unesp **UNIVERSIDADE ESTADUAL PAULISTA**

"JÚLIO DE MESQUITA FILHO"

FACULDADE DE CIÊNCIAS E LETRAS

CAMPUS DE ARARAQUARA – SP

LUCAS MIKAEL DA SILVA DOS SANTOS

UMA ANÁLISE CRÍTICA SOBRE OS FUNDAMENTOS DA ECONOMETRIA **CLÁSSICA**: VALOR DE PROBABILIDADE

ARARAQUARA – SP

2021

LUCAS MIKAEL DA SILVA DOS SANTOS

UMA ANÁLISE CRÍTICA SOBRE OS FUNDAMENTOS DA ECONOMETRIA CLÁSSICA: VALOR DE PROBABILIDADE

Dissertação de Mestrado, apresentada ao Conselho, Programa de Pós-graduação em Economia da Faculdade de Ciências e Letras – Unesp/Araraquara, como requisito para obtenção do título de Mestre em Economia.

Área de Concentração: Métodos quantitativos aplicados e modelo de simulação de agentes

Orientador: Prof. Dr. André Luiz Correa

Bolsa: CAPES

ARARAQUARA – SP

2021

S237a

Santos, Lucas Mikael da Silva dos
Uma análise crítica sobre os fundamentos da econometria clássica: valor de probabilidade / Lucas Mikael da Silva dos Santos. -- Araraquara, 2021
103 p.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Letras, Araraquara
Orientador: André Luiz Correa

1. Econometria. 2. Estatística. 3. Modelos econométricos. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Letras, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

LUCAS MIKAEL DA SILVA DOS SANTOS

UMA ANÁLISE CRÍTICA SOBRE OS FUNDAMENTOS DA ECONOMETRIA CLÁSSICA: VALOR DE PROBABILIDADE

Dissertação de Mestrado, apresentada ao Conselho, Programa de Pós em Economia da Faculdade de Ciências e Letras – UNESP/Araraquara, como requisito para obtenção do título de Mestre em Economia.

Linha de pesquisa: Métodos quantitativos aplicados e modelo de simulação de agentes
Orientador: Prof. Dr. André Luiz Correa
Bolsa: CAPES

Data da defesa: 28/10/2021

MEMBROS COMPONENTES DA BANCA EXAMINADORA:

Presidente e Orientador: Prof. Dr. André Luiz Correa
Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP

Membro Titular: Prof. Dr. Alexandre Sartoris Neto
Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP

Membro Titular: Prof. Dr. José Ricardo Fucidji
Universidade Estadual de Campinas - UNICAMP

Local: Universidade Estadual Paulista
Faculdade de Ciências e Letras
UNESP – Campus de Araraquara

AGRADECIMENTOS

Com essas palavras, tento manifestar minha felicidade pela sorte de ter pessoas que me acompanham e que me permitiram realizar este trabalho.

Em primeiro lugar, agradeço aos meus aos meus pais, Geraldo e Antônia, e a minha irmã, Michelly, por todo o carinho, amor e compreensão. Deles sou uma extensão, e devo absolutamente tudo que conquisto aos valores, ensinamentos e conhecimentos que me proporcionaram.

Agradeço também ao meu orientador, Prof. Dr. André Luiz Correa, por toda a paciência e dedicação junto ao meu trabalho. Sua presença foi imprescindível não apenas para a execução desta pesquisa, mas para a minha própria formação.

Agradeço, ainda, todos os professores e alunos do programa de pós-graduação com o qual convivi durante este período, pelas experiências pessoais e acadêmicas divididas e compartilhadas. Todos foram importantes para a minha formação e para o meu desenvolvimento pessoal.

Por fim, agradeço aos meus amigos que me acompanharam durante esses anos e ao longo da minha vida. Para que eu não esqueça de citar alguém, involuntariamente, deixo aqui este agradecimento de maneira generalizada.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

Esta dissertação teve como pretensão investigar de forma crítica todos os pressupostos da econometria referente ao valor de probabilidade (p-valor) e questionar se a sua utilização deve ser vista como uma metodologia absoluta dentro dos modelos estatísticos em Economia. Para tanto, analisou-se de maneira detalhada toda a literatura acerca do debate sobre a inferência estatística: em especial, sobre as estimações a partir da estatística frequentista e bayesiana, ao mesmo tempo, que se desenvolveu um paralelo com a teoria econométrica. A pesquisa inicialmente detalha a discussão Fisher x Neyman-Pearson, fato que será importante para entender os principais problemas do p-valor e como todos esses obstáculos metodológicos foram transportados para a econometria. Após isso, é apresentado de maneira resumida todo o desenvolvimento da teoria econometria e como o p-valor se tornou uma pedra angular para a metodologia tradicional da mesma. Por fim, da mesma forma, também é destacado de maneira sucinta os principais fundamentos e dilemas da teoria bayesiana. Conclui-se que apesar da estatística bayesiana também ter diversos pontos a serem questionados referente a sua utilização, não existe uma justificava plausível para o domínio metodológico da inferência frequentista - fato que é ilustrado pelo uso do p-valor - dentro do mainstream econométrico.

Palavras-chave: p-valor, teoria econométrica, econometria bayesiana, inferência frequentista, inferência bayesiana.

ABSTRACT

This dissertation aimed at critically investigating all econometric assumptions regarding the probability value (p-value) and questioning whether its use should be seen as an absolute methodology within statistical models in economics. To this end, a detailed analysis was made of all the literature on the debate about statistical inference: in particular, about estimations from frequentist and Bayesian statistics, and, at the same time, a parallel with econometric theory was developed. The research initially details the Fisher x Neyman-Pearson discussion, a fact that will be important to understand the main problems of the p-value and how all these methodological obstacles were transported to econometrics. After that, the whole development of econometric theory is briefly presented and how the p-value became a cornerstone for the traditional econometric methodology. Finally, the main foundations and dilemmas of Bayesian theory are also briefly highlighted. The conclusion is that although Bayesian statistics also has several points to be questioned regarding its use, there is no plausible justification for the methodological dominance of frequentist inference - a fact that is illustrated by the use of the p-value - within the econometric mainstream.

Keywords: p-value, econometric theory, bayesian econometrics, frequentist inference, bayesian inference.

SUMÁRIO

1. INTRODUÇÃO

Nos estudos empíricos em economia que utilizam os fundamentos da econometria tradicional, frequentemente encontra-se um valor de probabilidade (p-valor), acompanhado do termo “estatisticamente significante” que, de maneira geral, orienta as principais evidências e conclusões do trabalho.

O p-valor — também conhecido como nível descritivo — é o procedimento de avaliação de amostras que fornece, caso as suposições estejam corretas, uma medida de contradição em relação à hipótese estatística postulada, podendo ser definido como o instrumento estatístico que indica um nível de significância mínimo para rejeitar uma hipótese nula (H_0).

Em síntese, uma hipótese estatística é um pressuposto examinado a partir de um conjunto de dados, em que cada observação pode ser vista como uma variável aleatória em algum espaço de probabilidade e que possui apenas dois resultados: verdadeiro e falso. A H_0 é a hipótese que, normalmente, contraria a evidência que um pesquisador busca comprovar em um teste estatístico.

Habitualmente considera-se que um p-valor de 0.05 (5%) é o patamar ideal para avaliar a H_0: se o nível descritivo for inferior a 5% pode-se rejeitar a hipótese nula; em caso contrário (maior que 5%), não existem evidências que permitam não aceitar H_0 (o que não significa automaticamente que seja verdadeira).

Os resultados do p-valor costumam ser interpretados equivocadamente pela maior parte dos pesquisadores. Embora, se estima um nível descritivo de 0.05, isto apenas ilustra que se a hipótese nula for verdadeira, a chance de um evento como esse ocorrer é de apenas 5%. O valor de probabilidade não prova que a teoria proposta é válida e muito menos indica que H_0 é verdadeiro (sim que é inesperado).

Um exemplo dos problemas oriundos do mau uso do nível descritivo refere-se a uma deliberação da suprema corte dos Estados Unidos em 2011[1], que condenou uma empresa fabricante de um determinado remédio de gripe, por não informar aos seus acionistas que a utilização desse medicamento poderia gerar uma doença chamada de anosmia. Os seus investidores tiveram perda financeira após a correlação entre o uso do remédio e a enfermidade se tornar pública.

[1] SUPREME COURT OF THE UNITED STATES. Matrixx initiatives, Inc., et alii. V. Siracusano et alii. Certiorari to the United States Court of Appeals for the Ninth Circuit, 2011.

Como argumento de defesa, a empresa ressaltou que os indivíduos que tiveram anosmia durante a realização da pesquisa clínica do medicamento, isto é, antes de sua comercialização, eram insignificantes aos níveis usuais da teoria estatística. Sendo assim, esta situação eximiria o relato desse episódio aos seus acionistas, tendo em vista que este não existia estatisticamente.

Isso posto, o presente trabalho busca analisar os problemas e contradições presentes nas estimações do p-valor e sua relação com a teoria econométrica clássica. Para tanto, tem como objetivos gerais os seguintes pontos: a) revisar a literatura sobre o valor de probabilidade; b) analisar de forma crítica o método; c) apontar os principais procedimentos que podem contornar os problemas apresentados pelo p-valor.

Apesar de ser uma importante ferramenta na apresentação de resultados, há alguns cuidados imprescindíveis na utilização do p-valor como fundamento para as conclusões de uma investigação científica. Logo, este projeto de pesquisa tem como justificativa de estudo a necessidade de debater as características, os perigos e as controvérsias no que se refere ao seu uso.

Os objetivos específicos do trabalho procuram analisar a forma como a econometria, ao adotar novos métodos de estimações estatísticas, pode superar as barreiras impostas pela inferência e escolha de um nível descritivo, tendo como norte mostrar possíveis instrumentos que minimizem os erros de estimação do p-valor.

Finalmente, ressalta-se que esta dissertação está dividida em 7 capítulos: i) introdução; ii) teoria do p-valor; iii) econometria clássica e p-valor: o contexto histórico da sua simbiose; iv) crítica ao uso do p-valor: porque a econometria clássica deve se atentar a isso; v) abordagem bayesiana: abordagem alternativa a econometria *mainstream*; vi) econometria bayesiana versus econometria clássica: deve ser construído um novo paradigma em econometria?; e, vii) conclusão.

2. A TEORIA DO P-VALOR

Este capítulo apresenta o valor de probabilidade (p-valor) a partir de três tópicos: a) no primeiro, é delimitado o seu conceito; b) no segundo, é destacado o seu uso e suas principais discussões; e, c) por último, o terceiro explora a discussão Neyman/Pearson x Fisher sobre a sua utilização.

2.1 DEFINIÇÃO

O p-valor é o mais importante procedimento frequentista usado em testes de hipóteses e permite estabelecer a significância estatística de um resultado. Em um teste de hipótese pretende-se avaliar a validade de uma afirmação sobre uma população, denominada como hipótese nula (e frequentemente representado por H0), a partir dos dados de uma amostra. Pode-se ter também uma hipótese alternativa (H1) caso se conclua que H0 não seja verdadeira.

O valor de probabilidade possibilita verificar se a população analisada é consistente com a hipótese nula, dado um certo nível de significância. Tradicionalmente, o valor de corte para rejeitar a hipótese nula é de 0,05, o que significa que, quando não há nenhuma diferença, um valor tão extremo para a estatística de teste é esperado em menos de 5% das vezes.

De uma maneira mais formal, o p-valor pode ser definido a partir de uma amostra de dados x (x_1, x_2, x_3...,x_n), cujo a função densidade de probabilidade (FDP) é conhecida por meio de um parâmetro θ, e tem como interesse um determinado valor de $\theta 0$ e θ. No caso mais simples, deseja-se testar os dados que apoiam $\theta = \theta 0$ em vez de $\theta = \theta 1$, onde $\theta 1 \neq \theta 0$ é outro valor específico de θ (SELLKE *et al.*, 2001).

Uma situação mais comum ocorre quando $\theta 1$ não é especificado e se está interessado em teste H0(hipótese nula):$\theta = \theta 0$ versus H1(hipótese alternativa):$\theta = \theta 0$. Isto é conhecido como um teste de hipótese bilateral, uma vez que sob H1 o verdadeiro valor de θ poderia ser menor ou maior do que $\theta 0$.

De forma geral, a abordagem frequente para o estudo destes e outros problemas é encontrar uma estatística de teste T(X), ou seja, uma função conhecida dos dados X de tal forma que grandes valores de t = T(x), sendo x o valor observado de X, são provas contra a hipótese nula. Uma forma padrão de "calibrar" esta evidência é calcular a probabilidade para observar T = t ou um valor maior sob a hipótese nula; esta probabilidade é conhecida como o p-valor:

$$p \equiv \mathbb{P}r\,(T \geq t \mid H0) \quad (1)$$

em que, o p-valor (p) corresponde à probabilidade ($\mathbb{P}r$) de se observar valores tão ou mais extremos (contra H0) que o valor obtido na amostra, caso a hipótese nula H0 seja verdadeira, ou seja, P (valores mais extremos contra H0 | H0 é verdadeira).

2.2 VALOR DE PROBALIDADE: DESENVOLVIMENTO E DISCUSSÕES

A discussão sobre a utilização do p-valor tem sua origem vinculada à obra *Statistical methods for research workers*, de Ronald A. Fisher. Neste estudo de 1925, Fisher destaca que os trabalhos estatísticos devem possuir como principal fundamento a ideia da alocação de uma infinita população sobre uma distribuição de possibilidade.

Segundo Fisher (1925), por meio de um experimento delimitado, é possível inferir resultados do universo hipotético infinito do qual uma amostra é extraída e, assim, determinar as características da população. Se uma segunda amostra rejeitar essa inferência, no contexto estatístico, pressupõe-se que essa população é de outro universo (não é significativa em relação a primeira). Os testes críticos que seguem esse método podem ser denominados como testes de significância.

Para realizar um teste de significância (avaliar o ajuste entre as observações e a hipótese), o autor recomenda a utilização da distribuição χ^2: este exame analisa a aferição dos valores observados que delimitam a amplitude da classe com os números hipoteticamente esperados. De forma simplificada, o valor do qui-quadrado pode ser calculado da seguinte maneira:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x^2_i}{m_i}\right) \quad (2)$$

Em que m é o valor previsto e x a diferença entre os números observados e esperados.

A identidade (1) indica que, para qualquer grau de liberdade (n), sendo n um número inteiro, é possível calcular a proporção de casos em que o valor de qui-quadrado é rejeitado. Essa dimensão é representa por P (p-valor), que é a probabilidade de o χ^2 ultrapassar qualquer nível de significância estipulado.

Algebricamente a relação entre o p-valor e o qui-quadrado é complicada, uma vez que o primeiro diminui de 1 para 0 e o segundo aumenta de 0 ao infinito. Em virtude dessa problemática, Fisher (1925) recomenda a utilização de uma tabela de valores

correspondentes para que χ^2 e o valor da probabilidade sejam empregados de forma prática.

De acordo com Fisher (1925), uma importante tabela deste tipo foi elaborada por William Palin Elderton, na qual o p-valor a cada seis casas decimais referia-se a um valor integral do qui-quadrado entre 1 e 30, e depois de 10 a 70. O grau de liberdade foi utilizado na forma de n´(n=1), em razão de Elderton acreditar que esta fórmula poderia se harmonizar com os valores da distribuição de frequências.

Devido às limitações impostas pelos direitos autorais, Fisher decidiu construir a sua própria tabela de qualidade de ajuste. Assim, os valores do χ^2 foram relacionados a um p-valor específico e o valor da probabilidade não foi restringido a um padrão arbitrário de valores do qui-quadrado. Fato que, conforme o autor, tornou o experimento estatístico mais adequado quando comparado à tabela de Elderton.

Fisher (1925) salienta que não buscou inferir o número exato do p-valor para qualquer qui-quadrado observado, mas, sim, investigar se o valor analisado está ou não aberto à rejeição. Se o valor da probabilidade estiver entre 0,1 e 0,9, certamente não há motivo para suspeitar da hipótese analisada. Contudo, caso esteja em um padrão inferior a 0.02 é fortemente recomendada a rejeição da hipótese, um p-valor de 0,05 seria um limite satisfatório.

Por fim, o autor aponta que o termo "qualidade de ajuste" gerou uma crença de que quanto maior o p-valor, mais aceitável seria a hipótese. Valores de 0,999 estão sendo relatados em alguns estudos, no entanto, se esses resultados fossem coerentes, indicariam uma situação pouco provável (uma evidência em mil tentativas).

Ampliando as contribuições de Fisher (1925), Neyman e Pearson (1933) procuraram aprimorar a teoria sobre os testes de hipóteses e, por consequência, a utilização do valor da probabilidade. Para tanto, definem primeiramente que um evento observado pode ser retratado pelo ponto de amostra (Σ), em uma população de n dimensões, sendo ilustrado pelas coordenadas $x_1,x_2...x_n$.

Posteriormente, afirmam que a hipótese nula do teste (H0) indica o ponto de partida do evento e determina a probabilidade de ocorrência – p-valor – $p_0=p_0(x_1,x_2...x_n)$ de todos os conjuntos de resultados. Na terminologia estatística Σ representa uma ou mais amostras, e H_0 pode ser uma hipótese do universo do qual elas foram extraídas.

Além disso, ressaltam que em qualquer tipo de análise, deve-se assumir uma classe de hipótese admissível C (H) – $H_1,H_2...H_n$ – como alternativa a H_0, e que as

investigações estatísticas devem priorizar três resultados: a) aceitar H_0; b) rejeitar H_0; c) questionar a coerência das evidências fornecidas pelo teste.

Desenhar um teste de hipótese como uma regra do tipo em que se aceita ou rejeita a hipótese nula, pode gerar dois problemas: rejeitar H_0 quando ele é verdadeiro (erro do tipo I) e aceitar H_0 quando uma alternativa H_n é verídica (erro do tipo II). Segundo Neyman e Pearson (1933), para um pesquisador contornar essa situação é necessário que ele minimize esses possíveis erros de estimação. A fim de demonstrar essa sugestão, os autores exemplificam o caso de uma hipótese simples.

Neyman e Pearson (1933) supõem que as hipóteses admissíveis sobre um conjunto do espaço amostral sejam $H_0,H_1...H_n$, e que uma ou outra seja verdadeira. Simplificando o exemplo, adotam o valor n + 1 dessa alternativa como finito. Denotam, também, ϕ como a probabilidade a priori da não rejeição dos pressupostos, que satisfaça essa condição:

$$\sum_{i=0}^{n} (\phi) = 1 \tag{3}$$

A partir desse ponto, o p-valor será fundamental para o teste de hipótese. Por exemplo, se A indica um evento observado e B sua hipótese, então P (A|B) representa a probabilidade do evento A, presumindo a veracidade de B. Da mesma forma, P (Σ|H_0) configura a chance da amostra Σ aceitar a hipótese verdadeira (H_0). Caso seja selecionada uma região w (área de rejeição de Σ), em um universo W, como forma de teste para hipótese nula, então a probabilidade que os valores Σ definidos por $x_1,x_2...x_n$ estejam em w, se H_0 for verdadeiro, pode ser escrito como:

$$P(\Sigma|H_0) = \varepsilon \tag{4}$$

Conforme Neyman e Pearson (1933), a possibilidade de rejeitar H_0 é igual a ε, e w pode ser dimensionado de ε para hipótese nula. O segundo tipo de erro acontecerá quando alguma alternativa H_n é aceita e ε rejeita $\overline{w} = W - w$. A chance de erro para a rejeição da região crítica pode ser apresentada da seguinte maneira:

$$P_I(w) = \phi_0 P(0|H_0) \tag{5}$$

$$P_{II}(w) = \Sigma^{n}_{i=1} \phi_n P(\overline{w}|H_n) \tag{6}$$

$P(w) = P_I + P_{II}$ (7)

Em que: P_I (w), P_{II} (w) e P (w) são as possibilidades de erros do tipo I, tipo II e a chance total de erro, respectivamente.

Outra contribuição relevante de Neyman e Pearson (1933) refere-se à discussão de um teste independente utilizando o p-valor a priori. Para os autores, esse fato implica na escolha de uma região crítica de tal forma que, caso a probabilidade P(w) rejeite a hipótese nula, P(w) necessita apresentar valores independentes de ϕ.

Com o objetivo de verificar sua dimensão, os autores encontraram essa região crítica através de manipulações algébricas das expressões (5), (6) e (7). As identidades (8) e (9) representam de forma simplificada uma condição suficiente e necessária para que P(w) seja independente da probabilidade ϕ_n,

$P(w|H_0) = P(\overline{w}|H_n) = \varepsilon \qquad (n = 1,2,...,i)$ (8)

Dado equação (2), a regressão (7) pode ser representada em (8)

$P(w = \varepsilon)$ (9)

Para Neyman e Pearson (1933), mesmo que uma região satisfaça a expressão (8) e seja determinada por ela, o p-valor tende a ser nulo. Esse fato tem como ponto de origem os conjuntos de hipóteses alternativas numerosas (ou infinitas), que se diferem apenas de maneira insignificante de cada $H_1, H_2, .., H_n$.

Quando, por exemplo, examinamos uma hipótese retirada de uma população normal com média α_0 e desvio padrão σ_0, a H_n de $\sigma_0 = \sigma_0 + \delta\sigma$ e $\alpha_0 = \alpha_0 + \delta\alpha$ será normalmente uma alternativa aceita. Em geral, seleciona-se uma hipótese H_1, seja qual for a região de rejeição, de modo que $P(w|H_1)$ não divirja muito de $P(w|H_0)$. Essa relação pode ser expressa da seguinte maneira:

$P(w|H_1) = P(w|H_0)$ (10)

(10) somado a (11):

$P(\overline{w}|H_1) = P(w|H_0)$ (11)

Resulta em aproximadamente (12):

$P(w|H_0) = \frac{1}{2}$ (12)

No caso da alternativa ser infinita, encontra-se uma série $H^{(1)}, H^{(2)}, ..., H^{(n)}$:

$\lim P(w|H^{(k)}) = P(w|H_0)$ (13)

dado que para qualquer região que aceite $P(w = \varepsilon)$, a identidade (12) é exata.

Portanto, a utilização das probabilidades (10) e (13) tem uma chance de apresentar erro de decisões igual a $\frac{1}{2}$, o que confirma a nulidade do pressuposto. Neyman e Pearson (1933) acreditam que é inviável realizar testes independentes que usam o p-valor a priori, devido à ausência de ϕ_n que impossibilita o cálculo da probabilidade P(w) e impede a análise da magnitude dos dois tipos de erros.

Ainda de acordo com os autores, durante o processo de criação de uma definição alternativa, o pressuposto a ser desenvolvido deve passar por alguns julgamentos. Por exemplo, comprovar as diferenças essenciais dos erros tipo I e II: i) os erros do tipo I, quando uma hipótese verdadeira é rejeitada, apresentam resultados geralmente idênticos independente da amostra analisada; ii) os erros do tipo II, H_0 aceita quando uma ou mais H_n é admissível, têm suas evidências dependentes das características de H_n e da sua disparidade em relação à H_0.

Por fim, Neyman e Pearson (1933) destacam que, ao investigar um teste de hipótese e comparar seus resultados com o número de observações $x_1, x_2,...,x_n$, é possível ter valores da $p_1,p_2,...,p_n$ rejeitados devido à escolha do nível restritivo (p-valor). Os autores recomendam que para a aplicação de um teste qui-quadrado com x_8 e p_8, por exemplo, a H_0 pode ser fixada em um nível baixo de rejeição:

$P(\chi^2 \geq \chi_0^2) = P(w|H_0) = \varepsilon$ (14)

tendo como limite um p-valor de 0,01 ou menos (χ_0^2 é o suposto valor observado).

Por outro lado, os experimentos científicos para detectar possíveis fatores que mudem o funcionamento de uma teoria padrão podem ter um corte maior, o p-valor de 0,1 é indicado para que o risco de erro tipo II seja minimizado. Segundo Neyman e Pearson (1933), a relevância de encontrar um novo paradigma supera qualquer preciosidade metodológica.

2.3 DEBATE NEYMAN-PEARSON X FISHER

É importante destacar que as duas abordagens não são concorrentes, uma vez que abordam o problema seguindo uma perspectiva diferente. Nos métodos frequentistas,

usualmente realizam-se os cálculos e adaptam-se as propriedades da metodologia de Neyman-Pearson. A conclusão é dada segundo a perspectiva de Fisher, rejeitando a hipótese nula e tomando tal decisão de acordo com a comparação do nível de significância pré-determinado e o valor da probabilidade obtido (SALSBURG, 2009).

Esse procedimento pode ser considerado como confuso, uma vez que são abordagens desenvolvidas com princípios diferentes e visando respostas distintas, e podem até resultar em conclusões antagônicas. Conforme Salsburg (2009), as principais diferenças nestes dois métodos são as hipóteses em teste e a arbitrariedade no uso do p-valor.

O estudo de Fisher concentra-se essencialmente no erro de tipo I, isto é, na probabilidade de rejeitar a hipótese nula quando está e, de fato, verdadeira. Neyman-Pearson adicionaram à preocupação de Fisher o erro do tipo II. Para estes estatísticos, a probabilidade de aceitar a hipótese nula quando, na realidade, esta é falsa também deveria ser controlada.

Fisher criticava a metodologia Neyman-Pearson partindo do argumento que a teoria dos autores era essencialmente infundada dado que o nível de significância não pôde ser definido como uma frequência relativa de amostragem repetida a partir da mesma população. Como destaca sua biógrafa, Joan Fisher (1978), os intervalos de confiança de Neyman-Pearson não afirmam nada sobre a probabilidade de o parâmetro ter um resultado de uma determinada amostra de dados efetivamente obtido.

Três razões contribuíram para esta crítica (FISHER, 1935). Primeiro, não existe a possibilidade de repetir a mesma população, porque os valores em estatística não podem ser fixados a partir de uma amostra a outra em virtude de que a população - ou o conjunto de referência – muda de amostra em amostra.

Em segundo lugar, as provas contra a hipótese nula não correspondem necessariamente à frequência relativa com que esse valor de probabilidade é atingido. Terceiro, Fisher (1935) apontou uma contradição lógica: se a hipótese verdadeira não for especificável, é irrelevante definir o erro tipo II em relação à hipótese alternativa.

Para Fisher, a hipótese nula simplesmente não pode ser considerada verdadeira apenas porque falhou ao ser confrontada, e a noção de erro do segundo tipo é incomputável sempre que dependa de uma hipótese desconhecida. Além disso, Fisher era favorável a pesquisa aplicada frente às abstrações matemáticas e, nominalmente, rejeitou a estatística como parte do raciocínio dedutivo.

Conforme Joan Fisher (1978), este ponto é o principal fator que diferencia os métodos de pesquisa de Fisher e Neyman. Enquanto, Fisher era um pesquisador que usava conhecimentos matemáticos, Neyman era um matemático que aplicava conceitos de sua área para validar todos os seus experimentos.

Dessa forma, Fisher concebeu sua investigação estatística como a produção de uma declaração de probabilidade direta para a classe limitada de casos para os quais isso seria possível. O argumento refere-se à determinação do p-valor, como a estatística do teste comparada com a sua conhecida distribuição sob H0 verdadeiras, e atribuindo ao investigador uma decisão de rejeição ou aceitação com base na experiência (FISHER, 1922).

Embora Fisher tenha evitado a má interpretação do p-valor como probabilidades de erro - uma vez que o valor obtido da probabilidade é uma função de um único conjunto de dados -, Neyman e Pearson (1928) criticaram esta abordagem como uma violação do princípio da frequência e, pelo menos no início, conceberam o seu próprio esforço como uma tentativa de corrigir e alargar este método à escolha de um teste axiomático eficiente e com um protocolo objetivo de aceitação ou rejeição.

Por sua vez, Fisher (1966) criticava a dualidade de uma hipótese nula versus hipótese alternativa introduzida por Neyman e Pearson (1928) e formando uma parte integrante do seu modelo de teste de hipóteses no contexto de tomada de decisões. Fisher era inflexível contra qualquer implicação de testes ou compromisso de escolher uma hipótese alternativa para escolher resultados não explicados pela hipótese nula.

