

UNIVERSIDADE TÉCNICA DE LISBOA INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

MESTRADO EM MATEMÁTICA APLICADA À ECONOMIA E À GESTÃO

MODELOS DE REGRESSÃO PARA DADOS DE CONTAGEM

JOAQUIM JOSÉ DOS SANTOS RAMALHO

Orientação: Prof. Doutor João Manuel Caravana Santos Silva

Júri:

Presidente: Prof. Doutor João Manuel de Sousa Andrade e Silva

Vogais: Prof. Doutor João Manuel Caravana Santos Silva

Prof. Doutor José Pedro Andrade Portugal Dias

Agosto 1996

**UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO**

MESTRADO EM MATEMÁTICA APLICADA À ECONOMIA E À GESTÃO

**MODELOS DE REGRESSÃO
PARA DADOS DE CONTAGEM**

JOAQUIM JOSÉ DOS SANTOS RAMALHO

Orientação: Prof. Doutor João Manuel Caravana Santos Silva

Júri:

Presidente: Prof. Doutor João Manuel de Sousa Andrade e Silva

Vogais: Prof. Doutor João Manuel Caravana Santos Silva

Prof. Doutor José Pedro Andrade Portugal Dias

Agosto 1996

Resumo

Neste trabalho analisam-se os principais modelos de regressão para dados de contagem, na perspectiva habitual de especificação, estimação e avaliação. Após a descrição detalhada dos modelos tradicionais, Poisson e Binomial Negativo, explora-se pormenorizadamente o Poisson Generalizado, um modelo recentemente introduzido nesta área. Os três modelos são depois adaptados a casos em que a amostra recolhida é truncada em zero. De seguida, abordam-se alguns testes de especificação, começando-se por avaliar o modelo de Poisson face às duas alternativas referidas. Para tal, deduzem-se testes score, os quais provam que, para determinadas parametrizações, os modelos Binomial Negativo e Poisson Generalizado constituem alternativas localmente equivalentes. De modo a testar a adequabilidade destes últimos modelos, consideram-se ainda diversos testes de hipóteses não encaixadas, para alguns dos quais é realizado um estudo de simulação de Monte Carlo com o objectivo de examinar o seu comportamento quando aplicados a modelos truncados. Os resultados obtidos mostram que o desempenho dos testes é satisfatório, embora alguns deles sejam claramente afectados pelo nível de sobredispersão e pelo grau de truncagem admitidos.

Palavras-chave: dados de contagem; Poisson Generalizado; modelos truncados; sobredispersão; testes para hipóteses não encaixadas.

Abstract

In this study, the specification, estimation and evaluation of some count data regression models are discussed. After describing the most common models, Poisson and Negative Binomial, the Generalized Poisson model, recently included in this area, is analysed in detail. The three models are then extended to cases where the sample is truncated at zero. Furthermore, some specification tests are considered. First, the Poisson model is tested against the other two models. Some score tests are derived, and it is possible to prove that, with some parametrizations, the Negative Binomial and the Generalized Poisson are locally equivalent alternatives. In order to test the other count data models, some non-nested hypothesis tests are also discussed. Finally, a Monte Carlo simulation is done, with the objective of comparing the performance of non-nested tests, when they are applied to truncated models. The results obtained show that these tests perform satisfactorily, although some of them clearly depend on the magnitude of the overdispersion and truncation considered.

Key-words: count data; Generalized Poisson; truncated models; overdispersion; non-nested hypothesis tests.

ÍNDICE

Lista de Quadros	5
Agradecimentos	6
Introdução	7
I. Modelos de Regressão para Dados de Contagem	10
1. Modelo de Poisson	10
2. Modelo Binomial Negativo	15
2.1. Motivação	15
2.2. Parametrizações mais comuns	19
2.3. Estimação	23
3. Modelo de Poisson Generalizado	27
3.1. Distribuições Lagrangeanas	27
3.2. A Função de Distribuição de Poisson Generalizada	28
3.2.1. Motivação	28
3.2.2. Descrição	29
3.3. O Modelo de Regressão	32
3.3.1. Apresentação	32
3.3.2. Estimação	34
3.3.3. Abordagem de uma nova parametrização	36
4. Modelos para Amostras Truncadas em Zero	38
4.1. Distribuições Truncadas	39
4.2. Modelo de Poisson Truncado	41
4.3. Modelo Binomial Negativo Truncado	43
4.4. Modelo de Poisson Generalizado Truncado	46
II. Testes de Especificação	49
1. Testes de Sobredispersão	49
1.1. Poisson contra Binomial Negativo	50
1.2. Poisson contra Poisson Generalizado	56
1.3. Amostras Truncadas	58

2. Testes de Hipóteses Não Encaixadas.....	60
2.1. Testes baseados na combinação de Funções de Probabilidade.....	60
2.1.1. Testes Score para a combinação linear de Funções de Probabilidade....	61
2.1.2. Testes específicos para Modelos de Contagem	64
2.2. Testes baseados na realização de regressões artificiais	65
2.2.1. Testes de Davidson e MacKinnon	66
2.2.2. Testes de Fisher e McAleer e proposta de uma variante	69
III. Estudo de Simulação de Monte Carlo	71
1. Objectivos.....	71
2. Descrição	73
3. Resultados.....	76
3.1. Ensaios com diferentes graus de truncagem	76
3.2. Ensaios com diferentes níveis de sobredispersão	79
3.3. Ensaios com diferentes dimensões amostrais.....	82
Conclusão	84
Anexos.....	87
A. Formulário complementar	88
B. Programas-tipo utilizados no TSP	90
1. Teste P	91
2. Teste P_A	94
3. Teste P_{A1}	98
4. Teste T_k	102
Bibliografia.....	104

Lista de Quadros

1 - Dimensões estimadas para diferentes graus de truncagem e testes para a hipótese de que a dimensão real é 5% (H_1 : BNT2; $\bar{n} = 500$; $\alpha = 0.25$)	77
2 - Potências estimadas para diferentes graus de truncagem ($\bar{n} = 500$; $\alpha = 0.25$).....	78
3 - Dimensões estimadas para diferentes níveis de sobredispersão (H_1 : BNT2; $\bar{n} = 500$).....	80
4 - Potências estimadas para diferentes níveis de sobredispersão (H_2 : BNT2; $\bar{n} = 500$)	80
5 - Dimensões estimadas para diferentes tamanhos amostrais (H_1 : BNT2; $\alpha = 0.25$).....	82
6 - Potências estimadas para diferentes tamanhos amostrais (H_2 : BNT2; $\alpha = 0.25$).....	83

Agradecimentos

- Ao meu orientador, Prof. Doutor João Santos Silva, pela sugestão do tema, críticas, indicações bibliográficas... Enfim, pelo grande apoio prestado na elaboração deste trabalho;
- Ao Prof. Doutor Andrade e Silva, pela disponibilidade e atenção dispensadas e pelos comentários efectuados;
- À minha namorada e colega, Esmeralda, pela leitura atenta e cuidada desta tese e pelas suas úteis sugestões.

Introdução

Na última década, um novo campo de investigação tem vindo a conquistar o seu espaço na literatura econométrica: os modelos de regressão para dados de contagem. O elemento diferenciador deste tipo de modelos, que motivou o seu aparecimento, é a natureza especial da variável dependente, que assume apenas valores inteiros não negativos, correspondentes à ocorrência de um dado número de acontecimentos num certo intervalo de tempo.

Os trabalhos pioneiros nesta área devem-se a Hausman, Hall e Griliches (1984), Gourieroux, Monfort e Trougnon (1984b), Lee (1986) e Cameron e Trivedi (1986), tendo sido realizado, neste último artigo, o primeiro *survey* sobre dados de contagem. O tratamento destas variáveis evoluiu muito desde essa altura, estando os mais recentes desenvolvimentos sintetizados em Winkelmann e Zimmermann (1995).

Na ciência económica, a natureza dos dados disponíveis para retratar um determinado fenómeno exige, por vezes, o uso de modelos de contagem. Assim, não é de estranhar a proliferação, nos últimos tempos, de trabalhos práticos que recorrem à estimação deste género de modelos. Exemplos desta situação são os estudos sobre a procura de bens de lazer, de Shaw (1988), Creel e Loomis (1990), Hellerstein (1991), Grogger e Carson (1991) e Englin e Shonkwiler (1995), a entrada e saída de indústrias do mercado, de Chappell, Kimenyi e Mayer (1990), a mobilidade do factor trabalho, de Winkelmann e Zimmermann (1995), e o número de acidentes rodoviários e sua relação com as apólices de seguro, de Dionne e Vanasse (1992).

O modelo clássico de regressão linear não é adequado para explicar como uma variável discreta, que somente pode tomar um pequeno número de valores com probabilidades estritamente positivas, depende de um conjunto de regressores. Além disso, interessa calcular, após a estimação do modelo, a probabilidade de um acontecimento, dadas as variáveis explicativas, se verificar. Para tal, é necessário

pressupor que a variável endógena segue uma distribuição discreta específica.

A teoria económica sugere variáveis explicativas mas não fornece pistas sobre qual a distribuição mais indicada para modelar os dados de contagem. Por isso, tornou-se prática corrente começar por estimar o modelo de Poisson, devido à sua simplicidade. Contudo, cedo se tornou evidente que a maioria dos eventos económicos caracteriza-se por apresentar demasiada dispersão relativamente à que seria de esperar ao supor que respeitam uma distribuição de Poisson, onde a média e a variância são iguais.

Devido a este problema, habitualmente denominado de sobredispersão, a par do modelo de Poisson procede-se, geralmente, à estimação do modelo Binomial Negativo. A atractividade deste resulta do facto dele ser tanto uma alternativa de teste para o Poisson como um modelo apto para explicar ocorrências onde haja razões para esperar que a relação valor esperado - variância assuma uma determinada forma.

Em muitas circunstâncias, devido à maneira como a recolha da amostra é efectuada, não se consegue observar a distribuição completa da variável de contagem. Nomeadamente, é frequente a não observação dos zeros da população. Como este facto é de extrema importância, rapidamente os modelos citados foram adaptados à truncagem referida.

Nesta dissertação, toda esta teoria *standard* dos modelos de contagem é revista, na perspectiva usual de especificação, estimação e avaliação. Adicionalmente, estuda-se um modelo recentemente considerado neste âmbito, o Poisson Generalizado, e promove-se o recurso a testes de hipóteses não encaixadas.

Tal como o Binomial Negativo, o modelo de regressão de Poisson Generalizado, introduzido por Consul e Famoye (1992), é apropriado para modelizar variáveis caracterizadas por sobredispersão. Todavia, ao admitir outros padrões de heteroscedasticidade, o seu emprego pode-se tornar preferível em certas situações. Paralelamente, o modelo de Poisson Generalizado, como incorpora o Poisson, constitui mais uma opção para testar a adequabilidade deste. Assim, novos testes de sobredispersão serão aqui desenvolvidos.

Uma das lacunas actuais das aplicações práticas sobre dados de contagem é a não utilização de testes que permitam confrontar os vários modelos, com excepção do

caso em que o Poisson é a hipótese nula. Como os restantes modelos não são encaixados, nesta tese advoga-se o uso de testes próprios para os avaliar. Para lá de se tentar conjugar os principais testes para hipóteses não encaixadas existentes actualmente com a teoria dos modelos de contagem, propõem-se dois novos testes. Através da realização de uma análise de simulação de Monte Carlo, serão investigados os comportamentos de destes testes face a diferentes condições amostrais.

Além do cumprimento dos objectivos traçados, em cada tema específico indicam-se formas alternativas de abordagem e fornecem-se referências bibliográficas que permitirão aos eventuais interessados aprofundar não só os assuntos tratados neste trabalho, mas também outros tópicos da teoria dos dados de contagem.

A organização desta dissertação é descrita de seguida. Nas três primeiras secções do capítulo I são apresentados e formalizados os modelos de contagem de Poisson, Binomial Negativo e Poisson Generalizado, os quais são extendidos a amostras truncadas em zero na quarta secção. O capítulo II é dividido em duas partes: na primeira deduzem-se vários testes de sobredispersão; na segunda, expõem-se alguns testes de hipóteses não encaixadas. No capítulo III efectua-se o estudo de simulação de Monte Carlo.

Capítulo I:

Modelos de Regressão para Dados de Contagem

O objectivo deste capítulo é a formalização dos modelos de contagem que serão analisados ao longo de todo este trabalho. Começa-se por apresentar o modelo simples de Poisson para, introduzindo-se um tipo específico de heterogeneidade, se passar ao Binomial Negativo. De seguida, estuda-se o modelo de regressão de Poisson Generalizado, o qual é abordado com muito maior detalhe, em virtude da sua aplicação a dados de contagem ser bastante mais recente. Por último, procede-se à adaptação dos modelos referidos a situações em que a amostra é truncada em zero.

1. Modelo de Poisson

A tentativa de modelização de um conjunto de observações de uma dada variável de contagem começa, normalmente, pela estimação do modelo de Poisson. As razões fundamentais para este procedimento são as seguintes:

- a distribuição de Poisson adequa-se totalmente à principal característica dos dados de contagem, o facto destes assumirem apenas valores inteiros não negativos;
- após a sua estimação, é possível inferir relativamente à probabilidade de ocorrência futura de um dado acontecimento, admitindo que o modelo está bem especificado;
- o modelo de Poisson possui uma estrutura simples, podendo ser facilmente estimado.

Considerando um conjunto de observações independentes da variável aleatória discreta Y_i , $i = 1, \dots, n$, a função de probabilidade de Poisson para o indivíduo i é

$$P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad (I.1)$$

com $y_i = 0, 1, 2, \dots$ e $\lambda_i > 0$.

O modelo de regressão resulta da colocação do parâmetro λ_i em função de um conjunto de variáveis explicativas: $\lambda_i = f(x_i, \beta)$, sendo x_i um vector $1 \times m$ (contém observações das m variáveis exógenas para o indivíduo i) e β um vector $m \times 1$, composto por parâmetros desconhecidos. Como λ_i representa o valor esperado condicional de Y_i , só pode tomar valores não negativos, pelo que a função f deve ter como contradomínio apenas os números reais não negativos. Por este motivo, ela é geralmente expressa na forma exponencial,

$$\lambda_i = \exp(x_i \beta), \quad (I.2)$$

o que também acontecerá neste estudo.

No modelo de Poisson, a média e a variância condicionais são iguais:

$$E(Y_i | x_i) = V(Y_i | x_i) = \lambda_i \quad (I.3)$$

Uma vez que a variância depende dos regressores, este modelo é, por natureza, heterocedástico, característica partilhada, aliás, por todos os modelos de contagem. Se a forma de heterocedasticidade prevista, $V(Y_i | x_i) = \exp(x_i \beta)$, corresponder à real, não se registrarão quaisquer influências negativas sobre os estimadores, desde que estes sejam obtidos recorrendo a métodos que a tenham em conta. Caso contrário, as consequências serão semelhantes às verificadas aquando da aplicação do método dos mínimos quadrados a uma situação em que exista heterocedasticidade: sob certas condições, que serão enumeradas posteriormente, os estimadores continuarão a ser consistentes, mas a sua eficiência será menor.

Quando se conhece a distribuição condicional de uma dada variável endógena, para se poder tirar total partido dessa informação, deve-se usar o método da Máxima Verosimilhança (MV) na sua modelação. É o emprego deste método na estimação do modelo de regressão de Poisson que será seguidamente formalizado.

O Logaritmo da Função de Verosimilhança, $LL(\beta)$, é

$$LL(\beta) = \sum_{i=1}^n [-\lambda_i + y_i \ln(\lambda_i) - \ln(y_i!)], \quad (I.4)$$

de onde resulta o vector score:

$$s(\beta) = \frac{\partial LL(\beta)}{\partial \beta} = \sum_{i=1}^n \left(\frac{\partial \ln \lambda_i}{\partial \beta} \right)' (y_i - \lambda_i) \quad (I.5)$$

Admitindo a especificação I.2, as condições de primeira ordem que permitem maximizar $LL(\beta)$ correspondem a:

$$\sum_{i=1}^n x_i' (y_i - \lambda_i) = 0 \quad (I.6)$$

Esta expressão traduz a ortogonalidade entre as variáveis explicativas que formam o vector x_i e os erros $u_i = y_i - E(Y_i | x_i)$. Se o sistema de equações I.6 tiver solução, ela será única, visto que a matriz hessiana,

$$H(\beta) = \frac{\partial^2 LL}{\partial \beta \partial \beta'} = - \sum_{i=1}^n \lambda_i x_i' x_i, \quad (I.7)$$

é definida negativa, implicando que a função de verosimilhança seja estritamente côncava. Para a sua determinação, pode ser aplicado qualquer método de optimização não linear¹ pois, para expressões daquele tipo, a convergência geralmente é rápida.

Os estimadores obtidos terão, assintoticamente, as propriedades comuns num

¹ Ver, por exemplo, Greene (1993), pp. 346-350, que analisa resumidamente alguns destes métodos.

estimador da MV, isto é, serão consistentes, eficientes e normalmente distribuídos,

$$\sqrt{n}(\hat{\beta} - \beta) \approx N[0, nI^{-1}(\beta)], \quad (I.8)$$

representando $I(\beta)$ a matriz de informação que, por definição, é igual a $-E[H(\beta)]$, isto é,

$$I(\beta) = \sum_{i=1}^n \lambda_i x_i' x_i \quad (I.9)$$

Neste caso, como a matriz hessiana é constante, a estimação da matriz de informação pelo método de Newton, $I(\hat{\beta}) = -H(\hat{\beta})$, ou de Scoring, $I(\hat{\beta}) = -E[H(\hat{\beta})]$, conduz a resultados idênticos. Preferencialmente, serão estes os processos de estimação aplicados neste trabalho, embora, por vezes, se possa ter de recorrer, por razões práticas, ao método BHHH [Berndt, Hall, Hall e Hausman (1974)], onde $I(\hat{\beta}) = S'(\hat{\beta})S(\hat{\beta})$, sendo $S(\hat{\beta})$ a matriz contendo as contribuições individuais para o score. Com efeito, utilizando este último procedimento, a estimação da matriz de informação torna-se muito mais simples, pois não envolve o cálculo de segundas derivadas. Todavia, a primazia concedida aos métodos de Newton e de Scoring justifica-se por eles, geralmente, revelarem um melhor desempenho em pequenas amostras².

A validade das propriedades referidas estará dependente, naturalmente, do facto da distribuição condicional de Y_i ser, efectivamente, a de Poisson. No entanto, Gourieroux, Monfort e Trognon³ (1984a) provaram que, qualquer que seja a verdadeira distribuição de $Y_i | x_i$, se o modelo estimado pertencer à Família Exponencial Linear (FEL), então os estimadores da MV dos parâmetros do valor esperado condicional, se este estiver correctamente formalizado, serão consistentes, com excepção do associado ao regressor fixo. A distribuição assintótica também continuará a ser a normal, pelo que apenas a eficiência dos estimadores não fica garantida. Quanto aos parâmetros não incluídos no valor esperado condicional, eles não deverão ser estimados em simultâneo

² Consultar Godfrey (1988), p. 82, ou os estudos de simulação de Davidson e Mackinnon (1983) e Bera e McKenzie (1986).

³ Daqui para a frente, estes autores serão referidos apenas por GMT.

com os outros, de modo a que a consistência destes não seja prejudicada, a não ser que a distribuição em causa pertença à família exponencial quadrática.

Para uma distribuição pertencer à FEL, é necessário que a sua função de probabilidade possa ser expressa na forma

$$P(Y_i = y_i | x_i) = \exp[A(\mu_i) + B(y_i) + C(\mu_i)y_i], \quad (I.10)$$

onde $A(\cdot)$, $B(\cdot)$ e $C(\cdot)$ representam funções reais e $\mu_i = E(Y_i | x_i)$. Prova-se facilmente que a distribuição de Poisson verifica esta condição, com $A(\mu_i) = -\mu_i$, $B(y_i) = -\ln(y_i!)$ e $C(\mu_i) = \ln(\mu_i)$. Assim, admitindo que o valor esperado condicional está bem especificado, os estimadores da MV encontrados para este modelo serão consistentes para os parâmetros de interesse, mesmo que se esteja a assumir incorrectamente que a variável endógena segue uma distribuição de Poisson. Contudo, a variância dos estimadores será geralmente subestimada, impedindo a realização de testes de hipóteses válidos, a não ser que se utilize uma matriz de covariâncias corrigida, como a matriz de White (1982), por exemplo. Além disso, o modelo não poderá ser usado na estimação da probabilidade de ocorrência de um dado fenómeno.

GMT (1984a) propuseram métodos de estimação robustos contra pressupostos distribucionais, os quais apenas exigem que o valor esperado (Método da Pseudo Máxima Verosimilhança) e a variância (Método da Pseudo Máxima Verosimilhança Quase Generalizada) estejam correctamente especificados, podendo aplicar-se às distribuições pertencentes à FEL (no caso do segundo método, somente às que possuam dois tipos de parâmetros). Os estimadores obtidos por ambos os processos são consistentes, sendo os resultantes da aplicação da Pseudo MV menos eficientes, em geral, que os determinados pela Pseudo MV Quase Generalizada. Por sua vez, estes últimos são, normalmente, menos eficientes que os estimadores da MV, quando esta tem por base a verdadeira distribuição. Nenhum destes métodos, que podem ser estudados em GMT (1984a, b), será desenvolvido neste trabalho.

O facto do modelo de Poisson impor a restrição da média e da variância serem iguais obriga à consideração de modelos alternativos, pois a maior parte dos dados de

índole económica são caracterizados pela superioridade da segunda. O mais comum desses modelos é o Binomial Negativo (BN), o qual vai ser abordado em pormenor na próxima secção. Posteriormente, estudar-se-á o modelo de Poisson Generalizado.

Para outro tipo de modelos, ver Lawless (1987) e Winkelmann (1995), que desenvolvem modelos de duração supondo que o tempo de espera entre acontecimentos sucessivos não segue uma distribuição exponencial (pressuposto implícito na distribuição de Poisson), e Wedel, Desarbo, Bult e Ramaswamy (1993), que constróem um modelo que, embora baseado no de Poisson, permite obter coeficientes de regressão distintos para cada classe de indivíduos, sendo o número ideal de grupos também estimado.

2. Modelo Binomial Negativo

Nesta secção analisam-se as duas principais formas de justificar o emprego do modelo BN e apresentam-se as suas parametrizações mais comuns, bem como os resultados mais importantes para a sua estimação.

2.1. Motivação

Se uma variável aleatória discreta Y_i , $i = 1, \dots, n$, tem distribuição BN de parâmetros w e p , $Y_i \sim \text{BN}(w,p)$, a sua função de probabilidade é

$$P(Y_i = y_i) = \frac{\Gamma(y_i + w)}{\Gamma(y_i + 1)\Gamma(w)} p^w (1 - p)^{y_i}, \quad (\text{I.11})$$

onde $0 < p < 1$ e $w > 0$.

O valor esperado e a variância são iguais, respectivamente, a

$$E(Y_i) = \frac{w(1-p)}{p} \quad (\text{I.12})$$

$$V(Y_i) = \frac{w(1-p)}{p^2}, \quad (\text{I.13})$$

donde se pode estabelecer a seguinte relação:

$$V(Y_i) = \frac{1}{p} E(Y_i) \quad (I.14)$$

Como $0 < p < 1$, $V(Y_i) > E(Y_i)$, o que comprova a natural aptidão desta distribuição para modelar dados caracterizados pela existência de sobredispersão. A expressão I.14 parece impor uma relação linear entre a variância e a média mas, como se verá mais à frente, as diversas parametrizações efectuadas a p levarão a relações distintas entre aqueles dois momentos. Aliás, como Cameron e Trivedi⁴ (1986) salientaram, ocasionarão inclusivamente modelos de regressão distintos, pois conduzirão a diferentes hipóteses acerca da forma funcional da heterocedasticidade existente e, em geral, as estimativas de β não serão iguais.

Há muitas maneiras de derivar o modelo BN. Por exemplo, ele pode ter origem na soma de n variáveis aleatórias independentes com a mesma distribuição logarítmica, seguindo n uma distribuição de Poisson [Johnson, Kotz e Kemp (1992), p. 204]. Pode igualmente provir da adição de um conjunto de variáveis aleatórias de Poisson correlacionadas [Johnson, Kotz e Kemp (1992), p. 205]. Contudo, na literatura econométrica, unicamente duas razões têm sido vulgarmente apontadas para o uso do modelo BN, ambas partindo do modelo de Poisson.

A primeira, referida originariamente em Hausman, Hall e Griliches (1984), supõe que a distribuição de Y_i condicional também em λ_i é efectivamente Poisson. Todavia, admite-se que β não é constante na população, pelo que λ_i é um parâmetro estocástico, que varia aleatoriamente de acordo com uma distribuição Gama, daí resultando que a distribuição de Y_i , condicional apenas nos regressores, seja BN.

A segunda explicação, introduzida por GMT (1984b), baseia-se na aceitação de que o verdadeiro modelo é o de Poisson, mas que, devido à impossibilidade de se observarem todas as variáveis de interesse, existe um erro de especificação cuja distribuição é a Gama. Condicionando, então, Y_i apenas nos regressores observados, a

⁴ A partir deste momento referenciados simplesmente por CT.

distribuição obtida será a BN.

Em qualquer dos casos, o modelo de Poisson torna-se inadequado pois é negligenciada a heterogeneidade dos indivíduos. Embora ambas as justificações sejam equivalentes, demonstram-se, de seguida, as duas situações, considerando a seguinte parametrização para a função de densidade Gama:

$$Y \sim G(\tau, \omega) \Rightarrow f(y) = \frac{1}{\omega^\tau \Gamma(\tau)} y^{\tau-1} e^{-\frac{y}{\omega}} \quad , \quad y, \tau, \omega > 0 \quad (I.15)$$

1º Caso: λ_i não é fixo para o indivíduo i , sendo distribuído aleatoriamente segundo uma Gama.

Assumindo que

$$Y_i | x_i, \lambda_i \sim P(\lambda_i)$$

$$\lambda_i | x_i \sim G(\tau_i, \omega) \quad , \quad \tau_i, \omega > 0$$

e sabendo que $P(Y_i = y_i | x_i) = \int_0^\infty P(Y_i = y_i | x_i, \lambda_i) f(\lambda_i | x_i) d\lambda_i$, pode demonstrar-se que

$Y_i | x_i$ tem distribuição BN:

$$\begin{aligned} P(Y_i = y_i | x_i) &= \int_0^\infty \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \frac{1}{\omega^{\tau_i} \Gamma(\tau_i)} \lambda_i^{\tau_i-1} e^{-\frac{\lambda_i}{\omega}} d\lambda_i \\ &= \frac{1}{\omega^{\tau_i} \Gamma(\tau_i) y_i!} \int_0^\infty e^{-\left(1+\frac{1}{\omega}\right)\lambda_i} \lambda_i^{y_i+\tau_i-1} d\lambda_i \end{aligned}$$

Este último integral corresponde a $\frac{\Gamma(y_i + \tau_i)}{\left(1 + \frac{1}{\omega}\right)^{y_i+\tau_i}}$ e $y_i! = \Gamma(y_i+1)$, pelo que:

$$\begin{aligned}
P(Y_i = y_i | x_i) &= \frac{\Gamma(y_i + \tau_i)}{\Gamma(y_i + 1)\Gamma(\tau_i)} \left[\frac{1}{\omega \left(1 + \frac{1}{\omega}\right)} \right]^{\tau_i} \left(\frac{1}{1 + \frac{1}{\omega}} \right)^{y_i} \\
&= \frac{\Gamma(y_i + \tau_i)}{\Gamma(y_i + 1)\Gamma(\tau_i)} \left(\frac{1}{1 + \omega} \right)^{\tau_i} \left(\frac{\omega}{1 + \omega} \right)^{y_i} \tag{I.16}
\end{aligned}$$

Comparando I.16 com I.11, verifica-se que $Y_i | x_i \sim \text{BN}\left(\tau_i, \frac{1}{1 + \omega}\right)$ e, confrontando com I.12 e I.13, conclui-se que:

$$E(Y_i | x_i) = \omega \tau_i \tag{I.17}$$

$$V(Y_i | x_i) = \omega(1 + \omega)\tau_i \tag{I.18}$$

O modelo de regressão obtém-se estabelecendo:

$$\tau_i = \exp(x_i \beta) \tag{I.19}$$

A relação entre a variância e o valor esperado condicionais é linear, sendo dada pela expressão:

$$V(Y_i | x_i) = (1 + \omega)E(Y_i | x_i) \tag{I.20}$$

2º Caso: à partida, λ_i é fixo para todos os indivíduos mas, como existe um erro de especificação no modelo, passa a ser distribuído de forma aleatória.

