político, é possível aplicar os mecanismos da função poder para avaliar a aplicação do FBST para testar a nulidade dos parâmetros do Modelo da Liderança no FED frente ao Modelo Administração Presidencial no modelo expandido.¹³ Considerando também um nível de significância de 5% e simulando 1000 novos conjuntos de taxas de juros parametrizadas pelo EMV na hipótese nula e pelo EMV no espaço paramétrico, estima-se a chance do Erro Tipo II em 71%. Ao se modificar a hipótese de teste para a não pertinência do Modelo da Administração Presidencial, ou seja, os seus coeficientes são zero, a estimativa da probabilidade de Erro Tipo II decresce para menos de 1%.

De um modo geral, os resultados acima apresentados permitem constatar que as chances de se cometer um erro ao se optar por uma das hipóteses testadas é bastante semelhante nos

dois procedimentos de teste analisados, o J-Teste e o FBST. Fixando-se o nível de significância em 5%, a chance do Erro Tipo II no caso do teste da plausibilidade do Modelo A para o J-Teste é de 53% nos EMQs. Equivalentemente, a aplicação do FBST para testar a hipótese de nulidade dos coeficientes do Modelo B tem estimativa de Erro Tipo II de 51%, com as evidências calculadas a partir dos EMVs.

Tabela III.vii
Curva de Phillips – Dados Anuais : 1974-2003
Chances do Erro Tipo II para um nível de significância de 5%

Testes	Modelo A	Modelo B
$H_0: \alpha = 0$ (J-Teste)	19%	53%
$H_0: \omega = 0$ (FBST)	25%	51%

Notas: No caso do J-Teste, a chance do Erro Tipo II é calculada para o teste que avalia a hipótese nula de que o coeficiente dos valores preditos pelo modelo da coluna, no modelo geral, é nulo. Além disso, o erro é determinado considerando os EMQs como valores dos parâmetros. No caso do FBST, o teste se refere à nulidade dos parâmetros das variáveis do modelo da coluna no modelo expandido. A estimativa da chance do Erro Tipo II é feita a partir de evidências para novos conjuntos de dados simulados de uma verossimilhança parametrizada pelos EMVs.

Da mesma forma que no problema colocado pela Curva de Phillips, os resultados das chances do Erro Tipo II para os procedimentos de testes que avaliam a modelagem da taxa de juros real também se mostram semelhantes. O J-Teste, que analisa a adequação do Modelo da Administração Presidencial testando a hipótese de que as variáveis do Modelo do FED não agregam informação no modelo geral, obtém probabilidade de erro de 60% nos EMQs. A implementação do FBST para comparar esses dois modelos, ao testar a nulidade dos parâmetros do Modelo do FED, tem como estimativa da chance do Erro Tipo II um valor de 71%. Esses valores dos erros são calculados para um nível de significância de 5% em ambos os procedimentos de teste.

Tabela III.viii
Modelos Alternativos da Variação na Taxa de Juros Real:1961-1986
Chances do Erro Tipo II para um nível de significância de 5%

Testes	Modelo Presidencial	Modelo do FED
$H_0: \alpha = 0$ (J-Teste)	0%	60%
$H_0: \omega = 0$ (FBST)	1%	71%

Notas: No caso do J-Teste, a chance do Erro Tipo II é calculada para o teste que avalia a hipótese nula de que o coeficiente dos valores preditos pelo modelo da coluna, no modelo geral, é nulo. Além disso, o erro é determinado considerando os EMQs como valores dos parâmetros. No caso do FBST, o teste se refere à nulidade dos parâmetros das variáveis do modelo da coluna no modelo expandido. A estimativa da chance do Erro Tipo II é feita a partir de evidências para novos conjuntos de dados simulados de uma verossimilhança parametrizada pelos EMVs.

¹³ A avaliação da implementação do FBST para os demais modelos é apresentada no Anexo I.

Considerações Finais

Ao longo dos últimos anos o interesse dos pesquisadores econômicos sobre as técnicas estatísticas Bayesianas tem aumentado. Por um lado, essa postura pode ser explicada pela boa performance da metodologia nas implementações empíricas. No entanto, o diferencial da abordagem reside na sua capacidade de funcionar como um mecanismo operacional unificado para incorporação de informação a priori, assim como para atualização da crença dos pesquisadores sobre as proposições teóricas a medida em que as informações disponíveis se alteram. Essas premissas são o que Zellner (1971) definiu como o motor do progresso da ciência, uma vez que permitem a aprendizagem a partir da experiência.

