**Statistical Synthesis of Heckman-Based Selection with Conway-Maxwell-Poisson Probability Models**

**ABSTRACT**

This research elucidates the theoretical derivation and statistical foundations of the integrated HeckCOMPoisson model, which synthesizes the Heckman selection framework with the Conway-Maxwell-Poisson probability distribution. This unified statistical approach was formulated specifically to address truncated count data subject to sample selection mechanisms. The model exhibits superior statistical properties across multiple evaluation criteria: it demonstrates excellent performance in parametric Goodness-of-Fit assessments, yields consistent maximum likelihood estimates for count data with selection bias, and efficiently parameterizes both under-dispersed and over-dispersed count distributions. A distinguishing characteristic of the HeckCOMPoisson distribution is its parametric flexibility and its robust estimation properties when modeling discrete count data with heterogeneous dispersion parameters.

*Keywords: COMPoisson, HeckCOMPoisson, Integrated, Heckman*

1. **INTRODUCTION**

This paper explicitly explains how developed an integrated ‘HeckCOMPoisson’ model.

**1.2 Integrated HeckCOMPoisson Model**

The integrated HeckCOMPoisson model represents an advancement over the traditional Heckman model by incorporating a COMPoisson regression in the second stage to address count data outcomes while correcting for endogenous sample selection bias. While the standard Heckman model's linear approach can produce inappropriate negative or non-integer predictions for count data, the HeckCOMPoisson resolves this through its count-based framework. The model also tackles the critical issue of endogeneity that arises from correlation between variables in the probit and selection equations, which typically biases parameter estimates in the Heckman model. As noted by Breslow (1990), when error terms from selection and outcome equations are correlated, standard probit techniques yield biased results. This is addressed through the COMPoisson component's introduction of a decay parameter ν that governs probability ratios and maintains zero correlation with probit variables, thereby providing a more robust approach for analyzing count data with sample selection issues compared to traditional methods that rely on inverse Mills ratios and require manual error corrections.

**1.3 Integrated HeckCOMPoisson Stage One Probit Function**

The integrated HeckCOMPoisson model is structured with a dual-equation system: one equation handles the count outcome (y) while the other addresses the binary selection indicator (z), which can only be 0 or 1. To generate an observation from the HeckCOMPoisson distribution, probabilities are cumulatively added beginning with P(X=0) until their sum surpasses a randomly generated Uniform (0,1) value. At this point, the corresponding X value represents a draw from the HeckCOMPoisson distribution. The probabilities can be computed by the equation below

To calculate the starting probability, we initialize X to zero, resulting in an initial value of λ. The computation of HeckCOMPoisson probabilities requires determining the HeckCOMPoisson normalizing constant Z(λ) through an asymptotic approximation method

To begin calculating Z, we examine its asymptotic behavior by setting i equal to λ, where λ represents our starting point for the analysis.

This implies

This expresses Z(λ,v+1) as an integral in terms of Z(λ, v). When this formula is applied iteratively, beginning with Z(λ,1) = exp(λ), we can represent Z(λ, v) for any positive integer v as a multiple integral.

The behavior of this integral for large λ can be determined by making the change of variable  and applying Laplace’s method this results to,

When analyzing this integral's behavior at large λ values, we employ a variable substitution where and then apply Laplace's method to derive the result to get

The asymptotic formula mentioned is valid not only for integer values of v but extends to all real numbers. As λ increases or v decreases, there is a rapid growth in Z. Because the moments of HeckCOMPoisson are directly linked to the derivatives of the normalizing constant Z, we can derive approximate equations for these moments using the asymptotic approximation of the normalizing constant, which yields the equation below.

The count variable Yi follows a COMPoisson distribution based on the covariates Xi and their conditional mean, but this outcome y is only observable when z equals 1, specifically when a citizen engages in corruption. In the context of this study, a corruption count was recorded whenever a citizen either offered or accepted a bribe.

**1.4 Integrated HeckCOMPoisson Stage Two Outcome Function**

We have a situation where we only get to see the outcome variable Y for a particular observation i under specific conditions. Specifically, we only observe Y when a binary selection variable z equals 1 for that observation. This selection variable z is determined by a latent variable model that includes certain covariates (explanatory variables), that is;

The model examines corruption behavior using a binary variable (Zi) that indicates whether a citizen engages in corrupt practices. Several factors are hypothesized to influence an individual's decision to participate in corruption, represented by the variables 𝛩i. These factors encompass various aspects such as the underlying motivations for offering bribes, the specific services for which bribes were paid, whether individuals reported bribery incidents to the Ethics and Anti-Corruption Commission (EACC), the person's employment status, how effectively EACC's educational materials worked, and the particular form of corruption involved. The model incorporates error terms from two equations, which are assumed to follow a bivariate normal distribution centered at zero. The relationship between these error terms is characterized by a covariance matrix that specifies the variance of the first error term (σ²), the correlation between the errors (σρ), and sets the variance of the second error term to one for standardization purposes.

1. **Conclusion**

The HeckCOMPoisson model represents a sophisticated amalgamation of the Heckman selection framework and Conway-Maxwell-Poisson probability distribution, demonstrating remarkable efficacy in handling truncated count data subject to endogenous sample selection mechanisms. Through its incorporation of asymptotic approximation methods and parametric Goodness-of-Fit assessments, the model achieves consistent maximum likelihood estimates while effectively parameterizing both under-dispersed and over-dispersed count distributions via a bivariate normal error structure with specified covariance matrix parameters (σ², σρ). The model's distinctive feature lies in its utilization of a decay parameter ν governing probability ratios in conjunction with a dual-equation system, which not only maintains zero correlation with probit variables but also circumvents the limitations of inverse Mills ratios, thereby establishing itself as a statistically robust methodology for analyzing discrete count data characterized by heterogeneous dispersion parameters and sample selection bias.

**REFERENCES**

[1] Breslow, N. (1990). Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *J. Am. Statist. Ass.*, 85, 565–571