Nesta situação, há dois tipos de regressores: os observáveis, x_i , e os omitidos, z_i . As hipóteses iniciais são as seguintes:

$$Y_i | x_i, \lambda_i \sim P(\lambda_i)$$

$$\lambda_i = \exp(x_i \beta + z_i) = \exp(x_i \beta) \exp(z_i)$$

$$\mu_i = \exp(x_i \beta)$$

$$\varepsilon_i = \exp(z_i)$$

$$Y_i | x_i, \varepsilon_i \sim P(\mu_i \varepsilon_i)$$

$$\varepsilon_i | x_i \sim G(\tau, \omega) \quad , \quad \tau, \omega > 0,$$

representando ε_i , como é evidente, o erro de especificação. μ_i é considerado fixo para cada indivíduo.

Por um processo análogo ao realizado para o primeiro caso, pode demonstrar-se que $Y_i | x_i \sim \text{BN}\left(\tau, \frac{1}{1 + \omega\mu_i}\right)$ e, conseqüentemente,

$$E(Y_i | x_i) = \tau\omega\mu_i \tag{I.21}$$

$$V(Y_i | x_i) = \tau\omega\mu_i(1 + \omega\mu_i), \tag{I.22}$$

o que implica também, para cada indivíduo, uma relação linear entre a variância e a média condicionais:

$$V(Y_i | x_i) = (1 + \omega\mu_i)E(Y_i | x_i) \tag{I.23}$$

2.2. Parametrizações mais comuns

Vários autores têm proposto diferentes parametrizações da distribuição Gama. Os objectivos são, por um lado, justificar, no âmbito do modelo BN, relações não lineares entre a variância e a média de $Y_i | x_i$ e, por outro lado, incluir explicitamente no modelo BN, o de Poisson como caso particular, o que tem grande importância na formulação de testes de especificação. Também se pretende que em ambas as situações o valor esperado condicional seja definido simplesmente por $\mu_i = \exp(x_i\beta)$.

A parametrização mais corrente é aquela que origina uma relação quadrática entre a variância e o valor esperado de $Y_i | x_i$. Pegando novamente no caso que explicava o modelo BN como resultante da existência de variáveis omitidas, GMT (1984b) colocaram duas hipóteses adicionais:

$$E(\varepsilon_i | x_i) = 1 \quad (I.24)$$

$$V(\varepsilon_i | x_i) = \omega \quad (I.25)$$

A imposição destas condições implica a consideração de que ε_i segue uma distribuição Gama de parâmetros τ e ω , sendo $\tau = \frac{1}{\omega}$. Partindo de $\varepsilon_i | x_i \sim G\left(\frac{1}{\omega}, \omega\right)$, pode provar-se que $Y_i | x_i \sim \text{BN}\left(\frac{1}{\omega}, \frac{1}{1 + \omega\mu_i}\right)$ e que:

$$E(Y_i | x_i) = \mu_i \quad (I.26)$$

$$V(Y_i | x_i) = \mu_i + \omega\mu_i^2 \quad (I.27)$$

CT (1986) construíram um modelo mais geral, que contempla várias relações possíveis entre aqueles dois momentos.

Assumindo que

$$Y_i | x_i, \lambda_i \sim P(\lambda_i)$$

$$\lambda_i | x_i \sim G\left(v_i, \frac{\mu_i}{v_i}\right), v_i, \mu_i > 0,$$

mostra-se, nos moldes das demonstrações anteriores, que $Y_i | x_i \sim \text{BN}\left(v_i, \frac{v_i}{\mu_i + v_i}\right)$ e que:

$$E(Y_i | x_i) = \mu_i \quad (I.28)$$

$$V(Y_i | x_i) = \mu_i + \frac{1}{v_i} \mu_i^2 \quad (I.29)$$

O modelo de regressão obtém-se definindo:

$$\mu_i = \exp(x_i\beta) \quad (I.30)$$

$$v_i = \frac{1}{\alpha} [\exp(x_i\beta)]^a, \alpha > 0, a = 0, 1, 2, \dots \quad (I.31)$$

Substituindo I.30 e I.31 em I.29 resulta:

$$\begin{aligned}
 V(Y_i | x_i) &= \exp(x_i\beta) + \alpha [\exp(x_i\beta)]^{-a} [\exp(x_i\beta)]^2 \\
 &= \exp(x_i\beta) + \alpha [\exp(x_i\beta)]^{2-a} \\
 &= E(Y_i | x_i) + \alpha [E(Y_i | x_i)]^{2-a}
 \end{aligned} \tag{I.32}$$

A expressão I.32 origina variadíssimas formas para a relação variância-média. As mais usadas são aquelas em que se admite $a = 1$, que provoca

$$V(Y_i | x_i) = (1+\alpha)E(Y_i | x_i), \tag{I.33}$$

função linear tal como as encontradas em I.20 e I.23, e $a = 0$, que impõe a relação quadrática já determinada em I.27:

$$V(Y_i | x_i) = (Y_i | x_i) + \alpha [E(Y_i | x_i)]^2 \tag{I.34}$$

Assim, pode-se desde já passar a considerar dois tipos principais de modelos BN's, os quais serão estudados em pormenor ao longo de todo este trabalho:

- modelo BN1 (relação linear entre a variância e a média):

$$\begin{aligned}
 a = 1 \Rightarrow v_i &= \frac{1}{\alpha} \mu_i \Rightarrow y_i | x_i \sim \text{BN} \left(\frac{\mu_i}{\alpha}, \frac{1}{1+\alpha} \right) \\
 P(Y_i = y_i | x_i) &= \frac{\Gamma \left(y_i + \frac{\mu_i}{\alpha} \right)}{\Gamma(y_i + 1) \Gamma \left(\frac{\mu_i}{\alpha} \right)} \left(\frac{1}{1+\alpha} \right)^{\frac{\mu_i}{\alpha}} \left(\frac{\alpha}{1+\alpha} \right)^{y_i}
 \end{aligned} \tag{I.35}$$

- modelo BN2 (relação quadrática entre a variância e a média):

$$\begin{aligned}
a = 0 &\Rightarrow v_i = \frac{1}{\alpha} \Rightarrow y_i | x_i \sim \text{BN}\left(\frac{1}{\alpha}, \frac{1}{1 + \alpha\mu_i}\right) \\
P(Y_i = y_i | x_i) &= \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha\mu_i}\right)^\alpha \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i}
\end{aligned} \tag{I.36}$$

O modelo de Poisson é um caso particular destes pois, quando $\alpha \rightarrow 0$, $V(Y_i | x_i) \rightarrow E(Y_i | x_i)$ e, como isso implica que $v_i \rightarrow \infty$, a distribuição de λ_i será degenerada em μ_i .

A grande utilização, em modelos de contagem, de expressões do tipo da I.34 para representar a relação entre a variância e o valor esperado condicionais tem outra justificação, muito simples. Se se supuser que $Y_i | x_i, \varepsilon_i$ segue uma distribuição de Poisson de parâmetro $\lambda_i = \mu_i \varepsilon_i$ e se se colocarem novamente as hipóteses I.24 e I.25 então, qualquer que seja a distribuição de ε_i , a relação entre a média e a variância, condicionais apenas em x_i , será quadrática. Esta afirmação demonstra-se através do recurso à decomposição da variância condicional

$$\begin{aligned}
V(Y_i | x_i) &= E_{\varepsilon_i | x_i} [V(Y_i | x_i, \varepsilon_i)] + V_{\varepsilon_i | x_i} [E(Y_i | x_i, \varepsilon_i)] \\
&= E_{\varepsilon_i | x_i} (\mu_i \varepsilon_i) + V_{\varepsilon_i | x_i} (\mu_i \varepsilon_i) \\
&= \mu_i + \omega \mu_i^2,
\end{aligned}$$

sendo $\mu_i = E(Y_i | x_i)$.

É evidente que, por esta via, a hipótese que se coloca é muito forte: se não é difícil aceitar I.24, se um dos regressores for fixo, o mesmo não se passa com I.25, que exige que o erro seja homocedástico. Aliás, como se pode concluir pelo modo como foi deduzido, também o modelo BN é bastante exigente em termos de pressupostos, nomeadamente ao impor que o erro de especificação siga uma distribuição Gama. É por isso que, normalmente, se começa por estimar o modelo mais simples, o de Poisson, que é sujeito a uma série de testes, e somente depois se passa, caso se revele necessário, a outros modelos, os quais, naturalmente, também deverão ser avaliados.

Winkelmann e Zimmermann (1991, 1995) propuseram⁵ um modelo BN mais geral, o BN k ⁶, o qual assenta numa especificação para a variância similar à sugerida (I.32, substituindo 2-a por $k+1$) por CT (1986), mas com uma diferença fundamental: admite-se a possibilidade do parâmetro k ser estimado, não sendo necessariamente zero ou um. A análise pormenorizada do modelo BN k não será feita neste trabalho. No entanto, ele servirá de base, no próximo capítulo, à construção de testes de hipóteses não encaixadas que visarão testar o BN1 “contra” o BN2.

Mais recentemente, foi sugerido [Xue e Deddens (1992)] um modelo BN composto, designado pelos seus autores de BN sobredisperso. Como o próprio nome indica, trata-se de um modelo que incorpora ainda mais dispersão, podendo depois ser testada a necessidade de tal acto.

Para o estudo de qualquer um dos modelos referidos nos dois parágrafos anteriores, deverão ser consultados os artigos citados pois, a partir deste momento, no que respeita a modelos BN's, o BN1 e o BN2 merecerão tratamento exclusivo.

2.3. Estimação

A estimação dos parâmetros dos modelos BN's faz-se também através da MV. Tal como se procedeu em relação ao modelo de Poisson, apresentam-se nesta secção as expressões relativas à função de log-verosimilhança, ao vector score e aos elementos da matriz hessiana de cada um daqueles modelos. Embora seja quase certo que a globalidade do *software* econométrico existente necessite apenas da primeira daquelas expressões para poder determinar os estimadores da MV, resolveu-se formalizar por completo este processo de estimação de modo a facilitar, no capítulo II, a dedução de testes score considerando como alternativa ao modelo de Poisson, o BN1 ou o BN2. Além disso, reúne-se aqui um conjunto de expressões não disponíveis, julga-se, noutro local, as quais poderão servir de apoio a futuros trabalhos nesta área.

As funções de log-verosimilhança das duas variantes do modelo BN são as

⁵ Uma sugestão semelhante, mas não devidamente formalizada, surgiu sensivelmente ao mesmo tempo em Ruser (1991).

⁶ No primeiro daqueles dois artigos, estes autores sugeriam um modelo ainda mais vasto, onde também era possível a existência de subdispersão. O BN k era apenas um caso particular.

seguintes⁷:

$$\text{BN1: } LL(\beta, \alpha) = \sum_{i=1}^n \left[-\ln(y_i!) + \sum_{j=1}^{y_i} \ln(\alpha y_i + \mu_i - \alpha j) - \left(\frac{\mu_i}{\alpha} + y_i \right) \ln(1 + \alpha) \right] \quad (\text{I.37})$$

$$\text{BN2: } LL(\beta, \alpha) = \sum_{i=1}^n \left[-\ln(y_i!) + \sum_{j=1}^{y_i} \ln(\alpha y_i + 1 - \alpha j) + y_i \ln(\mu_i) - \left(\frac{1}{\alpha} + y_i \right) \ln(1 + \alpha \mu_i) \right] \quad (\text{I.38})$$

O modelo BN2 pertence à FEL⁸, com $A(\mu_i) = -\frac{1}{\alpha} \ln(1 + \alpha \mu_i)$, $B(y_i) = \ln \left[\frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \right]$ e $C(\mu_i) = \ln \left(\frac{\alpha \mu_i}{1 + \alpha \mu_i} \right)$. Assim, o vector de parâmetros

incluído na forma funcional do valor esperado condicional, β , pode ser estimado consistentemente⁹ por MV, seja ou não o BN2 o verdadeiro modelo. Talvez seja por este motivo que o BN2 é um modelo mais “popular” que o BN1, encontrando-se bastantes mais referências a ele: ver GMT (1984b), Collings e Margolin (1985), Dean e Lawless (1989) e Hellerstein (1991), por exemplo, que omitem a existência do BN1.

Além dos parâmetros de interesse β , a função de log-verosimilhança I.38 contém um parâmetro perturbador, α , o qual, se se pretender realizar inferências probabilísticas, deve ser igualmente estimado por MV. Por isso, vai-se também dedicar alguma atenção à sua estimação, mas realça-se novamente que quando se estimam, em simultâneo, os dois tipos de parâmetros, β e α , se a distribuição real não for a BN2, poderão ser obtidos estimadores inconsistentes [GMT(1984b)], pois uma das condições suficientes para a sua consistência [GMT(1984a)] não é cumprida: o BN2 não pertence à família

⁷ Devido à necessária derivação posterior da função gama, decidiu-se recorrer à igualdade $\frac{\Gamma(y+a)}{\Gamma(a)} = \frac{(y+a-1)!}{(a-1)!} = \prod_{j=1}^y (y+a-j)$, considerando-se que, para $y=0$, $\prod_{j=1}^0 (\cdot) = 1$. Como a maior parte

dos packages econométricos contempla o cálculo automático da derivada da função gama, seria perfeitamente admissível representar esta simplesmente por $\Gamma'(\cdot)$. Contudo, tendo em vista a dedução de testes score no próximo capítulo, achou-se melhor agir desta forma.

⁸ Só o modelo BN2 pertence a esta família porque, para tal, é condição necessária que o parâmetro v_i em I.31 seja fixo, não dependendo de μ_i [o que equivale, na perspectiva de GMT (1984b), a admitir que o erro de especificação é homocedástico], o que não acontece no modelo BN1.

⁹ Relembre-se que a exceção é $\hat{\beta}_0$, o estimador do parâmetro associado à constante.

exponencial quadrática.

Quanto ao modelo BN1, a aplicação da MV só produzirá estimadores garantidamente consistentes para β se ele for, efectivamente, o verdadeiro modelo. Todavia, se a finalidade de um dado estudo for apenas a estimação consistente do valor esperado condicional, há métodos que asseguram o cumprimento desse objectivo. Entre outros, pode-se recorrer ao método dos momentos generalizados [Hansen (1982)], que tem como único pressuposto a correcta especificação dos momentos que se querem estimar. Este processo pode ser empregue em qualquer modelo de contagem que os estimadores dos parâmetros contidos nas condições de momentos, desde que estas estejam bem formalizadas, serão consistentes e assintoticamente normais. No entanto, em nenhuma circunstância o modelo estimado poderá servir de base ao cálculo de probabilidades.

As condições de primeira ordem que permitem obter os estimadores que maximizam $LL(\beta, \alpha)$ são:

$$\text{BN1:} \begin{cases} \frac{\partial LL(\beta, \alpha)}{\partial \beta} = \sum_{i=1}^n x_i' \mu_i \left[\sum_{j=1}^{y_i} \frac{1}{\alpha y_i + \mu_i - \alpha j} - \frac{\ln(1 + \alpha)}{\alpha} \right] = 0 & \text{(I.39)} \\ \frac{\partial LL(\beta, \alpha)}{\partial \alpha} = \sum_{i=1}^n \left\{ \sum_{j=1}^{y_i} \frac{y_i - j}{\alpha y_i + \mu_i - \alpha j} + \frac{1}{\alpha} \left[\frac{\mu_i \ln(1 + \alpha)}{\alpha} - \frac{\mu_i + \alpha y_i}{1 + \alpha} \right] \right\} = 0 & \text{(I.40)} \end{cases}$$

$$\text{BN2:} \begin{cases} \frac{\partial LL(\beta, \alpha)}{\partial \beta} = \sum_{i=1}^n \left(x_i' \frac{y_i - \mu_i}{1 + \alpha \mu_i} \right) = 0 & \text{(I.41)} \\ \frac{\partial LL(\beta, \alpha)}{\partial \alpha} = \sum_{i=1}^n \left\{ \sum_{j=1}^{y_i} \frac{y_i - j}{\alpha y_i + 1 - \alpha j} + \frac{1}{\alpha} \left[\frac{\ln(1 + \alpha \mu_i)}{\alpha} - \frac{\mu_i (1 + \alpha y_i)}{1 + \alpha \mu_i} \right] \right\} = 0 & \text{(I.42)} \end{cases}$$

Repare-se que I.41 representa uma condição de ortogonalidade entre os regressores e os erros ponderados $u_i = p_i [y_i - E(Y_i | x_i)]$, sendo $p_i = (1 + \alpha \mu_i)^{-1}$.

Designando por ρ o vector $(m+1) \times 1$ contendo os parâmetros $\begin{pmatrix} \beta \\ \alpha \end{pmatrix}$, prova-se que a sua distribuição é assintoticamente normal,

$$\sqrt{n}(\hat{\rho} - \rho) \simeq N[0, nI^{-1}(\rho)], \quad \text{(I.43)}$$

$$\text{sendo } I(\rho) = -E \begin{bmatrix} \frac{\partial^2 LL(\beta, \alpha)}{\partial \beta \partial \beta'} & \frac{\partial^2 LL(\beta, \alpha)}{\partial \beta \partial \alpha} \\ \frac{\partial^2 LL(\beta, \alpha)}{\partial \alpha \partial \beta'} & \frac{\partial^2 LL(\beta, \alpha)}{\partial \alpha^2} \end{bmatrix} = \begin{bmatrix} I_{\beta\beta}(\beta, \alpha) & I_{\beta\alpha}(\beta, \alpha) \\ I_{\alpha\beta}(\beta, \alpha) & I_{\alpha\alpha}(\beta, \alpha) \end{bmatrix}.$$

Para o modelo BN1:

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \beta \partial \beta'} = \sum_{i=1}^n x_i' x_i \mu_i \left\{ \sum_{j=1}^{y_i} \frac{\alpha(y_i - j)}{[\alpha y_i + \mu_i - \alpha j]^2} - \frac{\ln(1 + \alpha)}{\alpha} \right\} \quad (I.44)$$

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \beta \partial \alpha} = \left[\frac{\partial^2 LL(\beta, \alpha)}{\partial \alpha \partial \beta} \right]' = \sum_{i=1}^n x_i' \mu_i \left[\frac{\ln(1 + \alpha)}{\alpha^2} - \sum_{j=1}^{y_i} \frac{y_i - j}{(\alpha y_i + \mu_i - \alpha j)^2} - \frac{1}{\alpha(1 + \alpha)} \right] \quad (I.45)$$

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \alpha^2} = \sum_{i=1}^n \left\{ - \sum_{j=1}^{y_i} \left(\frac{y_i - j}{\alpha y_i + \mu_i - \alpha j} \right)^2 + \frac{1}{\alpha^2} \left[\frac{2\mu_i + 3\alpha\mu_i + \alpha^2 y_i}{(1 + \alpha)^2} - \frac{2\mu_i \ln(1 + \alpha)}{\alpha} \right] \right\} \quad (I.46)$$

Para o modelo BN2:

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \beta \partial \beta'} = - \sum_{i=1}^n x_i' x_i \mu_i \frac{1 + \alpha y_i}{(1 + \alpha \mu_i)^2} \quad (I.47)$$

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \beta \partial \alpha} = \left[\frac{\partial^2 LL(\beta, \alpha)}{\partial \alpha \partial \beta} \right]' = - \sum_{i=1}^n x_i' \mu_i \frac{y_i - \mu_i}{(1 + \alpha \mu_i)^2} \quad (I.48)$$

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \alpha^2} = \sum_{i=1}^n \left\{ - \sum_{j=1}^{y_i} \left(\frac{y_i - j}{\alpha y_i + 1 - \alpha j} \right)^2 + \frac{1}{\alpha^2} \left[\frac{2\mu_i + 3\alpha\mu_i^2 + \alpha^2 \mu_i^2 y_i}{(1 + \alpha \mu_i)^2} - \frac{2 \ln(1 + \alpha \mu_i)}{\alpha} \right] \right\} \quad (I.49)$$

Calcular os valores esperados destas derivadas envolveria operações bastante complexas, principalmente no caso do modelo BN1, pelo que não se fará tal tarefa, até porque a estimação da matriz de informação, sempre que necessária, pode ser feita pelo método de Newton.

3. Modelo de Poisson Generalizado

A proposta de utilizar a função de distribuição de Poisson Generalizada (PG), como base de um modelo de regressão para dados de contagem, é muito recente na literatura econométrica. A própria distribuição só foi deduzida nos últimos vinte anos.

Assim, esta secção tem dois grandes objectivos: descrever a distribuição de PG; apresentar, e desenvolver, o novo modelo de regressão para variáveis de contagem.

3.1. Distribuições Lagrangeanas

A derivação de expressões para a função de probabilidade de certas distribuições discretas pode ser efectuada através de expansões lagrangeanas. O uso desta técnica está bem descrito em Consul e Shenton (1972), onde estes autores deduzem várias distribuições novas, determinadas como casos particulares daquela a que deram o nome de família de distribuições lagrangeanas ou generalizadas.

Em termos gerais, o seu método consiste no seguinte: considerando $G(t)$ e $F(t)$ como duas funções geradoras de probabilidade, definidas para o conjunto dos números inteiros não negativos, Consul e Shenton (1972) usam a transformação

$$t = u \cdot G(t) \tag{I.50}$$

e expandem $F(t)$, através da fórmula de Lagrange, em série de potências de u ,

$$F(t) = F(0) + \sum_{y=1}^{\infty} \frac{u^y}{y!} \frac{d^{y-1}}{dt^{y-1}} \left\{ [G(t)]^y \frac{dF(t)}{dt} \right\} \Big|_{t=0}, \tag{I.51}$$

encontrando a família de distribuições citada, cuja função de probabilidade característica é do tipo

$$P(Y = y) = \frac{1}{y!} \frac{d^{y-1}}{dt^{y-1}} \left\{ [G(t)]^y \frac{dF(t)}{dt} \right\} \Big|_{t=0} \quad (I.52)$$

De entre as distribuições lagrangeanas obtidas por eles, dedicar-se-á atenção exclusiva à PG, pois é a que se encontra mais divulgada, já tendo merecido, inclusivamente, a publicação de um livro [Consul (1989)] cujo texto é integralmente dedicado ao seu estudo.

3.2. A Função de Distribuição de Poisson Generalizada

Neste ponto, indicam-se algumas das razões que tornam a PG uma alternativa credível para a modelização de dados de contagem. Procede-se também à sua definição, em termos formais, e refere-se a reparametrização que, habitualmente, é aplicada a esta distribuição.

3.2.1. Motivação

Em princípio, qualquer fenómeno aleatório, para o qual interesse realizar uma contagem de qualquer tipo, é candidato a ser explicado através da distribuição de Poisson *standard*. Naturalmente, nem todas as contagens poderão ser justificadas deste modo. Todavia, se certas premissas acerca dos fenómenos em causa forem cumpridas, o modelo de Poisson será o mais adequado para tal. Um desses pressupostos diz respeito à independência dos acontecimentos em intervalos de tempo sucessivos, isto é, a ocorrência ou não de um acontecimento, num dado período, não influencia a posterior ocorrência (ou não ocorrência) de acontecimentos.

A distribuição de Poisson depende de um único parâmetro, λ , o qual representa tanto a média como a variância. Esta característica da distribuição de Poisson pode não ser muito apropriada quando se estudam determinadas situações, principalmente devido à violação frequente da premissa referida no parágrafo anterior. Com efeito, tem sido observado, em populações supostamente de Poisson, que a probabilidade de um

acontecimento se verificar pode não permanecer constante, variando com o tempo e/ou acontecimentos prévios, daí resultando uma desigualdade entre a média e a variância.

Numerosos investigadores, que efectuaram estudos exaustivos em diversas áreas, tentaram explicar aquela divergência modificando um pouco a distribuição de Poisson, tendo surgido várias derivações dela, podendo as mais relevantes serem encontradas em Johnson, Kotz e Kemp (1992). A maior parte destas distribuições foi desenvolvida para estudar casos concretos, sendo muito pouco flexíveis e dificilmente adaptáveis a outras situações.

Consul e Jain (1973) apresentaram a distribuição PG, a qual, contendo dois parâmetros, λ e θ , proporciona, segundo eles, bons ajustamentos a todo o tipo de dados de contagem, pois a variância desta distribuição é maior, igual ou menor do que a média, conforme θ seja positivo, nulo ou negativo. Nesta distribuição, a probabilidade de ocorrência de um dado acontecimento num determinado período de tempo está dependente dos acontecimentos verificados anteriormente, relação essa que é traduzida pelo parâmetro θ : ele será positivo se aquela probabilidade for aumentando à medida que um indivíduo vai realizando acontecimentos e será negativo se suceder o oposto.

3.2.2. Descrição

A distribuição PG é obtida a partir da expressão I.52, supondo que tanto $G(t)$ como $F(t)$ representam a função geradora de probabilidades da Poisson. Assim, fazendo $G(t) = e^{\theta(t-1)}$ e $F(t) = e^{\lambda(t-1)}$ e substituindo em I.52, resulta:

$$P(Y = y) = \frac{\lambda(\lambda + y\theta)^{y-1} e^{-\lambda - \theta y}}{y!} \quad (I.53)$$

Na última versão apresentada por Consul (1989), esta função de probabilidade é válida apenas para $y = 0, 1, 2, \dots$ ($\theta \geq 0$) \vee $y = 0, 1, 2, \dots, m$ ($\theta < 0$), sendo $\lambda > 0$, $\max\left(-1, -\frac{\lambda}{m}\right) < \theta < 1$ ($m =$ maior inteiro positivo para o qual $\lambda + \theta m > 0$ quando $\theta < 0$) e $m \geq 4$. Como rapidamente se nota, quando $\theta = 0$, a PG coincide com a Poisson.

Tem-se verificado alguma polémica na especificação do intervalo dos valores admissíveis para θ . Inicialmente, Consul e Jain (1973) consideraram $-1 < \theta < 1$ mas, tal como Nelson (1975) salientou, a truncagem efectuada nos valores de x ($x \leq m$), para evitar que a soma $\lambda + m\theta$ se tornasse negativa e originasse o aparecimento de probabilidades alternadamente positivas e negativas, levava a que a soma das probabilidades não fosse igual a um. Assim, Consul e Shoukri (1984) corrigiram os valores possíveis de θ , passando a admitir apenas $0 \leq \theta < 1$. Contudo, isto implicava que a PG deixasse de ser indicada para modelizar dados caracterizados por subdispersão, quando ela já tinha demonstrado um bom ajustamento a alguns conjuntos de observações deste tipo. Deste modo, os mesmos autores propuseram, em artigo de 1985, uma outra modificação.

Primeiro, fizeram a correcção habitual em modelos truncados, $P(Y = y | Y \leq m) = \frac{P(Y = y)}{F(m)}$, sendo $\theta < 0$, $P(Y = y)$ dada por I.53 e $F(m) = \sum_{y=0}^m P(Y = y)$. A soma das

probabilidades passou a ser um, mas o modelo revelou-se intratável em termos matemáticos e, por isso, de pouca utilidade.