Fisher (1966) destaca que é crível argumentar que uma experiência pode refutar a hipótese de que o sujeito não possui discriminação sensorial entre dois tipos diferentes de objeto, e, portanto, pode provar a hipótese contrária, que ela pode fazer alguma discriminação. Mas esta última hipótese, por mais razoável que seja ou verdadeira, é inelegível como hipótese nula a ser testada por experiência, porque é inexata.

Nisto, Fisher sustentou que a hipótese alternativa nunca pode ser encarada como uma hipótese a ser "anulada" porque não é precisa o suficiente para estar sob teste. A hipótese nula, como explicou Fisher (1966), deve ser exata, livre de imprecisões e ambiguidades, porque deve fornecer a base do "problema da distribuição", do qual o teste de significância é a solução.

Segundo Fisher (1966), uma vez que a hipótese alternativa não apresenta estas características, é inválido testá-la de qualquer forma com um teste de significância, e é questionável se pode inferir quando comparada com a hipótese nula. Quanto ao

tratamento da H0, Fisher ressalta a importar de notar que a hipótese nula nunca é provada ou estabelecida, mas é possivelmente refutada, no decurso da experimentação.

Portanto, para Fisher não importa quantas vezes uma hipótese nula não seja rejeitado, ela nunca em si mesmo está provada: uma H0 nunca pode ser demonstrada como verdadeira. Tudo o que o pesquisador pode esperar é possivelmente rejeitar a hipótese nula, que para Fisher era o propósito de usar testes de significância em um experimento.

Como apontado por Gigerenzer (1993), Fisher considera mais uma falácia concluir a partir de um teste de significância de que a hipótese nula é assim estabelecida; em que a maioria pode ser confirmada ou reforçada. A partir disto, parece que Fisher estava inclinado para uma confirmação da teoria da hipótese nula, mas esta inferência desvanece-se sobre a forma como se interpreta a sua utilização do termo "estabelecido" como sendo diferente do termo "confirmado".

Um outro componente que gerou discordância entre Fisher e Neyman/Pearson, ocorre pela escolha dos níveis de significância. Fisher foi vago quanto ao nível de importância que o investigador deveria adotar ao testar a sua hipótese nula. Esta ambiguidade não é uma característica surpreendente da sua contribuição sobre testes de significância.

Gigerenzer (1993), relata que os testes tinham uma qualidade notavelmente elusiva, mas no livro "*Statistical methods for research worker",* Fisher dá a possibilidade de os leitores interpretarem os testes de maneira bem diferente. Isso acontece muito por conta das recomendações de Fisher serem conflituosas, em que na mesma página (1928, p.13) é possível encontrar a sugestão da escolha de um nível descrito arbitrário e logo depois ele afirma que é habitual e conveniente que os experimentadores tomem 5%, como padrão do nível de significância.

Posteriormente, Fisher também argumentou, muito em resposta à definição "alfa" proposta por Neyman e Pearson (1928), que nenhum trabalhador científico tem um nível fixo de significância no qual, de ano para ano, e em circunstâncias difíceis, ele rejeita hipóteses; ele antes dá a sua mente a cada caso particular à luz das suas evidências e de suas ideias (FISHER, 1956).

Através destas passagens, Fisher deu uma instrução ambígua sobre quais os níveis de significância a utilizar e quando para os usá-los. Deve ser enfatizado, no entanto, que ele nunca deu a entender que o valor científico de um artigo deve ser julgado apenas com

base na escolha do nível descrito, ou que as decisões de publicação devem ser feitas ao cumprir este único critério.

Como relato em sua biografia, escrita por Joan Fisher (1978), o trabalho de Fisher foi dedicado ao que ele passou a considerar como o menor nível de inferência científica - a testes de significância que fazem uma dicotomia entre as hipóteses que são desacreditadas pelos dados e as que não o são.

Pode-se resumir a noção da escolha de níveis de significância de Fisher, por duas categorias. A primeira é a de um nível descrito padrão, que consiste em uma norma convencional (p-valor <0,05), que poderia ser adaptado pelos pesquisadores. Esta foi a posição inicial de Fisher.

A segunda posição tornou-se evidente perto do fim da sua carreira; o de um exato nível de significância, para o qual o nível (o nível exato, por exemplo, 0,01) foi anotado em suas últimas publicações (BERGER, 2003). Neyman e Pearson (1928), adotaram a visão inicial de Fisher, o que gerou mais um ponto discordância.

Para se defender das críticas de Fisher, Neyman, (1935) argumentou que é possível construir uma teoria da matemática e da estatística baseada unicamente na teoria da probabilidade e que a alicerce para tal teoria pode ser fornecido pela concepção da frequência dos erros em julgamento.

Esta era a abordagem que ele e Pearson tinham anteriormente descrito como "comportamento indutivo"; no caso de teste de hipóteses, o comportamento consistia em ou rejeitar a hipótese ou (provisoriamente) aceitá-la hipótese (NEYMAN;PEARSON,1928). Tanto Neyman como Fisher consideraram a distinção entre "comportamento indutivo" e "inferência indutiva" como mentira no centro do seu desacordo.

De fato, ao escrever retrospectivamente sobre a disputa, Neyman (1961) afirma que o objeto da disputa pode ser simbolizado pelos termos opostos "raciocínio indutivo" e "comportamento indutivo". O que Fisher (1973) aponta como algo horripilante no movimento ideológico representado pela doutrina que o raciocínio, propriamente falando, não deve ser aplicado a dados empíricos para levar a inferências válido no mundo real.

Outro ponto defendido por Neyman frente a Fisher, refere-se a ideia que a probabilidade é um conjunto de frequência de longo prazo em uma longa sequência de repetições em condições constantes (NEYMAN, 1957). Ampliando essa definição, Neyman (1977), ressalta que, pela lei dos números gerais, esta ideia permite uma extensão: se uma sequência de eventos independentes é observada, cada um com

probabilidade p de sucesso, então a frequência de sucesso a longo prazo será aproximadamente p mesmo se os eventos não forem idênticos.

Esta propriedade acrescenta muito aplicabilidade e coerência para uma probabilidade frequentista seguindo a lógica de Neyman/Pearson. Em particular, é a forma como Neyman veio a interpretar o valor de um nível de significância e contrapor o argumento de Fisher sobre o a arbitrariedade na escolha do nível descritivo.

Por outro lado, o significado da probabilidade é um problema com que Fisher se agarrou ao longo da sua vida. O conceito ao qual ele acabou por chegar é muito mais amplo do que o de Neyman. Fisher (1973) afirma que uma probabilidade é um conjunto de números (grandes), de entidades semelhantes, das quais uma proporção conhecida, P, tem alguma característica relevante, não possuída pelo resto. Mas nenhum subconjunto de todo o conjunto, tendo uma proporção diferente, pode ser reconhecida.

Em relação a hipótese alternativa, Neyman (1942) defende que é impossível ter testes de hipóteses apenas com a H0 e que, se os testes forem realmente concebidos sem consideração de qualquer coisa para além das hipóteses testadas, é porque os respectivos autores subconscientemente, tem em consideração certas circunstâncias relevantes, nomeadamente, a hipótese alternativa que pode ser verdadeira se a hipótese testada estiver errada.

Por isso, a teoria de Neyman-Pearson sugere, portanto, ao pesquisador escolher uma hipótese (geralmente) de ponto nulo e testá-la contra a hipótese alternativa. O seu enquadramento introduziu as probabilidades de se cometerem dois tipos de erros baseados em considerações relativas ao critério de decisão, tamanho da amostra e tamanho do efeito. Estes erros foram a falsa rejeição (erro Tipo I) e a falsa aceitação (erro Tipo II) da hipótese nula. A primeira probabilidade chama-se α, enquanto a segunda probabilidade é designada β (NEYMAN, 1950).

Neyman (1950), ainda afirma que este fato possibilita a minimização dos erros de estimação e em circunstâncias tão restritas, como em uma pesquisa de uma determina população aleatória infinita, a sua interpretação faz sentido: α representa a frequência a longo prazo dos erros de Tipo I e β é a contraparte para erros de Tipo II.

Em contrapartida, Fisher (1935,1966) aponta que os erros de Tipo II são cometidos apenas por aqueles que compreender mal a natureza e aplicação dos testes de significância. Além de que a noção de um erro do chamado “Tipo II”, devido à aceitação da H0 quando a hipótese é falsa, não tem significado no que diz respeito aos testes, em

que apenas as expectativas disponíveis são as que decorrem da hipótese nula ser verdadeira.

No entanto, Fisher (1966) dá pistas sobre a ideia do poder de um teste quando se refere à "sensibilidade" de uma experiência, ao indicar que ao aumentar o tamanho da população, pode-se torná-la mais sensível, permitir a detecção de um menor grau de discriminação sensorial, ou, por outras palavras, de um afastamento quantitativamente menor da hipótese nula.

Fisher (1966), ainda argumenta que uma vez que em todos os experimentos probabilísticos são cabíveis de refutação, mas nunca de provar a sua hipótese, é permitido dizer que o valor da probabilidade é aumentado sempre que permite que a hipótese nula seja mais facilmente rejeitada.

Como destacado inicialmente, a consideração do poder está ocasionalmente implícita nos escritos de Fisher. Essencialmente, o que Fisher considera como "sensibilidade" e Neyman-Pearson como “poder” é o mesmo conceito. Só que termina assim, puramente conceitual, esta convergência: o poder não tem qualquer papel metodológico na abordagem de Fisher, ao passo que tem um papel crucial em Neyman-Pearson.

Cabe ressaltar que diferente da visão de Fisher sobre a inferência indutiva centrada na rejeição da hipótese nula, Neyman e Pearson descartou toda a ideia de um raciocínio indutivo fora de controle. Em vez disso, o seu conceito de comportamento indutivo procurou estabelecer regras para a tomada de decisões entre duas hipóteses, independentemente de a crença do investigador em qualquer um deles.

Neyman (1950), explica que isso ocorreu, pois, o ato de aceitar uma hipótese significa apenas decidir tomar uma ação A em vez de uma ação B. Isto significa que se acredita necessariamente que a hipótese seja verdadeira, mas apenas que o teste prescreve a ação B e não implica que se acredita que a hipótese é falsa.

Assim, a teoria de Neyman-Pearson contrapõe a ideia de raciocínio indutivo utilizado por Fisher (1928) pelo comportamento indutivo. De acordo com Neyman (1971), a descrição da teoria da estatística envolvendo uma referência ao comportamento, por exemplo, estatística comportamentalista, foi introduzida para substituir o que tem sido denominado raciocínio indutivo.

O termo “comportamento indutivo” significa simplesmente o hábito de humanos e outros animais para ajustar as suas ações à frequência dos eventos notados, de modo a evitar consequências (NEYMAN,1961). Ao defender a sua preferência pelo

comportamento indutivo em detrimento da inferência indutiva, Neyman (1950) ressaltou que ao estabelecer esta regra, a teoria da probabilidade e estatísticas desempenham ambos um papel importante para denotar o ajustamento do comportamento humano a quantidades limitadas de informação.

Depois de ter em conta esses conselhos, o pesquisador conceberia uma oportunidade para controlar as probabilidades das taxas de erro α e β: o "melhor" teste é aquele que minimiza β sujeito a um limite em α. Ao determinar o que deveria ser este limite em α, Neyman declarou ainda que o controle dos erros de Tipo I eram mais importantes do que os erros de Tipo II.

Segundo Neyman (1950), o problema de testar hipóteses estatísticas é de selecionar regiões críticas. Ao tentar resolver este problema, é preciso lembrar que o objetivo de testar hipóteses é evitar erros na medida do possível. Assim, ao selecionar os testes, é preciso fazer um esforço para controlar a frequência dos erros Tipo I e depois pensar em erros do II.

O procedimento consiste em fixar arbitrariamente um pequeno número α e para exigir que a probabilidade de cometer um erro tipo I é necessária não exceder α. Neyman (1950), referiu-se a α como o nível de significância de um teste, um número pré-atribuído, tais como α = 0,05 ou 0,01, etc.

Finalmente, ressalta-se na opinião deste autor que a teoria estatística moderna passou por muitos grandes avanços: a extensão da probabilidade desde jogos de azar até ao conceito de medição e a sua extensão ao conceito da própria natureza. Como consequência, até ao final do século XIX, a estatística foi remodelada pela biologia, empenhada no desenvolvimento e não muito na redução do erro (ver mais em Gigerenzer *et al.*, (1989)).

Karl Pearson e Ronald A. Fisher emergiram como os construtores da estatística moderna, uma vez que eram ambos experimentalmente e teoricamente envolvidos na explicação da variação, primeiro sob a biometria e depois sob a exploração estatística na síntese de Darwin-Mendel.

A base empírica da biologia favoreceu a adoção de uma abordagem frequentista, mas foi dentro deste campo que surgiu uma fenda entre os testes fisherianos a teoria de Neyman-Pearson (em especial, os testes de hipóteses). Esta cisão gerou uma distinção entre os conceitos de probabilidade, de amostra e população, de inferência e como consequência, as definições dos próprios testes.

Ambos os concorrentes definiram claramente as suas teorias no campo em que procederam: para Fisher, postular uma população hipotética infinita era trivial, uma vez que a extração de uma pequena amostra de entidades biológicas não alteraria a natureza da população, enquanto que para as entidades matematicamente abstratas definidas por Neyman, a probabilidade pode ser concebida como a frequência das sucessivas extrações de amostras da mesma população.

Para os economistas, em vez disso, nenhum dos conceitos se aplica trivialmente e essa foi a razão de tanta resistência e erro de compreensão. Pois em qualquer contexto econômico realista, as extrações repetidas não são independentes, dada a dependência do tempo e, além disso, não há é uma população estável ao longo do tempo, dada a mudança estrutural.

Haavelmo (1944) abordado este problema com uma visão meta-histórica com a incorporação da abordagem de amostragem repetida, a fim de sustentar a aplicabilidade da teoria de Neyman-Pearson. Neste sentido, ele hibridizou as teorias de Fisher e Neyman-Pearson em estatística para a econometria. A consequência foi que as próprias variáveis econômicas deviam ser reconsideradas como processos estocásticos, se esta narrativa fosse coerente.

Mas o preço para isto é elevado suficiente e não foi imperceptível para os economistas contemporâneos, uma vez que isso requer o abandono do determinismo rigoroso: a economia já não podia ser escrita como a exploração em determinação mecânica, em encontrar leis de comportamento e de ação e consequência, mas deve antes consistir na descrição das leis de distribuição e da determinação das probabilidades derivadas das mesmas.

Toda esta problemática será tratada a partir dos próximos capítulos, com uma ligação direta entre a econometria e a inferência estatística.

3. ECONOMETRIA CLÁSSICA E P-VALOR: O CONTEXTO HISTÓRICO DA SUA SIMBIOSE

O termo econometria foi criado por Ragnar Frisch, co-vencedor do primeiro Prémio Nobel em Ciências Econômicas em 1969, juntamente com o Jan Tinbergen. Ambos foram fundadores da *Econometric Society* em 1933. Este grupo é uma sociedade internacional para o progresso da teoria econômica na sua relação com a estatística e a matemática, e tem como o objetivo promover estudos com o objetivo de unificar as abordagens teórico-quantitativa e empírico-quantitativa a problemas econômicos.

No primeiro volume da *Econometrica* (1933), a revista da *Econometric Society*, Ragnar Frisch destaca que a econometria é a unificação de três pontos de vista: estatístico, matemático e da teoria econômica. O seu conceito está relacionado a simbiose destas três áreas de estudos, que constitui toda a sua teoria e aplicação.

Atualmente, considera-se também que a econometria é o estudo combinado da estatística matemática, e modelos e dados econômicos. Dentro do seu espoco de análise, a teoria econométrica pode ser distinguida da econometria aplicada. A teoria econométrica diz respeito ao desenvolvimento de ferramentas e métodos, e o estudo de suas propriedades: pertence ao campo da estatística.

Já a econometria aplicada é um termo que descreve o desenvolvimento da estatística quantitativa, dos modelos econômicos e da aplicação de métodos econométricos a estes modelos utilizando dados econômico (toda a sua área de atuação está dentro principalmente no campo da economia aplicada).

Pode-se destacar 3 objetivos centrais da utilização da econometria: a) O conhecimento da economia real: permite estimar as magnitudes econômicas; b) Predição ou previsão: prever as variáveis ou relações econômicas no futuro; e, c) Política de simulação econômica: pode ser usa para simular os efeitos das políticas usuais ou alternativas.

3.1 O INÍCIO DAS PESQUISAS ECONOMÉTRICAS

Uma primeira tentativa de trabalho econométrico é encontrada já na década de 1870, com uma série de publicações de William Stanley Jevons e em 1914 com Henry L. Moore, "*Economic Cycles – Their Law and Cause*". Ambas as abordagens representam uma ideia de ruptura com a dominante metodologia econômica da época, uma vez que

ambos, Jevons e Moore, abandonaram o método bem estabelecido de introspecção e entraram na lógica da indução, confiando em dados externos (MORGAN, 1990).

Conforme Morgan (1990), apesar do fato de as teorias defendidas por Jevons e Moore terem sido largamente rejeitadas pelos seus contemporâneos, as contribuições metodológicas destes autores constituiu um grande avanço, que ditou importantes contornos para o trabalho econométrico a seguir.

Primeiro, Jevons (1866, 1874, 1884) previu uma combinação de causas endógenas e exógenas, que em conjunto determinaria os movimentos dos ciclos econômicos. As causas endógenas foram, naturalmente, devidas às forças econômicas internas, enquanto as causas exógenas foram marcadas pelas variações na atividade solar, que afetaram a colheita e isto, por sua vez, o ciclo da economia.

Enquanto muitos economistas do período estavam ocupados a estudar as crises de uma forma separada – e não prestou realmente muita atenção à ideia de Jevons sobre a explicação da "mancha solar do ciclo econômico" - Jevons estava à procura da existência de um ciclo exato subjacente: as análises separadas das crises não eram importantes. Os aspectos realmente significativos foram os padrões ou médias características dos dados a partir das quais se poderia obter uma explicação geral de um fenómeno cíclico geral.

A contribuição primordial de Moore (1914), por outro lado, reside no abandono da metodologia *a priori* e do raciocínio *ceteris paribus*, que consistia em comparar dois situações estáticas. Em vez disso, estava interessado na dinâmica da economia e na forma como a economia passou de uma situação para outra. Segundo Moore (1914), o mero exercício da estática comparativa nada diz ao economista sobre a verdadeira dinâmica da economia, que se alterna de forma constante e mudam como o mar.

Na sua abordagem, como na de Jevons, Henry L. Moore tenta demonstrar a existência de periodicidade nas flutuações econômicas, mas o seu foco estava nas interações econômicas envolvidas nesses ciclos e na análise dessas relações, e não na correlação entre o ciclo econômico e o fator causal exógeno.

Outros avanços notáveis foram realizados por George Udny Yule e Eugen Slutsky na década de 1920. Estes autores utilizaram procedimentos estatísticos para gerar artificialmente dados econômicos em condições teoricamente conhecidas e controladas, para fornecer uma norma para comparação e investigar o comportamento dos processos analisados sob certas condições.

Slutsky (1927) propôs um experimento para provar a sua hipótese de que os ciclos econômicos poderiam ser causados pela combinação de eventos ou choques aleatórios. É

interessante destacar, contudo, que o objetivo do Slutsky era mostrar que esse processo de estimação poderia gerar evidências muito semelhantes aos produzidos pela atividade econômica.

Já Yule (1927), construiu um modelo de regressão linear (de forma que $X_t = B_1X_{t-1} - B_2X_{t-2}$), a fim de simular o comportamento dos dados econômicos. Yule abordou o problema a partir da ideia de oscilações harmônicas como ponto de partida, em que se desenvolve a teoria do movimento de um pêndulo que está sujeito a uma força de flutuação contínua.

Em termo gerais, Slutsky e Yule abriram a porta para o uso de modelos de probabilidade na análise de séries de tempo em economia. A visão deles era ver as séries temporais como respostas lineares aos impulsos ou choques atuais e passados independentes e distribuídos de forma idêntica. Em contribuições distintas, eles mostraram como gerar ciclos aproximados com tais modelos.

Ao contrário da maioria dos autores anteriores, os interesses de Frisch centraram-se na metodologia pura da econometria, em vez da economia aplicada. O seu modelo de "cavalo de balanço" pretendeu ser um guia para economistas, embora não buscasse representar uma teoria do ciclo econômico (FRISCH, 1933).

O modelo de Frisch (1933) apresenta duas contribuições que foram aspectos importantes para pesquisa econométrica da primeira metade do século passado. Por um lado, ele desenvolveu um sistema de equações inter-relacionadas, que poderiam gerar um comportamento cíclico em função dos parâmetros estimados. Por outro, reconheceu a importância de choques aleatórios que foram vistos como perturbações reais em vez de erros de medição.

Esta linha de evolução termina com o importante trabalho de Tinbergen (1935), que deu origem aos primeiros modelos macroeconométricos. No qual, graças a uma combinação de avanços teóricos das obras de Yule, Slutsky e Frisch, a pesquisa econométrica passou da análise do ciclo econômico para a macroeconomia (após influência do Keynes) e expressou a teoria econômica em termos matemáticos.

Tinbergen tinha como maior preocupação investigar e considerar os impactos da economia política. Seguindo essa ideia, o autor percebeu que para construir modelo empírico passível de análise só seria possível se fosse relativamente simples, ou seja, um moderno com uma versão estilizada do sistema econômico. Este raciocínio soa muito familiar a análise econômica moderna, onde a realidade passa a uma fase secundária de importância (EPSTEIN, 2011).

O trabalho estatístico teria então a tarefa de avaliar a teoria econômica, não para dizer se estava correta, mas apenas para determinar se estava incorreta ou incompleta. De acordo com Epstein (2011), esta atitude em relação à capacidade avaliativa da teoria estatística sobre a economia, provou ser muito influente em futuros trabalhos econométricos, em particular nas obras dirigida pela *Cowles Commission*.

O centro de pesquisas dos Estados Unidos em economia, *Cowles Commission*, desempenhou um papel de liderança na promoção da teoria econômica formal e nas técnicas matemáticas e estatísticas em economia, estreitamente ligadas, especialmente nas suas primeiras décadas, com a *Econometric Society* and *Econometrica.*

3.2 A INCLUSÃO DO CÁLCULO DE PROBALIDADE NA TEORIA ECONOMÉTRICA

Ao mesmo tempo que tinha como foco os ciclos econômicos, a econometria encontrou na análise da demanda um outro campo de pesquisa. Só que ao contrário da outra vertente, os trabalhos relacionados a demanda já tinham um paradigma estabelecido dentro da teoria econômica (BLAUG, 1980).

Assim, os problemas enfrentados pelos economistas neste campo de investigação não foram os de descobrir as relações econômicas subjacente. Às questões levantadas seriam antes os problemas de identificação e correspondência, que implicaram em dois obstáculos principais: a) adequar os modelos teóricos e os resultados estimados; e, b) estimar uma curva de demanda.

Conforme Blaug (1980), as pesquisas sobre demanda foram um fator para o avanço da econometria, uma vez havia consenso sobre a sua veracidade na literatura econômica da época. Pois, foi o meio de expressar uma relação considerada como verdadeira em uma linguagem facilmente compreendida por todos, e mais importante, que permitiu os economistas tratar estas relações em modelos, e atribuir valores numéricos.

Esta possibilidade de expressar matematicamente a teoria econômica teve dois efeitos sobre a econometria. Por um lado, permitiu que os pesquisadores deste campo de pesquisa realizassem testes estatísticos para verificar teorias. Por outro lado, forneceu uma nova visão sobre os trabalhos empíricos, em que dois problemas foram tratados em profundidade. Primeiro, erros de medição, e segundo, a omissão de variáveis.

Ainda de acordo com Blaug (1980), a teoria era assumida como correta e a verificação seria o foco da análise no início das pesquisas econométricas. Por implicação,

então, para fechar a lacuna das evidências, os dados teriam de ser ajustados pela hipótese inicial. Mais tarde, uma vez que a investigação de verificação recuou sob a influência do operacionismo, o processo inverteria.

Em contrapartida a visão tradicional do período, cabe ressaltar que Moore (1914) já criticava o fato de a análise da demanda em economia ser concebida como sendo estática, enquanto os dados eram compostos por realizações de ponto único de interações de procura e oferta durante um longo período de tempo, quando outras coisas não eram constantes. Moore (1914) desenvolveu diferentes técnicas para limpar os dados, os quais provaram-se muito influente nos trabalhos econométricos posteriores, mas deixou a teoria intocada.

Estas técnicas consistiram em um único dispositivo estatístico que nada mais era do que a utilização de primeira diferença nos dados a fim de remover parcialmente ou, pelo menos, corrigir, os efeitos dinâmicos. Outro instrumento utilizado por Moore (1914), consistiu em remover tendências através de um "coeficiente de inclinação".

A crescente discussão dentro da econometria nas décadas de 1930 e 1940 de que a teoria econômica nem sempre poderia ser considerada como verdadeiro levou-os à necessidade de introduzir alguns critérios externos para testar a teoria em si. Não era apenas uma questão de transformar os dados, mas a teoria estava agora a tornar-se mais flexível e susceptível a mudança (MORGAN, 1991).

A estatística e os cálculos probabilísticos começaram a ser usados com o objetivo de obter estes critérios de teste para a teoria econômica. Introduzindo as técnicas estatísticas do tipo Neyman-Pearson, os economistas encontraram outro problema, uma vez que estes tipos de testes dependiam do cálculo de probabilidades. Econometristas e economistas em geral, de acordo com a Morgan (1991), não aceitaram totalmente o cálculo de probabilidade até ao final da década de 1940.

Era necessário que surgisse uma segunda técnica para solucionar essa problemática, dado que os erros de estimação precisavam ser considerados. Deste ponto de vista, a teoria econômica só era importante para determinar a existência da relação entre duas variáveis. Autores como Gini (1921), declararam que o modelo de regressão padrão apenas minimizava os efeitos dos erros de medição a partir do lado da variável dependente.

Mas a correção dos desvios quando encontrados em ambos os lados da regressão, na variável dependente e explicativa, não era possível de modo que outro tipo de estimação precisou ser utilizado para ultrapassar este problema. A regressão ortogonal foi

então proposta como um tipo de estimador que levaria em conta os erros em ambos os lados da regressão e assim, representaria uma melhor relação ajustada entre as variáveis (HENDRY;MORGAN, 1997).

Por outro lado, os econometristas descobriram que os efeitos de alguns fatores ocultos poderiam ser também disfarçados nos erros. Hendry e Morgan (1997) afirmam que isto deveu-se principalmente ao fato dos economistas estarem utilizando modelos simplificados para simular uma teoria mais complexa. Os erros não só continham os efeitos dos desvios de estimação, mas também os efeitos de variáveis que não tinham sido levadas em conta ao fazer a regressão.

A omissão de variáveis explicativas importantes conduziria a parâmetros enviesados. Esta questão levantou uma série de trabalhos tentando encontrar uma forma de desvendar e encontrar as relações ou componentes escondidas e interligadas nos dados observados. Alguns destas obras foram *Two Equations Systems developed* de Moore (1914) e *Causal Chain Models* de Tinbergen (1937).

Conforme Spanos (1998), embora o intenso trabalho estatístico em economia remonte ao início do século XIX, o cálculo de probabilidade teve de esperar até à primeira metade do século para ser introduzido. A estatística em si não tinha espaço nos termos em que os economistas pensavam, pois, o seu raciocínio era de um tipo determinista.

Assim, no início do século XX, os econometristas acreditavam que existiam leis reais e constantes de comportamento econômico à espera de ser descoberto e analisados. Além disso, os dados não constituíam uma matéria prima à qual se pudesse aplicar adequadamente o raciocínio probabilístico.