Embora a abordagem Bayesiana seja coerente com a lógica indutiva de geração de conhecimento científico, e conseqüentemente esteja apta a realizar os procedimentos inferenciais delineados nas Ciências Econômicas, dentro da modelagem econométrica a abordagem ainda desempenha um papel tímido em relação ao seu potencial. No que se refere especificamente a avaliação da adequação do ajuste de um modelo a um conjunto de dados, constata-se uma preponderância dos procedimentos inerentes a abordagem de Regressão Econômica Média, conforme conceituado por Gujarati (2000).

Paralelamente aos testes de especificação de modelos inerentes a abordagem da Regressão Econômica Média, é possível encontrar na literatura econométrica testes de diagnósticos para comparar e selecionar modelos estatísticos concorrentes. Particularmente, quando se trata de modelos não aninhados, o J-Teste aparece como uma ferramenta útil aos pesquisadores envolvidos com a aplicação estatística no estudo das variáveis econômicas.

O J-Teste foi inicialmente proposto por Davidson e MacKinnon (1981) como uma metodologia alternativa ao Teste F de restrições, na época o procedimento mais utilizado. Transcorridos alguns anos de sua formulação, o teste ocupa atualmente uma posição de destaque dentro das investigações econométricas. O teste tem sido utilizado como um mecanismo de testar a plausibilidade de um modelo frente a um ou mais modelos concorrentes, destacando-se o fato de que foi configurado para modelos não aninhados, ou seja, modelos que não podem ser expressos a partir de restrições lineares em outro modelo avaliado.

O J-Teste, mesmo aparecendo com frequência na modelagem econométrica, não é totalmente isento de críticas metodológicas. Um dos aspectos negativos do teste que pesa sobremaneira é o fato de que em alguns casos o resultado da aplicação do teste pode ser inconclusiva, tal como acontece quando os modelos avaliados são considerados ambos adequados ou no extremo oposto quando ambos são descartados. Além desse, outro aspecto de atenção

reside no fato de que o teste é formulado com base em propriedades assintóticas, podendo rejeitar o modelo verdadeiro com mais frequência para pequenas amostras.

Nesse sentido, dado a consistência teórica da abordagem Bayesiana com os procedimentos inferenciais utilizados nas Ciências Econômicas e também a existência de um potencial de desenvolvimento destas técnicas na econometria, o objetivo do trabalho consistiu em avaliar a aplicabilidade dos mecanismos da teoria Bayesiana para comparar e selecionar modelos estatísticos implementados em variáveis econômicas. Especificamente, o intuito do estudo era o de analisar a performance de um procedimento de teste Bayesiano, o FBST, frente ao J-Teste no que se trata da avaliação de modelos não aninhados.

No decorrer do trabalho observou-se que a abordagem Bayesiana tradicional procura responder às questões relacionadas aos testes de hipóteses, e conseqüentemente a comparação de modelos, através do Fator de Bayes e da utilização de procedimentos conceituais da Teoria da Decisão. No entanto, verificou-se que a introdução das hipóteses precisas coloca uma série de dificuldades metodológicas para a aplicação desses procedimentos, visto que a Fórmula de Bayes não permite atualizar uma probabilidade nula *a priori*.

Um caminho que as abordagens alternativas à ortodoxia têm trilhado a fim de contornar as dificuldades colocadas pelas hipóteses precisas é o abandono do cálculo do Fator de Bayes em direção da mensuração da plausibilidade da hipótese nula relativamente aos demais valores possíveis do parâmetro. Dentre essas abordagens encontra-se o FBST. O teste, inicialmente proposto por Pereira e Stern (1999), foi desenvolvido exatamente para tratar de hipóteses precisas sem a necessidade de pressupostos *ad hoc*, visto que a medida de evidência para a hipótese a ser tratada depende somente da distribuição *a posteriori*.

Pereira e Stern (2001) expandiram a fronteira de aplicação do FBST ao se voltarem para a determinação de medidas de evidência pertinentes à avaliação de modelos de regressão linear concorrentes. Particularmente, os autores utilizaram o teste com o intuito de comparar e selecionar modelos de regressão aninhados. Um diferencial positivo do FBST em relação a outros critérios de seleção de modelos consiste no mecanismo intrínseco de regularização que o teste possui, aspecto que lhe permite penalizar a parametrização excessiva dos modelos avaliados.