De seguida, através de um estudo de simulação, compararam as probabilidades obtidas aplicando a nova função, com as fornecidas pela expressão sem o factor de correcção $F(m)$ (I.53), chegando à conclusão de que o erro resultante, inferior a 0.05%, era negligenciável, desde que o número de acontecimentos distintos com probabilidades não nulas fosse pelo menos de cinco (daí a imposição arbitrária da restrição $m \geq 4$) ou que λ estivesse fora do intervalo $[0.7 ; 4.5]$, aproximadamente. Assim, sugeriram a utilização do modelo sem factor correctivo quando $m \geq 4$, devendo-se, nos restantes casos, aplicar o modelo corrigido, apesar da dificuldade de determinar por este meio estimativas razoáveis para λ e θ .

Por último, Consul (1989) propôs a restrição final: $\max\left(-1, -\frac{\lambda}{m}\right) < \theta < 1$.

As principais propriedades da distribuição PG encontram-se demonstradas em Consul (1989). Este autor prova que:

$$E(X) = \frac{\lambda}{1 - \theta} \quad (I.54)$$

$$V(X) = \frac{\lambda}{(1-\theta)^3} \quad (I.55)$$

Da análise destas duas expressões, podem ser comprovadas algumas das afirmações já realizadas: $\theta > 0 \Rightarrow V(X) > E(X)$; $\theta = 0 \Rightarrow V(X) = E(X)$ (distribuição de Poisson); $\theta < 0 \Rightarrow V(X) < E(X)$.

Quando $\theta = 0$, $V(X) = E(X) = \lambda$. Portanto, λ continua a representar o valor esperado da distribuição de Poisson. Quando θ é positivo, o seu aumento implica um acréscimo tanto na média como na variância, mas com esta última subindo mais rapidamente. O modelo generalizado torna-se, assim, apropriado para representar fenómenos de sobredispersão. Raciocínio oposto pode ser efectuado para θ negativo.

Em muitas situações, os parâmetros da distribuição PG vêm definidos um em função do outro. Consul (1989) sugeriu que, como em muitos problemas práticos se sabe à partida que θ é proporcional a λ , uma parametrização adequada seria $\theta = \alpha\lambda$. Fazendo esta substituição na expressão dada anteriormente para a distribuição PG (I.53), obtém-se a função de probabilidade

$$P(Y = y) = \frac{\lambda^y (1 + y\alpha)^{y-1} e^{-\lambda - \alpha\lambda y}}{y!}, \quad (I.56)$$

sendo $y = 0, 1, 2, \dots$ ($\alpha \geq 0$) \vee $y = 0, 1, \dots, m$ ($\alpha < 0$), $\lambda > 0$, $\max\left(-\frac{1}{\lambda}, -\frac{1}{m}\right) < \alpha < \frac{1}{\lambda}$ e $m \geq 4$.

A esta representação Consul (1989) deu o nome de distribuição PG restrita. As conclusões previamente tiradas de acordo com o valor de θ podem ser agora realizadas em termos de α : se $\alpha = 0$, esta distribuição reduz-se à de Poisson; se $\alpha > 0$, existirá sobredispersão; se $\alpha < 0$, haverá subdispersão.

Ao longo deste estudo, vão-se admitir como possíveis para θ (α), apenas os valores compreendidos no intervalo $[0;1)$. Com efeito, apenas nesta situação é que a PG é uma verdadeira função de distribuição, com todas as propriedades usuais. Além disso, com correcção para a truncagem ou sem ela, as probabilidades calculadas seriam sempre um pouco questionáveis para $\theta < 0$. Por outro lado, pretende-se construir a partir desta

distribuição um modelo de regressão apto para explicar fenómenos económicos, os quais, na maior parte das vezes, são caracterizados pela existência de sobredispersão. Aliás, num contexto de regressão, o facto dos parâmetros passarem a depender de um conjunto de variáveis explicativas, poderia inclusivamente originar a violação da restrição $\max\left(-1, -\frac{\lambda}{m}\right) < \theta$. Por todas estas razões, apesar da aplicação da PG a casos práticos com características de subdispersão ter revelado um bom ajustamento aos dados, a partir do próximo ponto, em que se começa a desenvolver o modelo de regressão, sempre que se falar na PG, estar-se-á a referir a função de probabilidade representada por I.53 mas sendo $y_i = 0, 1, 2, \dots, \lambda > 0$ e $0 \leq \theta < 1$.

3.3. O Modelo de Regressão

Descrita a distribuição PG, segue-se a exposição do modelo de regressão que se baseia nela.

3.3.1. Apresentação

Para se obter o modelo de regressão, basta impor que pelo menos um dos parâmetros da PG seja função de um conjunto de regressores. Pode-se fazer com que apenas um dos parâmetros, λ , dependa dos regressores ou permitir que θ também seja explicado por variáveis exógenas. Independentemente da opção tomada, é possível parametrizar o modelo de modo a que o valor esperado condicional passe a ser simplesmente $\mu_i = \exp(x_i\beta)$, à semelhança do que acontece em todos os modelos de contagem tratados até ao momento.

Na primeira situação, a reparametrização a efectuar, resultante de I.54, será

$$\lambda_i = \mu_i(1-\theta) \tag{I.57}$$

Foi este o procedimento adoptado por Consul e Famoye (1992), cujo modelo passará a ser designado por PG1.

No segundo caso, descrito por Famoye (1993), deve-se partir do modelo restrito e estabelecer

$$\lambda_i = \frac{\mu_i}{1 + \alpha\mu_i}, \quad (\text{I.58})$$

uma vez que $\theta_i = \alpha\lambda_i$. Este será o modelo PG2. De notar que a exigência, no modelo restrito, de $\alpha < \frac{1}{\lambda}$, não afecta o modelo de regressão, pois essa restrição é redundante, dado que, com a parametrização escolhida, verifica-se $\alpha < \alpha + \frac{1}{\mu_i}$, condição universal.

Assim, no modelo PG1, $Y_i | x_i \sim \text{PG}[\mu_i(1-\theta), \theta]$ ou, definindo $\varphi = \frac{1}{1-\theta}$, de modo a simplificar um pouco as expressões que se apresentarão no ponto 3.3.2., $Y_i | x_i \sim \text{PG}\left(\frac{\mu_i}{\varphi}, \frac{\varphi-1}{\varphi}\right)$, a que corresponde a função de probabilidade:

$$P(Y_i = y_i | x_i) = \frac{\mu_i [\mu_i + (\varphi - 1)y_i]^{y_i-1} \varphi^{-y_i} e^{-\frac{\mu_i}{\varphi}}}{y_i!} \quad (\text{I.59})$$

No modelo PG2, $Y_i | x_i \sim \text{PG}\left(\frac{\mu_i}{1 + \alpha\mu_i}, \frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)$, pelo que:

$$P(Y_i = y_i | x_i) = \frac{\left(\frac{\mu_i}{1 + \alpha\mu_i}\right)^{y_i} (1 + \alpha y_i)^{y_i-1} e^{-\frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}}}{y_i!} \quad (\text{I.60})$$

Em ambos os modelos, $E(Y_i | x_i) = \mu_i$, dependendo a opção por um ou outro do mesmo factor decisivo para a escolha do BN1 ou do BN2: a adequação da função variância aos dados observados. Com efeito, também os modelos PG1 e PG2 levam a diferentes padrões de heterocedasticidade.

No PG1, a variância é dada por

$$V(Y_i|x_i) = \frac{\mu_i}{(1-\theta)^2} = \varphi^2 \mu_i, \quad (I.61)$$

havendo, portanto, uma relação linear entre a média e a variância.

No PG2, a variância e a média verificam uma relação cúbica:

$$V(Y_i|x_i) = \mu_i(1 + \alpha\mu_i)^2 \quad (I.62)$$

Os dois modelos incluem o de Poisson como caso particular, o PG1 para $\varphi = 1$ ($\theta = 0$), o segundo para $\alpha = 0$.

3.3.2. Estimação

Como a distribuição PG satisfaz as condições de regularidade necessárias à aplicação da MV¹⁰, mais uma vez deverá ser este o método empregue na estimação dos parâmetros de interesse, pois os estimadores obtidos serão assintoticamente consistentes, eficientes e normalmente distribuídos, desde que, é claro, a verdadeira distribuição seja realmente a PG. Caso não o seja, mas o valor esperado condicional esteja bem especificado, os estimadores $\hat{\beta}$ da MV, referentes ao modelo PG2, embora ineficientes, ainda serão consistentes e terão distribuição assintótica normal, pois a distribuição PG2 pertence à FEL¹¹, ao contrário do que acontece com a PG1. Neste último modelo, existindo dúvidas sobre a sua adequabilidade, a estimação consistente de β só será assegurada recorrendo ao método dos momentos generalizados.

As funções de log-verossimilhança correspondentes a cada um dos modelos em estudo são as seguintes:

$$\text{PG1: } LL(\beta, \varphi) = \sum_{i=1}^n \left\{ \ln(\mu_i) + (y_i - 1) \ln[\mu_i + (\varphi - 1)y_i] - \right.$$

¹⁰ Consul e Famoye (1988) provaram que também para valores negativos se poderia aplicar a teoria *standard* da verossimilhança.

¹¹ Com $A(\mu_i) = -\frac{\mu_i}{1 + \alpha\mu_i}$, $B(y_i) = (y_i - 1) \ln(1 + \alpha y_i) - \ln(y_i!)$ e $C(\mu_i) = \ln\left(\frac{\mu_i}{1 + \alpha\mu_i}\right) - \frac{\alpha\mu_i}{1 + \alpha\mu_i}$.

$$-y_i \ln(\varphi) - \frac{\mu_i}{\varphi} - \frac{\varphi - 1}{\varphi} y_i - \ln(y_i!) \Big\} \quad (\text{I.63})$$

$$\text{PG2: LL}(\beta, \alpha) = \sum_{i=1}^n \left\{ y_i \ln(\mu_i) + (y_i - 1) \ln(1 + \alpha y_i) - \right. \\ \left. - y_i \ln(1 + \alpha \mu_i) - \frac{\mu_i(1 + \alpha y_i)}{1 + \alpha \mu_i} - \ln(y_i!) \right\} \quad (\text{I.64})$$

As primeiras derivadas parciais são definidas por:

$$\text{PG1: } \left\{ \begin{aligned} \frac{\partial \text{LL}(\beta, \varphi)}{\partial \beta} &= \sum_{i=1}^n x_i' \mu_i \left[\frac{1}{\mu_i} - \frac{1}{\varphi} + \frac{y_i - 1}{\mu_i + (\varphi - 1)y_i} \right] \end{aligned} \right. \quad (\text{I.65})$$

$$\left\{ \begin{aligned} \frac{\partial \text{LL}(\beta, \varphi)}{\partial \varphi} &= \sum_{i=1}^n \left[\frac{y_i(y_i - 1)}{\mu_i + (\varphi - 1)y_i} - \frac{y_i - \mu_i}{\varphi^2} - \frac{y_i}{\varphi} \right] \end{aligned} \right. \quad (\text{I.66})$$

$$\text{PG2: } \left\{ \begin{aligned} \frac{\partial \text{LL}(\beta, \alpha)}{\partial \beta} &= \sum_{i=1}^n x_i' \frac{y_i - \mu_i}{(1 + \alpha \mu_i)^2} \end{aligned} \right. \quad (\text{I.67})$$

$$\left\{ \begin{aligned} \frac{\partial \text{LL}(\beta, \alpha)}{\partial \alpha} &= \sum_{i=1}^n \left[-\frac{\mu_i y_i}{1 + \alpha \mu_i} + \frac{y_i(y_i - 1)}{1 + \alpha y_i} - \frac{\mu_i(y_i - \mu_i)}{(1 + \alpha \mu_i)^2} \right] \end{aligned} \right. \quad (\text{I.68})$$

Naturalmente, igualando a zero estas derivadas e resolvendo o sistema, serão obtidos os estimadores da MV. A sua matriz de covariâncias assintótica será dada, para o PG1, pelo inverso da matriz de informação (para o PG2, basta substituir φ por α):

$$I(\beta, \varphi) = -E \begin{bmatrix} \frac{\partial^2 \text{LL}(\beta, \varphi)}{\partial \beta \partial \beta'} & \frac{\partial^2 \text{LL}(\beta, \varphi)}{\partial \beta \partial \varphi} \\ \frac{\partial^2 \text{LL}(\beta, \varphi)}{\partial \varphi \partial \beta'} & \frac{\partial^2 \text{LL}(\beta, \varphi)}{\partial \varphi^2} \end{bmatrix}$$

Para o modelo PG1:

$$\frac{\partial^2 \text{LL}(\beta, \varphi)}{\partial \beta \partial \beta'} = - \sum_{i=1}^n x_i' x_i \mu_i \left\{ \frac{1}{\varphi} - \frac{(y_i - 1)[(\varphi - 1)y_i]}{[\mu_i + (\varphi - 1)y_i]^2} \right\} \quad (\text{I.69})$$

$$\frac{\partial^2 \text{LL}(\beta, \varphi)}{\partial \beta \partial \varphi} = \left[\frac{\partial^2 \text{LL}(\beta, \varphi)}{\partial \varphi \partial \beta} \right]' = \sum_{i=1}^n x_i' \mu_i \left\{ \frac{1}{\varphi^2} - \frac{(y_i - 1)y_i}{[\mu_i + (\varphi - 1)y_i]^2} \right\} \quad (\text{I.70})$$

$$\frac{\partial^2 LL(\beta, \varphi)}{\partial \varphi^2} = \sum_{i=1}^n \left\{ \frac{y_i}{\varphi^2} - \frac{y_i^2(y_i - 1)}{[\mu_i + (\varphi - 1)y_i]^2} + \frac{2(y_i - \mu_i)}{\varphi^3} \right\}, \quad (I.71)$$

Para o modelo PG2:

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \beta \partial \beta'} = - \sum_{i=1}^n x_i' x_i \mu_i \frac{1 + \alpha(2y_i - \mu_i)}{(1 + \alpha\mu_i)^3} \quad (I.72)$$

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \beta \partial \alpha} = \left[\frac{\partial^2 LL(\beta, \alpha)}{\partial \alpha \partial \beta} \right]' = - \sum_{i=1}^n 2x_i' \mu_i \frac{y_i - \mu_i}{(1 + \alpha\mu_i)^3} \quad (I.73)$$

$$\frac{\partial^2 LL(\beta, \alpha)}{\partial \alpha^2} = \sum_{i=1}^n \left[\frac{\mu_i^2 y_i}{(1 + \alpha\mu_i)^2} - \frac{y_i^2(y_i - 1)}{(1 + \alpha y_i)^2} + \frac{2\mu_i^2(y_i - \mu_i)}{(1 + \alpha\mu_i)^3} \right] \quad (I.74)$$

Tal como acontecia com o BN2, também a matriz de informação do modelo PG2 é diagonal por blocos.

3.3.3. Abordagem de uma nova parametrização

Uma das condições para o êxito de um modelo de contagem é a correcta especificação da distribuição que os dados seguem. Assim, é muito importante conseguir arranjar uma parametrização o mais geral possível, que permita adaptar uma determinada distribuição às características de um número elevado de conjuntos de dados.

O modelo PG2, ao pôr os dois parâmetros em função dos regressores, já é algo flexível. Contudo, se se colocar também o parâmetro α a depender das variáveis explicativas, conseguir-se-á obter um modelo ainda mais geral. É esta a proposta feita em Santos Silva (1996a)¹². Pegando novamente no modelo restrito e continuando a adoptar para λ a parametrização I.58, aquele autor definiu

$$\theta_i = \alpha_i \lambda_i = \frac{\alpha_i \mu_i}{1 + \alpha_i \mu_i}, \quad (I.75)$$

¹² Para o modelo BN2, uma sugestão semelhante encontra-se em Santos Silva e Andrade e Silva (1994).

onde $\alpha_i = \alpha(x_i, \delta)$, sendo δ um vector $m \times 1$ de parâmetros desconhecidos e $\alpha(x_i, \delta)$ uma função dos regressores (os mesmos de que depende μ_i) tal que $\alpha(x_i, \delta) > \max\left(-\frac{1}{2\mu_i}, -\frac{1}{m}\right)$. Adicionalmente, considera-se que $\alpha(x_i, \delta) > 0$, uma vez que o interesse deste trabalho centra-se apenas na modelização de observações caracterizadas pela existência de sobredispersão. Por isso, a formalização mais indicada para α_i será também do tipo exponencial:

$$\alpha_i = \exp(x_i \delta) \quad (I.76)$$

O modelo resultante, doravante denominado por PG3, inclui como casos particulares algumas das formulações abordadas anteriormente: se $\alpha_i = 0$, ele coincidirá com o modelo de Poisson; quando α_i for constante, não dependendo dos regressores, ter-se-á o PG2 (θ_i será proporcional a λ_i); se α_i for proporcional a μ_i^{-1} (θ constante), cair-se-á no PG1, pois λ e θ serão independentes. Ao englobar os modelos PG1 e PG2, o PG3 possibilitará, no próximo capítulo, a construção de testes de hipóteses não encaixadas com o objectivo de os testar, um “contra” o outro.

No modelo PG3, enquanto que o valor esperado condicional continua a ser μ_i , a variância passa a ser

$$V(Y_i | x_i) = \mu_i [1 + \alpha(x_i, \delta) \mu_i]^2, \quad (I.77)$$

permitindo a aderência deste modelo a conjuntos de dados caracterizados por padrões de heterocedasticidade muito diversos.

A estimação deste modelo envolve expressões muito semelhantes às deduzidas para o PG2. Assim, substituindo-se α por $\alpha(x_i, \delta)$, a função de log-verosimilhança e a primeira derivada parcial em ordem a β corresponderão, respectivamente, a I.64 e a I.67. Para se obter a derivada em ordem a δ , além daquela alteração, ter-se-á de pré-multiplicar a expressão I.68 por $\left[\frac{\partial \alpha(x_i, \delta)}{\partial \delta} \right]$. Quanto aos elementos da matriz hessiana,

eles serão dados por:

$$\frac{\partial^2 LL(\beta, \delta)}{\partial \beta \partial \beta'} = - \sum_{i=1}^n x_i' x_i \mu_i \frac{1 + \alpha_i (2y_i - \mu_i)}{(1 + \alpha_i \mu_i)^3} \quad (I.78)$$

$$\frac{\partial^2 LL(\beta, \delta)}{\partial \beta \partial \delta} = \left[\frac{\partial^2 LL(\beta, \delta)}{\partial \delta \partial \beta} \right]' = - \sum_{i=1}^n \left[2x_i' \mu_i \frac{y_i - \mu_i}{(1 + \alpha_i \mu_i)^3} \left(\frac{\partial \alpha_i}{\partial \delta} \right) \right] \quad (I.79)$$

$$\begin{aligned} \frac{\partial^2 LL(\beta, \delta)}{\partial \delta \partial \delta} = \sum_{i=1}^n \left\{ \left(\frac{\partial \alpha_i}{\partial \delta} \right)' \left(\frac{\partial \alpha_i}{\partial \delta} \right) \left[\frac{\mu_i^2 y_i}{(1 + \alpha_i \mu_i)^2} - \frac{y_i^2 (y_i - 1)}{(1 + \alpha_i y_i)^2} + \frac{2\mu_i^2 (y_i - \mu_i)}{(1 + \alpha_i \mu_i)^3} \right] + \right. \\ \left. + \frac{\partial^2 \alpha_i}{\partial \delta \partial \delta} \left[- \frac{\mu_i y_i}{1 + \alpha_i \mu_i} + \frac{y_i (y_i - 1)}{1 + \alpha_i y_i} - \frac{\mu_i (y_i - \mu_i)}{(1 + \alpha_i \mu_i)^2} \right] \right\} \quad (I.80) \end{aligned}$$

Adoptando a especificação $\alpha_i = \exp(x_i \delta)$, nas expressões anteriores $\frac{\partial \alpha_i}{\partial \delta}$ deve ser substituído por $x_i \alpha_i$ e $\frac{\partial^2 \alpha_i}{\partial \delta \partial \delta}$ por $x_i' x_i \alpha_i$.

Apesar das similitudes entre as estimações dos modelos PG2 e PG3, este último é bem mais difícil de maximizar, uma vez que contém quase o dobro de parâmetros (2m+1 contra m+1).

4. Modelos para Amostras Truncadas em Zero

A finalidade desta secção é a extensão dos modelos de regressão apresentados até aqui, a situações em que os zeros da amostra não são observados. Depois de, a título introdutório, se referirem algumas generalidades sobre distribuições truncadas, descrevem-se os dois modelos truncados habitualmente tratados na literatura econométrica [Creel e Loomis (1990), Grogger e Carson (1991), Gurmu (1991), Gurmu e Trivedi (1992) e Santos Silva e Andrade e Silva (1994)], o Poisson Truncado (PT) e o Binomial Negativo Truncado (BNT). De seguida, analisa-se o modelo Poisson Generalizado Truncado (PGT), até ao momento abordado apenas em Santos Silva (1996a).

4.1. Distribuições Truncadas

Quando se modeliza um conjunto de dados, muitas vezes depara-se com um problema relacionado com a forma de recolha da amostra, que impede a observação de todos os valores possíveis de encontrar na população. Nestes casos, o modelo a ser estimado deve ser corrigido, pois só assim se poderá inferir correctamente sobre a população em estudo.

A função de probabilidade de um modelo truncado é definida por

$$P(Y = y|a < Y \leq b) = \frac{P(Y = y)}{F(b) - F(a)}, \quad a < y < b, \quad (\text{I.81})$$

onde $F(\cdot) = \sum_{y=0}^{\cdot} P(Y = y)$.

Esta função pressupõe a existência de dupla truncagem. Contudo, a ocorrência prática mais comum é a não observação apenas dos zeros da amostra, isto é, somente se tem acesso aos valores positivos de y . Por isso, estudar-se-á unicamente este caso.

Nesta situação,

$$P(Y = y|Y > 0) = \frac{P(Y = y)}{1 - P(Y = 0)}, \quad y = 1, 2, \dots \quad (\text{I.82})$$

podendo os dois momentos principais ser formalizados do seguinte modo:

$$E(Y|Y > 0) = \frac{E(Y)}{1 - P(Y = 0)} \quad (\text{I.83})$$

$$V(Y|Y > 0) = \frac{1}{1 - P(Y = 0)} \left[V(Y) - \frac{E^2(Y)}{[P(Y = 0)]^{-1} - 1} \right] \quad (\text{I.84})$$

Como é evidente, para dados truncados em zero, o valor esperado condicional aumenta e a variância condicional diminui. Assim, para o primeiro, pode ser deduzida, entre os valores esperados das distribuições truncada e não truncada, uma relação do

tipo:

$$E(Y | Y > 0) = E(Y) + \eta, \quad \eta > 0 \quad (\text{I.85})$$

Conjugando com a expressão I.83, resulta:

$$\eta = \frac{E(Y)}{[P(Y = 0)]^{-1} - 1} \quad (\text{I.86})$$

O parâmetro η , denominado por Gurmu (1991) de factor de ajustamento de truncagem, desempenha um papel importante nos testes de sobredispersão derivados por aquele autor para distribuições truncadas, os quais serão expostos no capítulo II.

Tendo em conta I.84 e I.86, pode-se também estabelecer uma relação semelhante para a variância,

$$V(Y|Y > 0) = V(Y) - \eta \left[E(Y|Y > 0) - \frac{V(Y)}{E(Y)} \right], \quad (\text{I.87})$$

continuando η a ser definido por I.86. Também esta expressão terá a sua utilidade mais à frente.

Voltando a analisar I.83 e I.84, verifica-se que, se se quiser realizar inferências para o total da população, e não apenas para a observável, bastará resolver aquelas identidades em ordem a $E(Y)$ e a $V(Y)$, respectivamente.

Pode ser feita ainda outra constatação, bastante relevante: tanto o valor esperado como a variância condicionais, ao serem função de $P(Y = 0)$, dependem da correcta especificação da distribuição que os dados seguem. Em termos de validade do modelo de regressão, isto tem implicações muito importantes. Com efeito, referiu-se anteriormente que, desde que o modelo em causa pertencesse à FEL, era apenas necessário que o valor esperado condicional estivesse bem formalizado para os estimadores obtidos serem consistentes. O que acontece agora é que essa especificação está dependente de se estar ou não a considerar a verdadeira distribuição dos dados.

Assim, quando a amostra é truncada, um cuidado adicional deve ser colocado na escolha da distribuição que se pensa descrever apropriadamente os dados, pois o facto dela pertencer à FEL já não constitui garantia para a consistência dos estimadores. O próprio método dos momentos generalizados deixa de assegurar essa consistência.

Para além da não observação dos zeros da amostra, há outro motivo que leva à utilização de modelos truncados em zero, mesmo quando estes estão disponíveis. Isso acontece quando há “excesso de zeros”, isto é, quando o econometrista julga que parte dos zeros (ou a totalidade) da população é gerada por um mecanismo distinto do seguido pelos outros valores. Nestas circunstâncias, quando tal processo não é conhecido, pode optar-se por desprezar os valores referidos e empregar um modelo truncado em zero.

Alternativamente, poderão ser usados os modelos ZIP¹³ ou os *hurdle models*¹⁴. Os primeiros supõem a existência de dois tipos de zeros: uma parcela resultante do comportamento determinístico dos indivíduos; a outra explicada pela distribuição de Poisson. Os *hurdle models* dividem a escolha do indivíduo em duas fases: primeiro, cada indivíduo decide se realiza ou não um determinado acontecimento; depois, caso opte positivamente, escolhe o número de ocorrências a efectuar. A primeira etapa deste processo de decisão pode ser descrita através de um modelo binário; a segunda, recorrendo a uma distribuição truncada em zero. Nenhum destes dois modelos será desenvolvido neste trabalho.

Quando, além do problema da truncagem dos dados, se enfrenta o facto da amostra ter sido estratificada endogenamente (*on-site samples*), torna-se necessário fazer uma correcção adicional à função de probabilidade escolhida para justificar o comportamento dos indivíduos. Também este tipo de modelo não será aqui tratado, devendo-se consultar, para o seu estudo, os artigos de Shaw (1988), relativo à distribuição de Poisson, e de Englin e Shonkwiler (1995), referente à BN.

4.2. Modelo de Poisson Truncado

A função de probabilidade do modelo PT é definida por

¹³ *Zero Inflated Poisson*. Ver Lambert (1992).

¹⁴ Ver Mullahy (1986) e Winkelmann e Zimmermann (1995).

$$P(Y_i = y_i | x_i, Y_i > 0) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i! (1 - e^{-\lambda_i})}, \quad y_i = 1, 2, \dots \quad (I.88)$$

pelo que os momentos principais são:

$$E(Y_i | x_i, Y_i > 0) = \lambda_i^* = \frac{\lambda_i}{1 - e^{-\lambda_i}} \quad (I.89)$$

$$V(Y_i | x_i, Y_i > 0) = \frac{\lambda_i}{1 - e^{-\lambda_i}} \left[1 - \frac{\lambda_i}{e^{\lambda_i} - 1} \right] \quad (I.90)$$

É fácil de mostrar que a relação entre o valor esperado e a variância condicionais é a seguinte:

$$V(Y_i | x_i, Y_i > 0) = E(Y_i | x_i, Y_i > 0) \left[1 - P(Y_i = 0 | x_i) E(Y_i | x_i, Y_i > 0) \right] \quad (I.91)$$

Repare-se que, com truncagem, a variância condicional passa a ser inferior à média condicional. Deste modo, também quando os zeros da amostra não são observados, o modelo de Poisson não é apropriado para descrever dados demasiado heterogéneos.

Com base nas expressões I.85 a I.87 e I.89 a I.90, pode-se ainda afirmar que

$$E(Y_i | x_i, Y_i > 0) = \lambda_i + \eta_i \quad (I.92)$$

$$V(Y_i | x_i, Y_i > 0) = (1 - \eta_i) E(Y_i | x_i, Y_i > 0), \quad (I.93)$$

sendo $\eta_i = \frac{\lambda_i}{e^{\lambda_i} - 1}$.