Em resumo, também existiam um debate sobre as diferenças adequadas entre as observações e a teoria em termos de erros cometidos pelo investigador ou pelo modelo de regressão (variáveis omitidas) ou em termos de problemas estatísticos (erros de estimação). Na sua maioria, acreditavam que a teoria econômica corrente era supostamente ser verdadeira.

Vale a pena notar que mesmo os economistas que utilizavam a estatística rejeitaram a utilização do cálculo de probabilidade em economia. Eles justificavam esta rejeição porque acreditava que as observações econômicas não podiam preencher as características da probabilidade, pois eram não independentes uns dos outros (MORGAN,1990).

À medida que as questões de correlação e os testes teóricos se foram tornando cada vez mais importantes, os econometristas estavam se tornando mais sofisticados

sobre questões de inferência. No entanto, as suas atitudes em relação ao cálculo de probabilidade consistiam em uma contradição.

Esta incongruência é dada pelo fato de que a base teórica para a inferência estatística reside na teoria da probabilidade e os economistas utilizaram os métodos estatísticos, mas rejeitavam a probabilidade. Trygve Magnus Haavelmo, juntamente com outros pesquisadores da comissão cowles, criticou esta contradição na década de 1940.

Haavelmo (1944) argumentou que a independência e a homogeneidade dos problemas dos dados econômicos não impediram necessariamente a utilização da probabilidade, ou seja, não é necessário que as observações sejam independentes e que precisam seguir uma mesma lei unidimensional.

Pode-se assumir que todo o conjunto de, digamos n, unidades pode ser considerado como uma observação de n variáveis seguindo uma junta n-dimensional, cuja existência é puramente hipotética. Deste modo, é possível testar hipóteses relativas a esta distribuição de probabilidade conjunta, e fazer inferências quanto a sua forma, através de um ponto de amostra (em n dimensões).

Haavelmo (1944), afirma ainda que é necessário estabelecer as condições experimentais sob as quais o modelo proposto será válido, a fim de comparar a teoria com os dados. Isto significa, que os pesquisadores devem descrever como mediram a estrutura das variáveis ou objetos verdadeiros, que seriam identificados com os seus instrumentos teóricos correspondentes.

E como quase todos os dados econômicos são obtidos de experiências da natureza, em que os economistas são meros observadores passivos, que não podem estabelecer as condições experimentais. Haavelmo (1944), também sugere que caso não for possível isolar, manipular e controlar as observações é necessário introduzir estes fatos na suposição, a fim de obter um maior acordo entre a teoria e a realidade.

Outra contribuição importante de Haavelmo (1944) para literatura econométrica, refere-se ao desenvolvimento de uma estrutura de possibilidade para testes teóricos. A nova forma de construir a teoria econômica em termos probabilísticos de Haavelmo, não excluiu nenhum sistema de valores das variáveis, mas apenas deu diferentes pesos ou probabilidades para o mesmo.

Esta capacidade de reprodução de amostras implicava um novo critério de comparação de diferentes teorias: uma nova formulação probabilística da teoria econômica forneceu a possibilidade de realizar testes estatísticos do tipo Neyman-

Pearson. Esta mudança para testes estatísticos mais sofisticados, deu prioridade, de certa forma, a testes estatísticos para validar as suposições econômicas.

Com as dificuldades de aplicar a probabilidade aos dados econômicos sendo ultrapassadas, Haavelmo proporcionou aos economistas um quadro adequado para a condução das pesquisas e de testes rigorosos de teorias. Este novo quadro, no entanto, mudou a forma como os estudos econométricos devem ser apresentadas, pois tem que desenvolvidos de forma a poder representar hipóteses estatísticas (e, em sua maioria, validados pela medida do p-valor).

3.3 CONSOLIDACÃO DA ECONOMETRIA: TESTES DE HIPÓTESES E P-VALOR COMO PRINCIPAL SUPORTE

A Formulação de Fisher (1925) do valor de probabilidade e o aparato dos testes de hipóteses de Neyman-Pearson tornou-se conhecido por economistas no final da década de 1930.

Mas, a partir de Haavelmo (1944), começou a ser aplicado em séries econômicas ao considerar um conjunto de dados como uma única realização (amostra) a partir do número infinito de possíveis populações que a natureza poderia ter escolhido.

O trabalho pioneiro de Haavelmo sobre a abordagem da probabilidade tornou-se reconhecidamente como um marco crucial na construção da econometria. Foi a primeira tentativa sistemática de fazer a ponte entre a investigação empírica e a teoria de uma forma logicamente rigorosa: este processo facilitou a fertilização cruzada entre a teoria e os dados. O legado em termos da metodologia para identificar e estimar modelos continua a ser forte até os dias.

Após a grande contribuição de Haavelmo (1944), o processo de inovação metodológica em econometria continuou durante o período pós-guerra, mas, ao mesmo tempo, um processo de consolidação teve lugar. A consolidação foi feita principalmente em torno do modelo linear de regressão que se tornou a ferramenta econométrica dominante e através das pesquisas aplicadas, por meio da construção de modelos macroeconométricos patrocinados pelos governos e o crescimento do ensino da econometria nas universidades.

O método por mínimos quadrados ordinários (MQO) tinha sido amplamente utilizada por pesquisadores em todo o período entre guerras. Isto era particularmente significante para os economistas que trabalhavam na economia agrícola. Antes disso,

contudo, a sua estrutura e questões discutidas pela *Cowles Commission* levam a um declínio na sua aceitação (FOX, 1989).

Spanos (1989) destaca que o principal motivo que levou a essa desconfiança sobre o uso do MQO, se dava porque muitos pesquisadores acreditam que as estimativas por mínimos quadrados resultam geralmente em parâmetros com estimativas tendenciosas em modelos estruturais. O Estimador de Máxima Verossimilhança (EVM) seria uma abordagem alternativamente viável.

A consequência foi que, sob a influência da *Cowles Commission*, a regressão por MQO foi considerada uma técnica imperfeita. Stone (1946) resumiu esta posição ao afirmar que em exceto de casos muito especiais, as estimativas não podem ser obtidas considerando cada equação de forma isolada.

Só que as primeiras implementações do estimador de máxima verossimilhança não conseguiram mostrar muita diferença prática quando comparada com os resultados do MQO. Destaca-se o modelo Klein-Goldberger (1955), um modelo keynesiano da economia dos EUA durante o período 1929-1952, em que as estimações de Fox (1956) por mínimos quadrados apresentaram os mesmos resultados das regressões por EVM de Klein-Goldberger.

Outro trabalho relevante neste aspecto é de Christ (1960), no qual ao utilizar o método de Monte Carlo para comparar os estimadores de máxima verossimilhança com os mínimos quadrados em amostras finitas, também descobriu que os resultados diferiam pouco. Estas evidências contribuíram para a opinião generalizada na época de que os modelos de EVM oferecerem poucas melhorias em relação estimativas de mínimo quadrados.

Entre as contribuições de destaque contrárias ao uso do método EVM frente aos mínimos quadrados, havia o trabalho de Wold (1944) que defendia a teste que o método de máxima verossimilhança como um modelo causal mal especificado. Obviamente, uma vez que não há simultaneidade no modelo, o MQO torna-se legítimo. O argumento de Wold foi totalmente adoptado mais tarde por Liu (1969) no seu modelo experimental mensal dos Estados Unidos.

Já Stone (1954) acredita que a *Cowles Commission* tinha ressaltado de forma exagerada a importância da simultaneidade nas equações de demanda dos consumidores e que as regressões por MQO eram susceptíveis de ser mais precisas que os seus homólogos, as estimativas de máxima verossimilhança.

Além dessa discussão, o fim da primeira metade do século XX foi marcado também pela preocupação dos econometristas com a correlação residual nas séries econômicas. Como principal contribuição para diagnosticar este erro, tem o trabalho de Durbin e Watson (1950), ao desenvolver uma tese para delimitar os valores críticos da estatística de teste desta problemática.

Em relação ao tratamento, ressalta-se o artigo de Cochrane e Orcutt (1949) que demostrou que os erros residuais de uma equação seguem um processo autorregressivo de primeira ordem e seriam corrigidas a partir da utilização de séries integradas, o que ficou subsequentemente conhecido como o estimador Cochrane-Orcutt.

Na década de 1950, os pioneiros da econometria tiveram como um dos principais objetivos construir uma demarcação nítida entre a estatística, a econometria e a matemática. Para tanto, conforme Gilbert (1989), essa delimitação foi definida por meio das disciplinas específicas que foram criadas e pelos manuais escritos na época.

Assim, os docentes eram obrigados a cobrir estas disciplinas de forma mais especifica e os estudantes tiveram a possibilidade de ter um maior conhecimento em uma área delimitada do que a geração anterior que tinham desenvolvidos as ferramentas que agora ensinavam. A fronteira de conhecimento em economia tinha encontrado um divisor comum para o seu desenvolvimento (GILBERT, 1989).

No período pós-guerra, a econometria foi geralmente identificada com um campo empírico dentro da economia, com foco na mensurabilidade das relações econômicas. A partir de Klein (1952), inicia-se a procura para definir de forma mais cristalina os objetivos de econometria. Klein considerou como principal meta dar um conteúdo empírico ao *raciocínio a priori* em econometria.

Esta visão da econometria abrangeu questões de especificação, mensuração e a contabilidade da renda nacional, bem como às estimativas estatísticas. Em especial, nos Estados Unidos e na Inglaterra, o seu ensino concentrou estas questões mais vastas dentro de discussão de fronteira em economia (MORGAN, 1990).

Malinvaud (1964) é bastante explícito acerca deste enfoque. Ele afirma que a econometria pode ser interpretada em sentido *lato* para incluir a aplicação da matemática ou dos métodos estatísticos ao estudo dos fenômenos econômicos e, assim, ter como o objetivo ser uma determinação empírica do mesmo.

Por sua vez, Johnston (1960) fornece uma outra demarcação: a teoria econômica deveria consistir nas análises das suposições que descrevam o funcionamento de um

sistema econômico. A tarefa de uma pesquisa econométrica tem que ter como meta estimar estas relações estatisticamente.

A nova geração de livros textos também tinha como busca criar um padrão de conteúdos e, cada vez mais, adotar uma notação padrão. Isto pode ser notado na semelhança das notações de Stones (1954) e Johnston (1963) e como a identidade $y = X\beta + \varepsilon$ de Durbin e Watson (1950) se tornou o modelo de notação de uso padronizado em econometria (GILBERT, 1991).

Conforme Malinvaud (1987) esta síntese foi construída sobre as bases de uma modelação probabilística defendida por Haavelmo (1944) mas, ao mesmo tempo, tendia a relegar os problemas de simultaneidade para o papel de técnicas avançadas. O seu próprio manual de econometria (1964) é a representação deste processo.

Um outro elemento importante na divulgação da prática econométrica foi o desenvolvimento da tecnologia computacional. No período entre guerras, as regressões eram frequentemente realizadas de forma manual ou por uma mistura de calculadoras (utilizadas para cálculos de momento).

O trabalho dos econometristas começou a ser facilitado a partir da década de 60 com os avanços nas estimações dos modelos macroeconômicos e do progresso computacional (KLEIN, 1971), e atingiu o seu ápice com o desenvolvimento de softwares estatísticos no final da década de 1970. A possibilidade do uso de softwares para estimar os modelos econométricos também foi peça importante para a disseminação da econometria.

Enquanto esta mesma década assistiu à consolidação da econometria através de manuais e de softwares estatísticos, havia sinais crescentes e uma insatisfação maior com a abordagem padrão. Principalmente, quando os econometristas tentavam reforçar a ponte entre a teoria e os dados seguindo o paradigma estabelecido por Haavelmo e o *Cowles Commission*.

As estimações construídas pela comissão norte-americana foram concebidas como uma interface econométrica com teorias mais gerais (tanto simultâneas como dinâmicas). Mas a sua execução prejudicada pela escassez de dados e pressupostos testáveis. O trabalho de remendar essa ponte teve como norte várias direções, destacam-se: o desenvolvimento de hipóteses mais informativas, a melhor adaptação da teoria estatística à economia aplicada e a concepção de melhores meios para processar e melhorar os testes estatísticos (QIN, 1997).

Durante a década de 1970-1980, estas diversas estratégias de investigações passaram a representar um desafio crescente para a econometria tradicional. O tópico final deste capitulo esboça brevemente essa discussão.

3.4 A ECONOMETRIA MODERNA: CONSTRUÇÃO DE NOVOS MODELOS E SEPARAÇÃO DA MICRO E MACROECONOMETRIA

A necessidade de melhoria dos modelos teóricos, especialmente os dinâmicos, foi reconhecida pelos próprios pesquisadores da *Cowles Commission* pouco depois do seu trabalho fundacional sobre econometria (KOOPMANS, 1957). Conforme Weintraub (1991), uma parte substancial do grupo, incluindo Kenneth Arrow, Gérard Debreu, Leonid Hurwicz e Tjalling Charles Koopmans, buscaram construir novos modelos e lideraram o caminho para a ascensão de modelos de equilíbrio dinâmico e da teoria de crescimento.

. Esta linha de investigação, no entanto, contém uma mudança de agenda: a tarefa fundamental de caracterizar a dinâmica econômica foi reformulada como a de estabelecer condições de estabilidade ou equilíbrio de sistemas dinâmicos (WEINTRAUB, 1991). As analises teóricas dos percursos temporais das variáveis econômicas sob a forma de testar os modelos estruturais foram, na sua maioria, arquivados para uma possível atenção posterior.

Entretanto, a caracterização destes caminhos foi deixada para o tratamento *ad hoc* dos pesquisadores, e, particularmente, aqueles que trabalhavam com comportamento do consumidor e do investimento, conseguiram incorporar uma dinâmica comportamental em modelos estruturais *a priori*. Estes modelos dinâmicos têm implicações empiricamente testáveis e pode ser visto como uma extensão natural das regressões estruturais (BODKIN ET AL, 1991).

A maioria destas teorias dinâmicas baseadas numa única equação eram assimilados em modelos macroeconométricos de grande escala durante os anos de 1960. Só que a reforma sistemática da macroeconomia em dinamicamente testável aconteceu apenas depois da crise petrolífera de 1973.

A reforma ficou conhecida como a revolução das Expectativas Racionais (ER). A introdução da hipótese ER surgiu originalmente da investigação de Carnegie-Mellon no início da década de 1960. Mas, duas vertentes de pesquisas promoveram o seu desenvolvimento e colocaram na fronteira do conhecimento de economia.

A primeira tem origem no trabalho de Simon (1957) sobre "escolha racional" que pode ser definida como o modelo teórico de tomada de decisão e as implicações deste para a descrição do comportamento humano. Já o segundo, é o trabalho empírico em conjunto de Holt, Modigliani, Muth e Simon (1960) sobre a derivação dos modelos operacionais para controle de estoque: a sua abordagem implica um modelo de regressão dinâmica que pode ser racionalizado em termos das expectativas subjacentes à decisão empresarial.

As ideias de comportamento racional e as expectativas dinâmicas foram amalgamadas na hipótese de expectativas racionais proposta por Muth (1961) no contexto da generalização de modelos dinâmicos para as oscilações de preços. No modelo de Muth, as expectativas são formadas por preditivos imparciais e interpretadas como ER dos agentes das variáveis em questão.

O movimento iniciado da ideia de expectativa racional tinha como intenção de reforçar a abordagem da *Cowles Commission*, com a melhoria sistemática do seu lado teórico. No entanto, esta iniciativa resultou rapidamente em duas descobertas, o que iria minar gravemente a estrutura original das equações estruturais.

A primeira ficou famosa como a crítica de Lucas, no qual o autor argumentou que, sob a hipótese de RE, certas estruturas e os parâmetros não são invariantes para as mudanças de política (LUCAS, 1976). Como Lucas observou imediatamente, isto parece invalidar a utilização de equações estruturais na análise política.

A segunda aconteceu pela comparação modelos de regressões estruturais com os Modelos de Vetores Autorregressivos (VAR), quando Sargent (1976) elaborou uma base de testes gerais para os primeiros métodos. Esta descoberta, juntamente com um ressurgimento das preocupações de Liu (1960) sobre a identificação de Sims (1980), acabou por levar Sims a defender o abandono da abordagem estrutural e a sua substituição pelo VAR.

Por outro lado, embora a distinção entre microeconometria e macroeconometria é relativamente recente, a análise de dados microeconômicos, ou seja, séries medidas ao nível de agentes individuais, domésticos ou de empresa, estende-se durante este período: poucas distinções práticas foram feitas entre os métodos adequados à cada série (corte transversal ou de tempo).

Pode ser identificado vários precursores da microeconometria moderna, embora, no início, esta não tenha atraído grande atenção dos pesquisadores contemporâneos. De acordo com Cramer (1991), a primeira aplicação econômica do modelo PROBIT

aconteceu por meio de Farrell (1954), que era membro do departamento de economia aplicada de Cambridge. Anos mais tarde, Aitchison e Brown (1957), da mesma instituição, comparam os modelos LOGIT e PROBIT.

Do grupo da *Cowles Commission*, o estimador TOBIT foi concebido por Tobin (1958). Esta foi a contribuição final dos pesquisadores da *Cowles* para a metodologia econométrica e a única aplicação específica a dados microeconómicos. Além disso, ilustrou o poder da metodologia de Haavelmo para a construção de modelos econômicos com bases de distribuição segura.

Em relação a análise dos dados, destaca-se o trabalho de Orcutt (1952). Guy Orcutt procurou colocam em debate o problema de identificação nas séries agregadas e como às séries não eram extraídas a nível individual. O autor mostrou-se que estava à frente do seu tempo, uma não existiam esse tipo de dados e não havia ainda um poder computacional para estimar às séries.

. Estas preocupações eram reforçadas pela análise da demanda. Klein (1946) e Nataf (1953) sublinharam que a agregação exata das relações individuais para as relações agregadas era possível apenas sob hipóteses implausivelmente fortes. Por sua vez, Theil (1965) atingiu conclusões mais otimistas de maneira marginal, ao olhar para a agregação aproximada.

Nas décadas de 1960 e 1970, a literatura da demanda teve como foco destacar a importância da heterogeneidade entre indivíduos da forma prevista por Orcutt vinte e cinco anos antes. De acordo com Munnell (1987), isto implicava que as discussões políticas deveriam relacionar-se com agentes na plena diversidade das suas situações econômicas e sociais, em vez de supostamente ser analisadas de forma representativa.

Com os rendimentos constantes nos Estados Unidos durante esse período, gerou alguns dos mais extensos conjuntos de dados que deram origem a vários problemas, particularmente os de enviesamento de seleção de amostras e censura. A investigação sobre decisões de fornecimento de mão-de-obra deu origem à questão dos dados sistematicamente desaparecidos – pessoas desempregadas não reportam os seus níveis salarias (HECKMAN, 1974).

Além disso, conforme Heckman (2000), nestes anos as questões de especificação, particularmente as dinâmicas, que dominaram muito o debate macroeconométrico ainda estava ausente na microeconometria. Já as questões de identificação, que estavam presentes, assumiram uma forma diferente. Acima de tudo, a investigação microeconométrica ainda continuava fundamentada na ampla tradição estrutural da

Cowles Commission, enquanto a macroeconometria se encaminha para uma abordagem baseada em dados.

Vários fatores favoreceram o renascimento gradual da abordagem de modelização orientada por dados. Para além dos motivos já mencionados acima (o renascimento dos mínimos quadrados, insatisfação com abordagem estrutural e aumento da popularidade da abordagem VAR), dois outros pontos desenvolvimento tinham sido importantes - o progresso de testes de hipóteses como instrumento para a seleção de modelos e maior atenção à metodologia de previsão.

Klein (1971) aponta que as exigências comerciais e as demandas governamentais em relação as previsões econômicas proporcionaram um grande estímulo para a construção de grandes modelos macroeconométricos. Havia um otimismo visível de que a predição melhoraria com o aumento do tamanho do modelo e da complexidade estrutural.

No entanto, esta positividade foi amortecida pelos erros de previsão, incluindo simulações de políticas, especialmente em comparação com as previsões de modelos simples de séries temporais. Por exemplo, Nelson (1972) utilizou modelos ARIMA simples do tipo Box-Jenkins para comparar com a previsão do desempenho do modelo estrutural desenvolvido conjuntamente pelo *Federal Reserve*, MIT e a universidade da Pennsylvania. Os modelos de série temporal ARIMA tiveram um melhor desempenho.

A estratégia Box-Jenkins (1970), tinha como vantagem uma série de conceitos que foram negligenciados pela abordagem *Cowles Commission*, em particular o princípio da parcimônia na formação do modelo, a necessidade de verificações rigorosas sobre a adequação da regressão e um procedimento de identificação, que essencialmente conota uma estrutura de especificação diferente do sentido *Cowles* da terminologia.

A modelação de equações individuais foi o banco de ensaio que mais favoreceu para comparar teorias macroeconômicos, embora Griliches (1968) tivesse aconselhado os grandes macromodelos a submeter as experiências para “autópsia”. No debate mais destacado, Friedman e Meiselman (1963), confrontaram às abordagens monetárias e fiscais à política macroeconômica.

Isto deu uma nova discussão a uma série de problemas econométricos antigos: regressões que apresentam previsões erradas; as estimativas dos parâmetros não constantes; e, a abordagem da variável explicativa demasiadamente primitiva para a avaliação teórica. No entanto, segundo Dhrymes *et al*. (1972), a deficiência mais evidente

neste debate foi a natureza ad hoc do procedimento para o tratamento de hipóteses empiricamente testadas, em particular quando estas não exibiam uma estrutura aninhada.

Uma consequência do debate foi o modelo de Saint Louis, construído principalmente para fins de simulação de políticas no Banco da Reserva Federal de St, Louis, e que contrariou a tradição estrutural ao adoptar uma forma reduzida abordagem, baseada num pequeno número de equações com longos comprimentos de atraso em cada equação (ANDERSEN;CARLSON, 1974).

O progresso na teoria estatística também desempenhou um papel importante no renascimento da abordagem da modelização baseada em dados, em particular no que diz respeito aos testes de hipótese. Pesaran (1974) foi possivelmente o primeiro econometrista a aplicar a metodologia estatística de Cox (1962) para testes elementares. Da mesma forma, o teste de especificação RESET de Ramsey (1969), tornou-se amplamente implementado.

Um exemplo particular deste movimento foi o renascimento da análise de frequência, descartada como inútil quarenta anos antes (PHILLIPS, 1997). Graças a Tukey (1953), houve a extensão do exame frequentista de casos únicos para múltiplos casos, o que foi rapidamente adaptado em economia por Granger e Hatanaka (1964).

Esta investigação levou Granger (1969) a tomar consciência da falta de uma caracterização estatística adequada de causalidade em econometria, uma preocupação que acabou por resultar no que é agora conhecido como teste de causalidade de Granger. Este teste foi amplamente adotado pelo movimento de expectativas racionais.

Os testes de diagnóstico pós-estimativa se tornaram rotina em qualquer trabalho econométrico e foram implementados em pacotes econométricos padrão, tendendo a minar os modelos estruturais porque estes invariavelmente acabaram por ser mais simples. Os modelos eram adaptados, frequentemente de formas aparentemente arbitrárias, se não sobrevivessem ao confronto com os dados.

Algumas destas questões foram posteriormente esclarecidas através do desenvolvimento da análise de cointegração (ENGLE;GRANGER, 1987; JOHANSEN, 1988) no âmbito do VAR, uma vez que isto permite a interpretação da dinâmica dos coeficientes de ajustamento como parâmetros de incômodo.

Finalmente, ressalta-se que no final do século XX a econometria tinha uma conotação clara e tornou-se uma disciplina dentro da economia com as suas próprias revistas e com professores dedicados a ensinar uma gama de cursos especializados. O processo de se tornar uma matéria envolveu a formação de um paradigma padrão.

Embora nem todos estivessem de acordo com todo o paradigma, as posições foram definidas tendo como referência esse padrão. Este foi um processo que teve lugar nos anos cinquenta, sessenta e setenta, sobre as fundações lançadas no período entre guerras e a consolidação após a segunda guerra mundial.

Os avanços foram progressivos, no sentido em que, uma vez descoberto, este conhecimento não foi esquecido. Depois de meados da década 1970, foram acompanhados pelo desenvolvimento da computação que, com o tempo, eliminaram uma grande limitação da análise econométrica. Um tema importante que dominou grande parte do debate ao longo do século foi como, e de fato, se os modelos econométricos podem refletir estruturas económicas geradas por teorias.

Na fase da econometria moderna, os paradigmas tornaram-se mais soltos à medida que os economistas passaram a definir as suas posições em relação uns aos outros. O resultado foi uma maior diversidade tanto na teoria como na aplicação, mas com uma linguagem partilhada e uma história comum. Paradoxalmente, a tradição estrutural dos *Cowles,* que tinha surgido numa altura em que a vasta a maioria das aplicações eram em macroeconomia, sobreviveu mais eficazmente em microeconometria.

4. CRÍTICAS AO USO DO VALOR DA PROBABILIDADE: PORQUE A ECONOMETRIA CLÁSSICA DEVE SE ATENTAR A SUA UTILIZAÇÃO

Não há muito o que se possa dizer de novo sobre os perigos do p-valor e dos testes de significância que não são ditos há décadas (ZILIAN;MCCLOSKEY, 2008; HUBBARD, 2016). Só que saber o que não se deve fazer com o valor de probabilidade é de fato necessário, mas não é suficiente. Tendo este princípio como norte, o presente capítulo apresentará os principais problemas da utilização do p-valor e os motivos pelo qual esse paradigma dentro da teoria econométrica precisar ser discutido.

4.1 A UTILIZAÇÃO DO NÍVEL DE SIGNIFICÂNCIA ESTATÍSTICA

O primeiro ponto crítico da teoria do valor de probabilidade se dá pelo termo "estatisticamente significante", visto que seu uso se tornou extremamente questionável. O termo tem sua origem ligada a Fisher (1925), com a intenção original de simplesmente ser um instrumento para indicar quando um resultado justifica uma investigação mais aprofundada. Mas essa ideia foi irremediavelmente perdida.

A significância estatística nunca tinha como pretensão implicar a importância científica, e a confusão se iniciou logo após a sua utilização generalizada (GHOSE, 2013). No entanto, um século mais tarde, a confusão persiste. O problema não é simples utilizar a palavra "significativo", embora a estatística e os significados da palavra são agora desesperadamente confusos.

O problema é maior, no entanto: usando dessa ideia é possível justificar alegações ou conclusões científicas que pode levar a crenças erradas e a uma má tomada de decisões. A rotulação estatisticamente significante não acrescenta em nada ao que já é transmitido pelo p-valor; e, além disso, esta dicotomização torna a situação do valor de probabilidade ainda pior.

Por exemplo, nenhum p-valor pode revelar a plausibilidade, a presença, a verdade, ou a importância de uma associação. Por conseguinte, um rótulo de significativa estatística não significa ou implica que um efeito é altamente provável, real, verdadeiro ou importante. Nem um termo de não significância leva uma relação a ser improvável, ausente, falsa, ou sem importância.

No entanto, a dicotomização entre "significativo" e "não significativo" é tomado como uma sanção de autoridade sobre estas características. Como Gelman e Stern (2006) notoriamente destacaram, a diferença entre os dois termos não é, por si só,

estatisticamente significativo. Além do mais, esta divisão falsa em "digno" e "indigno" é, por si só, estatisticamente insignificante.

Os resultados conduzem à elaboração seletiva de relatórios e à publicação das evidências com base no seu significado estatístico. Assim, os problemas do uso do p-valor vão para além da simples publicação: quando os autores o utilizam de forma limiar para selecionar quais os resultados a discutir nos seus trabalhos, as suas conclusões e revisões serão tendenciosas. Criando não apenas uma distorção na literatura, mas inclina as evidências que devem ser publicadas na pesquisa.

Está dicotomização arbitrária dos resultados em "significativo" e "não significativo," em conjunto com o desejo dos próprios investigadores de obter resultados que possam ser declarados "significativo" é considerado uma das principais causas da crise de reprodutibilidade (GREENLAND, 2017; MACSHANE ET AL., 2017; TRAFIMOW ET AL., 2017).