Assim como uma série de outros estudos que se seguiram aos artigos de Pereira e Stern, o presente trabalho também buscou aprimorar as técnicas de implementação do FBST. Em específico, procurou-se criar mecanismos que permitissem utilizar o teste para avaliar modelos de regressão linear não aninhados. O aspecto fundamental nesse processo é a formulação de um modelo expandido, a partir de um aninhamento mecânico dos modelos concorrentes. Uma vez

determinado esse modelo, as técnicas de implementação para a comparação de modelos não aninhados se assemelham aquelas delineadas para os aninhados.

O teste prático da consistência da proposta de implementação do FBST para avaliação de modelos não aninhados é feito confrontando os seus resultados com aplicações relevantes do J-Teste na literatura econométrica. O interesse consistiu na análise de convergência dos resultados, dado tratar-se de mecanismos de seleção com princípios geradores antagônicos. Em linhas gerais, pode constatar-se que os dois procedimentos de teste apontaram para conclusões semelhantes no que se refere ao ajuste dos modelos aos dados.

No caso da modelagem da relação entre inflação corrente e desemprego, expresso pelo contexto da Curva de Phillips, ambos os procedimentos de teste apontam para o Modelo A como aquele com a melhor adequação aos dados. No que trata da determinação de um modelo para a taxa de juros real frente às modificações no cenário político, novamente o J-Teste e o FBST convergem para a conclusão de que o Modelo da Administração Presidencial é mais plausível do que o Modelo do FED. A única divergência de resultados está presente no teste entre o Modelo Presidencial e o Modelo Estatístico Puro, para o qual o J-Teste apresenta um resultado não conclusivo, enquanto para o FBST o primeiro modelo se ajusta melhor aos dados.

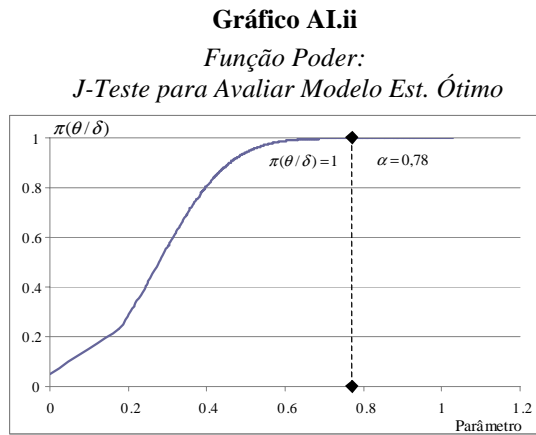
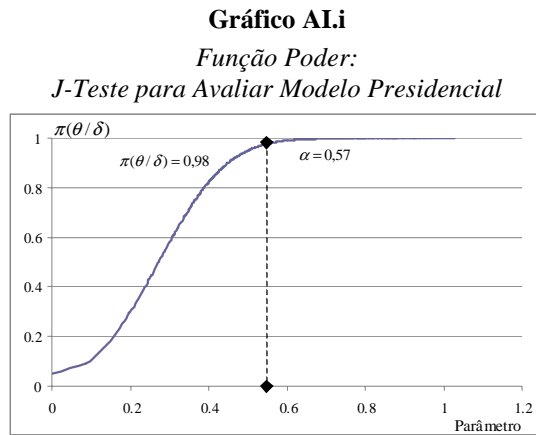
Por fim, utilizou-se a noção de que as chances de tomada de decisão errônea pelo pesquisador, Erro Tipo I e II, caracterizam um procedimento de teste para avaliar o J-Teste e o FBST. Com base na função poder e em um nível de significância de 5%, constatou-se que ambos os procedimentos possuem chances semelhantes de aceitação da hipótese nula quando essa não é verdadeira, o Erro Tipo II. Especificamente, no J-Teste observou-se esse último erro para valores dos parâmetros obtidos através dos EMQs, ao passo que no FBST o mesmo erro foi calculado a partir dos EMVs.

Nesse momento cabe argumentar que afirmações mais seguras sobre a performance do FBST na comparação e seleção de modelos econométricos não aninhados só são possíveis com a seqüência do estudo para outros casos práticos. No entanto, as constatações teóricas e empíricas acima apontam para a pertinência do teste para as inferências levadas a cabo na literatura econométrica.