Continuando a adoptar para λ_i uma forma exponencial, a função de log-verosimilhança é dada por

$$LL(\beta) = \sum_{i=1}^n \left[-\lambda_i + y_i \ln(\lambda_i) - \ln(y_i!) - \ln(1 - e^{-\lambda_i}) \right], \quad (I.94)$$

de onde resultam as condições de 1ª ordem:

$$\frac{\partial LL(\beta)}{\partial \beta} = \sum_{i=1}^n x_i' (y_i - \lambda_i^*) = 0 \quad (I.95)$$

A exemplo do que acontecia no modelo de Poisson não truncado, a expressão I.95 representa uma condição de ortogonalidade entre os regressores e os erros $u_i = y_i - E(Y_i | x_i, Y_i > 0)$.

A matriz hessiana corresponde a:

$$\frac{\partial^2 LL(\beta)}{\partial \beta \partial \beta'} = - \sum_{i=1}^n x_i' x_i \lambda_i^* (1 - \eta_i) \quad (I.96)$$

Supondo que o modelo descreve de um modo perfeito os dados observados, os estimadores da MV serão assintoticamente consistentes, eficientes e normalmente distribuídos, $\sqrt{n}(\hat{\beta} - \beta) \approx N[0, nI^{-1}(\beta)]$, sendo $I(\beta) = -\frac{\partial^2 LL(\beta)}{\partial \beta \partial \beta'}$ pois a matriz hessiana é constante.

Os estimadores obtidos para o modelo PT serão inconsistentes, caso a verdadeira distribuição dos dados não seja Poisson, isto é, se houver sobredispersão. Por outro lado, se se utilizar o modelo de Poisson *standard*, não atendendo ao facto da amostra ser truncada, os estimadores também não serão consistentes, pois o valor esperado condicional estará incorrectamente especificado. Assim, a opção que imediatamente se coloca é a estimação do modelo BNT, que tem em conta, tanto a existência de sobredispersão, como a truncagem dos dados.

4.3. Modelo Binomial Negativo Truncado

Tal como se fez para os não truncados, também aqui se consideram dois modelos principais: o BNT1, resultante da truncagem em zero do BN1, e o BNT2, correspondente ao BN2.

Dado que o seu tratamento individualizado seria muito semelhante e porque as

diferenças fundamentais entre eles já foram evidenciadas na secção dois, os dois modelos vão ser trabalhados em conjunto, usando a parametrização adoptada por Gurmu (1991). Por analogia com o BNk de Winkelmann e Zimmermann (1991, 1995), este modelo será denominado de BNTk, já que constitui a extensão lógica do anterior para amostras truncadas em zero. Tal como o BNk, para dados não truncados, o modelo BNTk servirá de base, no capítulo II, à elaboração de testes de hipóteses não encaixadas que permitirão testar as duas variantes principais dos modelos BNT's, uma "contra" a outra.

Gurmu (1991) apresenta como função de probabilidade do modelo BNTk a expressão

$$P(Y_i = y_i | x_i, Y_i > 0) = \frac{\Gamma\left(y_i + \frac{\mu_i^{-k+1}}{\alpha}\right)}{\Gamma(y_i + 1)\Gamma\left(\frac{\mu_i^{-k+1}}{\alpha}\right)} \left(\frac{1}{1 + \alpha\mu_i^k}\right)^{\frac{\mu_i^{-k+1}}{\alpha}} \times \\ \times \left(\frac{\alpha\mu_i^k}{1 + \alpha\mu_i^k}\right)^{y_i} \left[1 - \left(1 + \alpha\mu_i^k\right)^{-\frac{\mu_i^{-k+1}}{\alpha}}\right]^{-1}, \quad (I.97)$$

onde $\mu_i = \exp(x_i\beta)$, implicando $k = 0$ o modelo BNT1 e $k = 1$ o BNT2.

Os momentos principais são:

$$E(Y_i | x_i, Y_i > 0) = \mu_i^* = \frac{\mu_i}{1 - \left(1 + \alpha\mu_i^k\right)^{-\frac{\mu_i^{-k+1}}{\alpha}}} \quad (I.98)$$

$$V(Y_i | x_i, Y_i > 0) = \frac{\mu_i}{1 - \left(1 + \alpha\mu_i^k\right)^{-\frac{\mu_i^{-k+1}}{\alpha}}} \left[\left(1 + \alpha\mu_i^k\right) - \frac{\mu_i}{\left(1 + \alpha\mu_i^k\right)^{-\frac{\mu_i^{-k+1}}{\alpha}} - 1} \right] \quad (I.99)$$

Estas identidades podem ser reescritas, com base nas expressões I.85 a I.87, de modo a se relacionar directamente a variância e o valor esperado truncados¹⁵:

¹⁵ Chama-se a atenção para o facto da expressão I.101 estar incorrecta em Gurmu (1991), onde aparece sob a referência 1.17.

$$E(Y_i | x_i, Y_i > 0) = \mu_i + \eta_i \quad (I.100)$$

$$V(Y_i | x_i, Y_i > 0) = (1 + \alpha \mu_i^k - \eta_i)E(Y_i | x_i, Y_i > 0), \quad (I.101)$$

sendo $\eta_i = \frac{\mu_i}{(1 + \alpha \mu_i^k)^{\frac{\mu_i^{-k+1}}{\alpha}} - 1}$.

Se $\alpha = 0$, a relação entre a variância e a média condicionais, expressa por I.101, será equivalente à determinada para o modelo PT (I.93). Além disso, prova-se facilmente que, quando α tende para zero, a expressão I.98 tende para a I.89, que representa o valor esperado condicional do modelo PT. Consequentemente, também nos modelos truncados, o PT é um caso particular do BNT.

A função de log-verosimilhança do modelo BNTk é:

$$LL(\beta, \alpha) = \sum_{i=1}^n \left\{ -\ln(y_i!) + \sum_{j=1}^{y_i} \ln(\alpha y_i + \mu_i^{-k+1} - \alpha j) - \left(\frac{\mu_i^{-k+1}}{\alpha} + y_i \right) \ln(1 + \alpha \mu_i^k) + y_i \ln(\mu_i^k) - \ln \left[1 - \left(1 + \alpha \mu_i^k \right)^{\frac{\mu_i^{-k+1}}{\alpha}} \right] \right\} \quad (I.102)$$

Comparando com as funções de log-verosimilhança dos modelos BN1 e BN2 (expressões I.37 e I.38), verifica-se que a diferença reside apenas no termo $\ln \left[1 - \left(1 + \alpha \mu_i^k \right)^{\frac{\mu_i^{-k+1}}{\alpha}} \right]$, isto é, $LL(\beta, \alpha) = LL_0(\beta, \alpha) - \sum_{i=1}^n \ln \left[1 - \left(1 + \alpha \mu_i^k \right)^{\frac{\mu_i^{-k+1}}{\alpha}} \right]$, sendo $LL_0(\beta, \alpha)$ a função de log-verosimilhança do modelo não truncado. Então, para se obterem as condições de primeira ordem e os elementos da matriz hessiana, bastará derivar aquela parcela, em ordem ao(s) parâmetro(s) que interessem, e subtrai-la às expressões anteriormente derivadas para os modelos BN1 ou BN2, conforme a situação. De realçar apenas que o score para o BNT2, respeitante unicamente ao vector de parâmetros β ,

$$\frac{\partial LL(\beta, \alpha)}{\partial \beta} = \sum_{i=1}^n \left(x_i, \frac{y_i - \mu_i^*}{1 + \alpha \mu_i^k} \right), \quad (I.103)$$

representa, mais uma vez, uma condição de ortogonalidade entre as variáveis exógenas e os erros ponderados.

4.4. Modelo de Poisson Generalizado Truncado

Uma vez que os modelos PG1 e PG2 são casos particulares do PG3, somente este último será adaptado ao tipo de amostras em estudo.

De acordo com I.56, I.58 e I.82, a função de probabilidade do modelo PGT3 pode ser definida como

$$P(Y_i = y_i | x_i, Y_i > 0) = \frac{\left(\frac{\mu_i}{1 + \alpha_i \mu_i}\right)^{y_i} (1 + \alpha_i y_i)^{y_i - 1} e^{-\frac{\mu_i(1 + \alpha_i y_i)}{1 + \alpha_i \mu_i}}}{y_i! \left(1 - e^{-\frac{\mu_i}{1 + \alpha_i \mu_i}}\right)}, \quad (I.104)$$

com $y_i = 1, 2, \dots$, $\alpha_i = \alpha(x_i, \delta) > 0$ e $\mu_i = \mu(x_i, \beta)$.

O valor esperado condicional é dado por

$$E(Y_i | x_i, Y_i > 0) = \mu_i^* = \frac{\mu_i}{1 - e^{-\frac{\mu_i}{1 + \alpha_i \mu_i}}} \quad (I.105)$$

e a variância condicional por

$$V(Y_i | x_i, Y_i > 0) = \frac{\mu_i}{1 - e^{-\frac{\mu_i}{1 + \alpha_i \mu_i}}} \left[(1 + \alpha_i \mu_i)^2 - \frac{\mu_i}{e^{-\frac{\mu_i}{1 + \alpha_i \mu_i}} - 1} \right] \quad (I.106)$$

A exemplo do realizado para os outros modelos truncados, as expressões I.105 e I.106 podem ser reescritas como

$$E(Y_i | x_i, Y_i > 0) = \mu_i + \eta_i \quad (I.107)$$

$$V(Y_i | x_i, Y_i > 0) = \left[(1 + \alpha_i \mu_i)^2 - \eta_i \right] E(Y_i | x_i, Y_i > 0), \quad (I.108)$$

sendo $\eta_i = \frac{\mu_i}{e^{\frac{\mu_i}{1+\alpha_i\mu_i}} - 1}$.

Se $\alpha_i = 0$, a relação entre a variância e a média condicionais corresponderá à do modelo PT, se α_i for constante, ter-se-á o modelo PGT2, e se α_i for proporcional a μ_i^{-1} , resultará o modelo PGT1¹⁶.

A função de log-verosimilhança do modelo PGT3 é:

$$LL(\beta, \delta) = \sum_{i=1}^n \left\{ y_i \ln(\mu_i) + (y_i - 1) \ln(1 + \alpha_i y_i) - y_i \ln(1 + \alpha_i \mu_i) - \frac{\mu_i(1 + \alpha_i y_i)}{1 + \alpha_i \mu_i} - \ln(y_i!) - \ln\left(1 - e^{-\frac{\mu_i}{1 + \alpha_i \mu_i}}\right) \right\} \quad (I.109)$$

Como este modelo satisfaz as condições de regularidade inerentes à aplicação da teoria da verosimilhança, os estimadores da MV de β e δ obtêm-se mediante a resolução das condições de primeira ordem:

$$\frac{\partial LL(\beta, \delta)}{\partial \beta} = \sum_{i=1}^n x_i \left[\frac{y_i - \mu_i^*}{(1 + \alpha_i \mu_i)^2} \right] = 0 \quad (I.110)$$

$$\frac{\partial LL(\beta, \delta)}{\partial \delta} = \sum_{i=1}^n x_i \left[\alpha_i \left[-\frac{\mu_i y_i}{1 + \alpha_i \mu_i} + \frac{y_i(y_i - 1)}{1 + \alpha_i y_i} - \frac{\mu_i(y_i - \mu_i)}{(1 + \alpha_i \mu_i)^2} + \frac{\mu_i^2}{(1 + \alpha_i \mu_i)^2 \left(e^{\frac{\mu_i}{1 + \alpha_i \mu_i}} - 1 \right)} \right] \right] = 0 \quad (I.111)$$

Para se determinarem os elementos da matriz hessiana, é apenas necessário derivar duas vezes o factor $\sum_{i=1}^n \ln\left(1 - e^{-\frac{\mu_i}{1 + \alpha_i \mu_i}}\right)$, em ordem aos parâmetros que

interessam, e subtraí-lo às expressões correspondentes (I.78 a I.80) encontradas para o

¹⁶ Em termos rigorosos, considerando $\alpha_i = \delta_0 \exp(x_i \delta_1)$ e $\mu_i = \beta_0 \exp(x_i \beta_1)$, formulações que diferem das anteriormente propostas apenas porque o termo constante é colocado fora da função exponencial, o modelo PGT3 coincidirá com o PGT1 quando $\alpha_i = \delta_0 \beta_0 \mu_i^{-1}$.

modelo PG3.

Capítulo II:

Testes de Especificação

Uma das etapas fundamentais na construção de um modelo econométrico é aquela em que se procede à sua avaliação. O modelo deve ser sujeito a uma série de testes com a finalidade de verificar a sua adequação ao fenómeno que pretende explicar. Os modelos de regressão para dados de contagem não fogem a esta regra, pelo que este capítulo é inteiramente dedicado ao estudo de testes de especificação.

Vão ser abordados somente dois tipos de testes: testes de sobredispersão, em que a hipótese nula será sempre o modelo de Poisson, sendo o objectivo comprovar se a dispersão das observações é ou não maior que a esperada; e testes de hipóteses não encaixadas, em que estarão em discussão as diversas variantes dos modelos BN's e PG's.

Toda a análise que se segue é feita numa óptica de discriminação, isto é, só se deseja rejeitar um determinado modelo quando o teste efectuado detectar desvios significativos na direcção do alternativo. Assim, são privilegiados os testes de natureza unilateral mas a adaptação a outras perspectivas é imediata.

1. Testes de sobredispersão

Os testes que são desenvolvidos nesta secção exigem a explicitação da hipótese alternativa. Sob esta, como é usual, tem-se um modelo mais geral que, impondo-se um conjunto de restrições paramétricas, corresponderá à formulação referente à hipótese nula, o modelo de Poisson. Como alternativos serão contemplados, sucessivamente, os modelos BN's e PG's, sendo também tratado o caso em que as amostras são truncadas em zero.

1.1. Poisson contra Binomial Negativo

Na literatura econométrica existe uma variedade razoável de testes apresentando o modelo BN como alternativa ao Poisson. Contudo, a maioria das sugestões é dirigida apenas contra um tipo específico de modelo BN: o BN1, no caso de Lee (1986) e Kim e Park (1992), ou o BN2, em Collings e Margolin (1985), Lee (1986) e Dean e Lawless (1989). O próprio Lee (1986), apesar de investigar as duas vertentes principais do modelo BN, deriva, em separado, estatísticas de teste aparentemente não generalizáveis às duas formulações, dando a entender que os testes a realizar teriam de ser, forçosamente, bem distintos, conforme a alternativa admitida fosse a BN1 ou a BN2. A única exceção é o teste proposto por CT (1986), o qual, mediante uma simples adaptação, permite testar o Poisson contra aqueles dois modelos.

Os testes que se vão de seguida desenvolver são, em todos os casos, testes score¹⁷. Esta opção é justificada pelo facto de a restrição que é testada se situar no limite do espaço do parâmetro em causa, α , situação em que, segundo Godfrey (1988), pp. 92-96, os testes de Wald e do rácio de verosimilhanças não são válidos, pois não possuem a habitual distribuição de Qui-Quadrado.

O teste a aplicar deverá ser unilateral, pelo que o teste score será definido por

$$\hat{T} = \frac{s_{\alpha}(\hat{\beta}, 0)}{\sqrt{I^{\alpha\alpha}(\hat{\beta}, 0)}} \simeq N(0, 1), \quad (\text{II.1})$$

sendo $s_{\alpha}(\hat{\beta}, 0) = \left. \frac{\partial LL(\beta, \alpha)}{\partial \alpha} \right|_{\beta=\hat{\beta}, \alpha=0}$ e

$$I^{\alpha\alpha} = I_{\alpha\alpha} - I_{\alpha\beta} I_{\beta\beta}^{-1} I_{\beta\alpha}, \quad (\text{II.2})$$

¹⁷ Dedicar-se-á particular atenção aos testes derivados por Lee (1986) e CT (1986), uma vez que os outros autores citados deduziram testes $C(\alpha)$, que não serão abordados neste trabalho. Para outro tipo de testes, mais gerais, ver também CT (1990).

representando $I_{..}$, de acordo com o formalizado na página 25, as matrizes-bloco que constituem a matriz de informação¹⁸.

Lee (1986) e CT (1986) deduziram \hat{T} indirectamente: começaram por considerar como alternativa ao modelo de Poisson a família de distribuições de Katz¹⁹, o que implicava que a função de log-verosimilhança implícita em II.1 fosse a referente a este sistema, e só depois, restringindo o intervalo dos valores possíveis de um dos parâmetros, passaram aos modelos BN's.

A família de distribuições de Katz é caracterizada pela relação $\frac{P(Y_i = y_i + 1)}{P(Y_i = y_i)} = \frac{\mu_i + \nu y_i}{1 + y_i}$, com $y = 0, 1, 2, \dots$, $\mu_i > 0$ e $\nu < 1$. A média é dada por $E(Y) = \frac{\mu_i}{1 - \nu}$ e a variância por $V(Y) = \frac{\mu_i}{(1 - \nu)^2}$. Como casos particulares, estão incluídas a distribuição de Poisson ($\nu = 0$), a BN ($0 < \nu < 1$), a Binomial ($\nu < 0$ e $\frac{\mu_i}{\nu}$ inteiro) e outras distribuições ($\nu < 0$ e $\frac{\mu_i}{\nu}$ não inteiro).

Neste contexto, investigar a existência de sobredispersão equivale a testar se $\nu = 0$ (modelo de Poisson) contra $\nu > 0$ (modelos BN's), daí todos os testes desenvolvidos serem unilaterais. Após generalizar aquele sistema de distribuições de modo a incorporar regressores, Lee (1986) derivou duas versões do teste score apropriado para tal, cada uma delas dirigida especificamente contra um certo tipo de modelo BN: o teste \hat{T}_{L_A} ,

$$\hat{T}_{L_A} = \frac{\sum_{i=1}^n \frac{1}{\hat{\mu}_i} [(y_i - 1)y_i - \hat{\mu}_i^2]}{\sqrt{2n}}, \quad (II.3)$$

contra o BN1, e o teste \hat{T}_{L_B} ,

¹⁸ Para facilitar a exposição, suprimiu-se a indicação dos parâmetros de que as componentes desta matriz dependem, passando-se, a partir deste momento, a agir da mesma forma em relação ao score e à variável $I^{\alpha\alpha}$, excepto em situações que possam suscitar dúvidas.

¹⁹ Lee (1986) também derivou testes contra a família de distribuições de Pearson.

$$\hat{T}_{L_B} = \frac{\sum_{i=1}^n [(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i]}{\sqrt{2 \sum_{i=1}^n \hat{\mu}_i^2}}, \quad (\text{II.4})$$

contra o BN2, sendo ambos válidos somente se um dos regressores for a constante.

Posteriormente, CT (1986) apresentaram um terceiro teste, \hat{T}_{CT} ,

$$\hat{T}_{CT} = \frac{\sum_{i=1}^n \hat{\mu}_i^{k-1} [(y_i - \hat{\mu}_i)^2 - y_i]}{\sqrt{2 \sum_{i=1}^n \hat{\mu}_i^{2k}}}, \quad (\text{II.5})$$

que tinha a particularidade de ser adaptável aos dois modelos BN's: $k = 0$ define a alternativa BN1 e $k = 1$ a BN2.

Nos três casos, as estimativas utilizadas para μ_i são, evidentemente, oriundas da estimação do modelo de Poisson por MV e a distribuição assintótica é a $N(0,1)$.

Cada um dos testes tem a sua própria interpretação: \hat{T}_{L_A} compara o momento factorial de segunda ordem da amostra com o momento teórico correspondente ao modelo de Poisson, $E[(Y_i - 1)Y_i | x_i] = \mu_i^2$, \hat{T}_{L_B} as variâncias empírica e teórica, $E[(Y_i - \mu_i)^2 | x_i] = \mu_i$, e \hat{T}_{CT} a variância e o valor esperado condicionais dos dados observados. Apenas sob H_0 , os momentos confrontados serão iguais.

É curioso notar que nenhum autor procedeu à dedução do teste score adequado para testar o modelo de Poisson contra o BN recorrendo ao processo clássico: calculando os elementos da expressão II.1, s_{α} e $I^{\alpha\alpha}$, directamente²⁰, isto é, considerando a função de log-verosimilhança implícita em II.1 como a de um dos modelos BN's. É isso que se fará de seguida, tratando-se em simultâneo o BN1 e o BN2.

Sendo a hipótese nula $\alpha = 0$, os vectores score respeitantes aos modelos BN1 e BN2 podem ser encontrados a partir das expressões I.40 e I.42. Substituindo, então, α por zero e β pela sua estimativa $\hat{\beta}$, proveniente da maximização da função de

²⁰ Apenas para o caso truncado este procedimento foi seguido, em Gurmu (1991) e Gurmu e Trivedi (1992).

verossimilhança do modelo de Poisson, e aplicando a regra de Cauchy de modo a levantar algumas indeterminações, resulta:

$$s_\alpha = \frac{1}{2} \sum_{i=1}^n \hat{\mu}_i^{k-1} \left[(y_i - \hat{\mu}_i)^2 - y_i \right] \quad (\text{II.6})$$

Para se calcular $I^{\alpha\alpha}$, de acordo com a expressão II.2, é necessário primeiro construir a matriz de informação. Começando pelas expressões I.45 e I.48, referentes a um dos componentes da matriz hessiana dos modelos BN1 e BN2, respectivamente, e substituindo, mais uma vez, α por zero e β por $\hat{\beta}$, obtém-se, após novo recurso à regra de Cauchy,

$$\left. \frac{\partial^2 \text{LL}(\beta, \alpha)}{\partial \beta \partial \alpha} \right|_{H_0} = \left[\left. \frac{\partial^2 \text{LL}(\beta, \alpha)}{\partial \alpha \partial \beta} \right|_{H_0} \right]' = - \sum_{i=1}^n 2^{k-1} x_i' \hat{\mu}_i^{2k-1} \left\{ [y_i - (1-k)] y_i^{1-k} - \hat{\mu}_i^{2-k} \right\}, \quad (\text{II.7})$$

donde se pode concluir, usando-se o método de Scoring, que a matriz de informação é diagonal por blocos para os dois modelos,

$$\begin{aligned} \text{BN1 (k = 0): } I_{\beta\alpha} &= -E \left[\left. \frac{\partial^2 \text{LL}(\beta, \alpha)}{\partial \beta \partial \alpha} \right|_{H_0} \right] = - \sum_{i=1}^n \frac{x_i'}{2\hat{\mu}_i} E_{H_0} [(Y_i - 1)Y_i - \hat{\mu}_i^2] = 0 \\ \text{BN2 (k = 1): } I_{\beta\alpha} &= -E \left[\left. \frac{\partial^2 \text{LL}(\beta, \alpha)}{\partial \beta \partial \alpha} \right|_{H_0} \right] = - \sum_{i=1}^n x_i' \mu_i E_{H_0} (Y_i - \hat{\mu}_i) = 0, \end{aligned}$$

pois, para o Poisson, $E[(Y_i - 1)Y_i | x_i] = \mu_i^2$ e $E(Y_i | x_i) = \mu_i$.

Deste modo, a variância do score será dada simplesmente por

$$I^{\alpha\alpha} = I_{\alpha\alpha} = -E_{H_0} \left[\left. \frac{\partial^2 \text{LL}(\beta, \alpha)}{\partial \alpha^2} \right|_{H_0} \right].$$

Sendo

$$\left. \frac{\partial^2 \text{LL}(\beta, \alpha)}{\partial \alpha^2} \right|_{H_0} = - \frac{2y_i^3 - 3y_i^2 + y_i}{6\hat{\mu}_i^{2-2k}} + \hat{\mu}_i^{2k} y_i - \frac{2\hat{\mu}_i^{2k+1}}{3}, \quad (\text{II.8})$$

prova-se facilmente que:

$$I^{\alpha\alpha} = \frac{1}{2} \sum_{i=1}^n \hat{\mu}_i^{2k} \quad (\text{II.9})$$

Reunindo as expressões II.6 e II.9, encontra-se uma estatística de teste exactamente igual à proposta por CT (1986), apesar destes autores a terem deduzido partindo, como se referiu, da família de distribuições de Katz.

O score s_α pode ser apresentado de outra forma, equivalente sob certas condições. Trabalhando unicamente com o termo entre parêntesis rectos contido na expressão II.6, pode ser determinada a seguinte igualdade:

$$\begin{aligned} (y_i - \hat{\mu}_i)^2 - y_i &= (y_i - \hat{\mu}_i)^2 - y_i - \hat{\mu}_i + \hat{\mu}_i \\ &= (y_i - \hat{\mu}_i)^2 - \hat{\mu}_i - (y_i - \hat{\mu}_i) \end{aligned} \quad (\text{II.10})$$

Se um dos regressores for fixo, $\sum_{i=1}^n \hat{\mu}_i = \sum_{i=1}^n y_i$, de acordo com I.6. Assim, a expressão II.10 pode ser simplificada pois, ao aplicar-se-lhe somatórios, a última parcela desaparecerá. Com este procedimento, obtém-se o teste

$$\hat{T}_{L_B'} = \frac{\sum_{i=1}^n \hat{\mu}_i^{k-1} [(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i]}{\sqrt{2 \sum_{i=1}^n \hat{\mu}_i^{2k}}}, \quad (\text{II.11})$$

o qual constitui uma extensão da estatística \hat{T}_{L_B} (II.4) sugerida por Lee (1986), visto que esta só se podia utilizar contra o modelo BN2.

O teste $\hat{T}_{L_B'}$ apenas é válido, portanto, se um dos regressores for a constante. Caso contrário, o denominador de II.11 não estará correcto, pois a matriz de informação deixará de ser diagonal por blocos, passando a distribuição assintótica de s_α a estar dependente da de $\hat{\beta}$.

A computação de cada teste pode ser realizada directamente a partir das expressões II.3, II.5 ou II.11, mas existem processos mais simples. Por exemplo, seguindo uma indicação de CT (1990), comprova-se facilmente que a estatística \hat{T}_{CT} corresponde à raiz quadrada da soma de quadrados explicados da regressão linear:

$$\frac{1}{\sqrt{2\hat{\mu}_i}} \left[(y_i - \hat{\mu}_i)^2 - y_i \right] = \hat{\mu}_i^k \cdot \eta + \varepsilon_i \quad (\text{II.12})$$

A estatística \hat{T}_{LB} pode também ser determinada através de uma regressão deste tipo mas com variável dependente $\frac{1}{\sqrt{2\hat{\mu}_i}} \left[(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i \right]$. O mesmo se passa com \hat{T}_{LA} , devendo-se, neste caso, efectuar a regressão de $\frac{1}{\sqrt{2\hat{\mu}_i}} \left[(y_i - 1)y_i - \hat{\mu}_i^2 \right]$ na constante.

Alternativamente, os testes poderão ser computados através da estatística equivalente assintoticamente $t = \frac{\hat{\eta}}{\sigma_{\hat{\eta}}} \approx N(0,1)$. Actuando desta forma, os resultados dos testes \hat{T}_{LB} e \hat{T}_{CT} já não serão numericamente iguais, o que vem justificar a sua análise separada, até porque este é o método de cálculo mais utilizado, pois a maioria do *software* econométrico não fornece a soma de quadrados explicados.

É de referir ainda que os testes deduzidos contra o modelo BN2 podem também ser obtidos admitindo alternativas locais ao Poisson, de acordo com a sugestão de Cox (1983). Este autor coloca a possibilidade de existência de um erro de especificação no modelo de Poisson, mas não define a sua distribuição. A sua técnica exige a imposição da restrição I.25, $V(\varepsilon_i|x_i) = \omega$, e a realização de um teste score para a nulidade de ω . Como se viu anteriormente (p. 22), a consideração de um erro homocedástico leva à existência de uma relação quadrática entre a variância e o valor esperado, condicionais somente nos regressores, da variável dependente. Sendo também esta a principal característica do modelo BN2, não é então de estranhar que as estatísticas de teste encontradas pelas duas vias sejam iguais.