Isto vale tanto para os estudos experimentais como para os observacionais. O termo "p-hacking" foi cunhado para descrever o comportamento dos investigadores que tentam uma multiplicidade de alternativas analíticas e depois relatam apenas a que produziu o resultado desejado.

No p-hacking há várias características dignas de nota: primeiro, a lista de alternativas analíticas possíveis são quase infinitas na maioria dos contextos de investigação. Em segundo lugar, a consciência dos problemas entre os pesquisadores é frequentemente baixa, especialmente na análise de dados observacionais, devido à ambiguidade da especificação ser apropriada.

Terceiro, as múltiplas variantes analíticas que podem ser experimentadas muitas vezes parecem ter pouco em comum à primeira vista. No entanto, partilham uma qualidade nociva: a notificação seletiva de testes múltiplos encobertos podem inflacionar desastrosamente alegações de significância estatística.

A literatura relacionada com as armadilhas dos testes e declarações de significância discutiu extensivamente as várias formas de p-hacking[2]. Não se pretende resumir mais uma vez a mesma. No entanto, para facilitar a compreensão de que formas o p-hacking representa um dissimulado e, a partir daí, especialmente nocivo tipo de testes múltiplos, descrevemos brevemente as quatro principais formas de p-hacking que podem ser distinguidas.

[22] ver Hirschauer et al. 2016 para uma visão geral

A primeira forma está relacionada aos testes de conjuntos de dados: os investigadores podem ser tentados a explorar se os p-valores podem ser reduzidos quando o número de unidades de uma amostra é manipulado. Há várias possibilidades: a redução do tamanho da amostra, por exemplo, através de uma tentativa de eliminação de outliers, é uma possibilidade.

Outra é o aumento do tamanho da amostra, se uma população original produzisse valores de probabilidade "decepcionantes". Uma sensação geral de que amostras maiores são melhores pode impedir a consciência de que isto constitui um teste de múltiplos tamanhos. Finalmente, os pesquisadores podem tentar obter na análise de múltiplos subconjuntos de p-valores de acordo com o seu desejo.

Já a segunda forma de p-hacking refere-se a questão da possibilidade de testes de transformações de dados: os pesquisadores também podem ser tentados a verificar qual (combinação) de muitas transformações de dados concebíveis produz p-valores inferiores aos dados originais.

As possibilidades são abundantes: desclassificação das escalas de medição, a transformação de log, a quadratura, e a sintetização de várias variáveis, incluindo índices e termos de interação. Algumas destas manipulações podem ser alegadas como estatisticamente apropriadas à luz da teoria, questão de investigação e dados - por exemplo, quando investigadores perceber, depois de ver os dados, que as hipóteses de distribuição são violadas.

Por sua vez, a terceira forma está associada aos testes de variáveis: para além do número de unidades numa amostra, os pesquisadores podem também ser tentados a experimentar diferentes preditores e variáveis resposta. No estudo experimental, isto implica pronunciar significância estatística para resultados selecionados após uma multiplicidade de tratamentos foram testados para uma multiplicidade de resultados.

Finalmente, a última forma está correlacionada aos testes de modelos de estimação: a seleção de testes estatísticos também oferece uma ampla margem para decisões que "melhoram" os valores de probabilidade. Por exemplo, quando confrontados com uma escolha ambígua de utilizar ou não um estimador MQO ou um modelo de dados de painel, seria uma boa prática científica para comparar de forma transparente os resultados de ambos os modelos.

Entretanto, as regras de boas práticas científicas são ocasionalmente quebradas e as análises de dados não são realizadas como planejado em um desenho de estudo prévio, mas ajustada *ad hoc* de acordo com o critério do qual o modelo analítico produz p-valores

baixos. A transparência científica é perdida quando os resultados dos modelos concorrentes não são nem explicitamente relatado nem comparativamente discutido.

4.2 A FALTA DE REPLICABILIDADE

O segundo ponto crítico que merece ser ressaltado da teoria do p-valor refere-se ao fato de seus valores ser dificilmente replicáveis. Na maioria dos casos, o teste de significância de hipótese nula é usado para examinar o quão compatíveis alguns dados estão com a H0 de que o verdadeiro tamanho do efeito é zero.

O resultado do teste estatístico é um p-valor informando sobre a probabilidade dos dados observados, ou dados mais extremos, dado que H0 é verdadeira. Se $p \leq 0.05$, é consenso chamá-lo de significativo e rejeitar a hipótese nula, além de aceitar uma hipótese alternativa sobre algum efeito não nulo da população.

No entanto, o pesquisador não sabe e nem pode inferir que a hipótese nula ou uma hipótese alternativa é verdadeira. Com base em um único estudo, é logicamente impossível desenhar uma conclusão. É para aqueles possíveis efeitos de coincidência casual que Ronald Fisher escreveu: "nenhuma experiência isolada, por muito significativo que seja em si mesmo, pode ser suficiente para a demonstração experimental de qualquer fenômeno" (Fisher, 1937, p. 16).

Ao contrário de uma crença generalizada, o valor da probabilidade em si não indica quão replicáveis são os resultados de um modelo. O que um nível descrito pequeno significa é que os resultados são fiáveis e um estudo de replicação teria uma boa hipótese de encontrar novamente este valor (OAKES, 1986; MILLER, 2009).

Isto é mais evidente se a hipótese nula for verdadeira, porque então os p-valores são distribuídos uniformemente e, portanto, todos os valores são igualmente susceptíveis de ocorrer. Só que uma hipótese nula de um efeito exatamente zero é frequentemente improvável de ser verdade.

4.3 AMBIGUIDADE DA SUA FUNDAMENÇÃO TEÓRICA

Um dos principais problemas do p-valor pode ser atribuído à variação nas definições e terminologia entre os seus pioneiros. Existem pelo menos duas definições do valor de probabilidade observado que são largamente utilizadas. Na definição inferencial habitual (fisheriana), um p-valor é a probabilidade da cauda de seu resultado sob H0 de que uma estatística de teste seria tão grande ou maior do que o observado, dado o modelo incorporado.

Por outro, na definição de Neyman-Pearson, o valor de probabilidade é frequentemente definido como o nível descrito mais pequeno (normalmente, menor que 0,05) que permitiria a rejeição de uma regra de decisão de nível α (teste de hipótese de Neyman-Pearson) que contesta hipótese nula a ser testada pelo estudo.

Apesar da definição comum de p-valor ser a de Fisher, muitos pesquisadores centram-se apenas nas propriedades de amostragem repetida, e assim o define não como o valor observado, mas sim como uma variável aleatória cujo valor (realização em uma dada amostra é a sua probabilidade analisada (KUFFNER;WALKER, 2017).

Assim, configuram-se diversas definições logicamente distintas do p-valor. Mas, tal como a terminologia conflituosa, este conflito de conceitos raramente é notado - por isso é comum encontrar em manuais e estudos específicos a falta de distinção da probabilidade aleatória com a probabilidade observada.

A distinção entre as duas definições é importante, até porque os frequentistas definem a validade do p-valor em termos do p-valor aleatório: a probabilidade da variável aleatória é considerada válida para testar a hipótese dado um modelo X se for uniformemente distribuída quando H0 e α estiverem corretos; nesse caso, para cada α a regra "rejeitar a hipótese quando" descartará falsamente H0 com frequência.

Segundo Greenland (2016), a confusão do p-valor com a probabilidade da variável aleatória pode ser um grande contribuinte para algumas das falácias referente ao seu uso. Pior, porém, é que a definição do p-valor observado é frequentemente equacionada ou substituída por descrições totalmente incorretas tais como "p-valor é a probabilidade de que só a aleatoriedade produzir a associação", o que reflete um erro para além da mera terminologia.

Além disso, os pesquisadores que utilizar o p-valor são tipicamente demasiadamente confiantes nos seus juízos. Conforme O'Hagan (2019), as provas disso provêm principalmente de estudos em que se pediu aos sujeitos que dessem um intervalo de valores para uma quantidade incerta com uma probabilidade especificada, e em que a frequência com que os valores verdadeiros caíam dentro desses intervalos era menor do que a probabilidade especificada.

Por exemplo, em estudos em que os sujeitos deram intervalos de probabilidade de 95%, menos de 95%, e talvez tão pouco quanto 65% desses intervalos foram encontrados para conter os valores verdadeiros correspondentes. Os intervalos apresentavam um excesso de confiança e várias explicações possíveis podem ser avançadas para esta descoberta.

Embora não seja uma experiência heurística por direito próprio, o excesso de confiança pode estar relacionado com a ancoragem. Os pesquisadores procuram frequentemente um intervalo após terem dado uma primeira estimativa, que pode servir como âncora. Assim, esse intervalo ter como tendência ser demasiado estreito pode ser justificado por essa amarração.

Já quando precisam apresentar suas evidências, os pesquisadores podem se sentir pressionados a demonstrar um resultado válido, dando intervalos mais estreitos do que o seu trabalho realmente justificaria. No entanto, de igual modo, eles ainda podem dar intervalos mais alargados quando temem que possa haver consequências para "errar", expressando assim uma sub-confiança.

Tem sido sugerido que os pesquisadores desenvolvem o seu próprio atalho heurístico para alcançar respostas rápidas a questões que habitualmente surgem na sua área (O'HAGAN,2019). Uma confiança instintiva em tais investigações manifesta-se como excesso de convicção, porque, em um exercício de elicitação, é provável que lhes seja perguntado sobre evidências menos rotineiras.

O significado do último ponto é que quando se pede a um investigador que dê um intervalo de 95%, ele simplesmente pensa em termos de dar uma gama de valores em que julga que a quantidade de interesse é muito provável que se situe. Dariam o mesmo intervalo se lhes fosse pedido um intervalo de 99%, porque muito provavelmente é um termo impreciso.

4.4 A TAXA DE FALSOS POSITIVOS

Antes de iniciar a discussão deste tópico é necessário conceituar mais precisamente a probabilidade de os resultados serem devidos ao acaso. Isto deve ser definido como a probabilidade, à luz do p-valor que observa, um efeito é real, quando na realidade não o é: o risco de obter um falso positivo.

A este conceito foram atribuídos vários nomes. Por exemplo, foi chamada a taxa de falsos positivos por Wacholder et al. (2004) embora tenham utilizado a abordagem p-hacking de Colquhoun (2017). Também foi denominado como a taxa de falsos descobertos por Colquhoun (2014), uma escolha que de certo modo foi confusa por conta da utilização desse termo no contexto do problema de comparação múltipla (relacionado aos ensinos clínicos).

Em síntese, argumenta-se que o p-valor é quase sempre maior do que parece ("menos significativo") devido à taxa de falsos positivos, o que é uma das razões pelas

quais é possível afirmar que a maioria (e não apenas alguns) dos estudos que utilizar o valor de probabilidade podem ter resultados enviesados.

Isto indica que apesar de não ser difícil conceber um experimento com poder estatístico suficiente ao usar o aparato dos testes de hipóteses e do p-valor, existe um risco considerável de ter um resultado estatisticamente significativo que pode facilmente ter o sinal errado ou um tamanho do efeito exagerado.

O problema dos falsos positivos torna-se ainda pior porque os resultados significativos não são uma amostra aleatória das evidências possíveis — são resultados tendenciosos. Se forem feitos vários estudos sobre uma população com um tamanho de efeito fixo, os trabalhos que, devido à variação da amostragem, encontram um efeito maior são mais susceptíveis de ser significativos do que os que encontram efeitos menores (SCHMIDT, 1992).

Com a utilização do p-valor e da significância estatística como norte, esses resultados são demasiadamente positivos para serem verdadeiros. A consequência é que a maioria das associações verdadeiras descobertas são infladas. Este efeito foi denominado "inflação da verdade" por Reinhart (2015).

Naturalmente, a comunicação de resultados significativos leva a efeitos inflacionados não só nas meta-análises, mas em todos os estudos. Mesmo nos casos em que os autores relatam todos os testes realizados independentemente dos seus p-valores, mas depois selecionam o que interpretar e discutir com base em limiares de significância, os efeitos a partir dos quais os pesquisadores tirarem as suas conclusões serão enviesados.

O problema surge não só por descartar conscientemente descobertas não significativas. Também, porque, os procedimentos de seleção amplamente automatizados podem produzir efeitos inflacionados. Na simplificação do modelo estatístico, ou seleção de modelos, os preditores significativos terão estimativas de pontos aumentados, e a definição da importância de uma variável preditora baseada na significância estatística conduzirá assim a resultados distorcidos (IOANNIDIS, 2008).

4.5 AS HIPÓTESES NULAS TENDEM A SER IMPLAUSÍVEIS E PODEM SER IRRELEVANTES

Nas ciências sociais e na bioestatística, as associações são tipicamente pequenas e variam consideravelmente entre objetos e contextos. Além disso, as medições são frequentemente variáveis e apenas indiretamente relacionadas com construções subjacentes de interesse; assim, mesmo quando as dimensões das amostras são grandes,

as possibilidades de viés e variação sistemática podem resultar no equivalente de populações pequenas ou não representativas. Consequentemente, as estimativas de qualquer estudo único - a unidade fundamental típica de análise - são, em geral, ruidosas.

Fora que a hipótese nula empregada na esmagadora maioria das aplicações é a de efeito zero - ou seja, nenhuma diferença entre dois ou mais tratamentos ou grupos - e erro sistemático zero - que engloba tanto a adequação do modelo estatístico utilizado para calcular o p-valor, bem como todas e quaisquer formas de erro sistemático ou não amostral que variam por campo, mas incluem erro de medição; problemas de fiabilidade e validade; amostras enviesadas; e confusão.

A combinação destas características e este ponto agudo de hipótese nula e de erro sistemático de efeito zero é altamente problemática. Especificamente, porque os efeitos são geralmente pequenos e variáveis, a hipótese de efeito zero é falsa. Além disso, mesmo que a hipótese fosse verdadeira para algum fenômeno, o efeito em consideração em qualquer estudo concebido para examiná-lo não seria zero, porque as medições são geralmente ruidosas e sistemáticas.

Dessa forma, a hipótese nula e o erro sistemático de zero utilizada na esmagadora maioria das aplicações é implausível e, portanto, desinteressante. Isto ocorre porque a estimativas ruidosas que atingem a significância estatística são enviesadas em magnitude ascendente (potencialmente em grau maior) e frequentemente do sinal errado.

Em suma, várias características das ciências sociais — por exemplo, efeitos pequenos e variáveis, erro sistemático, medições ruidosas, uma regra de decisão lexicográfica para publicação, e práticas de investigação — podem tornar o usos dos testes de hipóteses e do p-valor, por consequência, para este domínio.

Outra crítica comum aos testes estatísticos é que obriga os seus utilizadores a concentrarem-se em hipóteses nulas irrelevantes. Não há dúvida de que muitas hipóteses nulas são de fato cientificamente irrelevantes (COHEN, 1994). Este problema de irrelevância não é, contudo, um defeito do p-valor, mas sim um produto da formação tradicional e de um ambiente académico que faz com que os utilizadores se concentrem em tais hipóteses.

Embora existam apelos em curso para abolir o jargão enganador envolvendo "significância", quase nenhuma tentativa foi feita para corrigir o erro de Fisher de usar H0 para qualquer hipótese testada. Esta tradição tem levado os pesquisadores a afirmar que a ciência estatística se limita a testar hipóteses nulas, onde "nulo" significa "sem

associação" ou "sem efeito" em vez de toda e qualquer hipótese de importância ou preocupação.

Tentando corrigir a termologia de Fisher, Neyman (1977) chamou, em vez disso, H0 de hipótese visada ou testada. Mas o jargão de Fisher prevaleceu com uma racionalização reversa de que H0 é a hipótese a ser "anulada" pelo teste, e foi sustentada por tentativas de distinguir hipóteses como $\beta = 0$ como "hipóteses nulas".

O efeitos nefastos continuam a ser visto nos testes de significância de hipótese nula (ZILIANK;MCCLOSKEY, 2008), no qual os p-valores são computados apenas para hipóteses sem efeito, quando também devem ser dadas para hipóteses alternativas de relevância. Um problema mais técnico disso é a concentração excessiva em hipóteses pontuais.

4.6 O SEU STATUS QUO INQUESTIONÁVEL

A história esclarece os debates que ocorrem hoje, em particular, muitas das objeções levantadas ao p-valor pelos pesquisadores modernos e a discussão que acompanha foram levantadas por contemporâneos de Fisher. Desde o início da utilização do valor da probabilidade, a sua teoria já foi questionada.

Um aspecto particular, a importância de considerar a dimensão do efeito e não apenas o significado estatístico, foi o ponto crucial da diferença entre a estrutura de Fisher e Gosset (ZILIAK e MCCLOSKEY, 2008). Ziliak (2008) reitera esta ligação demonstrando a relevância destes debates históricos para a discussão atual.

Um fio de discussão dos primeiros críticos de Fisher indica que a desvalorização do tamanho do efeito em favor do p-valor é um erro fácil de cometer e que precisa de ser tratado. Da mesma forma, os debates já tinham destacados o conflito do paradigma de Fisher com a abordagem Neyman-Pearson.

Estas diferentes questões inferenciais se hibridizaram, e as discussões sobre seu o poder e o papel dos estatísticos na concepção das experiências também já iniciado o debate desde o surgimento do p-valor. O próprio Fisher compreendeu esse questionamento e escreveu um livro inteiro sobre como conceber corretamente um experimento estatístico (FISHER, 1935).

O debate destas ideias, levantando argumentos semelhantes há um século, mostra que não existe uma justificativa clara que aponte o motivo pelo qual o p-valor ainda é um consenso e segue uma medida de estimação tradicional em econometria. Mas, olhando para o seu contexto histórico é possível encontrar uma resposta.

Em particular, se for analisado de forma mais atento obre como as ideias de Fisher se espalharam e aconteceu a hibridização dos paradigmas Fisher e Neyman-Pearson, pode-se ter uma ideia de como este método estatístico se fixou nas comunidades científicas, políticas e públicas em geral.

Benjamini (2016) observou que o valor da probabilidade foi bem-sucedido na ciência porque muito por conta de oferecer uma defesa contra qualquer tipo de aleatoriedade, ou seja, foi útil para os não matemáticos ao dar-lhes uma base quantitativa para enfrentarem a incerteza. Além disso, tem algum significado intuitivo, como se pode ver pelo fato de métodos semelhantes terem surgido repetidamente em vários campos mesmo antes de Fisher.

E teve defensores apaixonados que colocaram as ferramentas nas mãos dos cientistas de uma forma fácil de usar, como através das tabelas estatísticas de Fisher e Yates. Por fim, foi receptiva às condições da época. Estas abordagens abordavam questões sobre variância e desenho experimental que eram frequentemente levantadas na altura (GIGERENZER ET AL., 1989).

Outro fator que explica a consolidação do p-valor ocorre pela falta de capacidade computacional ter tornado as tabelas de Fisher tão valiosas e, portanto, muito influentes para os profissionais no período. E as limitadas capacidades computacionais dos anos de 1950 podem ter diminuído a capacidade dos métodos bayesianos de se aproximarem de um público mais vasto.

O efeito da facilidade do cálculo do p-valor para a criação do seu paradigma é um dos motivos que normalmente gera discordância. No entanto, não há razão para acreditar que as capacidades computacionais tenham um planalto, pelo que uma resposta adequada teria em conta não só as condições de hoje, mas também as susceptíveis de ocorrer no futuro.

Além disso, como destacado anteriormente, os métodos estatísticos nem sempre são utilizados com fidelidade e pressupostos originais, especialmente décadas após a sua formulação inicial. Várias das respostas do motivo pelo qual o valor da probabilidade ainda é utilizado no *mainstream*, são também susceptíveis de sua utilização indevida.

4.7 A SUA NATUREZA FREQUENTISTA

Como já vimos no contexto histórico dos p-valores, a sua natureza e explicação é totalmente frequentista. O p valor é a probabilidade de uma estatística de teste obter pelo

menos um determinado valor, assumindo que a experiência pode ser repetida um número infinito de vezes: isto tem duas implicações.

Primeiro, o investigador ao utilizar o p valor precisa aceitar a validade desta interpretação de probabilidade na sua experiência. Além disso, os valores de probabilidade devem ter uma justificação científica para embasar o problema que o pesquisador pretender responder. Com respeito à essas duas implicações, percebe-se que são utilizados sem que se perceba que isto significa que se concorda silenciosamente com a mentalidade de um frequentador.

Agora, não se pode dizer que isto por si só é um argumento contra a sua utilização. Significa apenas que antes de se utilizar um procedimento estatístico como o p-valor, deve-se chegar ao entendimento dos seus fundamentos e da sua utilização. Na realidade, isto parece faltar muitas vezes. Daí a sua natureza frequentíssima é um apelo a todos os que pensam em utilizá-lo, a pensar da natureza da probabilidade.

Briggs (2001) salienta que este argumento não contém uma disjunção lógica. O que realmente faz certo sentido, mas não completamente. A disjunção no argumento não é lógica uma vez que a primeira parte diz respeito à hipótese nula e a segunda parte está relacionado com o valor da estatística do teste.

Uma disjunção lógica seria do tipo "ou está chovendo ou não está". Briggs (2001), então, propõe uma correção da disjunção para torná-la uma disjunção e tenta mostrar com isto que um pequeno p-valor não tem qualquer relação com qualquer hipótese sem relação com o próprio p-valor. A sua versão fixa do argumento é: ou a hipótese é falsa e a estatística do teste atingiu um valor elevado ou a hipótese é verdadeira e a estatística do teste tem um elevado valor.

Isto, então, resume-se a “ou a hipótese é falsa ou é verdadeira e a estatística do teste atingiu um valor elevado". A tautologia não acrescenta qualquer informação e, portanto, resume-se com a afirmação "a estatística do teste atingiu um valor elevado". O argumento de Briggs (1998), no entanto, ignora quão provável ou improvável é que a estatística do teste atinja um valor elevado quando a hipótese alternativa - que é mutuamente exclusiva da hipótese nula - é verdadeira.

Portanto, a disjunção deve ser fixada de um modo como “ou a hipótese alternativa é verdadeira e a estatística do teste atingiu um valor provavelmente elevado, ou a hipótese nula é verdadeira e a estatística do teste atingiu um valor provavelmente elevado”. A disjunção original seria logicamente válida se declarasse “Ou a hipótese nula é falsa ou

ocorreu algo impossível". Não é certamente verdade que algo impossível ocorreu. Por conseguinte, é certamente verdade que a H0 é falsa.

Outra objeção contra esta justificativa está intimamente relacionada com a incapacidade do p-valor de falsificar uma declaração (próximo subitem). A teoria dos frequentadores afirma que, sob a hipótese nula, cada valor é igualmente susceptível de ser observado, ou seja, o valor da probabilidade é uniformemente distribuído ao longo do intervalo (0,1).

Isto significa que cada p-valor em (0,1) apoia a hipótese nula. Logicamente falando, para afirmar que a hipótese nula é falsa sempre que se observa um valor baixo de p, é portanto um *non sequitur*. Como testes de significância envolvem tanto uma hipótese nula como uma hipótese alternativa com a sua respectiva distribuição do valor da probabilidade, ela torna-se crucial para considerar o poder do teste. Se o poder do teste, ou seja, a probabilidade que um teste rejeita a hipótese nula quando a hipótese nula é falsa, é baixa, o teste torna-se menos fiável.

4.8 A IMPOSSIBILIDADE DE FALSEAR UM TESTE ESTATÍSTICO

Argumenta-se que os p-valores não podem descobrir a causa, mas também que isso não significa que os modelos de probabilidade não são úteis. Só que, em que medida são úteis e se devem ser utilizados? Nesta subsecção, procura-se mostrar que o p-valor não pode fazer o que alguns pensam que ele realiza, nomeadamente falsificar declarações.

Recordemos primeiro, quando é que de facto falsificamos uma declaração? Consideremos um modelo M e uma variável X. Suponhamos que M declara $P(X > 0) = 0$. Quando observamos um $X > 0$, isto contradiz o modelo M e, por conseguinte, o modelo é falsificado. A observação que era inconsistente falsificou o modelo. Quando, no entanto, teríamos um modelo M1 declarando que $P(X > 0) =$ para um certo pequeno, a situação seria diferente. Se depois observarmos um $X > 0$, isto não contradiz o modelo.

Dado o modelo M1, é improvável que um $X > 0$ seja observado, mas não é impossível. Portanto, A observação de $X > 0$ não é incompatível com o modelo e, portanto, o modelo não é falsificado. Assim, um modelo de probabilidade que faz uma declaração de probabilidade entre 0 e 1 - o intervalo aberto sem os valores limite - não podem ser falsificados uma vez que nenhuma observação é estritamente inconsistente com o modelo.

Passemos agora aos valores de probabilidade: sob a hipótese nula, o nível descritivo tem uma distribuição uniforme. Por conseguinte, cada valor é igualmente

susceptível de ser observado. Especificamente, podemos escrever isto como: o modelo M implica que o valor de p em (0; 1). Qualquer p-valor que observemos então, não é estritamente inconsistente com M. Isto significa que é impossível falsificar qualquer afirmação com qualquer p-valor.

Sempre que um pesquisador afirma que falsificou a hipótese nula quando se atinge um valor descritivo baixo, ele ou comete um erro ou tem uma definição muito fraca de falsificação - prática ou quase falsificada. Esta definição não é, no entanto, útil como falsificação no sentido popperiano e por isso não o é.

Além disso, não é possível utilizar valores de probabilidade para falsificar uma declaração, não é consistente com falsificação, como significa Popper. A concepção de um teste de significância envolve uma hipótese nula e uma hipótese alternativa. Quando rejeitamos a hipótese nula, aceitamos a sua contraparte, a hipótese alternativa.

Isto é verdade, uma vez que as hipóteses nulas e alternativas são mutuamente exclusivas. Aceitar uma teoria ou afirmação não está de acordo com a ideia popperiana de falsificabilidade, em que apenas rejeita teorias e não as verifica. A tentativa de usar os p- valores para falsificar qualquer é hipótese, portanto, uma violação da ideia de que as declarações não podem ser verificadas.

4.9 SEM CAPACIDADADE PREDITIVA, OS NÍVEIS DESCRITIVOS NÃO SÃO VERIFICADOS

Em algumas pesquisas parece haver uma tendência para ver os valores de probabilidade como o ponto final e o resultado final da investigação. Sempre que é encontrada uma ligação entre certas variáveis, ou seja, uma baixa ou "significativa": p-valor é encontrado, o resultado é considerado como publicável. Felizmente, já existem chamadas da comunidade estatística que critica este estado.

O ponto destacado aqui é que os valores descritos por vezes são aceitos com demasiada facilidade como o ponto final de uma investigação e que, em contraste para tal, os investigadores devem procurar encontrar provas adicionais sob a forma da capacidade de previsão dos modelos propostos. A capacidade de previsão é importante, pois um dos principais objetivos dos modelos é, muitas vezes, tornar reais previsões.

Quando os p-valores são considerados como o ponto final de uma pesquisa, muitas vezes são cometidos alguns erros. Primeiro, eles podem facilmente ser vistos como a prova de que a hipótese nula é falsa, o que não é, como já discutimos, uma afirmação

verdadeira. Além disso, uma conclusão baseada no p-valor pode transmitir a impressão de que o modelo é bom, no sentido de que se diz existir uma ligação entre variáveis.

Contudo, após o estabelecimento de um modelo, o método empírico deve ser verificado a fim de julgar a veracidade das ligações que são encontradas. Há duas etapas que são cruciais na verificação de um modelo. Em primeiro lugar, são importantes provas adicionais para além dos níveis descritos, como Trafimow *et al.* (1991), entre outros, salientam. Por exemplo, a utilização de razões de probabilidade foi defendida (ver Mogie (2004)). Em segundo lugar, uma vez estabelecidos os modelos, estes devem ser verificados utilizando dados que não foram vistos até esse momento.

Os testes em dados externos indicam até que ponto o modelo generaliza e, por conseguinte, mostra o que a sua capacidade preditiva. Uma vez que o objetivo de muitos modelos é fazer previsões, o método deve ser eficiente para atingir tal objetivo, e como pode-se saber melhor do que verificando a capacidade de previsão?