Anexo I

Avaliação dos Procedimentos de Teste

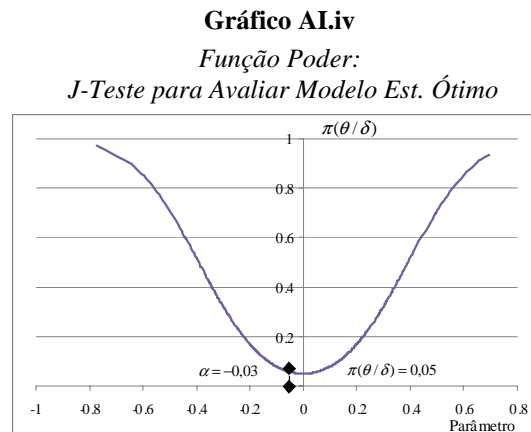
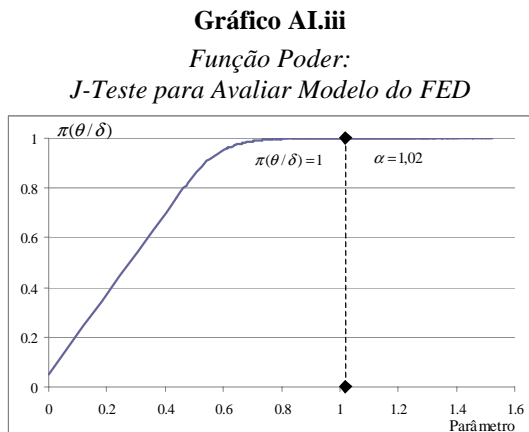
AI.I – FUNÇÃO PODER E J-TESTE: MODELO ESTATÍSTICO ÓTIMO VERSUS MODELO PRESIDENCIAL



Nota: Nos gráficos acima, o modelo avaliado equivale ao modelo de referência na implementação do J-Teste, ou seja, é neste que os valores previstos pelo modelo complementar do teste pareado são inseridos. Para maiores detalhes na metodologia do J-Teste vide Capítulo I.

As Funções Poder são especificadas para um nível de significância de 5%. O valor do parâmetro apresentado em cada gráfico se refere ao Estimador de Mínimos Quadrados para o coeficiente dos valores previstos.

AI.II – FUNÇÃO PODER E J-TESTE: MODELO ESTATÍSTICO ÓTIMO VERSUS MODELO DO FED



Nota: Nos gráficos acima, o modelo avaliado equivale ao modelo de referência na implementação do J-Teste, ou seja, é neste que os valores previstos pelo modelo complementar do teste pareado são inseridos. Para maiores detalhes na metodologia do J-Teste vide Capítulo I.

As Funções Poder são especificadas para um nível de significância de 5%. O valor do parâmetro apresentado em cada gráfico se refere ao Estimador de Mínimos Quadrados para o coeficiente dos valores previstos.

AI.III – PODER DO TESTE E FBST: MODELO ESTATÍSTICO ÓTIMO VERSUS MODELO PRESIDENCIAL

Tabela AI.i

*Modelos Alternativos da Variação na Taxa de Juros Real:1961-1986
Chance do Erro Tipo II para um nível de significância de 5%*

Testes	Modelo Presidencial	Modelo Est. Ótimo
$H_0: \omega = 0$ (FBST)	6,5%	73,5%

Notas: O teste se refere à nulidade dos parâmetros das variáveis do modelo da coluna no modelo expandido formado por ambos. A estimativa da chance do Erro Tipo II é feita a partir de evidências para 2000 novos conjuntos de dados da taxa de juros real simulados da verossimilhança parametrizada pelos Estimadores de Máxima Verossimilhança. Para maiores detalhes da metodologia de cálculo do Erro Tipo II para o teste FBST vide Capítulo II.

AI.IV – PODER DO TESTE E FBST: MODELO ESTATÍSTICO ÓTIMO VERSUS MODELO DO FED

Tabela AI.ii

*Modelos Alternativos da Variação na Taxa de Juros Real:1961-1986
Chance do Erro Tipo II para um nível de significância de 5%*

Testes	Modelo do FED	Modelo Est. Ótimo
$H_0: \omega = 0$ (FBST)	88,6%	0%

Notas: O teste se refere à nulidade dos parâmetros das variáveis do modelo da coluna no modelo expandido formado por ambos. A estimativa da chance do Erro Tipo II é feita a partir de evidências para 2000 novos conjuntos de dados da taxa de juros real simulados da verossimilhança parametrizada pelos Estimadores de Máxima Verossimilhança. Para maiores detalhes da metodologia de cálculo do Erro Tipo II para o teste FBST vide Capítulo II.