Num âmbito mais geral, Chesher (1984) desenvolveu uma abordagem deste género, provando que os testes determinados da maneira indicada por Cox (1983) são válidos contra um vasto lote de alternativas, desde que estas verifiquem, entre outras, a

condição I.25. Ele demonstrou igualmente a sua equivalência aos testes da matriz de informação, dando assim uma nova interpretação a estes últimos, em termos de heterogeneidade negligenciada. Por isso, em relação aos testes derivados contra o modelo BN2, percebe-se facilmente a razão porque eles são, tipicamente, testes da matriz de informação, baseados apenas na informação relativa à constante, correspondendo a condição de momentos que é testada a $E\left[(Y_i - \hat{\mu}_i)^2 - \hat{\mu}_i\right] = 0$.

1.2. Poisson contra Poisson Generalizado

Ao contrário do que acontece quando a hipótese alternativa é um dos modelos BN's, quase não há referências a testes de sobredispersão direccionados especificamente contra qualquer uma das formulações PG's, o que até não é de admirar, uma vez que estes modelos de regressão são bastante recentes. Os únicos testes propostos [Consul e Famoye (1992) e Famoye (1993)] baseiam-se no rácio de verosimilhanças, pelo que não são aplicáveis no contexto de sobredispersão em que este trabalho se insere, pois ao restringir-se o espaço do parâmetro perturbador, exclui-se a hipótese de recurso a testes desta natureza. Deste modo, a exemplo do realizado no ponto anterior, procede-se de seguida à dedução do teste score adequado para testar a existência de sobredispersão, considerando como alternativas os modelos PG1 e PG2.

Para o PG1, a determinação do vector score, avaliado em $\beta = \hat{\beta}$ (modelo de Poisson) e $\phi = 1$, pode ser feita a partir da expressão I.66, resultando:

$$s_\phi = \sum_{i=1}^n \frac{1}{\hat{\mu}_i} \left[(y_i - \hat{\mu}_i)^2 - y_i \right] \quad (\text{II.13})$$

Para o PG2, pegando em I.68 e substituindo μ_i por $\hat{\mu}_i$ e α por 0, obtém-se:

$$s_\alpha = \sum_{i=1}^n \left[(y_i - \hat{\mu}_i)^2 - y_i \right] \quad (\text{II.14})$$

As expressões II.13 e II.14 são muito parecidas, podendo ser agrupadas numa só,

$$s_{\psi} = \sum_{i=1}^n \hat{\mu}_i^{k-1} \left[(y_i - \hat{\mu}_i)^2 - y_i \right], \quad (\text{II.15})$$

onde $k = 0$ (e $\psi = \varphi$) origina o modelo PG1 e $k = 1$ (e $\psi = \alpha$) gera o PG2.

Comparando-se a expressão encontrada com a deduzida anteriormente para os modelos BN's (II.6), nota-se que a única diferença reside no termo $1/2$. Mas as analogias em relação ao caso precedente não ficam por aqui. A matriz de informação também é diagonal por blocos pois

$$\frac{\partial^2 \text{LL}(\beta, \psi)}{\partial \beta \partial \psi} \Big|_{H_0} = \left[\frac{\partial^2 \text{LL}(\beta, \psi)}{\partial \psi \partial \beta} \Big|_{H_0} \right]' = - \sum_{i=1}^n 2^k x_i' \hat{\mu}_i^{2k-1} \{ [y_i - (1-k)] y_i^{1-k} - \hat{\mu}_i^{2-k} \}, \quad (\text{II.16})$$

cujo valor esperado é nulo.

Como

$$\frac{\partial^2 \text{LL}(\beta, \alpha)}{\partial \alpha^2} \Big|_{H_0} = - \frac{y_i^3 - y_i^2}{\hat{\mu}_i^{2-2k}} + 3\hat{\mu}_i^{2k} y_i - 2\hat{\mu}_i^{2k+1}, \quad (\text{II.17})$$

aplicando o método de Scoring, a variância do score será estimada simplesmente por

$$\Gamma^{\alpha\alpha} = \sum_{i=1}^n \hat{\mu}_i^{2k}, \quad (\text{II.18})$$

expressão também muito semelhante com a calculada (II.9) para os testes construídos no ponto anterior.

Rearranjando s_{α} , de acordo com o realizado para os modelos BN's, e dividindo-o pelo seu desvio padrão, obtêm-se exactamente as mesmas estatísticas de teste. Assim, qualquer um dos dois testes, \hat{T}_{CT} e \hat{T}_{LB} , é válido para testar o modelo de Poisson, tanto contra as duas variantes BN's como contra os dois modelos PG's, nas condições enunciadas. A justificação desta situação deve-se à circunstância dos modelos BN1 e

PG1, por um lado, e BN2 e PG2, por outro, constituírem alternativas localmente equivalentes²¹.

1.3. Amostras Truncadas

Neste ponto, admite-se como hipótese nula o modelo PT, o qual será testado face às formulações BNT's e PGT's, que são, como se verificou no capítulo I, alternativas credíveis para a modelização de dados truncados demasiado heterogéneos para serem explicados pelo primeiro modelo. Deste modo, o objectivo continuará a ser a dedução de testes de sobredispersão, os quais, pelos motivos atrás apontados, terão, obrigatoriamente, de ser testes score.

A dedução do teste adequado para testar o modelo PT contra os BNT's foi efectuada por Gurmu (1991), que determinou a estatística de teste

$$\hat{T}_G = \frac{\sum_{i=1}^n \hat{\mu}_i^{k-1} [(y_i - 1)y_i - \hat{\mu}_i \hat{\mu}_i^* - 2\hat{\mu}_i (y_i - \hat{\mu}_i^*)]}{2\sqrt{I_{\alpha\alpha} - I_{\alpha\beta} I_{\beta\beta}^{-1} I_{\beta\alpha}}}, \quad (\text{II.19})$$

sendo os elementos da matriz de informação, que não é mais diagonal por blocos, dados por:

$$I(\hat{\beta}, 0) = \begin{bmatrix} \sum_{i=1}^n x_i' x_i \hat{\mu}_i^* (1 - \hat{\eta}_i) & \frac{1}{2} \sum_{i=1}^n \hat{\mu}_i^k x_i' \hat{\mu}_i^* \hat{\eta}_i \\ \frac{1}{2} \sum_{i=1}^n \hat{\mu}_i^k x_i \hat{\mu}_i^* \hat{\eta}_i & \frac{1}{2} \sum_{i=1}^n \hat{\mu}_i^{2k-1} \hat{\mu}_i^* \left(1 - \frac{1}{2} \hat{\eta}_i \hat{\mu}_i\right) \end{bmatrix} \quad (\text{II.20})$$

Como habitualmente, $k = 0$ define o modelo BNT1 e $k = 1$ o BNT2. Tendo em conta que $E[(Y_i - 1)Y_i | x_i, Y_i > 0] = \mu_i \mu_i^*$ e que $E(Y_i | x_i, Y_i > 0) = \mu_i^*$, o teste \hat{T}_G consiste na confrontação, simultânea, destes dois momentos teóricos com os correspondentes da amostra.

Este teste não é de fácil computação, tendo que ser calculado a partir da

²¹ Para o estudo deste tipo de alternativas, consultar Godfrey (1988), pp. 70-75.

expressão II.19. A sua distribuição assintótica, naturalmente, continua a ser a $N(0,1)$ e as estimativas utilizadas para μ_i e, portanto, para μ_i^* e η_i , resultam da aplicação da MV ao modelo PT.

Ao contrário do sucedido no ponto anterior, \hat{T}_G não é um teste da matriz de informação, pois a base deste último é a condição de momentos $E\left[(Y_i - \hat{\mu}_i^*)^2 - (1 - \hat{\eta}_i)\hat{\mu}_i^*\right] = 0$, enquanto que a presente em II.19, adaptada de modo a permitir a comparação, é $E\left[(Y_i - \hat{\mu}_i^*)^2 - (1 - \hat{\eta}_i)\hat{\mu}_i^* - (1 + 2\hat{\eta}_i)(Y_i - \hat{\mu}_i^*)\right] = 0$, que possui uma parcela adicional.

O teste \hat{T}_G pode também ser usado para testar o modelo PT contra o PGT1 ou o PGT2. Com efeito, prova-se facilmente²² que, para os modelos PGT's,

$$s_\psi = \sum_{i=1}^n \hat{\mu}_i^{k-1} \left[(y_i - 1)y_i - \hat{\mu}_i \hat{\mu}_i^* - 2\hat{\mu}_i (y_i - \hat{\mu}_i^*) \right] \quad (\text{II.21})$$

e

$$I(\hat{\beta}, 0) = \begin{bmatrix} \sum_{i=1}^n x_i' x_i \hat{\mu}_i^* (1 - \hat{\eta}_i) & \sum_{i=1}^n \hat{\mu}_i^k x_i' \hat{\mu}_i^* \hat{\eta}_i \\ \sum_{i=1}^n \hat{\mu}_i^k x_i \hat{\mu}_i^* \hat{\eta}_i & 2 \sum_{i=1}^n \hat{\mu}_i^{2k-1} \hat{\mu}_i^* \left(1 - \frac{1}{2} \hat{\eta}_i \hat{\mu}_i \right) \end{bmatrix}, \quad (\text{II.22})$$

sendo $k = 0$ para o modelo PGT1 e $k = 1$ para o PGT2.

Notando que,

$$\begin{aligned} (I^{\alpha\alpha})^{0.5} &= (I_{\alpha\alpha} - I_{\alpha\beta} I_{\beta\beta}^{-1} I_{\beta\alpha})^{0.5} \\ &= \left[4I_{\alpha\alpha}^{\text{bnt}} - 2I_{\alpha\beta}^{\text{bnt}} (I_{\beta\beta}^{\text{bnt}})^{-1} 2I_{\beta\alpha}^{\text{bnt}} \right]^{0.5} \\ &= \left\{ 4 \left[I_{\alpha\alpha}^{\text{bnt}} - I_{\alpha\beta}^{\text{bnt}} (I_{\beta\beta}^{\text{bnt}})^{-1} I_{\beta\alpha}^{\text{bnt}} \right] \right\}^{0.5} \\ &= 2 \left[I_{\alpha\alpha}^{\text{bnt}} - I_{\alpha\beta}^{\text{bnt}} (I_{\beta\beta}^{\text{bnt}})^{-1} I_{\beta\alpha}^{\text{bnt}} \right]^{0.5}, \end{aligned} \quad (\text{II.23})$$

²² Nas deduções efectuadas foi muito útil o recurso à identidade $E(Y^r | Y > 0) = \frac{E(Y^r)}{1 - P(Y = 0)}$.

representando $I_{\alpha\alpha}^{bnt}$ os vários componentes da matriz de informação correspondente aos modelos BNT's (II.20), ao se dividir s_{ψ} (II.21) por $(I_{\alpha\alpha})^{0.5}$ (II.23), comprova-se que, à semelhança do que se verificava para amostras não truncadas, os modelos em causa constituem, realmente, alternativas localmente equivalentes.

2. Testes de Hipóteses Não Encaixadas

Todos os testes analisados até ao momento envolvem apenas modelos encaixados, isto é, o modelo considerado sob a hipótese nula é um caso particular de uma formulação mais geral contra a qual é examinado. Mostrou-se, assim, como testar o modelo de Poisson face às alternativas BN's ou PG's, podendo acontecer, em virtude da existência de sobredispersão, a rejeição do primeiro. É verdade que se a forma funcional do valor esperado condicional estiver bem formalizada, os estimadores determinados para o modelo de Poisson serão sempre consistentes mas, querendo-se inferir relativamente à probabilidade de ocorrência de um dado acontecimento, ou sendo os modelos truncados, é essencial que a distribuição admitida seja efectivamente a correcta.

Torna-se relevante, portanto, testar os vários modelos BN's e PG's, uns “contra” os outros, de modo a averiguar até que ponto eles se adequam às observações em estudo. Dadas as características destes modelos, será necessário recorrer a testes próprios para hipóteses não encaixadas²³ pelo que, nesta secção, se apresentam alguns dos principais testes desta natureza aplicáveis a modelos de contagem.

2.1. Testes baseados na combinação de Funções de Probabilidade

O elemento primordial de distinção dos modelos abordados no capítulo anterior é a distribuição condicional que se assume para a variável dependente. Por isso, é natural que uma das formas de os testar incida directamente sobre as distribuições

²³ Para o seu estudo aprofundado, deve-se começar por consultar o *survey* de Gourieroux e Monfort (1994), que analisam um vasto lote de testes para hipóteses não encaixadas e fornecem uma extensa lista bibliográfica sobre o tema. Ver também Davidson e MacKinnon (1993) e o *survey* de MacKinnon (1983).

alternativas que se pensa ser possível os dados seguirem. De seguida, analisam-se dois tipos de testes que adoptam esta perspectiva: os primeiros, assentes na combinação linear de funções de probabilidade, são mais genéricos, podendo ser empregues para testar uma grande variedade de modelos, não necessariamente de contagem; os outros, servindo-se do modelo BNk (BNTk) ou do PG3 (PGT3) como auxiliar, só podem ser usados no âmbito dos modelos BN's e PG's.

2.1.1. Testes Score para a combinação linear de Funções de Probabilidade

Em termos formais, as hipóteses em confronto são

$$H_1: f(y_i|x_i; \rho), \rho \in \Omega_1$$

$$H_2: g(y_i|x_i; \gamma), \gamma \in \Omega_2,$$

representando $f(\cdot)$ e $g(\cdot)$ as respectivas funções de probabilidade de cada modelo e verificando-se, como as hipóteses são não encaixadas, $\Omega_1 \not\subset \Omega_2$ e $\Omega_2 \not\subset \Omega_1$.

Dos poucos testes elaborados com o objectivo de avaliar hipóteses deste género destaca-se, naturalmente, o de Cox (1961), que esteve na origem de todos os desenvolvidos posteriormente. Este autor sugeriu dois métodos alternativos cuja aplicação prática se afigura um pouco problemática. Por isso, optar-se-á pela utilização de um teste, referido inicialmente em Atkinson (1970) e estudado com algum detalhe em Quandt (1974), mas tornado operacional apenas em Santos Silva (1996b), baseado na construção de um modelo artificial a partir da combinação linear das duas funções de probabilidade,

$$H_c: h(y_i|x_i; \phi, \rho, \gamma) = (1-\phi)f(y_i|x_i; \rho) + \phi g(y_i|x_i; \gamma), \quad (\text{II.24})$$

sendo $\phi \in [0;1]$ o parâmetro de ligação, cujo valor determina se as duas distribuições necessitam realmente de ser conjugadas de forma a poderem constituir um modelo

satisfatório²⁴.

Quando $\phi = 0$, H_c corresponde a H_1 ; quando $\phi = 1$, H_c coincide com H_2 . De notar que $h(\cdot)$ representa uma função de probabilidade com todas as características usuais, pelo que até se poderá chegar à conclusão que é dela que provém a melhor descrição dos dados. Aliás, $h(\cdot)$ pode ser interpretada como a densidade representativa do comportamento de uma população que é uma mistura de duas sub-populações [Santos Silva (1996b)]. Deve-se ter em atenção, todavia, que desvios de uma hipótese na direcção oposta à da alternativa produzem valores de ϕ fora do intervalo $[0;1]$, podendo provocar probabilidades negativas [Quandt (1974)].

Considerando como hipótese nula H_1 , ela será examinada face a H_c , o que equivale a testar $H_0: \phi = 0$. Mais uma vez, a hipótese que é avaliada situa-se no limite do espaço do parâmetro, devendo-se recorrer a um teste score, de tipo unilateral.

O logaritmo da função de verosimilhança associada a II.24 é definido por

$$LL(\phi, \rho, \gamma) = \sum_{i=1}^n \ln[(1 - \phi)L_{1i}(\rho) + \phi L_{2i}(\gamma)], \quad (\text{II.25})$$

onde $L_{1i}(\cdot)$ e $L_{2i}(\cdot)$ representam as contribuições individuais para a verosimilhança de cada um dos modelos.

Sob a hipótese nula, γ não é identificado. Por isso, este vector de parâmetros é normalmente substituído pelos estimadores da MV referentes à maximização de $L_{2i}(\gamma)$, embora possa ser utilizado, de acordo com o exposto por Davidson e MacKinnon (1981), qualquer estimador que assintoticamente não seja estocástico (outro estimador que vulgarmente aparece neste género de testes é introduzido no ponto 2.2.2.).

A dedução do teste score poderia ser feita de acordo com o realizado na secção anterior: calculando os elementos da expressão II.1, tendo por base a função de log-verosimilhança II.25. Contudo, como esta tarefa já foi executada por Santos Silva (1996b), apresenta-se de imediato o estimador da estatística derivada por este autor,

²⁴ Um dos métodos propostos por Cox (1961), mas formalizado apenas em Atkinson (1970), baseia-se numa combinação exponencial das duas hipóteses, o que exige a presença de um factor de proporcionalidade, sem o qual $h(\cdot)$ deixa de constituir uma função de probabilidade; o problema é que a presença desse factor torna quase impraticável a computação do teste resultante.

$$\hat{T}_s = \frac{\sum_{i=1}^n \frac{L_{2i} - L_{1i}}{L_{1i}}}{\sqrt{\sum_{i=1}^n \left(\frac{L_{2i} - L_{1i}}{L_{1i}} \right)^2 - \left[\sum_{i=1}^n \frac{L_{2i} - L_{1i}}{L_{1i}} \left(\frac{\partial LL_{1i}}{\partial \rho} \right)' \right] \left[-\sum_{i=1}^n \left(\frac{\partial^2 LL_{1i}}{\partial \rho \partial \rho'} \right) \right]^{-1} \left[\sum_{i=1}^n \frac{L_{2i} - L_{1i}}{L_{1i}} \left(\frac{\partial LL_{1i}}{\partial \rho} \right) \right]}} \quad (\text{II.26})$$

tendo sido omitida, por questões de simplificação, a dependência de L_{2i} face a $\hat{\gamma}$ e de L_{1i} face a $\hat{\rho}$, provindo este último do uso da MV na estimação do modelo correspondente a H_1 . A distribuição assintótica de \hat{T}_s é a $N(0,1)$.

Se o objectivo for testar a validade de H_2 , o teste a empregar será também este, mas com a adaptação óbvia, trocando-se os papéis desempenhados por cada modelo.

Da aplicação do teste, podem surgir quatro resultados possíveis. Pode-se rejeitar H_1 e não rejeitar H_2 (ou vice-versa), situação em que tudo aponta para H_1 (H_2) como sendo o modelo mais adequado para representar os dados em estudo. Podem ser rejeitados ambos os modelos, o que indica que nenhum deles é satisfatório, havendo necessidade de tentar desenvolver melhores alternativas. Pode ainda acontecer que não se possa rejeitar nenhum dos modelos, ou porque eles são muito similares, ou porque os dados não são muito informativos.

Voltando ao âmbito dos dados de contagem, a estatística \hat{T}_s permitirá testar vários dos modelos abordados: as formulações BN1, BN2, PG1, PG2 e PG3, com ou sem truncagem, que originam hipóteses não encaixadas (com excepção do PG1 e do PG2, quando testados contra o PG3), podem agora ser testadas duas a duas. A computação do teste não é trivial mas também não será muito difícil de efectuar, pois $L_{1i}(\hat{\rho})$ e $L_{2i}(\hat{\gamma})$ representam simplesmente as funções de probabilidade associadas a

cada modelo, $\left[-\sum_{i=1}^n \frac{\partial^2 LL_{1i}(\rho)}{\partial \rho \partial \rho'} \Big|_{\rho=\hat{\rho}} \right]^{-1}$ é o estimador da matriz de covariâncias de $\hat{\rho}$

proveniente da utilização do método de Newton e $\frac{\partial LL_{1i}(\rho)}{\partial \rho} \Big|_{\rho=\hat{\rho}}$ é a contribuição de cada

indivíduo para o vector score do modelo considerado sob H_1 .

Por exemplo, se se pretender testar o modelo BN2 “contra” o PG2, $L_{1i}(\hat{\rho})$,

$\hat{\rho} = \begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix}$, será dado por I.36, $L_{2i}(\hat{\gamma})$, $\hat{\gamma} = \begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix}$, por I.60, $\sum_{i=1}^n \frac{\partial^2 LL_{1i}(\rho)}{\partial \rho \partial \rho'} \Big|_{\rho=\hat{\rho}}$ será uma

matriz com dimensão $(m+1) \times (m+1)$, composta pelas expressões I.47 a I.49, organizadas convenientemente, e $\frac{\partial LL_{1i}(\rho)}{\partial \rho} \Big|_{\rho=\hat{\rho}}$ será um vector coluna $(m+1)$ -dimensional, definido

por I.41 e I.42.

Ainda segundo Santos Silva (1996b), se for possível exprimir, para o modelo que é testado, as contribuições individuais para o vector score como condições de ortogonalidade entre uma função qualquer dos regressores e os erros, $h(x_i)'[y_i - E(Y_i|x_i)]$, então poderá ser construído um teste assintoticamente equivalente a \hat{T}_S , \hat{T}_{S^*} . Este resultará da substituição, em II.26, de $\frac{L_{2i} - L_{1i}}{L_{1i}} \left(\frac{\partial LL_{1i}}{\partial \rho} \right)$ por $h(x_i)'[E_2(Y_i|x_i) - E_1(Y_i|x_i)]$, onde o significado de $E_1(\cdot)$ e $E_2(\cdot)$ é evidente.

No contexto dos modelos de contagem, o BN2 e o PG2 cumprem o requisito necessário à aplicação do teste \hat{T}_{S^*} , mas apenas se o parâmetro perturbador, α , for conhecido, o que constitui uma grande limitação ao seu uso neste tipo de modelos. Retomando o exemplo apresentado, o novo termo presente em \hat{T}_{S^*} será estimado por $x_i' \left[\exp(x_i \hat{\beta}^{pg2}) - \exp(x_i \hat{\beta}^{bn2}) \right]$, representando $\hat{\beta}^{bn2}$ e $\hat{\beta}^{pg2}$ os estimadores da MV para os modelos BN2 e PG2, respectivamente. Como ambos os estimadores são sempre consistentes, assumindo que o valor esperado condicional está bem formalizado, aquela parcela será, em termos assintóticos, identicamente nula, o que torna o teste, neste caso específico, ainda mais simples de efectuar.

2.1.2. Testes Específicos para Modelos de Contagem

Se o objectivo for testar o BN1 “contra” o BN2, há outra maneira de o fazer. Recorrendo ao modelo BNk²⁵ de Winkelmann e Zimmermann (1991, 1995), este poderá desempenhar neste âmbito o papel que o modelo artificial dado por II.24 tem no teste anterior. O parâmetro de ligação será, naturalmente, o k : se $k = 0$, o melhor modelo será

²⁵ A formalização da função de probabilidade deste modelo corresponde à efectuada para o BNTk (expressão I.97), omitindo a última parcela.

o BN1; se $k = 1$, a preferência irá para o BN2.

Como k pode ser encarado como um parâmetro contínuo, sem restrições ao seu valor, qualquer uma das formas convencionais (Wald, rácio de verosimilhanças, score) é válida para computar este teste, doravante designado por T_k . Tendo em conta a elevada não linearidade das derivadas em ordem a k , desta vez não se formaliza o teste score, até porque o teste T_k pode ser estimado pelas três vias indicadas.

Independentemente do processo escolhido, deve-se ter em atenção que o teste a empregar deverá ser unilateral, de modo a direccioná-lo especificamente “contra” uma das alternativas BN’s. A distribuição assintótica do teste será a $N(0,1)$.

Também a avaliação do modelo PG1 “face” ao PG2 pode ser feita através de um teste deste género, usando uma versão restritiva do PG3 como auxiliar. Com efeito, adoptando, em vez de I.76, a parametrização mais limitativa

$$\alpha_i = \delta_0 \exp(\delta_1 x_i \beta_1), \quad (\text{II.27})$$

o modelo resultante engloba o PG1 e o PG2 de forma semelhante à que o BN k faz para os dois modelos BN’s. Nesta situação, $\delta_0 > 0$ e os vectores x_i e β_1 passam a ter $m-1$ elementos, nenhum deles referente ao termo constante.

Utilizando II.27, e reparando que δ_1 passa a ser o parâmetro de ligação, é fácil testar as hipóteses $H_1: \delta_1 = -1$ (modelo PG1) ou $H_1: \delta_1 = 0$ (modelo PG2). Como o objectivo é testar o PG1 “contra” o PG2, e vice-versa, deverá ser realizado, novamente, um teste unilateral.

De notar que todas as ideias expostas neste ponto podem ser igualmente aplicadas a amostras truncadas, tomando como modelos gerais o BNT k ou o PGT3 (com α_i parametrizado de acordo com II.27).

2.2. Testes baseados na realização de regressões artificiais

Os testes que se apresentam neste ponto baseiam-se na comparação dos valores esperados condicionais dos modelos que se querem testar. Considere-se a forma

funcional de dois modelos que pretendem descrever a mesma teoria:

$$H_1: E(Y_i | x_i) = f(x_i, \rho) \quad (\text{II.28})$$

$$H_2: E(Y_i | z_i) = g(z_i, \gamma) \quad (\text{II.29})$$

Estes modelos são não encaixados se for impossível impor restrições a ρ ou a γ que permitam estabelecer uma igualdade entre $f(x_i, \rho)$ e $g(z_i, \gamma)$ ou que possibilitem obter um deles como forma limite do outro. Nas formalizações lineares, é preciso que exista pelo menos um regressor em cada uma delas, que não esteja presente na alternativa, para os modelos serem daquele tipo. No caso não linear, em que se enquadram todos os modelos tratados neste trabalho, tal não é necessário, bastando que a forma funcional do valor esperado condicional seja diferente, o que só acontece, neste contexto, quando os dados são truncados. Consequentemente, apenas nesta situação os testes que se vão estudar nas páginas seguintes se poderão aplicar a modelos de contagem.

Ao contrário do que se verificava há pouco, para testar modelos desta natureza existem vários processos que, embora difiram numericamente, são equivalentes em termos assintóticos, possuindo a mesma distribuição. Os primeiros testes a serem analisados são os de Davidson e Mackinnon²⁶ (1981), os quais, de entre todos os testes próprios para hipóteses não encaixadas, são talvez os de mais fácil execução. Baseados nos testes de DM (1981), Fisher e McAleer²⁷ (1981) sugeriram uma alternativa, da qual será aqui proposta uma variante.

2.2.1. Testes de Davidson e MacKinnon

O método de DM (1981) assenta na realização de regressões artificiais, resultantes da junção dos dois modelos concorrentes num mais geral, devendo-se, de seguida, testar ambas as formulações originais contra o modelo artificial, o qual pode ser expresso da seguinte maneira:

²⁶ Daqui em diante referenciados apenas por DM.

²⁷ Referidos, posteriormente, somente por FM.

$$H_c: E(Y_i | x_i, z_i) = (1-\phi)f(x_i, \rho) + \phi g(z_i, \gamma) \quad (\text{II.30})$$

Novamente, se $\phi = 0$, H_c corresponderá a H_1 ; se $\phi = 1$, H_c coincidirá com H_2 . Esta abordagem é do mesmo tipo da referida no ponto 2.1.1. (ver II.24), mas com a importante diferença que $f(\cdot)$ e $g(\cdot)$ representam agora diferentes especificações do valor esperado condicional e não funções de probabilidade. O problema com H_c mantém-se pois, geralmente, não se conseguem identificar todos os parâmetros, uma vez que estes são em número superior aos regressores, não se podendo testar se aquelas restrições sobre ϕ se verificam ou não.