Felizmente, esta abordagem em que os modelos devem ser testados em dados externos tornou-se o padrão na comunidade de *machine learning*. Há apenas uma advertência: quando os modelos são construídos e subsequentemente testados com base em dados externos, o modelo pode nunca ser adaptado sobre esta amostra.

Quando se adapta novamente o modelo, este deve ser novamente validado assim a capacidade de previsão deve ainda ser provada. No entanto, em outros campos de trabalho, como a sociologia e psicologia, tais abordagens são menos comuns. Isto poderia resultar em conclusões que são também certo.

5. ECONOMETRIA BAYESIANA: ABORDAGEM ALTERNATIVA A ECONOMETRIA MAINSTREAM

A econometria bayesiana é considerada a principal inferência estatística que contrapõe a econometria tradicional. Ao contrário do método clássico, propõe estimar os parâmetros desde o ponto de partida de uma especificação explícita das distribuições *a priori* aos parâmetros em questão. Os estimadores de parâmetros são assim derivados das distribuições posteriores que são o resultado da combinação dos pressupostos antecedentes com as funções de probabilidade baseadas em dados representando as distribuições das amostras.

5. 1 INFÊRENCIA E TEOREMA DE BAYES

A inferência bayesiana recebe o seu nome do teorema de Bayes, uma teoria da probabilidade que é utilizado para atualizar as crenças sobre o valor de parâmetros ou outras quantidades aleatórias usando evidências a partir dos dados. O teorema decorre da fórmula da probabilidade condicional e mantém-se, independentemente de se tornar a visão frequentista ou bayesiana sobre a probabilidade.

Para derivar o teorema de Bayes, é preciso deixar que A e B sejam dois eventos definidos em relação a uma experiência aleatória, com probabilidades P(A) e P(B), respectivamente. A probabilidade de ambos, A e B que ocorre em uma repetição da experiência é denotado por P(A, B). Assumindo que $P(B) \neq 0$, a probabilidade de A ocorrer, dado que B ocorreu, é:

$$P(A|B) = \frac{P(A,B)}{P(B)} \tag{15}$$

Esta fórmula de probabilidade condicional é mais fácil de interpretar após algum rearranjo:

$$P(A, B) = P(A|B) \times P(B) \tag{16}$$

que pode ser explica da seguinte forma:

a probabilidade de ocorrência de A e B é igual à probabilidade de ocorrência de B

a probabilidade de A ocorrer dada a ocorrência de B

Pode-se pensar nesta identidade como uma forma de calcular a probabilidade conjunta de A e B — calculando a probabilidade de B a primeiro e depois examinando a probabilidade de A, enquanto trata B como tendo já ocorrido. Entretanto, a dimensão

temporal introduzida aqui, onde imagina que B ocorrendo antes de A, é utilizada apenas para dar alguma interpretação intuitiva da fórmula. Invertendo os papeis de A e B no lado direito, tem-se:

$$P(A, B) = P(B|A) \times P(A) \quad (17)$$

O teorema de Bayes segue-se igualando os lados direito de (16) e (17) e reordenando:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (18)$$

O teorema de Bayes inverte os papeis dos dois eventos no condicionamento e, ao fazê-lo, permite o cálculo de P(A|B) utilizando o pressuposto da P(B|A). Mais precisamente e no contexto da inferência bayesiana, conhecer a P(B|A) e a P(B) permite atualizar a crença de que A acontece após a certeza de que B ter ocorrido.

O teorema bayesiano foi apresentado acima através de eventos, mas também pode ser demonstrado por meio de variáveis aleatórias. Dado que X e Y sejam duas variáveis aleatórias com probabilidade funções de densidade p(x) e p(y), respectivamente, e usando estas funções de densidade de probabilidade de Bayes, o teorema pode ser expresso como:

$$p(x|y) = \frac{p(y|x) \times p(x)}{p\,(y)} \quad (19)$$

Na inferência bayesiana x desempenha o papel dos parâmetros de um modelo estocástico e y o papel dos dados. De maneira notacional, ao recolher todos os parâmetros em um vector θ e os dados num vetor y e renomeando algumas das densidades, o teorema torna-se (20):

$$\pi\,(\theta|y) = \frac{p(y|\theta) \times p(\theta)}{m\,(y)} \quad (20)$$

Naturalmente, se o modelo envolver mais do que um único parâmetro e mais do que um único dado ponto é utilizado, então todas as densidades na última expressão serão multivariadas. A última expressão envolve quatro densidades:

1. $\pi\,(\theta|y)$ é a densidade *a priori* e é a quantidade primária de interesse da inferência. Exprime o conhecimento sobre os valores dos parâmetros do modelo após os dados ser analisado.

2. p(y|θ) é a função de probabilidade e é a parte principal da especificação do modelo. A

   a função de probabilidade é a densidade da amostra, dados os valores dos parâmetros do modelo, e depende dos pressupostos impostos pelo investigador sobre a geração de dados do processo.
3. p(θ) é a densidade *a priori* dos parâmetros do modelo e é um elemento adicional do processo de especificação do modelo. A densidade *a priori* expressa conhecimentos ou crenças sobre os valores dos parâmetros antes de analisar os dados.
4. m (y) é a probabilidade marginal e, como o seu nome sugere, é a densidade dos dados marginalmente no que diz respeito aos parâmetros. A sua forma depende da especificação do modelo (função de probabilidade e densidade *a priori*) e pode ser obtido integrando θ a partir do numerador de (20).

Na maioria das aplicações será difícil realizar esta integração analiticamente do ponto, mas, dado que a quantidade primária de interesse é a densidade *a posteriori* deos parâmetros e que m (y) não envolve θ, o denominador na última expressão pode ser visto como uma constante de proporcionalidade para π (θ|y). Esta constante é irrelevante para fins de estimativa e podem ser ignorados nesta fase. Na prática, portanto, é mais frequentemente omitir (20) e expressar o teorema de Bayes assim:

$$\pi(\theta|y) \propto p(y|\theta) \times p(\theta) \qquad (21)$$

com o símbolo "∝" tomado para significar "proporcional à".
A identidade (21) indica que: a densidade *a posteriori* dos parâmetros do modelo é proporcional à probabilidade vezes a densidade *a priori*

O lado direito contém a especificação completa do modelo. É precisar salientar que um modelo de uma inferência bayesiana consiste tanto na função de probabilidade como na densidade *a priori* dos parâmetros. Devido ao papel fundamental que as três densidades que aparecem em (21) desempenham para esta teoria, cada uma delas é examinada em pormenor nas três subseções seguintes.

5.1.1 A FUNÇÃO DE PROBABILIDADE

A função de probabilidade constitui parte da especificação de um modelo estocástico e transmite os pressupostos sobre o processo que gera os dados. É expressa como uma densidade do formulário $p(y|\theta)$, onde y são os dados e θ os parâmetros do modelo. O significado de θ é simples, mas o significado de y merece alguma discussão.

Com um dado conjunto de dados definidos, y será povoado por valores numéricos. Em modelos estocásticos, estes valores são vistos como as realizações de variáveis aleatórias; as realizações são o que se observa (fixas), mas os processos subjacentes de geração de dados são aquilo em que se tem interesse.

Isto porque a inferência estatística não é preocupado em descrever simplesmente o conjunto de dados em questão, mas o seu objetivo principal é fazer afirmações sobre os valores de θ na população. E a única forma de o conseguir é considerando o processo que gera os dados na amostra. Evidentemente, os dados são utilizados no processo de inferência estatística para fornecer informações sobre os valores dos parâmetros.

Em outras palavras, e dada a discussão anterior, a função de probabilidade é a probabilidade de um conjunto de dados potencial, avaliado nos pontos de amostras observadas, tendo em conta os valores dos parâmetros. Estes ainda não são conhecidos e o condicionamento sobre eles pode parecer singular. No entanto, o teorema de Bayes pode ser utilizado para inverter os papéis de y e θ em condicionamento, de modo a obter a densidade dos parâmetros dos dados observados.

5.1.2 A DENSIDADE *A PRIORI*

A densidade *a priori* constitui a segunda parte da especificação de um modelo bayesiano de inferência e transmite crenças ou conhecimentos prévios sobre os valores dos parâmetros de um modelo. Estas crenças são *a priori* no sentido em que são formadas sem utilizar informações contidos no conjunto de dados em questão.

Tal como a função de probabilidade, a densidade *a priori* toma a forma de uma função de densidade de probabilidade, expressa em termos gerais como $p(\theta)$. Na prática, isto pertencerá a uma família paramétrica convenientemente escolhida e será aumentada pelos próprios parâmetros desta família, chamados de hiperparâmetros.

Tanto a família da densidade *a priori* como os valores dos hiperparâmetros são escolhidos pelo investigador e, como qualquer forma de especificação de um modelo, podem ter um impacto considerável sobre as conclusões retiradas da análise. Mais importante ainda, porque estas escolhas não são atualizadas no processo de estimativa ou inferência, têm o potencial de introduzir um grau de subjetividade na análise.

5.1.3 DENSIDADE *A POSTERIORI*

A densidade *a posteriori* é o produto final de um exercício de inferência bayesiana, pelo menos até gerar a estimativa dos parâmetros. Esta densidade toma a forma geral de π (θ|y) e expressa o conhecimento do pesquisar sobre θ depois de ter visto os dados. A densidade posterior é obtida a partir de uma aplicação do teorema de Bayes:

$$\pi(\theta|y) = \frac{p(y|\theta) \times p(\theta)}{m(y)} \propto p(y|\theta) \times p(\theta) \quad (22)$$

o que torna evidente que depende dos pressupostos de modelização que são incorporados tanto pela a probabilidade como pela suposição *a priori*.

Quando o modelo envolve um único parâmetro, a densidade *a posteriori* é a base deste parâmetro de função de densidade de probabilidade. Quando há mais parâmetros no modelo, a suposição *a posteriori* é a densidade de articular todos os parâmetros. Matematicamente, toma-se a figura de uma fórmula o que, na maioria dos casos, proporciona pouca intuição sobre os valores do(s) parâmetro(s).

A tarefa agora tem como objetivo de extrair a informação contida em π (θ|y) e apresentá-la de uma forma que seja fácil de compreender. Uma forma óbvia de proceder é traçar um gráfico de cada densidade *a posteriori* do parâmetro, marginalmente em relação ao resto. Contudo, esta abordagem torna-se impraticável se o modelo tiver mais do que alguns parâmetros. Por conseguinte, é habitual apresentar em uma tabela os dois primeiros momentos da densidade marginal *a posteriori* de cada parâmetro em θ.

5.2 O DESENVOLVIMENTO DO MÉTODO BAYESIANO EM ECONOMETRIA

O potencial da inferência bayesiana em econometria foi reconhecido primeiramente por Jacob Marschak na década de 1950. Para responder à pergunta se existe alguma relação entre as probabilidades subjetivas e o método estatístico, Marschak (1954) utilizou exemplos simples para ilustrar como a relação de duas das fórmulas de Bayes poderiam ser utilizadas para comparar e modificar graus de crença prévia através da sua combinação com as funções de probabilidade.

Mas o estudo de Marschak não inspirou nenhum discípulo econométrico a experimentar este novo campo de pesquisa. Levou uma década até que fosse seriamente explorada em econometria, motivada principalmente pela investigação bayesiana na

matemática estatística. As obras pioneiras incluem Fisher (1962), Hildreth (1963), Rothenberg (1963), Zellner e Tiao (1964).

As obras de Fisher (1962) e Hildreth (1963) foram exploratórias a um nível teórico. Fisher (1962) examinou os diferentes efeitos na estimativa de um modelo bayesiano induzidos por diferentes propósitos de utilização, como, por exemplo, para previsão ou para simulação de políticas.

A partir de duas funções de perda com correspondentes distintos para os dois diferentes fins, Fisher (1962) derivou dois conjuntos de estimativas de coeficientes que minimizariam as duas diferentes funções de perda, respectivamente, usando o teorema de Bayes (em termos de densidades anteriores assumidas).

O procedimento ligou efetivamente a etapa de estimativa com a função de bem-estar desejada para fins de controle da política. A discussão de Hildreth (1963) também considerou esta questão do ponto de vista da tomada de decisões. Ele mostrou como obter diferentes estimadores pontuais, substituindo os critérios estatísticos padrão por outros instrumentos de acordo com os requisitos do problema em questão.

Tiao e Zellner (1964) foram um dos primeiros pesquisadores a experimentar os métodos bayesianos para modelos de regressão. No seu trabalho (1964) os autores a partir do procedimento bayesiano padrão de um modelo de regressão simples, aplicaram o método a uma equação de investimento com os dados de séries temporal de duas empresas tratadas como dois subconjuntos, em que:

$$y_t = \beta_{xt} + \varepsilon_t; \qquad \varepsilon_t \sim \text{IID}(0,\sigma^2) \qquad (23)$$

em que $\varepsilon_t = \rho\varepsilon_{t-1} + \mu_t$, para contornar o dificuldade de derivar a desejado propriedade assimptótica dos estimadores clássicos nessa situação.

A ideia de Tiao e Zellner (1964) foi de dividir uma amostra de dados em duas, e tomar uma das estimativas do subconjunto como a parâmetros do anterior, que se presumiu seguir a função de distribuição uniforme local. As estimativas do coeficiente posterior foram então obtidas combinando a função de probabilidade, $I(\beta\ \sigma)$, com base nos dados do outro subconjunto.

Já Rothenberg (1963) combinou uma equação simultânea com uma função de perda que representa o comportamento político, a fim de estudar os efeitos dos diferentes antecedentes nas estimativas dos parâmetros posteriores sob os pressupostos alternativos de variação de erro do modelo.

O começo da década de 1970 foi marcada inicialmente pelo primeiro grande manual da economia bayesiana. O livro texto de Arnold Zellner em 1971, "*An Introduction to Bayesian Inference in Econometrics*", constituiu-se marco para inferência bayesiana em econometria ao reformular as obras padrões do método e apresentar de forma didática toda a estrutura da metodologia, passando desde de uma simples regressão linear a problemas especiais na análise de estimação (como modelos com autocorrelação ou com erros nas variáveis).

Nas pesquisas de fronteira da época, os economistas bayesianos se concentram sobre duas dificuldades técnicas na concepção de estimadores: a dificuldade de especificar distribuições *a priori* que eram ambas economicamente interpretável e matematicamente traçável, e o problema de integrar o cálculo de forma consistente para a derivação das distribuições *a posteriori* por meio da combinação com distribuição *a priori* a partir das funções de probabilidade.

As suas descobertas empíricas muitas vezes confirmam as do campo clássico, embora se acreditasse que os bayesianos tinham o poder de corrigir os erros de estimação, frequentemente observado em regressões clássicas, de ter errado economicamente os sinais ou a magnitude entre as estimativas de coeficientes (ROTHENBERG, 1975).

Segundo Press (1980), os pesquisadores bayesianos também tiveram que se envolver no desenvolvimento de software e de métodos de integração numérica, dado a complexidade que os cálculos exigiam. Um importante avanço foi feito por Kloek e van Dijk (1978), que exploraram a integração do procedimento de Monte Carlo para permitir o cálculo integral para uma gama mais vasta de distribuição *a priori* do que eram analiticamente solúveis.

Estes esforços para superar as dificuldades técnicas foram sustentados de forma vital pelo apelo filosófico da inferência Bayesiana (já naquele período era considerado o meio de superar os problemas da econometria *mainstream*). Mas a mera reformulação da econometria dos manuais clássicos pela via bayesiana foi inadequada para o benefício reclamado.

Qin (2011) relata que no que diz respeito aos estudos empíricos, nada realmente diferente do que tinha sido obtido pela análise clássica foi apresentado nos poucos resultados bayesianos, tais como o estudo do efeito da autocorrelação dos resíduos por Chetty (1968) e o trabalho sobre a influência da especificação do modelo de distribuição de defasagem por Zellner e Geisel (1970).

A situação indicava claramente um desacordo entre as informações de dados e as crenças *a priori* altamente restritivas. Esta discordância indicou a possibilidade de erro de especificação do modelo, consequentemente, o papel da distribuição *a prori* no exame da probabilidade. Como sugeriu Rothenberg (1973), era preciso fazer uma escolha entre utilizar a informação prévia (*a priori*) para melhorar as estimativas das amostras ou usar a amostra para testar a validade da informação. Esta última se tornou a abordagem de especificação bayesiana.

O primeiro economista bayesiano cuja investigação se centrou no efeito da especificação do modelo foi E. E. Leamer. Sua análise começou com o modelo de distribuição com *lag* (LEAMER, 1972). Ao contrário daqueles que tomaram o modelo como verdadeiro e que se concentraram na derivação dos estimadores, Leamer (1972) mergulhou na natureza da colinearidade. Ele descobriu que o problema nunca tinha sido rigorosamente examinado e começou a preencher a lacuna.

De um ponto de vista bayesiano, ele percebeu que a colinearidade decorrida das tentativas dos pesquisadores interpretar os resultados à luz de algumas incertezas *a priori* com informações incompletas, ao invés do problema das evidências fracas prescrito nos livros de texto de econometria.

Em seguida, reduziu o problema de caracterização e interpretação de uma função de probabilidade multidimensional para um problema de caracterização e interpretação de uma distribuição prévia multidimensional, e atribuiu a dificuldade de uma especificação inadequada de informação *a priori* para ajudar a conceder um conjunto de dados a coeficientes individuais.

Leamer propôs assim medir o grau de colinearidade através da sensibilidade da distribuição *a posteriori* a partir das alterações na distribuição *a priori* (LEAMER,1973). Tal análise de sensibilidade ensinou-lhe a lição de que a colinearidade cria assim um incentivo para utilizar as informações com mais cuidado.

Posteriormente, Leamer (1978) analisou a colinearidade juntamente com os problemas que envolvem a intepretações dos resultados (de acordo com certos postulados teóricos) e os esforços dos pesquisadores para lidar com estes problemas, e afirmou que estes erros deriva da forma que as hipóteses são alargadas e da existência de muitas "alternativas viáveis".

Para o autor, a inferência com dados econômicos pode ser descrita como uma busca discriminada de um modelo aceitável. Isto acontece devido (i) à natureza incontrolável e complexa dos processos econômicos, (ii) a ampla utilização de

informação não amostral por parte dos economistas, e (iii) a inclusão em modelos de apenas uma pequena fração do grande conjunto de variáveis econômicas.

Leamer (1978) ainda salientou que a sua investigação bayesiana não se destinava a lidar com a questão de como deve proceder a busca de novos modelos econômicos, mas para proporcionar um método para fazer inferências coerentes de modelos pós-dados. Na sua tentativa de tornar coerentes as inferências deste modelo, Leamer observou duas grandes dificuldades.

O primeiro empecilho foi causado por julgamentos ou decisões *a priori* implícitas sobre os quais foram selecionadas e apresentadas pelas evidências, uma vez que a seleção estava, portanto, obrigada a enviesar a procura de novas hipóteses. Leamer descreveu este viés como "descontado" ou "contaminado". Já a segunda dificuldade estava relacionada com o risco de dupla contagem nos resultados, ou seja, a utilização errada da mesma evidência tanto para a construção de novas hipóteses como para testar as hipóteses existentes.

Buscando uma solução para esta problemática, Leamer descobriu que nenhuma das hipóteses clássicas nem a inferência bayesiana teriam uma reposta pronta. Esta descoberta confirmou o argumento de Lempers (1971) sobre a impossibilidade de escolher diferentes variáveis explicativas no método bayesiano.

No entanto, Leamer (1974) reduziu as pesquisas de modelos pós-dados a dois tipos: os que acrescentam novas variáveis a um modelo existente e os que adicionam um método inteiramente novo para aquele modelo que já foi estimado. Independentemente do método, as buscas destinavam-se a começar com modelos teóricos muito simples e estendê-los até que vários critérios econométricos fossem cumpridos.

Vale salientar que essa discussão tinha como foco mostrar como uma medida bayesiana de custo-benefício pode avaliar o impacto da queda de uma variável explicativa aparentemente insignificante, ou marginalmente significativa, sobre as estimativas dos parâmetros de outras variáveis explicativas de retenção, uma questão que tinha sido examinado sob o rótulo de "polarização de variáveis omitidas" vinte anos antes por Griliches (1957).

Um fato que Leamer deixou de lado na sua busca de uma estratégia sistemática de escolha de modelos foi a exogeneidade, que foi introduzida na econometria bayesiana por Lubrano *et al.* (1986) e Zellner et al. (1988). Os testes de exogeneidade aproximou as pesquisas bayesianas com as da econometria tradicional.

A fusão dessas duas vias deu origem a novas ferramentas e critérios de seleção de modelos, tais como a comparação de métodos rivais ou de critério de informação *a posteriori* baseado em predições (PHILLIPS, 1995). Um catalisador importante dessa associação foi o avanço da abordagem VAR e especialmente o VAR Bayesiano como uma ferramenta útil para as pesquisas de séries temporais.

Semelhante à pesquisa de especificação e de seleção de modelo Leamer, o VAR surgiu como um remédio para a falta de modelos teóricos e empíricos *a priori*. Mas, enquanto, Leamer se concentrava na seleção de questões relativas a múltiplas variáveis explicativas em um contexto de uma única equação, a abordagem VAR preocupava-se principalmente com a especificidade dinâmica das equações estruturais.

Mais precisamente, a abordagem preconizava a utilização de um VAR geral como ponto de partida, em vez de qualquer equação estrutural específica, a fim de superar as dinâmicas determinadas da mesma. Não obstante, promoveu uma estratégia de modelização semelhante à de Leamer, embora de um ângulo diferente - o da dinâmica.

Tecnicamente, uma dificuldade que a estratégia encontrou inicialmente foi o problema da dimensionalidade, ou seja, o número de parâmetros aumentou drasticamente com o número de variáveis incluídas nos VARs e com a quantidade máxima de lag necessários no estágio inicial (SIMS, 1980).

Conforme Sims (1980), o problema da dimensionalidade também teve um efeito adverso sobre a predição dos modelos, uma vez que a melhoria do ajuste dentro da amostra através de VARs mais gerais tendia a dar origem a maiores desvios médios quadráticos fora da amostra. O método Bayesiano seria um possível candidato para ajudar a resolver o problema.

A aplicação de métodos bayesianos aos VARs foi primeiramente explorada por Robert Litterman no seu doutorado. Ao construir um modelo VAR de previsão mensal, Litterman (1986) empregou o princípio bayesiano de uma função de perda de erro média ao quadrado para controlar os desvios de previsão, juntamente com os fundamentos *a priori*, para resolver o problema de "sobreparametrização" dos VARs gerais.

A sua imposição dos pressupostos *a priori* no número de *lags* dos VARs foi motivada principalmente pelo estudo de Leamer (1972) sobre o modelo de defasagens distribuídas e da técnica de estimações mistas de Theil (1963). Os experimentos de Litterman foram realizados em pesquisa conjunta de Doan, Litterman e Sims (1984), que desempenharam um papel vital na promoção da abordagem BVAR (VAR Bayesiano).

A estrutura *a priori* que Doan *et al.* (1984) utilizaram sobre os termos desfasados equivaliam a impor certas restrições comuns sobre as variações ou nos números de *lags* dos coeficientes em questão. A relação entre as escolhas de informações antecedentes e os erros de previsão (de um passo à frente) foi monitorizada e foram selecionados uma distribuição *a priori* que minimizariam estes desvios.

Ressalta-se que os fundamentos *a priori* de *Doan et al.* (1984) utilizados nos BVARs não tinham qualquer base estrutural, ou seja, não exigiam qualquer conhecimento econômico inicial, bastante diferente da posição normal assumida pelos economistas bayesianos. Além disso, tornaram-se objeto de análise de sensibilidade e serviram efetivamente como um meio para alcançar uma simplificação nas equações dinâmicas.

A estimativa do modelo foi realizada recursivamente para estudar a constância dos coeficientes ao longo do tempo: os seus resultados mostraram que as estimativas dos parâmetros *a posteriori* eram insensíveis as variáveis *a priori* à medida que o tamanho da amostra aumentava, sugerindo que era possível obter modelos relativamente coerentes com os dados quando estes eram dinamicamente especificados de forma adequada.

A investigação bayesiana desviou-se ainda mais da econometria tradicional quando se envolveu no debate sobre a inferência estatística das raízes unitárias e das tendências estocásticas na análise de séries temporais (KOOP, 1994). A partir de meados dos anos 70, aumentou a discussão sobre o uso das séries temporais macroeconômicas.

Este debate teve início pelo alerta de Granger e Newbold (1974) sobre os erros que os pesquisadores cometiam ao realizarem regressões entre variáveis não estacionárias, que posteriormente foi amenizada pelo procedimento de teste unitário de Dickey-Fuller (1979) e pela descoberta de Nelson e Plosser (1982) de propriedades unitárias em séries do tempo.

As experiências BVAR, contudo, levaram Christopher Sims a ver essa solução de forma crítica. Usando as estimativas bayesianas com termos desfasados para permitir raízes unitárias, Sims (1988) mostrou que a inferência bayesiana não era afetada pela não estacionariedade de uma série temporal e podia gerar resultados finitos diferentes dos gerados pelo método clássico.

A contribuição de Sims sobre a discussão referente a raiz unitária foi criticada por Phillips (1991), que mostrou como a escolha de diferentes informações *a priori* afetaria as estimações bayesianas ao analisar as suas raízes unitárias e que seu uso era impróprio para o debate enquanto as predileções de seus fundamentos tivessem um impacto preponderante para a comparação de seus resultados com a inferência clássica.

A discussão sobre as raízes unitárias bayesianas levou a Phillips a explorar o uso dos métodos bayesianos para a seleção de modelos dinâmicos. Ele propôs um critério de informação *a posteriori* (PIC) para ajudar a determinar a ordem dos *lag*s e as propriedades de tendência das variáveis, além de ter concebido um teste abrangente de previsões para auxiliar uma pesquisa parcimoniosa para estes modelos (PHILLIPS, 1995).

Entretanto, a ideia de aplicar métodos bayesianos aos modelos VAR foi ampliado aos modelos de equilíbrio geral estocástico dinâmico (DSGE). O principal problema comumente reconhecido dos modelos DSGE era a utilização arbitrária de parâmetros "calibrados". Uma maneira útil de lidar com este problema é atribuir distribuições *a priori* a esses parâmetros e simular os resultados do modelo em termos probabilísticos.

No âmbito do modelo VAR, os métodos bayesianos foram também alargados para à produção de bandas de confiança para respostas de impulso — estas extensões ajudaram a revitalizar a econometria bayesiana. Notavelmente em relação ao revigoramento, a imagem subjetivista da abordagem bayesiana foi enfraquecida com o entendimento da importância da sua estimação em modelos extensos.

5.3 DESAFIOS DA INFERÊNCIA BAYESIANA

5.3.1 A ESCOLHA DA *DISTRIBUIÇÃO A PRIORI*

Como explicado na introdução deste capítulo, o núcleo da inferência bayesiana consiste em representar os graus de crenças por probabilidades, em mudá-los através da sua condicionalização, e basear as decisões em probabilidades posteriores. Em particular, o grau de crença *a posteriori* de uma hipótese H ao conhecer a evidência E pode ser escrita da seguinte forma:

$$p(H|E) = \frac{p(E|H) \times p(H)}{p\,(H)} \tag{24}$$

Com base na probabilidade *a posteriori* $p(H|E)$, um pesquisador que utiliza a metodologia bayesiano pode formar um julgamento teórico sobre H ou tomar uma decisão prática. Por exemplo, se H for a hipótese de que um novo medicamento é menos eficaz que um placebo, e se H for suficientemente provável tendo em conta os dados, então o desenvolvimento da droga não deve ser descontinuado.

A probabilidade posterior depende da probabilidade *a priori*, e muitas vezes não há evidências de base suficientes para estabelecer um consenso sobre este último. Os bayesianos subjetivos como o Finetti (1972) sublinharam que, em princípio, qualquer distribuição de probabilidade *a priori* coerente pode ser defendida como racional.