BIBLIOGRAFIAS DE REFERÊNCIA

- BAI, J. e PERRON, P. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, v.66, p.47-78, jan. 1998.
- BREMMER, D. S. J-Test: To Nest or Not to Nest, That Is the Question. *79th Annual Conference of the Western Economics Association International*. Vancouver, Canada, jul. 2003.
- BASU, D. Statistical Information and Likelihood. Edited by J. K. Ghosh. *Lect. Notes in Statistics*, 45, 1988.
- CASELLA, G. e BERGER, R. L. Reconciling Bayesian and Frequentist Evidence in the One-sided Testing Problem. *Journal of the American Statistic Association*, v.82, p.106-135, 1987.
- CAPORALE, T. e GRIER, K. B. Political Regime Change and the Real Interest Rate. *Journal of Money, Credit and Banking*, v.32, n.3, p.320-334, aug. 2000.
- DAVIDSON, R. e MACKINNON, J. G. Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica*, v.49, n.3, p.781-793, may 1981.
- DEGROOT, M. H. e SCHERVISH, M. J. *Probability and Statistics*. Addison-Wesley, 2000.
- EDDY, W. F. e GEISSER, S. A Predictive Approach to Model Selection. *Journal of the American Statistical Association*, v.74, n.365, p.153-160, mar. 1979.
- FOMBY, T. B.; HILL, R. C.; e JOHNSON, S. R. *Advanced Econometric Methods*. Springer-Verlag New York Inc., 1984.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; e RUBIN, D. B. *Bayesian Data Analysis*. Chapman & Hall, 1998.
- GUJARATI, D. N. *Econometria Básica*. São Paulo, Makron Books, 2000.
- LAURETTO, M.; PEREIRA, C. A. B.; STERN, J. M.; e SACKS, S. Full Bayesian Significance Test Applied to Multivariate Normal Structure Models. *Brazilian Journal of Probability and Statistics*, 17, p.147-168, 2003.
- MADRUGA, M. R.; ESTEVES, L. G. ; e WECHSLER, S. On the Bayesianity of Pereira-Stern Tests. *Test*, 10, p.291-299, 2001.
- MADRUGA, M. R.; PEREIRA, C. A. B.; e STERN, J. M. Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference*, 111, p. 185-198, 2003.
- MURTEIRA, B.; PAULINO, C. D.; e TURKMAN, M. A. *Estatística Bayesiana*. Lisboa, Fundação Calouste Gulbenkian, 2003.
- PEREIRA, C. A. B. e STERN, J. M. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy*, 1, p.99-110, 1999.
- PEREIRA, C. A. B. e STERN, J. M. Model Selection: Full Bayesian Approach. *Envirometrics*, n.12, p.559-568, 2001.

- PESARAN, M. H. e DEATON, A. S. Testing Non-nested Nonlinear Regression Models. *Econometrica*, v.46, p.677-694, 1978.
- PHILLIPS, A. W. The Relationship Between Unemployment and the Rate of Change of Money Wages in United Kingdom, 1981-1957. *Economica*, nov 1958.
- SACKS, S. e STERN, J. M. Sequential Estimation of Ratios, With Application to Bayesian Analysis. *Relatório Técnico*, MAC-IME-USP, dez. 2003.
- SAWA, T. Information Criteria for Discriminating Among Alternative Regression Models. *Econometrica*, v.46, n.6, p.1273-1291, nov. 1978.
- STERN, J. M. A Construtivist Epistemology for Sharp Statistical Hypotheses in Scientific Research. *Relatório Técnico*, MAC-IME-USP, out. 2004.
- SWAMY, P. A. V. B. A Comparison of Estimators for Undersized Samples. *Journal of Econometrics*, v.14, p.161-181, 1980.
- ZELLNER, A. *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons, 1971.
- ZELLNER, A. Bayesian Econometrics. *Economica*, v.53, n.2,p.253-269, 1985.