Tendo em conta esta questão, DM (1981) propuseram como solução substituir os parâmetros do modelo que não é testado por estimativas consistentes sob a hipótese da sua veracidade. Assim, pretendendo-se testar o modelo H_1 , pode-se reescrever II.30 como

$$H_c': E(Y_i | x_i, z_i) = (1-\phi)f(x_i, \rho) + \phi g(z_i, \hat{\gamma}), \quad (\text{II.31})$$

sendo $\hat{\gamma}$ obtido por qualquer método que permita assegurar a sua consistência sob H_2 . H_c' tem apenas mais um regressor que H_1 , mas, em contrapartida, também só é necessário estimar mais um parâmetro, pelo que todos eles poderão ser identificados.

Os regressores definidos por $g(z_i, \hat{\gamma})$ dependem de y_i através de $\hat{\gamma}$, estando correlacionados com o termo erro da regressão:

$$y_i = (1-\phi)f(x_i, \rho) + \phi g(z_i, \hat{\gamma}) + \varepsilon_{1i} \quad (\text{II.32})$$

DM (1981) estudaram as propriedades assintóticas do rácio t associado a ϕ , provando que, sob H_1 , a sua distribuição é a $N(0,1)$, podendo-se considerar $\hat{\gamma}$ não aleatório e ignorar a dependência existente entre $g(z_i, \hat{\gamma})$ e ε_{1i} .

A expressão II.31 é, então, a base do chamado teste J, sugerido por DM (1981), que principia, caso se queira testar H_1 , pela regressão II.32, ou, se se desejar testar H_2 , pela regressão

$$y_i = \phi f(x_i, \hat{\rho}) + (1-\phi)g(z_i, \hat{\gamma}) + \varepsilon_{2i}, \quad (\text{II.33})$$

visto que na equação II.32 a admissibilidade do rácio t para testar a nulidade de ϕ está condicionada pela verdade de H_1 e não de H_2 .

O teste termina com a análise do rácio t para a hipótese $\phi = 0$ (deve utilizar-se uma matriz de covariâncias robusta no cálculo de t , pois os erros das duas regressões são heterocedásticos). Também o teste J , tal como todos os testes de hipóteses não encaixadas, poderá originar quatro conclusões diferentes.

Se o modelo que se pretende testar não é linear, o teste J torna-se, por vezes, difícil de aplicar, o que levou DM (1981) a proporem uma alternativa, o teste P . Esta parte da linearização de II.32 e II.33 em torno de $\phi = 0$ e $\beta = \hat{\beta}$, da qual resulta a regressão (relativamente a II.32)

$$y_i - f(x_i, \hat{\rho}) = F(x_i, \hat{\rho})b + \phi[g(z_i, \hat{\gamma}) - f(x_i, \hat{\rho})] + \varepsilon_{3i}, \quad (\text{II.34})$$

onde $F(x_i, \hat{\rho}) = \left. \frac{\partial f(x_i, \rho)}{\partial \rho} \right|_{\rho=\hat{\rho}}$ e $b = \rho - \hat{\rho}$.

O teste P é assintoticamente equivalente ao J , sob a hipótese nula, sendo também completado com a análise do rácio t robusto, correspondente à hipótese $\phi = 0$, procedente da estimação de II.34 pelo método dos mínimos quadrados. A distribuição assintótica da estatística t , de acordo com a demonstração de DM (1981), é igualmente a $N(0,1)$.

Segundo estudos de simulação realizados por DM (1982) e Godfrey e Pesaran (1983), o comportamento do teste J em pequenas amostras não é o melhor, o que se deve ao facto de só assintoticamente ser legítimo afirmar a independência de $g(z_i, \hat{\gamma})$ e ε_{1i} em II.32. Ambos os trabalhos permitiram concluir que aquele teste leva a uma rejeição excessiva da hipótese nula, quando esta é verdadeira, principalmente se o número de regressores presentes no modelo alternativo for superior.

2.2.2. Testes de Fisher e McAleer e proposta de uma variante

FM (1981), seguindo uma sugestão de Atkinson (1970), argumentativa de que toda a estatística de teste deve ser avaliada apenas sob o modelo que é testado, aconselharam uma alternativa, o teste J_A . Tal como o J , ele baseia-se na regressão II.32, mas usa para γ uma estimativa consistente, sob H_1 , do seu limite em probabilidade. Na prática, este novo estimador, $\tilde{\gamma}$, pode ser obtido de acordo com o seguinte procedimento:

- 1º) estimar ρ por MV (ou qualquer outro método que permita encontrar estimadores consistentes para os parâmetros do modelo considerado sob H_1);
- 2º) estimar y_i para o modelo H_1 , $\hat{y}_i = f(x_i, \hat{\rho})$;
- 3º) determinar $\tilde{\gamma}$ através do método dos mínimos quadrados (não lineares, se necessário²⁸), admitindo como variável dependente \hat{y}_i .

Quando aplicado a modelos não lineares, em que unicamente se conhece a sua distribuição assintótica, $N(0,1)$, o teste J_A constitui um processo alternativo para testar hipóteses não encaixadas, embora seja mais fácil recorrer ao teste P_A , cuja extensão em relação a P é evidente.

DM (1981, 1982) provaram que os testes J , J_A , P e P_A são assintoticamente equivalentes, registando igual potência contra alternativas locais e sendo consistentes (à medida que a dimensão da amostra tende para infinito, a probabilidade de rejeitar a hipótese falsa tende para um).

O principal mérito do teste J_A é garantir a independência de $g(z_i, \tilde{\gamma})$ e ε_{ii} em II.32 (substituindo $\hat{\gamma}$ por $\tilde{\gamma}$), ao contrário do que se verificava no J . Em princípio, isto deveria levar a que o comportamento do teste J_A em pequenas amostras fosse superior. Contudo, os estudos de simulação referidos há pouco mostraram que tal só acontece quando o modelo que se testa é o verdadeiro. Aliás, se nenhum dos dois modelos em análise for o correcto, o seu desempenho será bastante pobre. O motivo essencial para esta fraca *performance* do teste J_A é a inclusão no modelo alternativo de demasiada

²⁸ Em modelos não lineares, o uso da MV não permite obter estimadores com as características de $\tilde{\gamma}$.

informação sobre aquele que se testa, implicando que este seja rejeitado em menor escala do que deveria acontecer.

De seguida, propõe-se um teste que procura ser um compromisso entre o J e o J_A : pretende-se reduzir a dependência do estimador de $g(z_i, \gamma)$ face a ε_{i1} em II.32, mas evitando o radicalismo de J_A . A sugestão que aqui se faz é muito simples: partir do estimador utilizado na computação do teste J, $\hat{\gamma}$, e dar um passo na direcção de $\tilde{\gamma}$, o estimador empregue no teste J_A , realizando uma única iteração. Obviamente, isto só faz sentido para modelos não lineares. O estimador assim obtido, $\tilde{\gamma}_1$, continuará a ser consistente para γ .

Assimptoticamente, os testes J e J_A são equivalentes: no limite, $\hat{\gamma}$ e $\tilde{\gamma}$ serão iguais. O estimador $\tilde{\gamma}_1$ deverá situar-se entre os outros dois sendo, também no limite, idêntico a eles. Assim, é de esperar que o teste baseado em $\tilde{\gamma}_1$, que será denominado de J_{A_1} (P_{A_1}), registre um melhor comportamento em amostras de pequenas dimensões, relativamente aos de DM (1981) e FM (1981).

Uma vez que $\tilde{\gamma}_1$ é um estimador consistente de γ , todo o desenvolvimento teórico de DM (1981) pode ser aplicado a J_{A_1} (P_{A_1}).

Capítulo III:

Estudo de Simulação de Monte Carlo

Neste capítulo procede-se a um estudo de simulação de Monte Carlo com vista a avaliar o desempenho em diversas condições amostrais de alguns dos testes expostos nas páginas anteriores, nomeadamente os aplicáveis a situações em que as amostras são truncadas e as hipóteses não encaixadas. Após a justificação e descrição das experiências efectuadas, apresentam-se os resultados obtidos e as principais conclusões extraídas.

1. Objectivos

Na impossibilidade de realizar uma análise de simulação abrangendo todos os tipos de testes desenvolvidos no capítulo anterior, apenas são tratados os referentes a hipóteses não encaixadas. Esta opção explica-se pelo facto de dois dos testes de sobredispersão (\hat{T}_{CT} , \hat{T}_{LB}) já terem sido alvo de uma investigação desta natureza em CT (1990), não suscitando a abordagem dos restantes grande interesse, em virtude dela ser, presumivelmente, muito semelhante.

Por outro lado, como grande parte dos testes de hipóteses não encaixadas considerados não podem ser usados em modelos relativos a amostras casuais simples de dados de contagem, resolveu-se orientar este estudo exclusivamente para os modelos truncados. Além disso, neste género de modelos, a obtenção de estimadores consistentes está totalmente dependente da veracidade da distribuição admitida, pelo que assume extrema importância dispor de um conjunto de testes capazes de detectarem incorrecções nos modelos.

De entre os testes indicados para hipóteses não encaixadas, dar-se-á especial

relevância aos baseados em regressões artificiais. Como os modelos que vão ser examinados são não lineares, só o teste P e as suas duas variantes serão tidos em conta. Assim, a comparação das *performances* dos testes P, P_A e P_{A_1} será o objectivo central desta pesquisa, o que, julga-se, a torna inédita, pois, até ao momento, não se tem conhecimento da elaboração de trabalhos com esta intenção²⁹. Adicionalmente, sempre que as alternativas em jogo o permitirem, computa-se o teste T_k .

O teste P e seus derivados constituem somente uma aproximação assintótica às respectivas versões J, as quais, por sua vez, também são válidas unicamente naquelas condições. Ora, nesta dissertação, os três testes P vão ser aplicados a modelos não lineares, onde, como é reconhecido, a teoria assintótica é mais lenta a produzir efeito, necessitando de um número de observações superior ao exigido noutras situações. Por este motivo, outras metas que se pretendem aqui atingir são, por um lado, a análise do comportamento dos testes de hipóteses não encaixadas em pequenas amostras; por outro, caracterizar o desempenho desses testes face a diferentes dimensões amostrais.

Ainda em relação a estes testes, a sua potência depende bastante da diferença registada entre os valores esperados condicionais estimados dos modelos em avaliação [por exemplo, no teste P, $g(z_i, \hat{\gamma}) - f(x_i, \hat{\rho})$, ver expressão II.34], o que é grave no caso concreto que vai ser tratado. Efectivamente, a especificidade dos modelos de contagem implica que, quando não há truncagem, a formalização do valor esperado condicional seja idêntica para todos eles. Quando as amostras são truncadas, ela já difere de modelo para modelo, sendo a amplitude dessa divergência função do grau de truncagem: se este for reduzido, o diferencial entre os valores esperados condicionais também o será. Deste modo, é provável que à medida que a truncagem diminua, isto é, $P(Y_i = 0|x_i)$ tenda para zero, $g(z_i, \hat{\gamma})$ e $f(x_i, \hat{\rho})$ se tornem praticamente indistinguíveis, prejudicando enormemente os resultados dos testes. Por isso, neste estudo, procura-se identificar a existência de uma relação entre a potência dos testes e o grau de truncagem da amostra.

Quando a sobredispersão é ínfima, todos os modelos de contagem abordados tendem para o Poisson. Logo, é previsível que a potência dos testes esteja directamente

²⁹ No que se refere aos testes de DM (1981) e FM (1981), apenas o comportamento dos testes J e J_A , aplicados a formulações lineares, foi objecto de uma investigação desta natureza [ver DM (1982) e Godfrey e Pesaran (1983), por exemplo].

relacionada com a magnitude daquela, justificando-se também, portanto, uma investigação neste sentido.

2. Descrição

O cumprimento dos objectivos delineados na secção anterior levou a que as experiências fossem agrupadas em três conjuntos principais:

- conjunto A: ensaios com diferentes graus de truncagem;
- conjunto B: ensaios com diferentes níveis de sobredispersão;
- conjunto C: ensaios com diferentes dimensões amostrais.

Para cada experiência, o modelo adoptado como verdadeiro foi sempre o BNT2. O processo de geração da variável dependente respeitante a este modelo foi dividido em duas fases, gerando-se primeiro o modelo BN2 e só depois se procedendo à truncagem da amostra.

O modelo BN2 foi obtido seguindo um raciocínio análogo ao exposto anteriormente (pp. 18-19), assente na suposição de que aquela distribuição resulta da existência de um erro de especificação no modelo de Poisson. Assim, para cada indivíduo gerou-se uma variável dependente com distribuição de Poisson de parâmetro:

$$\lambda_i = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}) \varepsilon_i \quad (\text{III.1})$$

Condicionando Y_i apenas em x_i , a distribuição da variável endógena será a BN2 se $\varepsilon_i \sim G\left(\frac{1}{\alpha}, \alpha\right)$. Este erro foi gerado recorrendo ao *package* TSP 4.3 [Hall (1995a, b)],

o qual foi igualmente utilizado na realização de toda a restante análise simulatória. Como este programa não gera automaticamente a distribuição Gama, cada ε_i correspondeu à soma de $\frac{1}{\alpha}$ variáveis exponenciais de parâmetro α .

O modo de truncagem dos dados não diferiu do usual em aplicações deste género [Grogger e Carson (1991)]: desprezar todos os zeros gerados para a variável dependente.

Com este procedimento, correu-se o risco de se ficar com amostras de tamanhos desiguais em situações em que tal não era desejado. Em réplicas pertencentes à mesma experiência, como o valor dos parâmetros é idêntico, a quantidade esperada de observações geradas também o é, pelo que este problema não é grave. Quando as amostras integram ensaios distintos, a questão é mais complicada. Contudo, adoptando uma dimensão amostral adequada para cada experiência, conseguiu-se alcançar, após a exclusão dos zeros, um número médio de observações semelhante (excluindo os ensaios do conjunto C, é claro).

De maneira a evitar o mais possível a excessiva dependência de todo este estudo face a um dado conjunto de regressores, as variáveis exógenas X_1 e X_2 e o termo ε_i foram alterados de réplica para réplica. Enquanto que X_1 representa uma variável dummy, com 66% de probabilidade de ser zero, obtida a partir da geração auxiliar de uma uniforme, X_2 foi gerado como sendo uma variável normal estandardizada. Aos parâmetros β_1 e β_2 atribuiu-se o valor 1.

Nas experiências de tipo A, a fixação de diferentes graus de truncagem efectuou-se fazendo variar β_0 , para o qual se admitiram os seguintes valores: -0.5, 0, 0.5, 1, 2 e 3. Nos ensaios de tipo B, o parâmetro α assumiu os valores 0.125, 0.25 e 0.5, o que possibilitou a construção de modelos com níveis de sobredispersão distintos. Nestes dois conjuntos de experiências, as amostras foram constituídas, em média, depois da eliminação dos zeros, por 500 observações, dimensão que pode ser encarada como relativamente pequena, uma vez que os modelos aplicados são microeconómicos. Nos restantes ensaios, geraram-se amostras compostas, em média, por 250, 500, 1000 e 2000 indivíduos.

Cada experiência foi baseada num conjunto de 2000 réplicas, estimando-se, para cada uma delas, três modelos, o BNT1, o BNT2 e o PGT2. Consequentemente, computaram-se os testes P , P_A e P_{A_1} para os seguintes pares de hipóteses:

- a) H_1 : BNT1; H_2 : BNT2;
- b) H_1 : BNT1; H_2 : PGT2;
- c) H_1 : BNT2; H_2 : BNT1;
- d) H_1 : BNT2; H_2 : PGT2.

e) H_1 : PGT2; H_2 : BNT1;

f) H_1 : PGT2; H_2 : BNT2.

A consideração dos pares b) e e) teve como objectivo avaliar o comportamento dos testes na presença de dois modelos falsos. O teste T_k apenas foi feito, obviamente, para os pares a) e c), tendo-se optado por estimá-lo através de um rácio de verosimilhanças³⁰. Em todos os casos, efectuaram-se testes unilaterais para um nível de significância de 5%, pelo que o valor crítico que serviu como critério para a rejeição da hipótese nula foi 1.645.

No teste P, os parâmetros dos modelos referentes à hipótese H_2 foram estimados por MV, enquanto que no P_A e no P_{A_1} se recorreu ao método dos mínimos quadrados não lineares. Na regressão inerente a cada um destes testes, ao invés de se usar, como indicado anteriormente, uma matriz de covariâncias robusta à heterocedasticidade, adoptou-se um processo análogo ao proposto por DM (1993), p. 525, para modelos binários. Assim, em relação ao teste P, em lugar de se estimar a equação II.34, multiplicaram-se ambos os membros desta expressão por $[V_{H_0}(Y_i|x_i, Y_i > 0)]^{-\frac{1}{2}}$, pelo que a regressão artificial realizada foi:

$$\begin{aligned} \{\hat{V}[f(x_i, \rho)]\}^{-\frac{1}{2}}[y_i - f(x_i, \hat{\rho})] = \{\hat{V}[f(x_i, \rho)]\}^{-\frac{1}{2}}F(x_i, \hat{\rho})b + \\ + \phi\{\hat{V}[f(x_i, \rho)]\}^{-\frac{1}{2}}[g(z_i, \hat{\gamma}) - f(x_i, \hat{\rho})] + u_i \quad (\text{III.2}) \end{aligned}$$

Sob a hipótese nula, a distribuição assintótica da estatística de teste assim encontrada continua a ser a $N(0,1)$. Para os testes P_A e P_{A_1} , a equação de regressão base foi também esta mas, naturalmente, utilizando para γ os estimadores característicos de cada teste.

Em anexo, fornecem-se cópias dos programas-tipo empregues (anexo B), bem como as expressões necessárias à execução dos testes que ainda não foram expostas

³⁰ Embora qualquer uma das outras duas vias (score, Wald) também fosse admissível, experiências preliminares revelaram claramente que o rácio de verosimilhanças constituía a melhor opção para efectuar o teste T_k . Como o objectivo deste estudo não é a avaliação dos diversos modos de computar este teste, resolveu-se apresentar apenas os resultados relativos à situação referida.

neste trabalho (anexo A).

3. Resultados

Estabelecidas as principais directrizes deste estudo de simulação, passa-se à descrição e interpretação dos resultados obtidos.

3.1. Ensaio com diferentes graus de truncagem

As primeiras experiências levadas a cabo disseram respeito ao nível de truncagem das amostras. As percentagens de rejeições da hipótese nula constam dos quadros 1 e 2.

No quadro 1, são apresentadas as dimensões estimadas dos vários testes para uma situação padrão em que o número médio de observações por amostra (\bar{n}) é de 500 e o parâmetro perturbador, α , assume o valor 0.25. Só o parâmetro β_0 e, conseqüentemente, o grau de truncagem, varia de ensaio para ensaio. O nível de truncagem que se indica para cada modelo foi determinado comparando o número total de observações que foi necessário gerar para alcançar a dimensão média amostral pretendida, com a quantidade de zeros eliminados. Não se consideraram modelos com maior truncagem dado que, nos ensaios desse tipo efectuados, a convergência dos estimadores tornou-se muito difícil de conseguir, talvez porque a variabilidade dos regressores deixou praticamente de se reflectir nos valores assumidos pela variável dependente³¹.

Também no quadro 1, entre parêntesis encontra-se o valor da estatística t para a hipótese de que a dimensão real dos vários testes computados é 5%. Com efeito, encarando cada ensaio como um conjunto de 2000 experiências de Bernoulli³², é

³¹ Para se conseguirem obter amostras caracterizadas por altos graus de truncagem, o valor atribuído ao parâmetro β_0 tem de ser tal que implique que $P(Y_i = 0|x_i)$ seja muito elevada. O que acontece é que, nestas condições, a maior parte das observações da variável dependente são geradas com base em médias inferiores à unidade, levando a que os valores de Y_i , que têm de ser inteiros, fiquem compreendidos num intervalo de amplitude bastante reduzida, não expressando convenientemente as diferenças existentes ao nível das variáveis explicativas.

³² Ver DM (1993), p. 739.

possível testar a dimensão empírica, \hat{p} , de cada teste, dada a sua dimensão nominal, p :

$$t = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{2000}}} \approx N(0,1) \quad (\text{III.3})$$

Considerando um nível de significância de 5%, o valor crítico para este teste é 1.96.

Quadro 1: Dimensões estimadas^a para diferentes graus de truncagem e testes para a hipótese de que a dimensão real é 5% (H_1 : BNT2; $\bar{n} = 500$; $\alpha = 0.25$)

β_0	Grau Trunc.	H_2 : BNT1				H_2 : PGT2		
		T_k	P	P_{A1}	P_A	P	P_{A1}	P_A
-0.5	45.8%	5.90 (1.85)	6.05 (2.15)	5.95 (1.95)	5.95 (1.95)	5.70 (1.44)	5.50 (1.03)	5.50 (1.03)
0.0	34.1%	5.95 (1.95)	4.90 (-0.02)	4.65 (-0.72)	4.70 (-0.62)	5.45 (0.92)	5.45 (0.92)	5.30 (0.62)
0.5	23.6%	5.80 (1.64)	5.30 (0.62)	5.25 (0.51)	5.25 (0.51)	4.60 (-0.82)	4.25 (-1.54)	4.15 (-1.74)
1.0	15.0%	5.25 (0.51)	4.80 (-0.41)	4.70 (-0.62)	4.75 (-0.51)	5.10 (0.21)	4.35 (-1.33)	4.40 (-1.23)
2.0	4.5%	6.45 (2.98)	5.20 (0.41)	5.20 (0.41)	5.10 (0.21)	6.60 (3.28)	5.55 (1.13)	5.65 (1.33)
3.0	0.9%	5.10 (0.21)	5.55 (1.13)	5.75 (1.54)	5.90 (1.85)	7.70 (5.54)	5.35 (0.72)	5.10 (0.21)

^a Valores em percentagem para um nível de significância nominal de 5%.

Como se verifica, os resultados produzidos por este primeiro conjunto de experiências foram mais ou menos os esperados. De acordo com o previsto, o teste P_A registou um bom desempenho: em nenhuma ocasião se pôde rejeitar que a sua dimensão fosse realmente 5%. O mesmo se passou com o teste P_{A1} , que começa assim a confirmar as esperanças nele depositadas. O comportamento destes dois testes foi muito similar em todos os ensaios realizados nesta fase, o mesmo acontecendo com o teste P mas somente para o caso em que o modelo BNT1 era a alternativa.

Quando a hipótese H_2 passou a ser o modelo PGT2, o teste P revelou uma certa tendência para a rejeição excessiva do verdadeiro modelo, embora unicamente para graus de truncagem muito reduzidos. Também aqui não há surpresas, uma vez que todos

os estudos de simulação empreendidos até hoje indicam que a versão J do teste P possui, em muitas circunstâncias, dimensão real superior à nominal, pelo que não é de estranhar que o mesmo se passe com este.

Quanto ao teste T_k , a dimensão estimada foi satisfatória, rejeitando-se a hipótese dela ser 5% uma única vez. Portanto, aparentemente, o único teste cuja dimensão real parece ser afectada pelo nível de truncagem é o P.

Quadro 2: Potências estimadas^a para diferentes graus de truncagem ($\bar{n} = 500$; $\alpha = 0.25$)

β_0	Grau Trunc.	H_2	H_1 : BNT1				H_2	H_1 : PGT2		
			T_k	P	P_{A1}	P_A		P	P_{A1}	P_A
-0.5	45.8%	BNT2	99.75	25.10	23.85	22.60	BNT1	5.35	5.00	5.10
0.0	34.1%		100.00	25.05	23.60	22.65		4.80	4.60	4.65
0.5	23.6%		100.00	24.11	22.91	22.60		5.50	5.40	5.35
1.0	15.0%		100.00	25.05	24.05	24.70		4.90	4.85	4.85
2.0	4.5%		100.00	19.15	21.55	22.85		5.80	5.40	5.55
3.0	0.9%		100.00	12.10	16.85	18.40		7.00	7.20	6.95
-0.5	45.8%	PGT2	---	26.00	24.55	^b	BNT2	4.85	4.65	4.70
0.0	34.1%		---	26.25	24.50	^b		4.75	4.60	4.60
0.5	23.6%		---	25.36	23.81	^b		4.90	4.35	4.35
1.0	15.0%		---	26.55	25.10	^b		5.45	4.85	4.85
2.0	4.5%		---	20.30	20.65	^b		6.75	4.75	4.70
3.0	0.9%		---	12.70	13.75	^b		9.45	6.30	6.50

^a Valores em percentagem para um nível de significância nominal de 5%.

^b Não determinada.

A leitura do quadro 2 permite chegar a algumas conclusões surpreendentes. Começando pela abordagem conjunta dos testes P, P_A e P_{A1} , verifica-se que a sua potência depende bastante da hipótese que é testada. Quando a hipótese nula é o modelo BNT1 e a alternativa é o BNT2, aqueles testes comportam-se razoavelmente pois, embora a potência empírica não seja muito elevada, ela é significativamente superior às dimensões estimadas. Substituindo o BNT2 pelo modelo PGT2, as potências determinadas praticamente não se alteram. Em qualquer dos casos, comprova-se que a truncagem tem o efeito esperado, mas só até certa altura: a potência é reduzida quando quase não existe truncagem, aumenta até aos 15% de truncagem, estagnando a partir daí.

Admitindo como hipótese nula o modelo PGT2, facilmente se conclui que, de modo algum, os testes baseados em regressões artificiais devem ser usados como critério de escolha. Quer a alternativa seja o BNT1, quer seja o BNT2, a potência dos

testes é muito baixa. Inclusivamente, comparando os quadros 1 e 2, observa-se que, nalguns dos ensaios, a proporção de rejeições do modelo verdadeiro (BNT2) é superior à do falso (PGT2). Além disto, repare-se que é sempre praticamente indiferente testar o modelo BNT1 “contra” o BNT2 ou “contra” o PGT2 e vice-versa. Como estes testes consideram apenas as diferenças ocorridas entre os modelos ao nível do primeiro momento estimado, tudo leva a crer que, em todas as circunstâncias, o valor esperado condicional seja muito semelhante para os modelos BNT2 e PGT2.

Individualizando agora o estudo, constata-se que o desempenho dos três testes P é muito similar. A potência empírica é sempre bastante parecida, embora o teste P pareça levar uma ligeira vantagem para truncagens superiores e o P_A na situação oposta. De realçar as dificuldades sentidas na computação deste último teste para o par de hipóteses H_1 : BNT1, H_2 : PGT2. Foi impossível, em mais de metade das réplicas, efectuar a regressão necessária à realização do teste, pelo que não se mostram aqui os resultados obtidos.

Relativamente ao teste T_k , presumivelmente por assentar na comparação de todos os momentos implícitos na função de probabilidade de cada modelo, as potências estimadas atingiram níveis de tal ordem que não há a mínima dúvida da sua capacidade para detectar a inadequabilidade do modelo BNT1. Infelizmente, este teste não pode ser empregue para testar a veracidade do modelo PGT2.

As potências empíricas do teste T_k parecem apontar para a sua independência do grau de truncagem. Todavia, esta conclusão pode ser apenas aparente, pois a construção do exemplo que serve de suporte a este terceiro capítulo teve como primeira preocupação a caracterização dos testes P, P_A e P_{A_1} , não sendo forçosamente o mais indicado para estudar pormenorizadamente o T_k .

Terminada esta análise, investiga-se de seguida a influência do nível de sobredispersão no comportamento de todos os testes de hipóteses não encaixadas que se têm vindo a examinar neste capítulo.

3.2. Ensaio com diferentes níveis de sobredispersão

As experiências incluídas nesta secção basearam-se na atribuição de diferentes valores a α , parâmetro que determina o nível de sobredispersão de cada modelo. Devido à forma de geração dos dados (ver p. 72), a escolha de α ficou muito limitada, pelo que apenas se puderam admitir três níveis de sobredispersão, correspondentes aos valores 0.125, 0.25 e 0.5 daquele parâmetro. Com o intuito de averiguar se o efeito provocado pelas variações de α é o mesmo na presença de diferentes graus de truncagem, adoptaram-se para β_0 os valores 0, 1, 2 e 3. A dimensão média amostral continuou a ser 500.