Esta atitude parece pôr em risco quaisquer reivindicações de objetividade que os bayesianos subjetivos poderiam eventualmente fazer. Este fato, gera uma grande questão: que tipo de determinação epistêmica que uma inferência bayesiana proporciona? Afinal de contas, a escolha *a priori* pode esconder todo o tipo de valores perniciosos.

Em outras palavras, quem utiliza os fundamentos bayesianos podem enviesar o resultado final na sua direção preferida escolhendo um dado inicial adequado. O primeiro desafio baseia-se no ideal que a atividade principal do raciocínio científico, com o objetivo de avaliar e aceitar teorias, deve estar livre de valores não cognitivos e preconceitos individuais - um requisito que a inferência bayesiana parece violar de forma flagrante.

Aderente a liberdade de escolha de valores dos pesquisadores é, no entanto, de uma forma ou de outra, mantida como uma marca comercial de caráter científico para a estatística bayesiana, desempenhando um papel ainda mais importante devido as restrições regulamentares e aos conflitos de interesses. Mesmo que se duvide que o possa ser atingido na prática, não devem ser permitidos valores que substituam os fundamentos científicos. Como a inferência bayesiana pode evitar esse perigo?

A primeira linha de defesa observa que a opinião subjetiva não precisa de ser separada do pré-conceito individual. Por exemplo, dois médicos podem com base na sua experiência de trabalho dar um julgar de forma diferente sobre o que pode ser uma boa terapia para um paciente dado um conjunto de sintomas. O fato de não concordarem não significa que um deles ou ambos sejam tendenciosos: eles podem ter desfrutado de uma formação diferente, viram disciplinas ou têm experiências distintas ao lidar com esses sintomas.

As distribuições *a priori* fornecem uma forma de tornar explícito um julgamento que seja alimentado por especialistas individuais e registo de percurso. Esta é também uma razão pela qual muitos modelos de julgamento e tomada de decisão utilizam inferências bayesianas – mesmo quando são necessárias avaliações de risco objetivos (COOKE, 1991).

A segunda linha de defesa observa que as probabilidades iniciais estão abertas a críticas racionais. Sempre que uma distribuição *a priori* é utilizada, seja a sua forma

convencional ou peculiar, o investigador deve justificar a sua escolha particular e explicar quais as considerações (teóricas e empíricas) que o levaram a esta escolha.

Dado que não se pode justificar uma distribuição *a posteriori* simplesmente escolhendo um dado inicial adequado, uma vez que faz parte da construção do modelo bayesiano justificar a escolha da amostra *a priori*. A afirmação acima proporciona uma segunda exigência no fundamento bayesiano: realizar uma análise de sensibilidade sobre a escolha *a priori* e verificar se o resultado principal da investigação permanece intacto sob diferentes pressupostos iniciais.

Esta análise contribui também para a sua objetividade científica em termos de convergência (DOUGLAS, 2004), visto que o resultado científico pode ser considerado como objetivo quando é validado a partir de diferentes pressupostos e perspectivas. Verificando como uma variação *a priori* afeta os resultados do modelo, os fundamentos resultadistas da estatística bayesiana satisfazem este sentido de objetividade.

Finalmente, a terceira linha de defesa observa que a escolha explícita de uma distribuição expõe mais claramente as hipóteses de modelação do que a disputa de paradigmas. Na inferência frequentista, por exemplo, tais pressupostos são mais implícitos e mais difíceis de identificar.

O resultado final deste conjunto de questionamentos teóricos é que a escolha dos dados iniciais como qualquer outro pressuposto cientifico precisa ser confrontado com uma crítica racional. De fato, senão fossem questionados e julgados de forma crítica, não haveria correção dos mecanismos de aferição até ao ponto do pré-conceito pessoal influenciassem os resultados através da escolha *a priori*.

Mas o mesmo é válido para os resultados científicos: as inferências em geral e de paradigmas estatísticos concorrentes em particular. Invalidar uma análise bayesiana subjetiva com um dado inicial enviesado é tão fácil como invalidar uma análise não bayesiana com pressupostos relacionados a modelação tendenciosa. Portanto, este desafio não é mais assustador para os bayesianos do que para qualquer outro quadro de inferência indutiva.

5.3.2 CRENÇA VERSUS PROVAS

O segundo desafio sustenta que o raciocínio científico e a análise estatística, em particular, não se tratam de avaliar o grau de crença numa hipótese, mas sobre descobrir se um determinado efeito é real ou devido ao acaso. Nesta perspectiva, o estatístico bayesiano comete um erro de categoria: eles tentam responder a uma pergunta que os

cientistas não são (e não deveria) estar interessado em, nomeadamente, responder quão plausível é uma hipótese a partir de um ponto de vista subjetivo.

O raciocínio estatístico deve ser independente de tais julgamentos; é tarefa da ciência declarar a evidência objetiva para a verdade da hipótese. Ronald A. Fisher, um dos pais das estatísticas modernas, articulou energicamente este ponto de vista: defensores das probabilidades inversas são forçados a considerar a probabilidade matemática, não como uma quantidade objetiva medida por frequências observáveis, mas como a medição de tendências meramente psicológicas, teoremas que são inúteis ou fins científicos.

Subjacente a este desafio está a ideia de "objetividade desprendida" (DOUGLAS, 2009, 459): “as alegações ao conhecimento científico devem ser destacadas de crença pessoal e pensamento desejoso”. Os bayesianos também lutam para alcançar a objetividade “concordante" que é expressa de forma intersubjetiva em avaliações de provas acordadas: a exigência de intersubjetividade é o que torna a ciência objetiva.

Contudo, as "tendências psicológicas" que correspondem a graus pessoais de crença não preenchem este requisito. Muitos filósofos e cientistas partilham a opinião de que a inferência bayesiana fica aquém de alcançar uma objetividade concordante. Williamson (2007) observa que a objetividade plena - ou seja, uma única função de probabilidade que se enquadra na evidência disponível não pode ser alcançada no quadro bayesiano subjetivo.

Para abordar esta preocupação, ressalta-se as medidas bayesianas mais populares de apoio probatório com algum detalhe. O fator de Bayes, que é encontrado em variações *a priori*, exprime o apoio para H0 sobre a alternativa H1 em termos da relação *a posteriori* e probabilidades iniciais. Da mesma forma, o fator de bayes pode ser expresso como a razão de probabilidade (integrada) de H0 e H1:

$$B_{01}(D) = \frac{p(HO \mid D)}{p(H1 \mid D)} \, x \, \frac{p(H1)}{P(H0)} = \frac{\int_{\theta \epsilon \vartheta 0}^{1} p(D|\theta) p\,(\theta) d(\theta)}{\int_{\theta \epsilon \vartheta 1}^{1} p(D|\theta) p\,(\theta) d(\theta)} \qquad (25)$$

É importante notar que o fator de Bayes não é afetado pelo simplificador p(H0) e p(H1). Para duas hipóteses pontuais H0 e H1, ele é totalmente independentemente da distribuição *a priori* da probabilidade: é apenas a razão do probabilidade p(D │ HO) / p(D │ H1), indicando o quanto D favorece H0 em relação a H1: nada depende da crença pessoal.

Para hipóteses compostas (e.g., H1: $\theta \epsilon$ [a,b], $\theta \neq \theta_0$) as coisas são mais complicadas.

O valor do fator de bayes depende de a probabilidade das provas observadas serem inferiores os vários componentes de H0 e H1, ponderados com a sua probabilidade relativa *a priori*. É importante perceber que esta dependência é benigna e não perniciosa no contexto de testes de hipóteses nulas.

De fato, a inferência frequentista também com o Teste de Significância de Hipótese Nula (NHST) precisa de tais julgamentos de plausibilidade. Em NHST, a hipótese nula H0:θ = 0 de efeito zero é colocado contra a hipótese alternativa H1:$\theta \neq 0$ que existe algum efeito; Enquanto um erro do tipo I corresponde à rejeição errónea da hipótese nula, um erro do tipo II significa a aceitação errada do nulo, ou mais precisamente, do fracasso para rejeitar o nulo. Convencionalmente, as taxas de erro de tipo I aceitáveis são fixadas em um nível de 5%, 1% ou 0,1%, dependo da experiência. Ao escolher um tamanho de amostra N apropriado, tenta-se minimizar o erro de tipo II.

No entanto, quando H1 é uma hipótese composta, o poder de uma experiência tem de ser calculado em relação a tamanhos de efeitos específicos. Normalmente, escolher-se-iam tamanhos de efeito que correspondessem às expectativas teóricas e que implicassem uma diferença cientificamente significativa em relação à hipótese nula. Sem ter um julgamento sobre quais os tamanhos de efeito prováveis de serem esperados, a escolha do tamanho da amostra N equivale a tatear no escuro.

Pode-se recolher muito mais provas do que as necessárias ou, no caso contrário, acabam por ser objeto de um estudo severamente subestimado. Por conseguinte, a relativa plausibilidade das diferentes alternativas e a da hipótese nula inicial afeta a concepção e a avaliação das experiências em inferência dos frequentistas.

A mesma dependência é ainda mais evidente quando a inferência frequentista é utilizado para apoiar decisões práticas. Como já argumentado por Rudner (1953), escolher e equilibrar os níveis de erro de tipo I e tipo II envolve juízos de valor não cognitivos: os pesquisadores revelam implicitamente quão severos e como é provável que se encontra estes erros.

Uma decisão sobre se deve rejeitar ou não uma hipótese e agir com base na mesma tem a necessidade de trocar a plausibilidade pela força das provas. Por conseguinte, considera-se que os graus de crença não devem desempenhar qualquer papel na avaliação. Se aplicado rigorosamente, isto significaria que seria necessário parar de fazer inferências desde os dados até à teoria.

Assim, também Douglas (2004) salienta que a objetividade no raciocínio científico não implica a eliminação de perspectiva pessoal; isto seria de fato uma deturpação grosseira de como a ciência funciona. Assim, pode-se afirmar que o segundo desafio também é mal orientado: as provas científicas não podem, nem devem ser ordenadamente separados dos julgamentos de plausibilidade e graus de crença.

5.3.3 NEGLIGÊNCIA DO DESENHO EXPERIMENTAL

O terceiro desafio da teoria bayesiana diz respeito ao problema do enviesamento em julgamentos com uma análise provisória dos dados. De fato, críticos da inferência bayesiana, como Deborah Mayo (1996) queixam-se que desacoplar a inferência estatística do protocolo de amostragem pode levam a uma elevada probabilidade de erro e essa elevada probabilidade de erro não é refletida na interpretação dos dados.

Por exemplo, o pesquisador pode gerar dados até obter um resultado convincente, sem mencionar o tendencioso procedimento tendencioso na submissão e publicação de seu trabalho. Afinal de contas, se a amostra foi obtida por meios de um protocolo de dados manipulados, um protocolo imparcial ou nenhum não afeta de modo algum a probabilidade *a posteriori* nem o fator de Bayes. Mais uma vez, a percepção da ameaça à objetividade da inferência bayesiana vem da intrusão oculta de preconceitos e valores não cognitivos no raciocínio estatístico.

Quatro respostas podem ser dadas a esta crítica. Em primeiro lugar, o fenômeno em que a crítica se baseia também pode ser descrita de forma diferente: mais não são surpreendentes, mas previsíveis (GOODMAN et al., 2010). De todos os tratamentos, os eficazes serão mais propensos a rescisão antecipada a fim de beneficia-lo, ou seja, quando a dimensão real do efeito é grande, é mais provável que também se um grande efeito na amostra e decidir terminar o experimento.

Assim, a diferença observada entre ensaios truncados e terminados é precisamente o que devesse analisar. Comparando os montantes das experiências incompletas com as concluídas, conforme salientado por Berry et al. (2010), para selecionar os ensaios a serem confrontados com base no seu resultado. A essa luz, é questionável se as diferenças de tamanho do efeito entre ensaios truncados e não truncados é realmente problemática.

Em segundo lugar, o conhecimento ou as expectativas prévias são altamente relevante para lidar com efeitos sobrestimados. A distribuição *a posteriori* visualiza estas diferenças de uma forma intuitiva que pode ser diretamente utilizada para a tomada de

decisões, ou seja, tem uma salvaguarda automática contra conclusões precipitadas que outras escolas de inferência não possuem.

Em terceiro lugar, que os fatores de bayes não dependem do protocolo de amostragem, mas não implica que os bayesianos devam ignorar as questões de concepção experimental. A objetividade processual sob a forma de seguir certas restrições regulamentares e procedimentos padrão podem ser úteis para eliminar certas formas de preconceitos institucionais.

As diretrizes para a utilização das estatísticas bayesianas salientar que os pesquisadores devem ser tão conscientes e diligentes em questões de concepção experimental como frequentistas. A título de exemplo, também de uma perspectiva bayesiana, um teste com erros elevados de tipo I e tipo II é evidentemente um mau experimento.

O ponto de desacordo é diferente: enquanto o frequentador baseia a sua avaliação pós-experimental das provas no desenho pré-experimental e as propriedades de toda a experiência, o bayesiano considera estas propriedades como essenciais para a obtenção de dados válidos, mas como ortogonais para a questão de como interpretá-las uma vez que são intuitivos (SPRENGER, 2009).

Em quarto lugar, o fato de questionar as suas regras segue imediatamente de um princípio da inferência bayesiana: o Princípio da Probabilidade. Segundo este princípio, todas as provas experimentais sobre um parâmetro q desconhecido está contido na probabilidade de função $L(\theta) = p(D \mid \theta)$ para dados observados. Este princípio é uma das pedras angulares da inferência Bayesiana.

Como Birnbaum (1962) mostra num célebre artigo, pode ser derivado de mais dois princípios fundacionais: o Princípio da Suficiência e a Condicionalidade. Começamos com o primeiro, uma estatística T(X) é suficiente se a distribuição dos dados X não depender de o parâmetro desconhecido q, condicional a T. Por outras palavras, são compressões dos conjuntos de dados que não perdem qualquer informação relevantes sobre q.

Um exemplo é uma experiência sobre o enviesamento de uma moeda. Assumindo que os lançamentos são independentes e distribuídos de forma idêntica, a número total de caras e coroas, essas evidências são suficientes para uma inferência sobre o enviesamento da moeda. Assim, pode-se negligenciar a ordem precisa em que os resultados ocorreram.

Formalmente, o Princípio da Suficiência declara que qualquer duas observações x_1 e x_2 são evidentemente equivalentes no que diz respeito parâmetro de interesse q desde

que $T(x_1) = T(x_2)$ para um T estatístico suficiente. Por conseguinte, o princípio é geralmente aceito pelos bayesianos e frequentista da mesma forma.

Na medida em que o Princípio da Suficiência e da Condicionalidade são encontrados com requisitos imperativos para uma inferência objetiva, orientada para a verdade e livre de valores não cognitivos, o Princípio da Probabilidade limita os juízos de apoio probatório de uma forma que é incompatível com a inferência dos frequentadores, por exemplo, valores de p.

Em particular, o Princípio da Probabilidade implica o Princípio da Regra de Probabilidade (BERGER;BERRY,1988). Uma vez que o Princípio da Probabilidade implica que apenas as informações contidas na função de probabilidade são evidentes relevantes, e uma vez que os valores dos parâmetros sob regras de paragem diferentes são proporcionais umas às outras (prova omitida).

Esta posição tem vantagens práticas substanciais: se os julgamentos forem encerrados por razões imprevistas, por exemplo, porque os fundos estão esgotados ou porque se ocorrerem efeitos secundários inesperados, os dados observados podem ser interpretados corretamente num quadro Bayesiano, mas não num quadro frequentista.

No total, a alegação de que a inferência bayesiana em ensaios sequenciais contém um pre-conceito implícito ou invalida o raciocínio científico pode ser rebatido de forma sólida. O problema particular da análise sequencial e do de controle dos ensaios em curso não constitui um desafio à inferência bayesiana que não é igualmente urgente para os seus concorrentes, tais como a inferência frequentista.

6. ECONOMETRIA BAYESIANA VERSUS A ECONOMETRIA CLÁSSICA: DEVE SER CONSTRUÍDO UM NOVO PARADIGMA NA TEORIA ECONOMÉTRICA?

Em primeiro lugar, faz necessário definir o que é um paradigma científico. O termo está vinculado a obra de Thomas Kuhn, *The Structure of Scientific Revolutions*, de 1962; no qual, Kuhn conceitua-o como um quadro, uma forma dominante de pensar e de fazer as coisas, de expectativas e regras partilhadas.

A noção central de um paradigma para Kuhn, volta ao significado original da palavra em que é definida como um excelente exemplo, um modelo ao qual outros aspiram. Este, para Kuhn (1989), foi o aspecto mais novo e menos compreendido de *The Structure of Scientific Revolutions.*

Kuhn (1962) estava trabalhando contra uma tradição filosófica que sustentava que o processo de descoberta, ou pelo menos a tarefa de avaliar uma teoria, é uma questão de seguir um método de regras ou de lógica indutiva. Embora a utilização de tais regras pelos cientistas possa ser em grande medida irrefletida ou inconsciente, pensava-se que era a tarefa da filosofia da ciência desvendar estes fatos.

As regras conduziriam à verdade e, por conseguinte, promoveriam o progresso da ciência (concebido como ficar mais próximo da verdade). Kuhn (1962) não só rejeitou este quadro da ciência como visando a verdade, ele também questionou todas questões relacionadas ao conhecimento cientifico que era discutido até então.

A noção de paradigma de Kuhn destina-se a explicar como a ciência funciona sem tais regras. Em vez de segui-las, os cientistas procuram adequar o seu trabalho ao paradigma de certa forma, depende de verem as semelhanças entre o seu trabalho e o que está vigente como *mainstream* na literatura.

Ver as semelhanças são uma capacidade que não pode ser reduzida a regras, tal como reconhecer um rosto ou qualquer objeto familiar: a semelhança não é redutível a regras. Kuhn sentiu que o funcionamento dos paradigmas poderia explicar todo o processo de desenvolvimento científico, sem te que utilizar à verdade ou a regra.

Na explanação kuhniana, a ciência normal forma um binômio indissociável com o paradigma. A ciência entra em uma fase normal justamente quando é guiada sob a égide de um paradigma. Nas palavras de Kuhn "ciência normal" significa a pesquisa firmemente baseada em uma ou mais realizações científicas passadas (paradigmas). Essas realizações são reconhecidas durante algum tempo por alguma comunidade científica específica como proporcionando os fundamentos para a sua prática posterior.

Em função das críticas de polissemia, equivocidade, entre outras, dirigidas ao conceito de paradigma, Kuhn (1989) procurou em diversas ocasiões responder a elas, chegando mesmo a substituir o termo pela expressão matriz disciplinar, composta por quatro elementos principais: exemplares, generalizações simbólicas, modelos e valores. Em seus últimos trabalhos, é comum encontrar o uso do termo léxico.

Os paradigmas propiciam o advento do consenso - visível nas revistas especializadas, bem como nos manuais de ensino — acerca dos fundamentos da prática científica. Sob sua posse, cessam os debates de ordem metodológica (quais os meios adequados de investigação), de ordem epistemológica (o que deve ser investigado e quais soluções devem ser alcançadas) e de ordem ontológica (qual a natureza das entidades investigadas).

Isto posto, pode-se considerar que os paradigmas centrais em econometria têm como norte a medida do p-valor. Toda a sua estrutura da teoria econométrica clássica está vinculada a utilização dos testes de hipóteses e do valor de probabilidade para validar e testar os seus modelos empíricos.

Como demostrado no segundo capítulo desta dissertação, todo o processo de desenvolvimento e de discussão de metodológica em econometria, mantém ainda como alicerce a utilização do p-valor desde da contribuição de Haalvemo (1944). Isto vai desde os primeiros modelos de estimação por mínimos quadrados ordinários até a fase moderna da econometria, com os modelos VAR.

Apesar da econometria bayesiana existir desde a década de 1950, o método ainda é visto como uma alternativa para a econometria clássica. Para este autor, faz necessário iniciar um debate e até um novo direcionamento para o mainstream da teoria econométrica sobre esse fato, visto que o paradigma consolidado da área não é justificável.

A medida do p-valor que fundamenta a econometria clássica, apresenta diversos questionamentos e críticas. É um método que como destacado no capítulo 4 apesenta uma série de problemas, como: ausência de falseabilidade, as hipóteses nulas tendem ser implausíveis e podem ser irrelevantes, excesso de taxas de falsos positivos, apresenta ambiguidade na sua fundamentação teórica e falta de replicabilidade.

Isto, por si só, já se constitui um motivo relevante para ter toda essa discussão sobre a necessidade da econometria bayesiana não ser vista apenas pelo prisma de ser concebida simplesmente como uma alternativa a econometria tradicional, mas também a

possibilidade ser um novo paradigma e, eventualmente, um meio de substituir a utilização do p-valor dentro da metodologia econométrica.

A vantagem da econometria bayesiana deriva inicialmente da possibilidade de incorporar crenças prévias sobre os parâmetros de interesse no processo de estimação. Este mérito é valioso, particularmente em macroeconomia, onde muitas vezes apenas o número limitado de observações está à disposição do investigador.

A título de exemplo, os modelos de vetores autorregressivos são fortemente utilizados em macroeconomia para previsão, mas também para análise estrutural. Estes modelos sofrem de problema da parametrização excessiva, como o número de coeficientes nos modelos de crescimento muito precipite quando variáveis adicionais são incluídas.

Os modelos VAR bayesianos superam este modelo, impondo informação prévia sobre os parâmetros. Essa informação pode ter origem no conhecimento sobre o comportamento da dinâmica da inflação, por exemplo, que tende a ser estável entre dois períodos ou aproxima-se do objetivo da inflação a médio prazo.

As previsões destes modelos tendem a ter intervalos credíveis mais precisos e também apresentam erros de previsão menores, em comparação com os modelos VAR tradicionais. Uma vez que os modelos são normalmente simulados utilizando técnicas Monte-Carlo, que extraem inferências em torno de quantidades de interesse, tais como funções de resposta ao impulso ou previsões, sem a necessidade de utilizar *bootstrapping* ou outros métodos mais complicados.

No caso de modelos macroeconômicos baseados em micro fundamentações, tem-se frequentemente conhecimentos prévios sobre vários parâmetros, que provém ou da teoria econômica ou de estudos empíricos. Impondo estimativas com parâmetros mais significativos sobre os quais a informação está disponível e muitas vezes também para melhores previsões e mais plausibilidade.

Outra vantagem das técnicas bayesianas na macroeconomia tem sido a sua capacidade de simular modelos de espaço-estado, que na econometria clássica se baseiam em métodos de otimização muitas vezes instáveis. Estas técnicas permitem aos investigadores estimar processos mais complexos não observados, incluindo modelos estocásticos de volatilidade, modelos de coeficiente variável no tempo, ou modelos de fatores.

No entanto, as diferenças entre os métodos bayesianos e a econometria tradicional na estimativa e inferência, não está apenas relacionado a utilização de informação prévia

por parte dos primeiros e a falta de adoção de tais informações por parte destes últimos. Embora a econometria bayesiana desempenha um papel formal para a adoção de informação *a priori* e coerente no quadro de atualização de crenças na extração dos dados, os métodos econométricos clássicos não ignora completamente qualquer tipo de informações antecedentes.

Na visão deste autor, a principal distinção entre a inferência bayesiana e a inferência da econometria *mainstream* é mais profunda do que a questão da informação prévia, mesmo que se argumente que as pesquisas a *priori* estão provavelmente em jogo em ambas metodologias. A diferença ocorre, na verdade, porque o primeiro é um meio para a fim da obtenção de resultados que condicionam os dados observados. Enquanto, o método econométrico tradicional envolve o cálculo da média sobre conjuntos de dados que podem ter sido potencialmente observados, mas não o foram.

Cabe ressaltar, que a busca para que os métodos bayesianos substituam o paradigma atual da econometria, não tem como base o apego a ideologia de Bayes, mas sim, por conta de que em algumas áreas de pesquisas aplicadas em economia, tais como os ramos macroeconomia empírica e finanças (que já foram a citados), as inferências bayesianas possuem todas as ferramentas necessárias para minimizar os erros e lacunas de estimação da econometria tradicional, e ser um quadro dominante da área.

A economia do trabalho e as disciplinas relacionadas a saúde também se apresentam como um campo de pesquisa podem serem explorados de uma melhor forma pela econometria bayesiana. Por exemplo, vários estudos em economia da saúde, incluindo Li e Poirier (2003), Munkin e Trivedi (2003), Bretteville-Jensen e Jacobi (2009), Hu, Munkin e Trivedi (2015) e Jacobi e Sovinsky (2016), entre outros, utilizaram a metodologia bayesiana e deram um novo prisma de análise dessa temática em economia.

Da mesma forma, trabalhos como de Koop e Tobias (2004), Kline e Tobias (2008), Li e Tobias (2011), Hoogerheide, Block e Thurik (2012) e Fruwirth-Schnatter et al. (2012) e (2018) analisaram temas relacionados com o trabalho a partir da inferência bayesiana e também possibilitaram uma outra visão sobre assunto na literatura econômica.

O problema da endogeneidade desempenha um papel central em muitas, se não na maioria, das aplicações em economia do trabalho e disciplinas relacionadas. Estas aplicações partilham frequentemente de uma estrutura comum: o efeito causal de um variável-chave, digamos x, sobre algum resultado y é procurado, mas reconhece-se que x é susceptível a ser endógeno.

Enquanto a maioria dos tratamentos de endogeneidade nessa literatura são de natureza clássica, centradas ou empregando variáveis instrumentais, mínimos quadrados em dois estágios ou outras abordagens tradicionais, estudos tais como Dreze (1976), Geweke (1996) e Conley et al (2008) marcam importantes avanços bayesianos a esta problemática.

Além disso, várias aplicações têm sido abordadas a partir de um ponto de vista bayesiano, salientando frequentemente a facilidade com que os métodos de Monte Carlo via Cadeias de Markov (MCMC) podem ser adaptados para lidar com problemas de endogeneidade em muitos tipos de modelos diferentes.

Como destacado no capítulo anterior, as primeiras aplicações dos métodos econométricos bayesianos, particularmente antes do início dos anos de 1990, era comum assumir que os termos de erro se seguiam a uma distribuição paramétrica particular. Esta prática pode, de fato, continuar a ser comum, pelo menos para fases de análise de dados exploratórios.

Assumir que os resíduos são normalmente distribuídos pode ser uma suposição apelativa, já que muitas vezes se revela matematicamente conveniente para análise subsequente quando associado à adoção de exame *a priori* normal. Tais formas rígidas, contudo, baseiam-se tipicamente em conveniência computacional em vez de aplicação mais profunda.

A tendência geral em econometria é um movimento em direção à robustez e permanecer o mais agnóstico possível em relação aos pressupostos de modelação. A econometria bayesiana seguiu o exemplo e parece bastante razoável acreditar que o futuro desta área continuará a ser caracterizada pela adoção de estimações semi-paramétricas flexíveis e não paramétricas.

O Processo de Dirichlet é um exemplo dessa modelagem flexível de distribuições dentro do paradigma bayesiano. De acordo com Heckman e Vytlacil (1999), as questões advindas deste processo em econometria neste campo são a presença de regressores endógenos, variáveis latentes e a utilização de muitos parâmetros para aumentar robustez dos resultados estimados.

Outro ponto de destaque da econometria Bayesiana refere-se ao seu potencial para combinar diferentes fontes de informação sobre um modelo econométrico, bem como para rastrear as consequências da estimativa e da incerteza aos campos de decisão que são conhecidos como análise de cenários políticos em macroeconomia, finanças e microeconomia.

As decisões políticas macroeconômicas precisas necessitam de uma compreensão detalhada da dinâmica da economia. Assim, a estimativa dos processos econômicos utilizando modelos probabilísticos é crucial para às decisões políticas. Isto já foi reconhecido por Tinbergen (1939) e formalizado por Haavelmo (1944).

Os economistas só nos anos de 1960 conseguiram determinar o impacto da incerteza nas estimativas sobre as políticas econômicas em modelos simples. Brainard (1967) estudou a distribuição do multiplicador em um modelo keynesiano simples, a fim de determinar a eficácia da análise política em uma situação de incerteza sobre os valores dos parâmetros. O seu artigo é claramente um dos mais antigos sobre esse tema.

Um outro exemplo é apresentado em Van Dijk e Kloek (1980), onde a incerteza prévia sobre parâmetros estruturais e um modelo keynesiano simples foi simulada até ao valor preditivo *a priori* implícito do multiplicador e de oscilação do ciclo econômico e, numa etapa seguinte, este período prévio e a informação preditiva foi combinada, usando métodos bayesianos baseados em simulação, com a probabilidade para obter resultados preditivos *a posteriori* consistente.

Neste contexto, existem questões fundamentais. Sims (2012) destaca como o tratamento formal das intervenções políticas de Haavelmo era uma grande fraqueza no seu programa de investigação: Uma vez que os parâmetros do modelo são tratados como não aleatórios (em oposição aos seus estimadores), as incertezas sobre eles não podem transitar para distribuições preditivas, limitando substancialmente a tomada de decisão sob incerteza no modelo.

Em segundo lugar, ainda conforme Sims (2012), o comportamento político não foi incorporado na obra de Haavelmo e subsequentes na *Cowles Commission*. Enquanto para um agente político, uma política aleatória na equação do modelo de comportamento pode parecer estranha, uma vez que as decisões políticas são conhecidas por ele, para o setor privado e para o economista essas informações são desconhecidas.

A inferência bayesiana fornece soluções claras para estas questões. Do ponto de vista bayesiano, a incerteza, ou seja, o que é "aleatório" e o que é "não aleatório", depende do que é "observado" e do que não é "não observado". Os parâmetros do modelo, sendo desconhecidos, têm uma distribuição de probabilidade que é atualizada por informação de dados utilizando a regra de bayes. Isto fornece uma ferramenta probabilística clara para uma tomada de decisão em tempo real sob incerteza.

Além disso, a interpretação bayesiana elimina também o paradoxo do comportamento político aleatório na modelização econômica. Tanto a visão do tomador

de decisão política, como de um economista ou de um profissional do setor privado, pode existir conjuntamente pela perspectiva bayesiana.

Um aspecto importante da análise política de ponta é que o comportamento político tem de ser acomodado por um modelo económico onde a tomada de decisão dos agentes econômicos é especificada. Este modelo é diferente em diferentes campos da economia e depende da decisão típica que tem de ser tiradas.

Na macroeconomia, a incorporação das preferências dos consumidores e das características tecnológicas é importante para além de várias outras fontes de constrangimentos. Isto é necessário a fim de analisar os agentes econômicos. As decisões em resposta à mudança das políticas governamentais também são conhecidas como a crítica de Lucas (1976).

Os primeiros modelos macroeconômicos eram relacionados a teoria dos ciclos reais (RBC). Em que são modelos totalmente articulados, contudo, requerem mais parâmetros que vão desde os que se relacionam com as decisões políticas para aquelas que se referem às decisões dos agentes econômicos. Um método inferencial precoce para os modelos de RBC, chamado "calibração", implica a ligação de alguns valores plausíveis para os parâmetros.

Uma vez que foram criados os valores dos parâmetros, são gerados dados artificiais utilizando o modelo e o seu ajuste pode ser avaliado para verificar se a captura artificial da amostra estilizou fatos reais da mesma. Estes modelos evoluíram para a os chamados modelos de Equilíbrio Geral Estocástico Dinâmico (DSGE), incorporando vários mecanismos, em vez de apenas choques exógenos, para gerar flutuações do ciclo econômico. As técnicas de inferência bayesiana fornecem hoje em dia ferramentas que substituem a calibração.

No domínio das finanças, os pressupostos comportamentais dos agentes econômicos estão naturalmente ligados à forma de lidar com a previsibilidade, em particular a da incerteza e do risco. A ligação entre a previsão e a tomada de decisões tem sido sempre um tema central em finanças. Mais especificamente, pode-se mencionar: a relevância da previsão do retorno dos ativos e do seu efeito na gestão de carteiras, medição das volatilidades para o Value-at-Risk, preço das opções e negociação algorítmica moderna como tópicos recentes de simulação com base na investigação econométrica bayesiana.

No campo da microeconomia, a tomada de decisões com base em modelos de escolha de marketing é uma área importante onde as vantagens da abordagem bayesiana

acima enumeradas pode ser aplicada de forma mais consistente quando comparada com o arcabouço teórico e prático da econometria clássica.

As vantagens relativamente recentes da disponibilidade de grandes conjuntos de dados e a proposta de estruturas de modelos flexíveis para captar a heterogeneidade no comportamento dos consumidores dá aos pesquisadores e aos profissionais um estímulo para tirarem vantagens adicionais da abordagem bayesiana, nomeadamente a facilidade de cálculo utilizando métodos computacionais avançados.

Apesar do seu potencial de aplicabilidade, os métodos bayesianos não têm sido amplamente utilizados no campo da microeconometria. Mesmo fornecendo uma solução clara para incorporar incerteza nos parâmetros e em aplicações centradas na tomada de decisão baseada sobre que lidam com efeitos casuais.

Isto se tornou recentemente um importante tema de investigação para a econometria bayesiana e tem sua origem no artigo seminal de Rubin (1978). Vários trabalhos nas últimas décadas, por exemplo, Imbens e Rubin (1997); Li et al. (2004), entre outros, fornecem contribuições importantes sobre a análise bayesiana para os modelos casuais.

Em alguns casos, as previsões dos agentes podem ser reveladas não através de um modelo econométrico, mas sob a forma de mercados de previsão, onde os contratos baseados nas previsões dos agentes económicos são comercializados. Berg et al. (2010) deriva uma distribuição de probabilidade usando métodos não paramétricos bayesianos para o evento específico das eleições presidenciais americanas que é consistente com a previsão dos preços de mercado destes contratos.

Como as distribuições preditivas são os elementos-chave da tomada de decisão formal, estes também fornecem uma ferramenta para resumir a informação nos mercados de previsão a utilizar na tomada de decisões. Em resumo, a análise sistemática da inclusão da incerteza dos parâmetros e dos modelos na análise de decisão requer tanto uma estrutura econômica clara que contenha o comportamento dos agentes econômicos como sofisticados algoritmos estocásticos de simulação em conjunto com os recentes desenvolvimentos de hardware.

Finalmente, sobre o tema da seleção de métodos e combinação de modelos como contributos para a tomada de decisões. A análise bayesiana padrão faz uso do fator Bayes, definido como o índice de probabilidade marginal de dois modelos concorrentes. É bem sabido que isto é sensível à escolha de distribuições *a priori*. Uma alternativa mais robusta e atraente é a probabilidade preditiva (GEWEKE;AMISANO, 2011).

Se os dados são de fato gerados por um modelo específico então esta abordagem dá uma orientação para o "melhor" modelo. Mais realista parece ser o conceito de combinação do modelo com informação de diferentes fontes. No entanto, a tomada de decisões também pode ser transmitida através da combinação de modelos concorrentes, em vez de se escolher o melhor.

Embora a média de modelos Bayesianos indique um quadro estatisticamente básico de combinação de modelos, a sua combinação eficiente continua a ser uma área ativa de investigação. Durham e Geweke (2014) propõem a utilização propostas em Geweke e Amisano (2011) para combinar modelos de previsão de retorno de ativos que estão intimamente relacionados com a decisão sob incerteza para um agente avesso ao risco.

Além disso, tal combinação admite também que todos os modelos considerados podem ser falsos, o que não é o caso por exemplo, no modelo bayesiano padrão. An e Schorfheide (2007) utilizam modelos diferentes, mas os modelos semelhantes de DSGE concentrar-se nas dificuldades potenciais como a especificação, identificação de erros de especificação do modelo e a multimodalidade das distribuições de parâmetros.

Tais comparações revelam que as restrições impostas são corretamente especificadas. Os autores combinam os fatores a *priori* baseados no modelo DSGE com os VAR's para formar os chamados modelos DSGE-VAR. Del Negro e Schorfheide (2009) centram-se na análise da política monetária utilizando modelos DSGE potencialmente mal especificados.

Leeper et al. (1996) argumentam que os modelos para análise política e de previsão não são nitidamente distintas. Estes autores mostram que os efeitos de tamanho atribuídos às mudanças da política monetária variam de acordo com as especificações do comportamento econômico. Uma conclusão robusta, comum em vários modelos, é que uma grande fração da variação dos instrumentos de política monetária é atribuível à reação sistemática das autoridades políticas ao estado da economia.

Como estudo de caso final relativo a uma combinação de modelos em que os métodos individuais são falsos, ressalte-se o trabalho de Billio *et al.* (2013). Onde é realizada uma experiência financeira de investimento em ativos com e sem risco utilizando uma combinação de métodos que constitui um modelo de passeio aleatório sobre retornos de ações e uma previsão dos analistas profissionais.

Os investidores ditos profissionais na bolsa de valores teriam encontrado perdas substanciais quando tinham seguido o modelo de previsão tradicional em econometria

enquanto os investidores que utilizassem uma combinação de modelo de previsões e um passeio aleatório teriam obtidos resultados muito melhores.

Mudando o foco para análise propriamente da metodologia estatística, o primeiro ponto que é favorável à inferência bayesiana frente a inferência frequentista (econometria tradicional) refere-se ao fato de seu método ser prescritivo — dada a especificação de um modelo, só existe uma forma de obter a resposta apropriada.

A inferência Bayesiana não requer soluções *ad hoc* para remediar procedimentos que produzam resultados internamente inconsistentes. Ela é imune a tais problemas porque se baseia num pequeno conjunto de axiomas para uma tomada de decisão racional e eles levam à mesma conclusão: o raciocínio sob incerteza só pode ser coerente se obedecer às leis da teoria da probabilidade.

Um dos famosos métodos para provar esta conclusão de grande alcance deve-se a Bruno de Finetti e envolve um cenário de apostas. Finetti (1974) assumiu que existe um bilhete juridicamente vinculativo que garante o pagamento de 1 euro caso uma proposta se revele verdadeira. Por exemplo, a proposta poderia ser “em 2022, a seleção brasileira de futebol ganhará copa do mundo".

Agora um indivíduo qualquer tem que determinar o preço que está disposto a pagar por este bilhete: este preço é a probabilidade subjetiva operacional que atribui à proposta. A complicação é que este cenário também apresenta um adversário. O oponente pode decidir, com base no preço que foi escolhido, comprar este bilhete do primeiro indivíduo ou fazer com que ele compre o dele.

Neste exemplo, é obviamente irracional fixar o preço mais alto do que 1 euro, porque o adversário vai obrigá-lo a comprar este bilhete e tem a garantia de ter lucro; e, é também irracional fixar o preço inferior a 0 euros, porque o adversário vai comprá-lo a um preço negativo (ou seja, ganhar dinheiro) e tem novamente a garantia de ter lucro.

Isto também pode ser ampliado para a determinação do preço de três bilhetes individuais. Por exemplo, o bilhete I e II declaram que o vencedor da copa do mundo será Argentina e Alemanha, respectivamente, e o bilhete III considera que ou os argentinos ou os alemães vencerão o mundial: pode-se fixar os preços como o indivíduo quiser, em particular, não há nada que o impeça para defini-los tais que o preço (bilhete I) + preço (bilhete II) ≠ preço (bilhete III).

No entanto, quando se estabelecem os preços desta forma, é garantido alguém vai perder dinheiro em comparação com o adversário; a título de exemplo, podemos supor

que os preços dos bilhetes I, II, III sejam 0,5, 0,3, e 0,6 euros, respectivamente. Então, quem comprar o bilhete III tem a garantia de sair à frente.

Usando cenários de apostas como os referidos acima, Finetti (1974) mostrou que a única forma de determinar os valores subjetivos e maximizar a utilidade individual é fazer que os valores obedecem às regras da teoria das probabilidades (ou seja, a regra de que as probabilidades residem entre 0 e 1, que os eventos mutuamente exclusivos são aditivos e a existência da probabilidade condicional).

Outro ponto relevante da teoria bayesiana está relacionado ao seu teste de hipótese. Enquanto, na metodologia tradicional, o modelo ideal captura toda a estrutura replicável e ignora o ruído idiossincrático e produz as melhores previsões para dados não vistos provenientes da mesma fonte.

Cabe destacar, quando um modelo é demasiado complexo, afirma-se que ele se sobrepõe aos dados; o modelo trata erroneamente o ruído idiossincrático como se fosse uma estrutura replicável; por sua vez, quando um modelo é demasiado simples, diz-se que ele *overfitting* (sub-ajusta) a amostra, o que significa que o método não capta toda a estrutura replicável nos dados. Os modelos que sub ou sobre-ajustam os dados fornecem previsões sub-ótimas e é dito que são generalizados de forma errada (JAYNES, 2003).

Assim, o principal desafio do teste de hipóteses ou seleção de modelos é identificar o método com o melhor desempenho preditivo. Só que não é evidente como isto deve ser feito de imediato; modelos complexos geralmente fornecerão um melhor ajuste aos dados observados do que modelos mais simples, e, portanto, não se pode simplesmente preferir o método com o melhor ajuste pois tal estratégia levaria a um sobre-ajustamento maciço.

Conforme Jaynes (2003), a intuição dos pesquisadores sugere que essa tendência para o *overfitting* deve ser neutralizada ao valorizar os modelos mais simples. Esta intuição é consistente com a lei da parcimônia ou da navalha de Ockham que afirma que, quando *ceteribus paribus* (tudo o mais é igual), os modelos simples devem ser preferidos aos modelos complexos.

Os métodos formais de seleção de modelos tentam quantificar o *tradeoff* entre boa aderência e a parcimônia. Muitos destes medem um modelo de desempenho global pela soma de dois componentes, um que mede a precisão descritiva e outro que coloca um prêmio sobre a parcimónia, que também é conhecido como o fator Ockham.

Uma das características atrativas do teste de hipóteses bayesianos é que determina automaticamente o modelo com o melhor desempenho preditivo. O teste incorpora, portanto, o que é conhecido como navalha automática de Ockham. A fim de mostrar por que razão isto ocorre, será apresentado a seguir duas linhas de raciocínio conforme MacKay (2003).

A primeira considera que que a seleção do modelo bayesiano se baseia na probabilidade marginal de um modelo *t*, $m(y|H_t)$, e denota-se uma sequência de ordem n como $y^n = (y_1,...,y_n)$, isto é, y_{i-1} denota o (i - 1)n ponto de dados individuais, enquanto y^{i-1} denota toda a sequência de observações, desde y1 até y_{i-1} inclusive.

Além disso, quantifica o desempenho preditivo para um único ponto de dados pela função de perda logarítmica $\ln\hat{\rho}_i(y_i)$: quanto maior for a probabilidade que $\hat{\rho}_i$ (determinada com base nas observações *a priori* y^{i-1}) atribui ao resultado y_i observado, menor será a perda. Da definição de probabilidade condicional, isto é, $p(y_i|y^{i-1}) = p(y^i)/p(y^{i-1})$, segue-se que a probabilidade marginal dos dados pode ser decomposta como uma série de previsões probabilísticas sequenciais, com um passo *a priori*.

Já a segunda linha de raciocínio, confirma que cada modelo estatístico faz predições iniciais. Os modelos complexos têm um espaço de parâmetros relativamente grande, e são, portanto, capazes de fazer mais previsões e cobrir mais eventualidades que os modelos simples. No entanto, estes modelos precisam distribuir a sua probabilidade *a priori* por todos os seus parâmetros.

No limite, um modelo que prediz quase tudo tem de espalhar a sua probabilidade inicial de forma tão ténue que a ocorrência de qualquer evento não contribuirá grandemente para a credibilidade desse modelo. Formalmente, a probabilidade marginal dos dados é calculada através da média da probabilidade $f(y \mid \theta, H_t)$ sobre a distribuição *a priori* $p(\theta \mid H_t)$. Quando a probabilidade inicial está muito dispersa, ocupará uma parte relativamente do espaço de parâmetros em que a probabilidade é quase zero, e isto diminui grandemente a probabilidade média ou marginal.

Além do mais, o teste de hipóteses bayesiano não faz uma distinção fundamental entre modelos aninhados e não aninhados. Isto significa que pode ser aplicado em mais situações do que o teste de hipóteses frequentistas: para a inferência bayesiana, as questões substantivas podem ser testadas estatisticamente exatamente da mesma forma; na inferência frequentista, contudo, o fato dos modelos não estarem aninhados causa graves complicações.

Uma terceira característica relevante da metodologia bayesiana, se dá pelo fato dela permitir a implementação flexível de técnicas estatísticas relativamente complicadas como as que envolvem modelos hierárquicos não lineares. Nos modelos hierárquicos, assume-se que os parâmetros para pessoas individuais são extraídos de uma distribuição a nível de grupo.

Tais estruturas multiníveis incorporam naturalmente tanto as diferenças como os pontos comuns entre as pessoas e, por conseguinte, fornecem ao experimento os meios para resolver o antigo problema de como lidar com as diferenças individuais. Entre os dois extremos de assumir que os participantes são completamente iguais e que são completamente diferentes encontra-se o compromisso de modelação hierárquica.

Embora as análises hierárquicas possam ser realizadas utilizando metodologia ortodoxa, existem fortes vantagens filosóficas e práticas razões para preferir a metodologia Bayesiana. Por outro lado, a estatística bayesiana também facilita a concentração nas variáveis relevantes, integrando as chamadas variáveis incômodas.

Para ilustrar essa situação, segue-se novamente o exemplo de MacKay (2003) e propõe conjugar antecedentes impróprios para μ e σ. Diz-se que uma distribuição *a priori* é conjugado quando está na mesma família de distribuição que *a posteriori*. Por exemplo, quando a probabilidade inicial para μ é normal, *a posteriori* para μ também é normal. A conjugação de distribuições iniciais são frequentemente as únicas que permitem a derivação analítica do efeito causal.

A distribuição *a priori* é considerado impróprio quando não se integra a um número infinito, ou seja, quando a amostra inicial para μ é uma distribuição normal com média $\mu_0 = 0$ e desvio padrão $\sigma_\mu \rightarrow \infty$, isto produz um conjunto de dados que é plano ao longo de toda a linha real. Para o presente exemplo, utiliza-se os conjugados impróprios em μ e σ porque levam a resultados analíticos que correspondem aos resultados da inferência frequentista.

Em particular, assumimos aqui que a distribuição sobre μ é normal com média $\mu_0 = 0$ e desvio padrão $\sigma_\mu \rightarrow \infty$. Este plano simplesmente declara que todos os valores de μ são igualmente prováveis a priori. Porque σ é sempre superior a 0, mas log σ cobre toda a linha real, um padrão não-informativo se torna o plano em a escala de tronco, que se transforma na distribuição *a priori* $p(\sigma) = 1/\sigma$.

Usando estes antecedentes, pode-se derivar analiticamente a distribuição *a posteriori* conjunta de μ e σ dado as amostras, ou seja, $p(\mu, \sigma|y)$. Agora que foi definido

e conhecido a distribuição conjunta de μ e σ, considera-se dois cenários em que é necessário eliminar um parâmetro de incômodo.

No primeiro cenário, deseja-se conhecer a média μ de uma distribuição normal com desvio padrão desconhecido σ. Assim, μ é o parâmetro de interesse, enquanto que σ é um parâmetro que se gostaria de ignorar (i.e., um parâmetro de incômodo). Usando a lei da probabilidade total é fácil de marginalizar, ou integrar, σ, como $p(\mu|y) = \int p(\mu, \sigma|y)d\sigma$.

O fato de que esta equação pode ser reescrito como $p(\mu|y) = \int p(\mu|\sigma, y)p(\sigma)d\sigma$ destaca o fato de que o parâmetro de perturbação σ só pode ser integrado depois de ter sido atribuído uma distribuição *a priori.* Depois de integrar σ, a probabilidade marginal *a posteriori* resultante em p(μ|y) acaba por ser a distribuição Student-t, a famosa distribuição freqüente para uma estatística de teste que envolve a média de um normal distribuição com variância desconhecida.

Na segunda situação, ressaltar o desvio padrão μ de uma distribuição normal com média desconhecida μ. Isto significa que σ é a parâmetro de interesse, enquanto μ é agora o parâmetro do incômodo. A partir do parâmetro distribuição *a posteriori* conjunta de μ e σ, pode-se novamente aplicar a lei da probabilidade total, desta vez para integrar μ, como se segue: $p(\sigma|y) = \int p(\sigma, \mu|y)d\mu = \int p(\sigma|\mu, y)p(\mu)d\mu$.

Tal como antes, esta equação mostra que o parâmetro de incômodo μ só pode ser integrado quando lhe tiver sido atribuída uma distribuição inicial. Após o cálculo da distribuição marginal a posteriori p(σ|y), o valor mais provável para σ (dados y) acaba por ser e $\sigma_{MP} = \sqrt{\frac{S}{n-1}}$, onde n é igual ao número de observações e $S^2 = \sum_{i=1}^{n}(y_1 - \overline{y})^2$. O fator n−1 (em vez de n) também ocorre na inferência frequentista, em que e $S^2/(n-1)$ é o estimador imparcial para a variação de uma distribuição normal com distribuição desconhecida média.

Em suma, a inferência bayesiana permite ao utilizador concentrar-se em parâmetros de interesse, integrando parâmetros de incômodo de acordo com a lei da probabilidade total. As distribuições marginais posteriores resultantes podem ter correspondência com as frequentistas, mas isto só é válido em alguns casos especiais.

Um quinto fator, reside no fato da inferência bayesiana produzir resultados que se ligam estreitamente ao que os investigadores querem saber. Para esclarecer esta afirmação por analogia, Gerd Gigerenzer sugeriu que para muitos investigadores a inferência estatística envolve uma luta interna freudiana entre o Superego, o Ego, e o Id.

Na analogia de Gigerenzer (1993), o Superego promove o teste de hipóteses Neyman-Pearson, no qual um nível α é determinado antes da experiência. O Ego promove o teste de hipóteses de Fisher, no qual o valor preciso de p supostamente mede a evidência de força contra a hipótese nula. Finalmente, a Id deseja que às hipóteses em consideração sejam atribuídas as probabilidades, algo que o Superego e o Ego são incapazes e não estão dispostos a fazer.

Como resultado deste inconsciente conflito interno, os investigadores relatam frequentemente resultados dos procedimentos dos frequentadores, mas muitas vezes acreditam (implicitamente ou mesmo explicitamente) que aprenderam algo sobre a probabilidade das hipóteses em consideração.

Gigerenzer (1993) tem certa razão, lá no fundo, o que os pesquisadores realmente desejam é chegar às conclusões bayesianas. Esta asserção é apoiada pelo fato de que os investigadores frequentemente interpretar mal os conceitos frequentistas e interpretá-los de uma forma que é decididamente bayesiana (ou seja, a interpretação teria sido correta se o método de inferência fosse o segundo).

Para ilustrar o precedente com uma ilustração concreto, considere um intervalo frequentista de confiança para a média normal μ:μ ϵ [-0,5,1,0]. Como já vimos, a interpretação correta, mas, contra-intuitiva deste resultado é que quando o procedimento é aplicado muitas vezes a todos os tipos de conjuntos de dados possíveis, os intervalos diferentes cobrem o valor real de μ em 95% dos casos.

Mas porque é que isto seria relevante para o investigador que quer aprender sobre μ os seus dados? Em contraste, considere o mesmo [-0,5;1,0] intervalo para μ, mas assuma agora que é um intervalo Bayesiano 95% credível. Consistente com a intuição, e consistente com o que os investigadores querem saber, este intervalo bayesiano transmite que existe uma probabilidade de 95% de μ se situar em [-0,5;1,0].

Por fim, uma objeção comum à inferência Bayesiana é que ela é subjetiva e, portanto, não tem lugar na comunicação científica. Por exemplo, num artigo intitulado Why Isn't Everyone a Bayesian?", Bradly Efron argumentou que "a objetividade restrita é um dos fatores cruciais que separam o pensamento científico do pensamento desejoso" e concluiu que "o elevado terreno da objetividade científica foi apreendido pelos frequentistas"(1986, p. 4).

As alegações da Efron (1986) podem ser rebatidas por várias razões. Primeiro, a partir de uma perspectiva bayesiana, não existe a "objetividade estrita", pois o raciocínio sob incerteza é sempre relativo a algum tipo de conhecimento inicial. Nesta perspectiva,

a busca da "objetividade estrita" é um ideal quixotesco. Assim, os bayesianos podem querer mudar a reivindicação de Efron (1986) para o terreno elevado da objetividade científica que é um conceito que não pode ser apreendido por ninguém, porque não existe.

Em segundo lugar, existe uma escola de bayesianos objetivos, que especificam os antecedentes de acordo com certas regras pré-determinadas. Dada uma regra específica, o resultado da inferência estatística é independente da pessoa que efetua a análise. Exemplos de distribuições *a priori* objetivas incluem as informações unitárias (ou seja, que transportam tanta informação como uma única observação), que são invariantes sob transformações e que maximizam a entropia (KASS;WASSERMAN, 1996).

As distribuições iniciais objetivas são geralmente vagas ou pouco informativas, ou seja, dispersos por toda a gama para a qual são desprovidos. Assim, os bayesianos podem querer mudar a alegação de Efron para: embora o elevado terreno de objetividade científica possa parecer ser apreendido pelos frequentistas, os bayesianos também têm uma legítima alegação da mesma.

Terceiro, a inferência dos frequentistas é tão objetiva como se pode (desejavelmente) pensar. Como ilustrado no capítulo 2, a intenção com que uma experiência é realizada pode ter um impacto profundo em sua estimação. As ideias e pensamentos não revelados que orientaram a experimentação são cruciais para as medidas de evidência dos adeptos a estatística frequentista.

Berger e Berry (1988) concluem que a percepção da objetividade da inferência frequentista é largamente ilusória. Assim, os seus críticos podem querer mudar a pretensão de Efron (1986) para “embora o elevado terreno da objetividade científica possa parecer ser apreendido pelos frequentistas, após uma inspeção mais atenta, esta objetividade é apenas faz-de-conta, como na realidade os mesmos têm de confiar na honestidade e na capacidade introspectiva dos investigadores que recolheram os dados.

Em contraste com a inferência frequentista, as inferências bayesianas geralmente não dependem de intenções subjetivas, nem de dados que nunca foram observados (LINDLEY, 1993). A distribuição posterior dos parâmetros μ é escrito $p(\theta|y)$ e a probabilidade marginal de um modelo, digamos H0, é dada por $m(y|H_0)$ - em ambos os casos, y são os dados observados, e é irrelevante que outros os dados poderiam ter sido observados, mas não o foram.

Na inferência Bayesiana, a subjetividade a que Efron (1986) aludiu através da especificação da distribuição prévia para os parâmetros do modelo. Independentemente de esta especificação ocorrer automaticamente, como no caso de objetivos *a priori*, ou se

ocorre através da incorporação de conhecimento iniciais, como no caso dos subjetivos, o ponto crucial é que a distribuição prévia está formalmente especificada e disponível para todos os outros investigadores para inspecionar e criticar.

Isto também significa que a subjetividade bayesiana pode ser analisada por métodos formais que quantificam a robustez *a priori*. Note-se como a noção de subjetividade é diferente para os dois paradigmas: na estatística bayesiana está aberta à inspeção, enquanto que a frequentista está escondida da vista, cuidadosamente trancada na mente dos investigadores que recolheu os dados.

Por conseguinte, um ajustamento final da declaração da Efron (1988) poderia ler que a objetividade científica é ilusória, e tanto a inferência bayesiana como a frequentista tem elementos subjetivos: a diferença é que na primeira está aberta à inspecção, enquanto que a frequentista não está.

7. CONSIDERAÇÕES FINAIS

Este estudo propôs discutir o *status quo* que a inferência frequentista – representado pelo p-valor – possui dentro da teoria econométrica tradicional. O valor de probabilidade se tornou ao longo do desenvolvimento da econometria o principal instrumento para definir a validade dos modelos estatísticos e de seus resultados, sendo o fator preponderante para qualquer trabalho empírico da área.