Quadro 3: Dimensões estimadas^a para diferentes níveis de sobredispersão (H_1 : BNT2; $\bar{n} = 500$)

β_0	Grau Trunc.	α	H_2 : BNT1				H_2 : PGT2		
			T_k	P	P_{A1}	P_A	P	P_{A1}	P_A
0	32.4%	0.125	6.80	6.30	6.20	6.15	5.05	5.00	4.95
	34.1%	0.250	5.95	4.90	4.65	4.70	5.45	5.45	5.30
	37.5%	0.500	4.80	5.45	5.25	5.30	4.70	4.50	4.45
1	13.3%	0.125	5.50	5.85	5.85	5.80	6.30	5.00	4.95
	15.0%	0.250	5.25	4.80	4.70	4.75	5.10	4.35	4.40
	18.4%	0.500	4.60	5.50	5.45	5.35	6.35	5.55	5.40
2	3.5%	0.125	5.95	5.60	5.30	5.20	6.45	4.95	4.75
	4.5%	0.250	6.45	5.20	5.20	5.10	6.60	5.55	5.65
	6.9%	0.500	5.40	5.50	5.60	5.65	6.65	4.70	4.70
3	0.6%	0.125	5.85	5.95	6.20	6.45	8.60	5.35	5.20
	0.9%	0.250	5.10	5.55	5.75	5.90	7.70	5.35	5.10
	1.9%	0.500	5.70	5.25	5.30	5.45	6.50	5.35	4.80

^a Valores em percentagem para um nível de significância nominal de 5%.

Quadro 4: Potências estimadas^a para diferentes níveis de sobredispersão (H_2 : BNT2; $\bar{n} = 500$)

β_0	Grau Trunc.	α	H_1 : BNT1				H_1 : PGT2		
			T_k	P	P_{A1}	P_A	P	P_{A1}	P_A
0	32.4%	0.125	98.45	17.10	15.90	16.05	4.80	4.45	4.45
	34.1%	0.250	100.00	25.05	23.60	22.65	4.75	4.60	4.60
	37.5%	0.500	100.00	33.90	33.00	29.65	4.80	4.75	4.85
1	13.3%	0.125	100.00	16.40	16.20	15.75	4.65	4.70	4.75
	15.0%	0.250	100.00	25.05	24.05	24.70	5.45	4.85	4.85
	18.4%	0.500	100.00	32.55	32.10	^b	6.60	5.30	5.15
2	3.5%	0.125	100.00	12.35	15.60	15.10	6.75	4.85	4.90
	4.5%	0.250	100.00	19.15	21.55	22.85	6.75	4.75	4.70
	6.9%	0.500	100.00	27.15	27.40	^b	10.20	5.50	5.55
3	0.6%	0.125	100.00	5.80	13.10	11.70	7.50	5.80	5.85
	0.9%	0.250	100.00	12.10	16.85	18.40	9.45	6.30	6.50
	1.9%	0.500	100.00	22.90	24.20	^b	9.60	8.25	6.65

^a Valores em percentagem para um nível de significância de 5%.

^b Não determinada.

Os principais resultados estimados constam dos quadros 3 e 4, só se apresentando os referentes aos pares de hipóteses em que o modelo BNT2 intervém. Efectivamente, tal como aconteceu na pesquisa anterior, também aqui cedo se tornou evidente que testar o BNT1 “contra” o BNT2 ou “contra” o PGT2, e vice-versa, produzia consequências muito parecidas³³.

Em termos gerais, ressalta imediatamente da análise destes dois quadros a manutenção das características fundamentais dos vários testes, isto é, não foi o nível de sobredispersão escolhido no ponto anterior que determinou os resultados então encontrados. Mais, reforçaram-se algumas ideias que o estudo realizado no ponto anterior já sugeria mas não mostrava com total clareza:

- a tendência do teste P para sobre-rejeitar a hipótese nula quando a alternativa é o modelo PGT2 e a truncagem é relativamente reduzida (o intervalo de confiança a 95% para \hat{p} , calculado a partir de III.3, é [0.0404 ; 0.0596]);
- a supremacia dos testes P_{A_1} e P_A face ao P para provar a falsidade do modelo BNT1 quando o grau de truncagem é baixo, situação completamente inesperada à partida;
- as dificuldades de computação do teste P_A (novamente quando a hipótese nula era o modelo BNT1, mas agora também “contra” o BNT2, quando $\alpha = 0.5$).

Adicionalmente, podem fazer-se duas constatações importantes, ambas relativas aos testes baseados em regressões artificiais. A primeira refere-se ao acréscimo claro que as potências dos testes P, P_A e P_{A_1} sofrem com o aumento da sobredispersão. Esta influência é a esperada pois, à medida que o valor de α aumenta, acentuam-se as diferenças entre os vários modelos, sendo lógica a melhoria da *performance* dos testes.

A segunda ilacção que pode ser retirada diz respeito à dimensão destes testes. Também ela, quando se avalia o modelo BNT2 face ao BNT1, parece depender um pouco do valor de α : quando este assume o valor 0.125, há dúvidas sobre o correcto comportamento dos testes, independentemente do grau de truncagem; ao aumentar a

³³ Os eventuais interessados nestes resultados poderão contactar o autor.

sobredispersão, elas desaparecem por completo.

Assim, tendo em conta o exposto, pode-se afirmar, sem reservas, o efeito benéfico da sobredispersão sobre o desempenho de todos os testes de hipóteses não encaixadas quando aplicados a modelos de contagem deste tipo.

3.3. Ensaios com diferentes dimensões amostrais

Para terminar este estudo de simulação, falta apenas apresentar e interpretar os dados relativos às características dos vários testes na presença de diferentes dimensões amostrais. Os quadros 5 e 6 resumem os resultados obtidos para um nível de sobredispersão médio e para três graus de truncagem distintos.

Quadro 5: Dimensões estimadas^a para diferentes tamanhos amostrais
(H_1 : BNT2; $\alpha = 0.25$)

β_0	Grau Trunc.	\bar{n}	H_2 : BNT1				H_2 : PGT2		
			T_k	P	P_{A1}	P_A	P	P_{A1}	P_A
0	34.1%	250	6.60	5.80	5.50	6.15	4.90	4.55	4.95
		500	5.95	4.90	4.65	4.70	5.45	5.45	5.30
		1000	6.20	4.50	4.50	4.55	5.30	5.00	5.00
		2000	5.30	4.95	4.85	4.95	5.10	5.10	5.05
1	15.0%	250	6.30	6.00	5.90	5.95	6.65	5.45	5.45
		500	5.25	4.80	4.70	4.75	5.10	4.35	4.40
		1000	4.10	5.10	5.00	5.00	5.55	5.45	5.50
		2000	5.10	5.45	5.25	5.35	4.60	4.40	4.50
3	0.9%	250	6.05	4.80	5.35	5.55	8.15	4.95	4.35
		500	5.10	5.55	5.75	5.90	7.70	5.35	5.10
		1000	5.10	5.85	5.75	6.00	6.00	4.50	4.65
		2000	5.55	5.45	5.60	5.85	6.00	4.80	4.95

^a Valores em percentagem para um nível de significância nominal de 5%.

Os valores estimados para cada teste não deixam margem para dúvidas: logicamente, à medida que a dimensão da amostra aumenta, as *performances* dos vários testes evoluem de forma marcadamente positiva. A única exceção continua a ser o caso em que se testa o modelo PGT2 “contra” o BNT2, onde não há mesmo nada a fazer.

Repare-se que para uma amostra de 250 observações, metade da considerada anteriormente, surgem novos problemas pois rejeita-se sempre a hipótese que a dimensão real do teste T_k seja 5%, o mesmo se passando com o teste P para $\beta_0 = 1$, o que

até aqui nunca sucedera.

Quadro 6: Potências estimadas^a para diferentes tamanhos amostrais
(H_2 : BNT2; $\alpha = 0.25$)

β_0	Grau Trunc.	\bar{n}	H_1 : BNT1				H_1 : PGT2		
			T_k	P	P_{A1}	P_A	P	P_{A1}	P_A
0	34.1%	250	97.85	22.35	20.20	14.10	4.65	4.35	4.45
		500	100.00	25.05	23.60	22.65	4.75	4.60	4.60
		1000	100.00	32.30	31.40	30.75	5.15	5.15	5.15
		2000	100.00	39.65	38.65	37.80	5.30	5.25	5.30
1	15.0%	250	99.90	21.30	20.50	20.10	5.10	4.55	4.65
		500	100.00	25.05	24.05	24.70	5.45	4.85	4.85
		1000	100.00	32.15	31.55	31.35	5.50	5.05	5.00
		2000	100.00	39.90	39.95	39.50	5.15	4.95	4.85
3	0.9%	250	100.00	9.65	13.55	14.95	10.40	6.85	5.80
		500	100.00	12.10	16.85	18.40	9.45	6.30	6.50
		1000	100.00	20.40	24.75	24.95	8.65	6.90	7.15
		2000	100.00	29.45	32.50	31.10	7.00	6.30	6.10

^a Valores em percentagem para um nível de significância nominal de 5%.

Aumentando o número de observações, aquelas questões desaparecem. Todavia, mesmo quando a amostra é de 2000 indivíduos, continua a verificar-se um desempenho menos bom: o teste P, quando a truncagem é quase nula e a alternativa é o modelo PGT2, é novamente caracterizado por apresentar dimensões estimadas superiores às nominais, mas agora só por muito pouco se rejeita a hipótese delas serem iguais.

Em simultâneo com o aumento do número de indivíduos, é de notar o lento mas seguro incremento da potência daquele teste, e dos seus derivados, que atinge níveis razoáveis mesmo quando praticamente não existe truncagem (somente para H_1 : BNT1, é claro). De qualquer modo, só em amostras de muito maior dimensão, a teoria assintótica dos testes baseados em regressões artificiais se deverá poder aplicar, na plenitude, aos modelos de contagem, pelo menos aos que foram abordados neste estudo.

A finalizar, refira-se o facto de mais uma vez ser evidente que a truncagem exerce um efeito positivo sobre a potência dos testes P's, mas apenas numa fase inicial, sendo depois indiferente o grau daquela.

Conclusão

Neste trabalho foram expostos os principais modelos de regressão para dados de contagem. Para lá da análise das formulações tradicionais, deu-se especial atenção ao modelo de Poisson Generalizado, por se encontrar ainda pouco divulgado na literatura econométrica, e aos modelos truncados em zero, pois nos fenómenos económicos é frequente a observação apenas de valores positivos.

Quando a truncagem não está presente na amostra recolhida, é possível usar métodos semi-paramétricos, robustos contra pressupostos distribucionais, que garantam a obtenção de estimadores consistentes para os parâmetros do valor esperado condicional. Além do clássico método dos momentos generalizados, nos modelos pertencentes à FEL, Poisson, BN2 e PG2, pode ser empregue o método da Pseudo Máxima Verosimilhança. Em todos os casos, o pressuposto inerente à sua aplicação é a correcta especificação do primeiro momento condicional, tornando-se importante testar se este está bem formalizado. Um teste próprio para isso, não incluído neste estudo, é o teste RESET de Ramsey (1969), extendido aos modelos microeconómicos por Pagan e Vella (1989).

A partir do momento em que se tem necessidade de recorrer a modelos truncados, a acertada especificação do valor esperado condicional passa a estar dependente do facto do modelo admitido ser realmente o verdadeiro. Nesta situação, e também sempre que se pretender realizar inferências probabilísticas com base num dado modelo estimado, é fundamental testar a adequabilidade da distribuição condicional considerada para a variável endógena. Deste modo, uma das grandes preocupações desta tese foi fornecer meios para tal.

Para testar o modelo de Poisson, os testes mais populares são os de sobredispersão. Devido à restrição que é testada se situar no limite do espaço do parâmetro perturbador, só testes score podem ser utilizados. As deduções efectuadas mostraram que o BN2 e o PG2, por um lado, e o BN1 e o PG1, por outro, quer sejam ou

não truncados, constituem alternativas localmente equivalentes, pelo que é indiferente testar o modelo de Poisson contra qualquer um dos dois modelos integrantes de cada conjunto.

Avaliar a qualidade dos outros modelos de contagem é um pouco mais difícil, uma vez que se trata de formulações não encaixadas. Por conseguinte, propôs-se o recurso a testes próprios para hipóteses desta natureza. Resumiram-se alguns dos testes deste género aplicáveis aos modelos de contagem e propuseram-se dois novos testes: o primeiro, usando como modelo geral o BNk de Winkelmann e Zimmermann (1991, 1995) ou o PG3 de Santos Silva (1996a), permite testar as duas formas principais quer dos modelos BN's quer dos PG's; o segundo, consiste numa pequena modificação dos testes baseados em regressões artificiais.

O estudo de simulação de Monte Carlo empreendido teve como finalidade examinar as características deste último tipo de testes quando empregues em modelos truncados. Embora seja certo que os resultados obtidos estão directamente ligados aos exemplos construídos, foi possível tirar algumas conclusões que se julgam facilmente generalizáveis a outros casos.

Assim, um aspecto não deixa dúvidas: nenhum dos testes considerados serve para testar o modelo PGT2 “contra” os BNT's. Pelo contrário, no âmbito exclusivo dos modelos BNT's, ou quando o PGT2 funciona como H_2 , o panorama é bem mais agradável, pois qualquer um dos testes abordados revelou um comportamento razoável, devendo-se, no entanto, ter um pouco de cuidado na utilização dos testes P's nalgumas circunstâncias (sobredispersão reduzida e/ou truncagem diminuta associadas a dimensões amostrais relativamente pequenas). Por isso, o teste a que se aconselha a recorrer, sempre que tal for possível, é, naturalmente, o T_k .

De entre os vários testes P's, o novo teste sugerido, P_{A_1} , revelou-se promissor, parecendo possuir, em alguns factores, vantagens em relação aos outros:

- o teste P_A apresenta um comportamento similar, mas a sua execução é muito mais demorada, além de, por vezes, não se ter conseguido estimá-lo, se bem que, suspeita-se, devido a insuficiências do *software* usado;
- o teste P é mais simples e rápido de executar mas sobre-rejeita a hipótese nula

em determinadas condições.

Em suma, seria interessante estudar mais aprofundadamente as características do teste P_{A_1} , nomeadamente em situações semelhantes àquelas onde o desempenho do J_A não foi famoso (não esquecer que o P_{A_1} só se aplica a modelos não lineares, não existindo aquilo a que se chamaria teste J_{A_1}), visto que também a *performance* da sua versão P_A excedeu o previsto.

Anexos

Anexo A

Formulário complementar

1) Modelo BNT1:

$$\left. \frac{\partial \mu_i^*}{\partial \beta} \right|_{\substack{\beta=\hat{\beta} \\ \alpha=\hat{\alpha}}} = x_i' \hat{\mu}_i^* \left[1 - \frac{\hat{\eta}_i \ln(1 + \hat{\alpha})}{\hat{\alpha}} \right]$$

$$\left. \frac{\partial \mu_i^*}{\partial \alpha} \right|_{\substack{\beta=\hat{\beta} \\ \alpha=\hat{\alpha}}} = \frac{\hat{\mu}_i^* \hat{\eta}_i}{\hat{\alpha}} \left[\frac{\ln(1 + \hat{\alpha})}{\hat{\alpha}} - \frac{1}{1 + \hat{\alpha}} \right]$$

2) Modelo BNT2:

$$\left. \frac{\partial \mu_i^*}{\partial \beta} \right|_{\substack{\beta=\hat{\beta} \\ \alpha=\hat{\alpha}}} = x_i' \hat{\mu}_i^* \left(1 - \frac{\hat{\eta}_i}{1 + \hat{\alpha} \hat{\mu}_i} \right)$$

$$\left. \frac{\partial \mu_i^*}{\partial \alpha} \right|_{\substack{\beta=\hat{\beta} \\ \alpha=\hat{\alpha}}} = \frac{\hat{\mu}_i^* \hat{\eta}_i}{\hat{\alpha}} \left[\frac{\ln(1 + \hat{\alpha} \hat{\mu}_i)}{\hat{\alpha} \hat{\mu}_i} - \frac{1}{1 + \hat{\alpha} \hat{\mu}_i} \right]$$

3) Modelo PGT2:

$$\left. \frac{\partial \mu_i^*}{\partial \beta} \right|_{\substack{\beta=\hat{\beta} \\ \alpha=\hat{\alpha}}} = x_i' \hat{\mu}_i^* \left[1 - \frac{\hat{\eta}_i}{(1 + \hat{\alpha} \hat{\mu}_i)^2} \right]$$

$$\left. \frac{\partial \mu_i^*}{\partial \alpha} \right|_{\substack{\beta=\hat{\beta} \\ \alpha=\hat{\alpha}}} = \frac{\hat{\mu}_i^* \hat{\mu}_i \hat{\eta}_i}{(1 + \hat{\alpha} \hat{\mu}_i)^2}$$

Anexo B:

Programas-tipo utilizados no TSP

1. Teste P

```
OPTIONS CRT, DOUBLE;
SUPRES @SMPL @LOGL @NOB @SES @COEF;
SET REP=2000;
SET N=588;
SET C0V=1; SET C1V=1; SET C2V=1; SET A=0.25;
SET VALA=0.75; SET VALB=-1.5; SET VALC=-2.5;
SET P=1/A;
FRML MIU MI=EXP(C0+C1*DUM+C2*X2);
FRML ALFA ALF=EXP(AL);
FRML FO1 LGAMFN(Y+(MI**(1-K))/ALF)-LGAMFN((MI**(1-K))/ALF);
FRML FO2 Y*LOG(ALF*(MI**K))-(Y+(MI**(1-K))/ALF)*LOG(1+ALF*(MI**K));
FRML FO3 LOG(1-(1+ALF*(MI**K))**(-(MI**(1-K))/ALF));
FRML BNTK FOBNT2=FO1+FO2-FO3;
EQSUB BNTK FO1 FO2 FO3 MIU ALFA;
FRML FO4 Y*LOG(MI)+(Y-1)*LOG(1+ALF*Y)-Y*LOG(1+ALF*MI);
FRML FO5 MI*(1+ALF*Y)/(1+ALF*MI)+LOG(1-EXP(-MI/(1+ALF*MI)));
FRML PGT2 FOPGT2=FO4-FO5;
EQSUB PGT2 FO4 FO5 MIU ALFA;
SMPL 1 REP;
DOT 1-6; T.=@MISS; ENDDOT;
SMPL 1 N;
Y=0;
SET NOB=0;
smpl 1 1; random (seedin=0603952200) asd;
DO I=1 TO REP;
?
?                               Geração do modelo BNT2
?
SMPL 1 N;
RANDOM (UNIFORM) X1;
DUM = X1>0.66;
RANDOM X2;
DO J=1 TO N;
  SMPL 1 P;
  RANDOM (EXPON,LAMBDA=P) Z;
  MAT E1=SUM(Z);
  SET E=E1(1);
  SMPL J J;
  SET M=EXP(C0V+C1V*DUM+C2V*X2)*E;
  RANDOM (MEAN=M, POISSON) Y;
ENDDO;
SMPL 1 N;
SMPLIF Y;
SET NOB=NOB+@NOB;
?
?                               Estimação do modelo BNT1
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALA;
PARAM C0 C1 C2 AL K 0; CONST K;
ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) BNTK;
PARAM C0 C1 C2 AL K 0; CONST K;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
```

```

if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0A=C0;
SET B1A=C1;
SET B2A=C2;
SET PPA=EXP(AL);
MA=EXP(B0A+B1A*DUM+B2A*X2);
YEA=MA/(1-(1+PPA)**(-MA/PPA));
AJA=YEA-MA;
PONDA=(YEA*(1+PPA-AJA)**0.5;
VDA=(Y-YEA)/PONDA;
F0A=(YEA*(1-AJA*LOG(1+PPA)/PPA))/PONDA;
F1A=DUM*F0A;
F2A=X2*F0A;
F3A=(YEA*AJA/PPA*(LOG(1+PPA)/PPA-1/(1+PPA)))/PONDA;
?
?                               Estimação do modelo BNT2
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALB;
PARAM C0 C1 C2 AL K 1; CONST K;
ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) BNTK;
PARAM C0 C1 C2 AL K 1; CONST K;
ML(SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0B=C0;
SET B1B=C1;
SET B2B=C2;
SET PPB=EXP(AL);
MB=EXP(B0B+B1B*DUM+B2B*X2);
YEB=MB/(1-(1+PPB*MB)**(-1/PPB));
AJB=YEB-MB;
PONDB=(YEB*(1+PPB*MB-AJB)**0.5;
VDB=(Y-YEB)/PONDB;
F0B=(YEB*(1-AJB/(1+PPB*MB)))/PONDB;
F1B=DUM*F0B;
F2B=X2*F0B;
F3B=(YEB*AJB/PPB*(LOG(1+PPB*MB)/(PPB*MB)-1/(1+PPB*MB)))/PONDB;
?
?                               Estimação do modelo PGT2
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALC;
PARAM C0 C1 C2 AL;
ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) PGT2;
PARAM C0 C1 C2 AL;
ML(SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) PGT2;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0C=C0;
SET B1C=C1;
SET B2C=C2;
SET PPC=EXP(AL);
MC=EXP(B0C+B1C*DUM+B2C*X2);
YEC=MC/(1-EXP(-MC/(1+PPC*MC)));
AJC=YEC-MC;
PONDC=(YEC*((1+PPC*MC)**2-AJC)**0.5;
VDC=(Y-YEC)/PONDC;
F0C=(YEC*(1-AJC/((1+PPC*MC)**2)))/PONDC;
F1C=DUM*F0C;

```

```

F2C=X2*F0C;
F3C=(YEC*AJC*MC/((1+PPC*MC)**2))/PONDC;
?
?                               Testes P
?
? H1: BNT1   H2: BNT2
?
DIF=(YEB-YEA)/PONDA;
OLS (SILENT) VDA F0A F1A F2A F3A DIF;
SET T1(I)=@COEF(5)/@SES(5);
?
? H1: BNT1   H2: PGT2
?
DIF=(YEC-YEA)/PONDA;
OLS (SILENT) VDA F0A F1A F2A F3A DIF;
SET T2(I)=@COEF(5)/@SES(5);
?
? H1: BNT2   H2: BNT1
?
DIF=(YEA-YEB)/PONDB;
OLS (SILENT) VDB F0B F1B F2B F3B DIF;
SET T3(I)=@COEF(5)/@SES(5);
?
? H1: BNT2   H2: PGT2
?
DIF=(YEC-YEB)/PONDB;
OLS (SILENT) VDB F0B F1B F2B F3B DIF;
SET T4(I)=@COEF(5)/@SES(5);
?
? H1: PGT2   H2: BNT1
?
DIF=(YEA-YEC)/PONDC;
OLS (SILENT) VDC F0C F1C F2C F3C DIF;
SET T5(I)=@COEF(5)/@SES(5);
?
? H1: PGT2   H2: BNT2
?
DIF=(YEB-YEC)/PONDC;
OLS (SILENT) VDC F0C F1C F2C F3C DIF;
SET T6(I)=@COEF(5)/@SES(5);
100 ENDDO;
SMPL 1 REP;
DOT 1-6; R.=T.>1.645; ENDDOT;
MSD T1 T2 T3 T4 T5 T6 R1 R2 R3 R4 R5 R6;
WRITE (FILE='P_11_A.TXT') T1 T2 T3;
WRITE (FILE='P_11_B.TXT') T4 T5 T6;
SET NMED=NOB/REP;
SET GT=1-NMED/N;
PRINT NMED GT;

```

2. Teste P_A

```
OPTIONS CRT, DOUBLE;
SUPRES @SMPL @LOGL @NOB @SES @COEF;
SET REP=2000;
SET N=588;
SET C0V=1; SET C1V=1; SET C2V=1; SET A=0.25;
SET VALA=0.75; SET VALB=-1.5; SET VALC=-2.5;
SET P=1/A;
FRML MIU MI=EXP(C0+C1*DUM+C2*X2);
FRML ALFA ALF=EXP(AL);
FRML FO1 LGAMFN(Y+(MI**(1-K))/ALF)-LGAMFN((MI**(1-K))/ALF);
FRML FO2 Y*LOG(ALF*(MI**K))-(Y+(MI**(1-K))/ALF)*LOG(1+ALF*(MI**K));
FRML FO3 LOG(1-(1+ALF*(MI**K))**(-(MI**(1-K))/ALF));
FRML BNTK FOBNT2=FO1+FO2-FO3;
EQSUB BNTK FO1 FO2 FO3 MIU ALFA;
FRML FO4 Y*LOG(MI)+(Y-1)*LOG(1+ALF*Y)-Y*LOG(1+ALF*MI);
FRML FO5 MI*(1+ALF*Y)/(1+ALF*MI)+LOG(1-EXP(-MI/(1+ALF*MI)));
FRML PGT2 FOPGT2=FO4-FO5;
EQSUB PGT2 FO4 FO5 MIU ALFA;
FRML BNTKA YE=MI/(1-(1+ALF*(MI**K))**(-(MI**(1-K))/ALF));
EQSUB BNTKA MIU ALFA;
FRML PGT2A YE=MI/(1-EXP(-MI/(1+ALF*MI)));
EQSUB PGT2A MIU ALFA;
SMPL 1 REP;
DOT 1-6; T.=@MISS; ENDDOT;
SMPL 1 N;
Y=0;
SET NOB=0;
smpl 1 1; random (seedin=0603952200) asd;
DO I=1 TO REP;
?
?                                     Geração do modelo BNT2
?
SMPL 1 N;
RANDOM (UNIFORM) X1;
DUM = X1>0.66;
RANDOM X2;
DO J=1 TO N;
  SMPL 1 P;
  RANDOM (EXPON,LAMBDA=P) Z;
  MAT E1=SUM(Z);
  SET E=E1(1);
  SMPL J J;
  SET M=EXP(C0V+C1V*DUM+C2V*X2)*E;
  RANDOM (MEAN=M, POISSON) Y;
ENDDO;
SMPL 1 N;
SMPLIF Y;
SET NOB=NOB+@NOB;
?
?                                     Estimação do modelo BNT1
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALA;
PARAM C0 C1 C2 AL K 0; CONST K;
```

```

ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) BNTK;
PARAM C0 C1 C2 AL K 0; CONST K;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0A=C0;
SET B1A=C1;
SET B2A=C2;
SET PPA=EXP(AL);
MA=EXP(B0A+B1A*DUM+B2A*X2);
YEA=MA/(1-(1+PPA)**(-MA/PPA));
AJA=YEA-MA;
PONDA=(YEA*(1+PPA-AJA)**0.5;
VDA=(Y-YEA)/PONDA;
F0A=(YEA*(1-AJA*LOG(1+PPA)/PPA))/PONDA;
F1A=DUM*F0A;
F2A=X2*F0A;
F3A=(YEA*AJA/PPA*(LOG(1+PPA)/PPA-1/(1+PPA)))/PONDA;
?
?                               Estimação do modelo BNT2
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALB;
PARAM C0 C1 C2 AL K 1; CONST K;
ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) BNTK;
PARAM C0 C1 C2 AL K 1; CONST K;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0B=C0;
SET B1B=C1;
SET B2B=C2;
SET PPB=EXP(AL);
MB=EXP(B0B+B1B*DUM+B2B*X2);
YEB=MB/(1-(1+PPB*MB)**(-1/PPB));
AJB=YEB-MB;
PONDB=(YEB*(1+PPB*MB-AJB)**0.5;
VDB=(Y-YEB)/PONDB;
F0B=(YEB*(1-AJB/(1+PPB*MB)))/PONDB;
F1B=DUM*F0B;
F2B=X2*F0B;
F3B=(YEB*AJB/PPB*(LOG(1+PPB*MB)/(PPB*MB)-1/(1+PPB*MB)))/PONDB;
?
?                               Estimação do modelo PGT2
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALC;
PARAM C0 C1 C2 AL;
ML (SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) PGT2;
PARAM C0 C1 C2 AL;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) PGT2;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0C=C0;
SET B1C=C1;
SET B2C=C2;
SET PPC=EXP(AL);
MC=EXP(B0C+B1C*DUM+B2C*X2);
YEC=MC/(1-EXP(-MC/(1+PPC*MC)));
AJC=YEC-MC;
PONDC=(YEC*((1+PPC*MC)**2-AJC)**0.5;