A sua origem remete ao trabalho de Ronald A. Fisher e, posteriormente, as contribuições de Jerzy Neyman e Karl Pearson – especialmente, com o desenvolvimento dos testes de hipótese. Este fato se tornou uma grande discussão dentro da teoria estatística, dado que o p-valor concebido por Fisher não é compatível com o teste da hipótese de Neyman-Pearson na qual se tornou incorporado.

O valor de probabilidade foi criado para ser uma medida inferencial flexível, enquanto que o teste de hipótese é uma regra de comportamento, não de inferência. Logo, a combinação dos dois métodos levou a uma reinterpretação do p-valor simultaneamente como uma taxa de erro observada e uma medida de evidência.

Ambas as interpretações são problemáticas, e a sua combinação obscureceu as importantes diferenças entre Neyman e Fisher – que, por sinal, essas distinções teóricas geraram um grande debate entre os dois autores – sobre a natureza do método científico e inibe a compreensão das implicações filosóficas dos métodos básicos em uso hoje em dia.

A teoria econométrica por meio dos trabalhos de Trygve Magnus Haavelmo, adotou em sua metodologia tradicional a utilização de uma mistura entre as contribuições de Neyman-Pearson e Fisher. O que por si já é um grande problema, uma vez que são fundamentações teóricas diferentes e em que os seus próprios autores veem uma distinção enorme entre as duas fundamentações.

Este fato já demonstra como é questionável a posição em que se encontra atualmente o valor de probabilidade na econometria *mainstream*, não existe justificativa plausível para que ele seja um instrumento fundamental para validar a maior parte dos modelos estatísticos em Economia.

Além desta questão em torno do debate Fisher x Neyman, se for analisado propriamente o seu arcabouço teórico, o p-valor apresenta outros problemas: falta de replicabilidade, ausências de falseabilidade, fundamentos ambíguos e a má utilização do nível de significância, entre outras questões que já foram citadas nesta dissertação.

Assim, faz necessário discutir dentro da teoria econométrica outros instrumentos estatísticos para validar os modelos estatísticos com o objetivo de minimizar os possíveis erros da econometria tradicional. Neste estudo, foi destacado a inferência bayesiana como um dos meios que podem ser utilizados como contraponto ao p-valor.

A vantagem da utilização da econometria bayesiana tem como origem possibilidade de adotar crenças *a priori* sobre os parâmetros de interesse no processo de estimação. Esta capacidade é de grande valia, particularmente em macroeconomia, onde os investigadores na maior das pesquisas trabalham com um número limitado de observações.

Outro ponto positivo das inferências bayesianas tem sido a sua capacidade de simular modelos de espaço-estado, que na econometria clássica se baseiam em métodos de otimização muitas vezes instáveis. Estes conjuntos de métodos possibilitam aos pesquisadores, estimar processos mais complexos não observados, incluindo modelos estocásticos de volatilidade, modelos de coeficiente variável no tempo, ou modelos fatoriais.

A econometria bayesiana já tem alcançado grandes avanços dentro da teoria econométrica, mas é necessário se discutir de forma mais contundente a sua incorporação dentro do arcabouço tradicional, tanto no ensino como nas pesquisas. Mesmo que também possua seus dilemas, como: a escolha da distribuição *a priori* e a negligência do desenho experimental, as técnicas bayesianas se apresentam como a melhor opção para superar os problemas apresentados pela econometria clássica.

Por fim, cabe ressaltar que a busca para que os métodos bayesianos possam eventualmente substituir o paradigma atual da econometria *mainstream*, não tem como base o apego a teoria de Bayes, mas devido à inferência bayesiana possuir todas as ferramentas necessárias para minimizar os erros e lacunas de estimação da econometria tradicional.

Como sugestão para as próximas pesquisas sobre o tema, recomenda-se uma análise empírica com o objetivo de comparar as evidências econométricas das duas técnicas (utilizando os mesmos dados e problemas), a fim de avaliar as suas principais diferenças e se um dos métodos consegue ter os resultados mais coesos estatisticamente.

REFERÊNCIAS BIBLIOGRÁFICAS

AITCHISON, J.; BROWN, J. A. C. *The Lognormal Distribution with Special Reference to Its Uses in Economics*. Cambridge: Cambridge University Press, 1956.

AN, S.; SCHORFHEIDE, F. Bayesian Analysis of DSGE Models. *Econometric Reviews*, v. 26, n. 2-4, p. 113-172, 2007.

ANDERSEN, L. C.; CARLSON, K. M. St. Louis Model Revisited. *International Economic Review*, v. 15, n. 2, p. 305-327, 1974.

BENJAMINI, Y. The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, v. 70, n. 2, p. 129–133, 2016.

BERGER, J. O. Could Fisher, Jeffreys and Neyman Have Agreed on Testing?. *Statistical Science*, v. 18, n. 1, p. 1-31, 2003.

BERGER, J.; GOTTSCHALK, J.; PORTILLO, R.; ZANNA, L. The Macroeconomics of Medium-Term Aid Scaling-Up Scenarios. IMF Working Paper, p. 1- 45, 2010.

BERRY, S.; ISHAK, K. J.; RUCE, B. R.; BERRY, D. A. Bayesian Meta-Analyses for Comparative Effectiveness and Informing Coverage Decisions, *Medical Care*, v. 48, n. 6, p. 137-144, 2010.

BLAUG, M. *The Methodology of Economics: Or How Economists Explain*. Cambridge: Cambridge University Press: 1980.

BODKIN, R. G.; KLEIN, L. R.; MARWAH, K. *A History of Macroeconometric Model-Building*. Cheltenham: Edward Elgar Pub, 1991.

BOX, G.; JENKINS, G. Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day, 1970.

BRIGGS, A. H. A Bayesian approach to stochastic cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care*, v. 17, n. 1, p. 69–82, 2001.

BRAINARD, W. C. Uncertainty and the Effectiveness of Policy. *The American Economic Review*, v. 57, n. 2, p. 411-425, 1967.

CHETTY, V. K. Bayesian Analysis of Haavelmo's Models. *Econometrica*, v. 36, n. 3-4, p. 582-602, 1968.

CHRIST, C. F. A Symposium on Simultaneous Equation Estimation: Simultaneous Equation Estimation: Any Verdict Yet?. *Econometrica*, v. 28, n. 4, p. 835-845, 1960.

COCHRANE, D.; ORCUTT, G. H. Applications of Least Square Regression to Relationships Containing Autocorrelates Error Term. *Journal of the American Statistical Association*, v. 44, n. 245, p. 32-61, 1949.

COHEN, J. The earth is round (p < .05). *American Psychologist*, v. 49, n. 12, p. 997-1003,1994.

COLQUHOUN, D. An investigation of the false discovery rate and the misinterpretation of p-value. *Royal Society*, v. 1, n. 3, 2014.

COLQUHOUN, D. The reproducibility of research and the misinterpretation of p-values. *Royal Society*, v. 4, n. 12, 2017.

CONLEY, T.; HANSEN, C.; MCCHULLOCH, R. E.; ROSSI, P. E. A semi-parametric Bayesian approach to the instrumental variable problem. *Journal of Econometrics*, v. 144, n. 1, p. 276-305, 2008.

COOKE, R. M. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford: Oxford University Press, 1991.

COX, D. R. Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society*, v. 24, n. 2, p. 406-424, 1962.

CRAMER, P. *The development of defense mechanisms: Theory, research, and assessment*. Amsterdã: Springer-Verlag Publishing, 1991.

DEL NEGRO, M.; SCHORFHEIDE, F. Monetary Policy Analysis with Potentially Misspecified Models. *American Economic Review*, v. 99, n. 4, p. 1415-1450, 2009.

DHRYMES, P; HORWEY, E. P.; HYMANS, S. H.; KMENTA, J.; LEAMER, E. E.; QUANDT, R. E; RAMSEY, J. B.; SHAPIRO, H. T.; ZARNOWITZ, V. Criteria for Evaluation of Econometric Models. *Annals of Economic and Social Measurement*, v. 1, n. 3, p. 291-324, 1972.

DICKLEY, D.; FULLER, W. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, v. 74, n. 366, p. 427-431, 1979.

DURHAM, G.; GEWEKE, J. *Adaptively sequential posterior simulators for massively parallel computing environments*. In: JELIAZKOV, I.; POIRIER, D. J. (orgs.), Bayesian Model Comparison, 2014. p. 1-44.

DOAN, T.; LITTERMAN, R.; SIMS, C. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, v. 3, n. 1, p. 1-100, 1984.

DOUGLAS, H. The irreducible complexity of objectivity. Synthese, v. 138, n. 3, p. 453-473, 2004.

DREEZE, J. Some Theory of Labor Management and Participation. *Econometrica*, v. 44, n. 6, 1125-1139, 1976.

DURBIN, J.; WATSON G. S. Testing for serial correlation in least squares regression I. *Biometrika*, v. 37, n. 3-4, p. 409–428, 1950.

EFRON, B. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, v. 1, v. 1, p. 54-75, 1986.

ENGLE, R. F.; GRANGER, C. W. L. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, v. 55, n. 2, p. 251-276, 1987.

EPSTEIN, R. J. *Econometric Methodology in Historical Perspective*. Amsterdã: Elsevier, 2011.

FARREL, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society*. v. 120, n. 3, p. 253-290, 1957.

FINETTI, B. *Probability, Induction and Statistics: The Art of Guessing.* New York: John Wiley and Sons, 1972.

FISHER, J. O. *R. A. Fisher — The Life of a Scientist*. New York: John Wiley and Sons, 1978.

FISHER, R.A. On the Mathematical Foundations of Theoretical Statistics. *Royal Society*, v. 222, n. 594-604, p. 309-368, 1922.

FISHER, R.A. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd, 1925.

FISHER, R.A. *The design of experiments*. Edinburgh: Oliver and Boyd, 1935.

FISHER, R.A. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd, 1956.

FISHER, R.A. Some Examples of Bayes' Method of the Experimental Determination of Probabilities A Priori. *Journal of the Royal Statistical Society*, v. 24, n. 1, p. 118–124, 1962.

FOX, K. A. Econometric Models of the United States. *Journal of Political Economy*, v. 64, n. 2, p. 128-142, 1958.

FOX, K. A. Agricultural economists in the econometric revolution: institutional background, literature and leading figures. *Oxford Economic Papers*, v. 41, n. 1, p. 53–70, 1989.

FRIEDMAN, M.; MEISELMAN, D. *The relative stability of monetary velocity and the investment multiplier in the United States, 1897-1958*. In: STABILIZATION policies. Englewood Cliffs, N.J.: Prentice Hall, 1963, p. 165-268.

FRISCH, R. *Propagation Problems and Impulse Problems in Dynamic Economics.* In: CASSEL, G. (Org.). Economic Essays in Honour of Gustav Cassel, Londres: Allen and Unwin, 1933. p. 171–205.

FRUHWIRTH-SCHNATTER, S.; PAMMINGER, C.; WEBER, A.; WINTER-EBMER, R. Labor Market Entry and Earnings Dynamics: Bayesian Inference Using Mixtures-of-Experts Markov Chain Clustering, *Journal of Applied Econometrics*, v. 27, n. 7, p. 1116–1137, 2012.

FRUHWIRTH-SCHNATTER, S.; PAMMINGER, C.; WEBER, A.; WINTER-EBMER, R. "Mothers' Long-Run Career Patterns After First Birth, *Journal of the Royal Statistical Society*, v. 179, n. 3, p. 707–725, 2016.

GELMAN, A.; STERN, H. The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant. *The American Statistician*, v. 60, n. 4, 2006.

GEWEKE, J. Bayesian Reduced Rank Regression in Econometrics. *Journal of Econometrics*, v. 75, n. 1, p. 121-146, 1996.

GEWEKE, J.; AMISANO, G. GEWEKE, J. Optimal prediction pools. *Journal of Econometrics*, v. 164, n. 1, p. 130-141, 2011.

GHOSE, T. Just a Theory: 7 Misused Science Words. *Scientific American*, 2013. Disponível em:https://www.scientificamerican.com/article/just-a-theory-7-misused-science-words/. Acesso em: 10 de jun. 2021.

GIGERENZER, G. *The Superego, the Ego, and the Id in Statistical Reasoning*. In: KEREN, G; LEWIS, C. A. (Orgs). A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues, Hillsdale: Erlbaum, 1993, p. 311-339.

GIGERENZER, G.; SWIJTINK, Z; PORTER, T.; DASTON, L.; BEATTY, J.; KRUGER, L.

The empire of chance: How probability changed science and every day life. Cambridge: Cambridge University Press, 1989.

GILBERT, C. L. Economic Theory and Econometric Models. *The Economic and Social Review*, v. 21, n. 1, p. 1-25, 1989.

GOODMAN, N.; HENDERSON, L; TENENBAUM, J.; WOODWARD, J. The structure and dynamics of scientific theories: A hierarchical Bayesian perspective. *Philosophy of Science*, v. 77, n. 2, p. 172-200, 2010.

GRANGER, C. W. L. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, v. 37, n. 3, p. 424-438, 1969.

GRANGER, C. W. L.; HATANAKA, M. *Spectral Analysis of Economic Time Series*. Princeton: Princeton University Press, 1964.

GRANGER, C. W. L.; NEWBOLD, P. Spurious Regressions in Econometrics. *Journal of Econometrics,* v. 2, n. 2, p. 111-120, 1974.

GRILICHES, Z. Production Functions in Manufacturing: Some Additional Results. *Southern Economic Journal*, v. 35, n. 2, p. 151-156, 1968.

GREELAND, S. The Need for Cognitive Science in Methodology. *American Journal of Epidemiology*, v. 186, n. 6, p. 639-645, 2017.

HAAVELMO, T. The Probability Approach in Econometrics. *Econometrica,* v. 12, n.4, p. 1-115, 1944.

HECKMAN, J. Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, v. 42, n. 4, p. 679-694, 1974.

HECKMAN, J. VYTLACIL, E. Local instrumental variables and latent variable models for identifying and bounding treatment effects. PNAS, v. 96, n. 8, p. 4730-4734, 1999.

HECKMAN, J. Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective. *Quarterly Journal of Economics*, v. 115, n.1, p. 45-97, 2000.

HENDRY, D. F.; MORGAN, M. S. *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press, 1997.

HILDRETH, C. Bayesian Statisticians and Remote Clients, *Econometrica*, v. 31, n. 3, p. 422–438, 1963.

HOLT, C. C.; MODIGLIANI, F.; MUTH, J. F.; SIMON, H. A.; BONINI, C. P.; WINTERS, P. R. *Planning Production, Inventories, and Work Force*. Nova Jersey: Prentice Hall, 1960.

HOOGERHEIDE, L. F.; BLOCK J. H.; THURIK, A. R. Family Background Variables as Instruments for Education in Income Regressions: A Bayesian Analysis. *Economics of Education Review*, v. 31, n. 5, p. 515–523, 2012.

HU, X.; MUNKIN, M. K.; TRIVEDI, P. Estimating Incentive and Selection Effects in the Medigap Insurance Market: An Application with Dirichlet Process Mixture Model, Journal of Applied Econometrics, v. 30, n. 7, p. 1115-1143, 2015.

HUBBARD, R. *Corrupt Research: The Case for Reconceptualizing Empirical Management and Social Science.* Califórnia: SAGE Publications, 2016.

IMBENS, G. W.; RUBIN, D. B. Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *The Review of Economic Studies,* v. 64, n. 4, p. 555–574, 1997.

IOANNIDIS, J. P. A. Why most discovered true associations are inflated. *Epidemiology*, v. 19, n. 5, p. 640-648, 2008.

JAYNES, E. T. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press, 2003.

JACOBI, L.; SOVINSKY, M. Marijuana on Main Street? Estimating Demand in Markets with Limited Access. *American Economic Review,* v. 106, n. 8, p. 2009- 2045, 2016.

JEVONS, W. S. *Pure Logic: Or, the Logic of Quality Apart Form Quality: with Remarks on Boole's System and on the Relation of Logic and Mathematics*. Londres: Edward Stanford, 1864.

JEVONS, William Stanley. *The Principles of Science: A Treatise on Logic and Scientific Method*. Londres: MacMillan, 1877.

JOHANSEN, S. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, v. 12, n. 2-3, p. 231-254, 1988.

JOHNSTON, J. *Econometric Methods.* New York: McGraw-Hill, 1960.

KASS, R. E.; WASSERMAN, L. The Selection of Prior Distributions by Formal Rules. *Journal of the American Statistical Association*, v. 91, n. 435, p. 1343-1370, 1996.

KLINE, B.; TOBIAS, J. L. The Wages of BMI: Bayesian Analysis of a Skewed Treatment Response Model with Nonparametric Endogneity. *Journal of Applied Econometrics*, v. 23, p, 767–793, 2008.

KLEIN, L. R. Macroeconomics and the Theory of Rational Behavior. *Econometrica*, v. 14, n. 2, p. 93-108, 1946.

KLEIN, L. R. *Textbook of Econometrics*. Nova Jersey: Prentice-Hall, 1958.

KLEIN, L. R. Whither Econometrics?. *Journal of the American Statistical Association*, v. 66, n. 334, p. 415–421, 1971.

KLEIN, L. R.; GOLDBERGER, A. S. *An Econometric Model for the United States, 1929-1952*. Amsterdã: Elsevier, 1955.

KLOEK, T.; VAN DIJK, H. K. Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica,* v. 46, n. 1, p. 1-19, 1978.

KOOP, G. Recent progress in applied Bayesian econometrics. *Journal of Economic Surveys*, v. 8, n. 1, p. 1-34, 1994.

KOOP, G.; TOBIAS, J. Learning about heterogeneity in returns to schooling. *Journal of Applied Econometrics*, v. 19, n. 7, p. 827-884, 2004.

KOOPMANS, T. C. *Three Essays on the State of Economic Science*. New York: McGraw-Hill, 1957.

KUFFNER, T. A.; WALKER, S. G. Why are p-Values Controversial*?*. *The American Statistician*, v. 73, n. 1, p. 1-3, 2017.

KUHN, T. *The Structure of Scientific Revolutions*. Chicago: University Of Chicago Press, 1962.

KUHN, T. A tensão essencial. Lisboa: Edições 70, 1989.

LEAMER, E. E. Multicollinearity: A Bayesian interpretation. *The Review of Economics and Statistics*, v. 55, n. 3, p. 371–380, 1973.

LEAMER, E. E. False models and post-data model construction. *Journal of the American Statistical Association*, v. 69, n. 345, p. 122–131, 1974.

LEAMER, E. E. *Specification Searches: Ad Hoc Inference With Nonexperimental Data*. New York: John Wiley and Sons, 1978.

LEEPER, E.; SIMS, C.; ZHA, T. What Does Monetary Policy Do?. *Brookings Papers on Economic Activity*, v. 27, n. 2, p. 1-78, 1996.

LEMPERS, F. B. *Posterior Probabilities of Alternative Linear Models*, Rotterdam: Rotterdam University Press, 1971.

LI, K.; POIRIER, D. J. Relationship between Maternal Behavior During Pregnancy, Birth Outcome, and Early Childhood Development: An Exploratory Study. *CESifo Working Paper Series*, n. 1030, 2003.

LI, M.; TOBIAS, J. L. Bayesian Inference in a Correlated Random Coefficients Model: Modeling Treatment Effect Heterogeneity and Heterogeneous Returns to Schooling, *Journal of Econometrics*, v. 162, p. 346–361, 2011.

LINDLEY, D. V. Bayesian Statistics: A Review. Londres: Society for Industrial and Applied Mathematics, 1993.

LIU, T. C. A monthly recursive econometric model of United States: a test of feasibility. *The Review of Economics and Statistics*, v. 51, n. 1, p. 1-13, 1969.

LUBRANO, M.; PIERSE, R. G.; RICHARD, L. F. Stability of a U. K. Money Demand Equation: a Bayesian Approach to Testing Exogeneity. *Review of Economic Studies*, v. 53, n. 4, p. 603-634, 1986.

LUCAS, R. E. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy- Elsevier*, v. 1, n. 1, p. 19-46, 1976.

MACKAY, D. *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press, 2003.

MCSHANE, B. B.; GAL, D.; GELMAN, A.; ROBERT, C.; TACKETT, J. L. Abandon Statistical Significance. *The American Statistician*, v. 73, n. 1, p. 235-245, 2017.

MOORE, H. L. *Economic Cycles: Their Law And Cause.* Montana: Kessinger Publishing, 1914.

MORGAN, M. S. *The History of Econometric Ideas*. Cambridge: Cambridge University Press, 1989.

MUNNELL, A. H. Why has productivity growth declined? Productivity and public investment. *New England Economic Review*, p.3-22, 1990.

NELSON, C. The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy. *The American Economic Review*, v. 62, n. 5, p. 902-917, 1972.

NELSON, C.; PLOSSER, C. R. Trends and random walks in macroeconmic time series: Some evidence and implications. *Journal of Monetary Economics*, v. 10, n. 2, p. 139-162, 1982.

NEYMAN, J. Basic Ideas and Some Recent Results of the Theory of Testing Statistical Hypotheses. *Royal Statistical Society*, v. 105, n. 4, p. 292-327, 1942.

NEYMAN, J. *First Course in Probability and Statistics*. New York: Holt, 1950.

NEYMAN, J. "Inductive Behavior" as a Basic Concept of Philosophy of Science. *Review of the International Statistical Institute*, v. 25, n. 1-3, p. 7-22, 1957.

NEYMAN, J. Silver Jubilee of My Dispute with Fisher. *Journal of the Operations Research Society of Japan*, v. 3, n. 4, p. 145–154, 1961.

NEYMAN, J. Frequentist Probability and Frequentist Statistics. *Synthese*, v. 36, n.1, p. 97-131, 1977.

NEYMAN, J.; PEARSON, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, v. 20A, n. 1-2, p. 263–294, 1928.

NEYMAN, J.; PEARSON, E. S. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London*, v. 231, p. 289-337, 1933.

MALINVAUD, E. *The Statistical Methods of Econometrics*. Amsterdã: North Holland, 1964.

MALINVAUD, E. The Challenge of Macroeconomic Understanding. *Banco Nazionale Del Lavoro Quarterly Review*, n. 169, p. 219–238, 1987.

MAYO, D. *Error and the Growth of Experimental Knowledge*. Chicago: University Of Chicago Press, 1996.

MILLER, J. What is the probability of replicating a statistically significant effect?. *Psychonomic Bulletin and Review*, v. 16, n. 1, p. 617–640, 2009.

MOGIE, M. In support of null hypothesis significance testing. *Proceedings of the Biological Sciences*, v. 271, n. 3, p. 582-584, 2004.

MUTH, J. F. Rational Expectations and the Theory of Price Movements. *Econometrica*, v. 29, n. 3, p. 315-335, 1961.

OAKES, M. *Statistical inference: A commentary for the social and behavioral sciences*. Chicester: John Wiley & Sons, 1986.

O´HAGAN, A. Expert Knowledge Elicitation: Subjective but Scientific. *The American Statistician*, v. 73, n. 1, p. 69-81, 2019.

ORCUTT, G. H. Toward Partial Redirection of Econometrics. *The Review of Economics and Statistics*, v. 34, n. 3, p. 195-200, 1952.

PESARAN, M. H. On the General Problem of Model Selection. Review of Economic Studies, v. 41, n. 2, p. 153-171, 1974.

PHILLIPS, P. To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends. *Journal of Applied Econometrics*, v. 6, n. 4, p. 333-364, 1991.

PRESS, S. J. Bayesian Computer Programs. In: ZELLNER, A (org,), Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys. Amsterdã: North-Holland, 1980. p. 429-442.

QIN, D. *The Formation of Econometrics: A Historical Perspective*. Oxford: Oxford University Press, 1997.

QIN, D. *A History of Econometrics: The Reformation from the 1970s*. New York: Oxford University Press, 2011.

RAMSEY, J. B. Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society*, v. 31, n. 2, p. 350-371, 1969.

REINHART, A. *Statistics done wrong.* San Francisco: No Starch Press, 2015.

ROTHENBERG, T. J. A Bayesian Analysis of Simultaneous Equations Systems. Econometric Institute report 6315, Netherlands School of Economics, 1963.

ROTHENBERG, T. J. *Bayesian Analysis of Simultaneous Equations Models*. In: FIENBERG, S. E.; ZELLNER, A. (orgs.), Studies in Bayesian Econometrics and Statistics. Amsterdã: North-Holland, 2015.

RUBIN, D. B. Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, v. 6, n. 1, p. 34-58, 1978.

RUDNER, R. The scientist qua scientist makes value judgments. *Philosophy of Science*, v. 20, n. 1,p. 1-6, 1953.

SALSBURG, D. *UMA SENHORA TOMA CHÁ...: como a estatística revolucionou a ciência no século XX*. Rio de Janeiro: Jorge Zahar, 2009.

SARGENT, T. J. A Classical Macroeconometric Model for the United States. *Journal of Political Economy*, v. 84, n. 2, p. 207-238, 1976.

SCHMIDT, F. L. What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, v. 47, n. 10, p. 1173-1181, 1992.

SELLKE, T; BAYARRI, M. J.; BERGER, J. O. Calibration of p Values for Testing Precise Null Hypotheses. *The American Statistician*, v. 55, n. 1, p. 62-71, 2001.

SIMON, H. A. *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: John Wiley and Sons, 1957.

SIMS, C. A. Macroeconomics and Reality. *Econometrica*, v. 48, n. 1, p. 1-48, 1980.

SLUTZKY, E. The Summation of Random Causes as the Source of Cyclic Processes. Econometrica, v.5, n. 2, p. 105-146, 1937.

SPANOS, A. Probability Theory and Statistical Inference: econometric modeling with observational data. Cambridge:Cambridge University Press, 1998.

SPRENGER, J. *Bayesian Philosophy of Science.* New York: Oxford University Press, 2010.

STONES, R. The review of Haavelmo (1944). *The Economic Journal*, v. 56, n. 222, p. 265–269, 1946.

THEIL, H. The Information Approach to Demand Analysis. Econometrica, v. 33, n. 1, p. 67-87, 1965.

TINBERGEN, J. Annual Survey: Suggestions on Quantitative Business Cycle Theory. *The Econometric Society*, v. 3, n. 3, p. 241-308, 1935.

TOBIN, J. Estimation of relationships for limited dependent variables. *Econometrica*, v. 26, n. 1; p. 24-36, 1958.

TRAFIMOW, D.; TRIANDIS, H.; GOTO, S. Some tests of the distinction between the private self and the collective self. *Journal of Personality and Social Psychology*, v. 60, p. 649-655, 1991.

TUKEY, J.W. *The problem of multiple comparisons*. Princeton: Princeton University, 1953.

VAN DIJK, H.; KLOEK, T. Further experience in Bayesian analysis using Monte Carlo integration. *Journal of Econometrics*, v. 14, n. 3, p. 307-328, 1980.

WACHOLDER, S.; CHANOCK, S.; GARCIA-CLOSAS, M.; EL GHORMLI, L.; ROTHMAN, N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute*, v. 96, n. 6, p. 434-442, 2004.

WEINTRAUB, E. R. *Stabilizing Dynamics: Constructing Economic Knowledge*. New York: Cambridge University Press, 1991.

WILLIAMSON G. Social Theory and Applied Health Research. Journal of Advanced Nursing, V. 57, p. 114-114, 2007.

WOLD, H. Causality and Econometrics. *Econometrica*, v. 22, n. 2, p. 162-177, 1954.

YULE, G. U. On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers. *Royal Society of London*, v. 226, n. 636-646, p. 267-298, 1927.

ZELLNER, A.; TIAO, G. C. Bayesian analysis of the regression model with autocorrelated errors. *Journal of the American Statistical Association*, v. 59, n. 307, p. 763-778, 1964.

ZELLNER, A.; GEISEL, M. S. Analysis of Distributed Lag Models with Applications to Consumption Function Estimation. *Econometrica*, v. 38, n. 6, p. 865-888, 1970.

ZILIAK, S. T.; MCCLOSKEY, D. N. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Michigan: University of Michigan Press, 2008.