```



```

VDC=(Y-YEC)/PONDC;
F0C=(YEC*(1-AJC/((1+PPC*MC)**2)))/PONDC;
F1C=DUM*F0C;
F2C=X2*F0C;
F3C=(YEC*AJC*MC/((1+PPC*MC)**2))/PONDC;
?
?                               Testes PA
?
? H1: BNT1   H2: BNT2
?
YE=YEA;
SET C0=B0B; SET C1=B1B; SET C2=B2B; SET AL=LOG(PPB);
PARAM C0 C1 C2 AL K 1; CONST K;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=500,MAXSQZ=100) BNTKA;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 200;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-(1+PP*MPA)**(-1/PP));
DIF=(YPA-YE)/PONDA;
OLS (SILENT) VDA F0A F1A F2A F3A DIF;
SET T1(I)=@COEF(5)/@SES(5);
?
? H1: BNT1   H2: PGT2
?
200 YE=YEA;
SET C0=B0C; SET C1=B1C; SET C2=B2C; SET AL=LOG(PPC);
PARAM C0 C1 C2 AL;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=500,MAXSQZ=100) PGT2A;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 300;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-EXP(-MPA/(1+PP*MPA)));
DIF=(YPA-YE)/PONDA;
OLS (SILENT) VDA F0A F1A F2A F3A DIF;
SET T2(I)=@COEF(5)/@SES(5);
?
? H1: BNT2   H2: BNT1
?
300 YE=YEB;
SET C0=B0A; SET C1=B1A; SET C2=B2A; SET AL=LOG(PPA);
PARAM C0 C1 C2 AL K 0; CONST K;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=500,MAXSQZ=100) BNTKA;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 400;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-(1+PP)**(-MPA/PP));
DIF=(YPA-YE)/PONDB;
OLS (SILENT) VDB F0B F1B F2B F3B DIF;
SET T3(I)=@COEF(5)/@SES(5);
?
? H1: BNT2   H2: PGT2
?
400 YE=YEB;
SET C0=B0C; SET C1=B1C; SET C2=B2C; SET AL=LOG(PPC);
PARAM C0 C1 C2 AL;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=500,MAXSQZ=100) PGT2A;

```

```

if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 500;
  SET PP=EXP(AL);
  MPA=EXP(C0+C1*DUM+C2*X2);
  YPA=MPA/(1-EXP(-MPA/(1+PP*MPA)));
  DIF=(YPA-YE)/PONDB;
  OLS (SILENT) VDB F0B F1B F2B F3B DIF;
  SET T4(I)=@COEF(5)/@SES(5);
  ?
  ? H1: PGT2   H2: BNT1
  ?
500 YE=YEC;
  SET C0=B0A; SET C1=B1A; SET C2=B2A; SET AL=LOG(PPA);
  PARAM C0 C1 C2 AL K 0; CONST K;
  LSQ (SILENT,HITER=G,HCOV=G,MAXIT=500,MAXSQZ=100) BNTKA;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 600;
  SET PP=EXP(AL);
  MPA=EXP(C0+C1*DUM+C2*X2);
  YPA=MPA/(1-(1+PP)**(-MPA/PP));
  DIF=(YPA-YE)/PONDC;
  OLS (SILENT) VDC F0C F1C F2C F3C DIF;
  SET T5(I)=@COEF(5)/@SES(5);
  ?
  ? H1: PGT2   H2: BNT2
  ?
600 YE=YEC;
  SET C0=B0B; SET C1=B1B; SET C2=B2B; SET AL=LOG(PPB);
  PARAM C0 C1 C2 AL K 1; CONST K;
  LSQ (SILENT,HITER=G,HCOV=G,MAXIT=500,MAXSQZ=100) BNTKA;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
  SET PP=EXP(AL);
  MPA=EXP(C0+C1*DUM+C2*X2);
  YPA=MPA/(1-(1+PP*MPA)**(-1/PP));
  DIF=(YPA-YE)/PONDC;
  OLS (SILENT) VDC F0C F1C F2C F3C DIF;
  SET T6(I)=@COEF(5)/@SES(5);
100 ENDDO;
SMPL 1 REP;
DOT 1-6; R.=T.>1.645; MSD T. R.; ENDDOT;
WRITE (FILE='PA_11_A.TXT') T1 T2 T3;
WRITE (FILE='PA_11_B.TXT') T4 T5 T6;
SET NMED=NOB/REP;
SET GT=1-NMED/N;
PRINT NMED GT;

```

3. Teste P_{A1}

```
OPTIONS CRT, DOUBLE;
SUPRES @SMPL @LOGL @NOB @SES @COEF;
SET REP=2000;
SET N=588;
SET C0V=1; SET C1V=1; SET C2V=1; SET A=0.25;
SET VALA=0.75; SET VALB=-1.5; SET VALC=-2.5;
SET P=1/A;
FRML MIU MI=EXP(C0+C1*DUM+C2*X2);
FRML ALFA ALF=EXP(AL);
FRML FO1 LGAMFN(Y+(MI**(1-K))/ALF)-LGAMFN((MI**(1-K))/ALF);
FRML FO2 Y*LOG(ALF*(MI**K))-(Y+(MI**(1-K))/ALF)*LOG(1+ALF*(MI**K));
FRML FO3 LOG(1-(1+ALF*(MI**K))**(-(MI**(1-K))/ALF));
FRML BNTK FOBNT2=FO1+FO2-FO3;
EQSUB BNTK FO1 FO2 FO3 MIU ALFA;
FRML FO4 Y*LOG(MI)+(Y-1)*LOG(1+ALF*Y)-Y*LOG(1+ALF*MI);
FRML FO5 MI*(1+ALF*Y)/(1+ALF*MI)+LOG(1-EXP(-MI/(1+ALF*MI)));
FRML PGT2 FOPGT2=FO4-FO5;
EQSUB PGT2 FO4 FO5 MIU ALFA;
FRML BNTKA YE=MI/(1-(1+ALF*(MI**K))**(-(MI**(1-K))/ALF));
EQSUB BNTKA MIU ALFA;
FRML PGT2A YE=MI/(1-EXP(-MI/(1+ALF*MI)));
EQSUB PGT2A MIU ALFA;
SMPL 1 REP;
DOT 1-6; T.=@MISS; ENDDOT;
SMPL 1 N;
Y=0;
SET NOB=0;
smpl 1 1; random (seedin=0603952200) asd;
DO I=1 TO REP;
?
?                                     Geração do modelo BNT2
?
SMPL 1 N;
RANDOM (UNIFORM) X1;
DUM = X1>0.66;
RANDOM X2;
DO J=1 TO N;
  SMPL 1 P;
  RANDOM (EXPON,LAMBDA=P) Z;
  MAT E1=SUM(Z);
  SET E=E1(1);
  SMPL J J;
  SET M=EXP(C0V+C1V*DUM+C2V*X2)*E;
  RANDOM (MEAN=M, POISSON) Y;
ENDDO;
SMPL 1 N;
SMPLIF Y;
SET NOB=NOB+@NOB;
?
?                                     Estimação do modelo BNT1
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALA;
PARAM C0 C1 C2 AL K 0; CONST K;
```

```

ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) BNTK;
PARAM C0 C1 C2 AL K 0; CONST K;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0A=C0;
SET B1A=C1;
SET B2A=C2;
SET PPA=EXP(AL);
MA=EXP(B0A+B1A*DUM+B2A*X2);
YEA=MA/(1-(1+PPA)**(-MA/PPA));
AJA=YEA-MA;
PONDA=(YEA*(1+PPA-AJA)**0.5;
VDA=(Y-YEA)/PONDA;
F0A=(YEA*(1-AJA*LOG(1+PPA)/PPA))/PONDA;
F1A=DUM*F0A;
F2A=X2*F0A;
F3A=(YEA*AJA/PPA*(LOG(1+PPA)/PPA-1/(1+PPA)))/PONDA;
?
?                               Estimação do modelo BNT2
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALB;
PARAM C0 C1 C2 AL K 1; CONST K;
ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) BNTK;
PARAM C0 C1 C2 AL K 1; CONST K;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0B=C0;
SET B1B=C1;
SET B2B=C2;
SET PPB=EXP(AL);
MB=EXP(B0B+B1B*DUM+B2B*X2);
YEB=MB/(1-(1+PPB*MB)**(-1/PPB));
AJB=YEB-MB;
PONDB=(YEB*(1+PPB*MB-AJB)**0.5;
VDB=(Y-YEB)/PONDB;
F0B=(YEB*(1-AJB/(1+PPB*MB)))/PONDB;
F1B=DUM*F0B;
F2B=X2*F0B;
F3B=(YEB*AJB/PPB*(LOG(1+PPB*MB)/(PPB*MB)-1/(1+PPB*MB)))/PONDB;
?
?                               Estimação do modelo PGT2
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALC;
PARAM C0 C1 C2 AL;
ML (SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) PGT2;
PARAM C0 C1 C2 AL;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) PGT2;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET B0C=C0;
SET B1C=C1;
SET B2C=C2;
SET PPC=EXP(AL);
MC=EXP(B0C+B1C*DUM+B2C*X2);
YEC=MC/(1-EXP(-MC/(1+PPC*MC)));
AJC=YEC-MC;
PONDC=(YEC*((1+PPC*MC)**2-AJC)**0.5;

```

VDC=(Y-YEC)/PONDC;
F0C=(YEC*(1-AJC/((1+PPC*MC)**2)))/PONDC;
F1C=DUM*F0C;
F2C=X2*F0C;
F3C=(YEC*AJC*MC/((1+PPC*MC)**2))/PONDC;
?
? Testes PA1
?
? H1: BNT1 H2: BNT2
?
YE=YEA;
SET C0=B0B; SET C1=B1B; SET C2=B2B; SET AL=LOG(PPB);
PARAM C0 C1 C2 AL K 1; CONST K;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=1,MAXSQZ=100) BNTKA;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-(1+PP*MPA)**(-1/PP));
DIF=(YPA-YE)/PONDA;
OLS (SILENT) VDA F0A F1A F2A F3A DIF;
SET T1(I)=@COEF(5)/@SES(5);
?
? H1: BNT1 H2: PGT2
?
SET C0=B0C; SET C1=B1C; SET C2=B2C; SET AL=LOG(PPC);
PARAM C0 C1 C2 AL;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=1,MAXSQZ=100) PGT2A;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-EXP(-MPA/(1+PP*MPA)));
DIF=(YPA-YE)/PONDA;
OLS (SILENT) VDA F0A F1A F2A F3A DIF;
SET T2(I)=@COEF(5)/@SES(5);
?
? H1: BNT2 H2: BNT1
?
YE=YEB;
SET C0=B0A; SET C1=B1A; SET C2=B2A; SET AL=LOG(PPA);
PARAM C0 C1 C2 AL K 0; CONST K;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=1,MAXSQZ=100) BNTKA;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-(1+PP)**(-MPA/PP));
DIF=(YPA-YE)/PONDB;
OLS (SILENT) VDB F0B F1B F2B F3B DIF;
SET T3(I)=@COEF(5)/@SES(5);
?
? H1: BNT2 H2: PGT2
?
SET C0=B0C; SET C1=B1C; SET C2=B2C; SET AL=LOG(PPC);
PARAM C0 C1 C2 AL;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=1,MAXSQZ=100) PGT2A;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-EXP(-MPA/(1+PP*MPA)));
DIF=(YPA-YE)/PONDB;
OLS (SILENT) VDB F0B F1B F2B F3B DIF;

```

SET T4(I)=@COEF(5)/@SES(5);
?
? H1: PGT2   H2: BNT1
?
YE=YEC;
SET C0=B0A; SET C1=B1A; SET C2=B2A; SET AL=LOG(PPA);
PARAM C0 C1 C2 AL K 0; CONST K;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=1,MAXSQZ=100) BNTKA;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-(1+PP)**(-MPA/PP));
DIF=(YPA-YE)/PONDC;
OLS (SILENT) VDC F0C F1C F2C F3C DIF;
SET T5(I)=@COEF(5)/@SES(5);
?
? H1: PGT2   H2: BNT2
?
SET C0=B0B; SET C1=B1B; SET C2=B2B; SET AL=LOG(PPB);
PARAM C0 C1 C2 AL K 1; CONST K;
LSQ (SILENT,HITER=G,HCOV=G,MAXIT=1,MAXSQZ=100) BNTKA;
SET PP=EXP(AL);
MPA=EXP(C0+C1*DUM+C2*X2);
YPA=MPA/(1-(1+PP*MPA)**(-1/PP));
DIF=(YPA-YE)/PONDC;
OLS (SILENT) VDC F0C F1C F2C F3C DIF;
SET T6(I)=@COEF(5)/@SES(5);
100 ENDDO;
SMPL 1 REP;
DOT 1-6; R.=T.>1.645; ENDDOT;
MSD T1 T2 T3 T4 T5 T6 R1 R2 R3 R4 R5 R6;
WRITE (FILE='PA1_11_A.TXT') T1 T2 T3;
WRITE (FILE='PA1_11_B.TXT') T4 T5 T6;
SET NMED=NOB/REP;
SET GT=1-NMED/N;
PRINT NMED GT;

```

4. Teste T_k

```
OPTIONS CRT, DOUBLE;
SUPRES @SMPL @LOGL @NOB @SES @COEF;
SET REP=2000;
SET N=588;
SET C0V=1; SET C1V=1; SET C2V=1; SET A=0.25;
SET VALA=0.75; SET VALB=-1.5; SET VALC=-2.5;
SET P=1/A;
FRML MIU MI=EXP(C0+C1*DUM+C2*X2);
FRML ALFA ALF=EXP(AL);
FRML FO1 LGAMFN(Y+(MI**(1-K))/ALF)-LGAMFN((MI**(1-K))/ALF);
FRML FO2 Y*LOG(ALF*(MI**K))-(Y+(MI**(1-K))/ALF)*LOG(1+ALF*(MI**K));
FRML FO3 LOG(1-(1+ALF*(MI**K))**(-(MI**(1-K))/ALF));
FRML BNTK FOBNT2=FO1+FO2-FO3;
EQSUB BNTK FO1 FO2 FO3 MIU ALFA;
SMPL 1 REP;
DOT 1-6; T.=@MISS; ENDDOT;
SMPL 1 N;
Y=0;
SET NOB=0;
smpl 1 1; random (seedin=0603952200) asd;
DO I=1 TO REP;
?
?                               Geração do modelo BNT2
?
SMPL 1 N;
RANDOM (UNIFORM) X1;
DUM = X1>0.66;
RANDOM X2;
DO J=1 TO N;
  SMPL 1 P;
  RANDOM (EXPON,LAMBDA=P) Z;
  MAT E1=SUM(Z);
  SET E=E1(1);
  SMPL J J;
  SET M=EXP(C0V+C1V*DUM+C2V*X2)*E;
  RANDOM (MEAN=M, POISSON) Y;
ENDDO;
SMPL 1 N;
SMPLIF Y;
SET NOB=NOB+@NOB;
?
?                               Estimação do modelo BNT1
?
SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALA;
PARAM C0 C1 C2 AL K 0; CONST K;
ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) BNTK;
PARAM C0 C1 C2 AL K 0; CONST K;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET LLA=@LOGL;
MFORM S=@GRAD;
MFORM INF=@VCOV;
MAT VS=INF*S;
```

```

SET SIN1=VS(5);
SET SINB1=SIGN(SIN1);
?
?
?
Estimação do modelo BNT2
300 SET C0=C0V; SET C1=C1V; SET C2=C2V; SET AL=VALB;
PARAM C0 C1 C2 AL K 1; CONST K;
ML(SILENT,HITER=B,HCOV=B,MAXIT=20,MAXSQZ=20,TOL=0.1,TOLG=0.1) BNTK;
PARAM C0 C1 C2 AL K 1; CONST K;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
SET LLB=@LOGL;
MFORM S=@GRAD;
MFORM INF=@VCOV;
MAT VS=INF*S;
SET SIN1=VS(5);
SET SINB2=SIGN(SIN1);
?
?
?
Estimação do modelo BNTk
PARAM C0 C1 C2 AL K;
ML (SILENT,HITER=B,HCOV=B,MAXIT=2000,MAXSQZ=100) BNTK;
PARAM C0 C1 C2 AL K;
ML (SILENT,HITER=N,HCOV=N,MAXIT=500,MAXSQZ=100) BNTK;
if @ses(1)=0 | @ses(2)=0 | @ses(3)=0 | @ses(4)=0 | @ifconv=0; then; goto 100;
?
?
?
Testes Tk (RV)
? H1: BNT1 H2: BNT2
?
SET T1(I)=-SINB1*(-2*(LLA-@LOGL))**0.5;
?
? H1: BNT2 H2: BNT1
?
SET T2(I)=SINB2*(-2*(LLB-@LOGL))**0.5;
100 ENDDO;
SMPL 1 REP;
DOT 1-2; R.=T.>1.645; ENDDOT;
MSD T1 T2 R1 R2;
WRITE (FILE='E_11.TXT') T1 T2;
SET NMED=NOB/REP;
SET GT=1-NMED/N;
PRINT NMED GT;

```


BIBLIOGRAFIA

- Atkinson, A. (1970), "A method for discriminating between models", *Journal of the Royal Statistical Society*, **B32**, pp. 323-353.
- Bera, A. e McKenzie, C. (1986), "Alternative forms and properties of the score test", *Journal of Applied Statistics*, **13**, pp. 13-25.
- Berndt, E., Hall, B., Hall, R. e Hausman, J. (1974), "Estimation and inference in nonlinear structural models", *Annals of Economic and Social Measurement*, **3/4**, pp. 653-665
- Cameron, A. e Trivedi, P. (1986), "Econometric models based on count data: comparisons and applications of some estimators and tests", *Journal of Applied Econometrics*, **1**, pp. 29-53.
- Cameron, A. e Trivedi, P. (1990), "Regression-based tests for overdispersion in the Poisson model", *Journal of Econometrics*, **46**, pp. 347-364.
- Chappel, W., Kimenyi, M. e Mayer, W. (1990), "A Poisson probability model of entry and market structure with an application to U. S. industries during 1972-77", *Southern Economic Journal*, **56(4)**, pp. 918-927.
- Chesher, A. (1984), "Testing for neglected heterogeneity", *Econometrica*, **52(4)**, pp. 865-872.
- Collings, B. e Margolin, B. (1985), "Testing goodness of fit for the Poisson assumption when observations are not identically distributed", *Journal of the American Statistical Association*, **80(390)**, pp. 411-418.

- Consul, P. (1989), "Generalized Poisson distributions", Marcel Dekker, New York.
- Consul, P. e Famoye, F. (1988), "Maximum likelihood estimation for the generalized Poisson distribution when sample mean is larger than sample variance", *Communications in Statistics - Theory and Methods*, **17(1)**, pp. 299-309.
- Consul, P. e Famoye, F. (1992), "Generalized Poisson regression model", *Communications in Statistics - Theory and Methods*, **21(1)**, pp. 89-109.
- Consul, P. e Jain, G. (1973), "A generalization of the Poisson distribution", *Technometrics*, **15(4)**, pp. 791-799.
- Consul, P. e Shenton, L. (1972), "Use of lagrangian expansion for generating discrete generalized probability distributions", *SIAM Journal on Applied Mathematics*, **23(2)**, pp. 239-248.
- Consul, P. e Shoukri, M. (1984), "Maximum likelihood estimation for the generalized Poisson distribution", *Communications in Statistics - Theory and Methods*, **13(12)**, pp. 1553-1547.
- Consul, P. e Shoukri, M. (1985), "The generalized Poisson distribution when the sample mean is larger than the sample variance", *Communications in Statistics - Theory and Methods*, **14(3)**, pp. 667-681.
- Cox, D. (1961), "Tests of separate families of hypotheses", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, University of California Press, Berkeley, pp. 105-123.
- Cox, D. (1983), "Some remarks on overdispersion", *Biometrika*, **70**, pp. 269-274.

- Creel, M. e Loomis, J. (1990), "Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California", *American Journal of Agricultural Economics*, **72**, pp. 434-441.

- Davidson, R. e MacKinnon, J. (1981), "Several tests for model specification in the presence of alternative hypotheses", *Econometrica*, **49**, pp. 781-793.

- Davidson, R. e MacKinnon, J. (1982), "Some non-nested hypothesis tests and the relations among them", *Review of Economic Studies*, **49(3)**, pp. 551-565.

- Davidson, R. e MacKinnon, J. (1983), "Small sample properties of alternative forms of the lagrange multiplier test", *Economics Letters*, **12**, pp. 269-275.

- Davidson, R. e MacKinnon, J. (1993), "Estimation and inference in econometrics", Oxford University Press, New York.

- Dean, C. e Lawless, J. (1989), "Tests for detecting overdispersion in Poisson regression models", *Journal of the American Statistical Association*, **84(406)**, pp. 467-472.

- Dionne, G. e Vanasse, C. (1992), "Automobile insurance ratemaking in the presence of asymmetrical information", *Journal of Applied Econometrics*, **7**, pp. 149-165.

- Englin, J. e Shonkwiler, J. (1995), "Estimating social welfare using count data models: an application to long-run recreation demand under conditions of endogenous stratification and truncation", *The Review of Economics and Statistics*, **77(1)**, pp. 104-112.

- Famoye, F. (1993), "Restricted generalized Poisson regression model", *Communications in Statistics - Theory and Methods*, **22(5)**, pp. 1335-1354.

- Fisher, G. e McAleer, M. (1981), "Alternative procedures and associated tests of significance for non-nested hypotheses", *Journal of Econometrics*, **16**, pp. 103-119.

- Godfrey, L. (1988), "Misspecification tests in econometrics: the lagrange multiplier principle and other approaches", Cambridge University Press.

- Godfrey, L. e Pesaran, M. (1983), "Tests of non-nested regression models: small sample adjustments and Monte Carlo evidence", *Journal of Econometrics*, **21**, pp. 133-154.

- Gourieroux, C. e Monfort, A. (1994), "Testing non-nested hypotheses", in Engle, R. e McFadden, D. (ed.), *Handbooks of Econometrics*, vol. 4, North-Holland, pp. 2585-2637.

- Gourieroux, C., Monfort, A. e Trognon, A. (1984a), "Pseudo maximum likelihood methods: theory", *Econometrica*, **52(3)**, pp. 681-700.

- Gourieroux, C., Monfort, A. e Trognon, A. (1984b), "Pseudo maximum likelihood methods: applications to Poisson models", *Econometrica*, **52(3)**, pp. 701-720.

- Greene, W. (1993), "Econometric analysis", Macmillan Publishing Company, New York.

- Grogger, J. e Carson, R. (1991), "Models for truncated counts", *Journal of Applied Econometrics*, **6**, pp. 225-238.

- Gurmu, S. (1991), "Tests for detecting overdispersion in the positive Poisson regression model", *Journal of Business & Economic Statistics*, **9**, pp. 215-222.

- Gurmu, S. e Trivedi, P. (1992), "Overdispersion tests for truncated Poisson regression

models”, *Journal of Econometrics*, **54**, pp. 347-370.

- Hall, B. (1995), “TSP version 4.3 reference manual”, TSP International, Stanford, California.

- Hall, B. (1995), “TSP version 4.3 user’s guide”, TSP International, Stanford, California.

- Hansen, L. (1982), “Large sample properties of generalized method of moments estimators”, *Econometrica*, **50(4)**, pp. 1029-1054.

- Hausman, J., Hall, B. e Griliches, Z. (1984), “Econometric models for count data with an application to the patent-R&D relationship”, *Econometrica*, **52(4)**, pp. 909-938.

- Hellerstein, D. (1991), “Using count data in travel cost analysis with aggregate data”, *American Journal of Agricultural Economics*, **73**, pp. 860-867.

- Johnson, N., Kotz, S. e Kemp, A. (1992), “Univariate discrete distributions”, John Wiley & Sons, New York.

- Kim, B. e Park, C. (1992), “Some remarks on testing goodness of fit for the Poisson assumption”, *Communications in Statistics - Theory and Methods*, **21(4)**, pp. 979-995.

- Lambert, D. (1992), “Zero-inflated Poisson regression, with an application to defects in manufacturing”, *Technometrics*, **34(1)**, pp. 1-14.

- Lawless, J. (1987), “Regression methods for Poisson process data”, *Journal of the American Statistical Association*, **82(399)**, pp. 808-815.

- Lee, L. (1986) “Specification test for Poisson regression models”, *International*

Economic Review, **27(3)**, pp. 689-706.

- MacKinnon, J. (1983), “Model specification tests against non-nested alternatives”, *Econometric Reviews*, **2(1)**, pp. 85-158.

- Mullahy, J. (1986), “Specification and testing of some modified count data models”, *Journal of Econometrics*, **33**, pp. 341-365.

- Nelson, D. (1975), “Some remarks on generalizations of the negative binomial and Poisson distributions”, *Technometrics*, **17(1)**, pp. 135-136.

- Quandt, R. (1974), “A comparison of methods for testing nonnested hypotheses”, *Review of Economics and Statistics*, **56**, pp. 92-99.

- Pagan, A. e Vella, F. (1989), “Diagnostic tests for models based on individual data: a survey”, *Journal of Applied Econometrics*, **4**, pp. S29-S59.

- Ramsey, J. (1969), “Tests for specification errors in classic linear least squares regression analysis”, *Journal of the Royal Statistical Society*, **B31**, pp. 350-371.

- Ruser, J. (1991), “Workers’ compensation and occupational injuries and illnesses”, *Journal of Labor Economics*, **9(4)**, pp. 325-350.

- Santos Silva, J. (1996a), “Generalized Poisson regression for positive count data”, Department of Economics, University College London e ISEG/Universidade Técnica de Lisboa, trabalho não publicado.

- Santos Silva, J. (1996b), “A score test for non-nested hypotheses with applications to discrete data models”, ISEG/Universidade Técnica de Lisboa, trabalho não publicado.

- Santos Silva, J. e Andrade e Silva, J. (1994), “Misspecification in models for positive count data”, Departamento de Matemática, ISEG/Universidade Técnica de Lisboa, trabalho não publicado.

- Shaw, D. (1988), “On-site samples’ regression: problems of non-negative integers, truncation, and endogenous stratification”, *Journal of Econometrics*, **37**, pp. 211-223.

- Wedel, M., Desarbo, W., Bult, J. e Ramaswamy, V. (1993), “A latent class Poisson regression model for heterogeneous count data”, *Journal of Applied Econometrics*, **8**, pp. 397-411.

- White, H. (1982), “Maximum likelihood estimation of misspecified models”, *Econometrica*, **50(1)**, pp. 1-25.

- Winkelmann, R. (1995), “Duration dependence and dispersion in count data models”, *Journal of Business & Economic Statistics*, **13(4)**, pp. 467-474.

- Winkelmann, R. e Zimmermann, K. (1991), “A new approach for modeling economic count data”, *Economics Letters*, **37**, pp. 139-143.

- Winkelmann, R. e Zimmermann, K. (1995), “Recent developments in count data modelling: theory and application”, *Journal of Economic Surveys*, **9(1)**, pp. 1-24.

- Xue, D. e Deddens, J. (1992), “Overdispersed negative binomial regression models”, *Communications in Statistics - Theory and Methods*, **21(8)**, pp. 2215-2226.