

Estimation of hurdle models for overdispersed count data

Helmut Farbmacher
Department of Economics
University of Munich, Germany
helmut.farbmacher@lrz.uni-muenchen.de

Abstract. Hurdle models based on the zero-truncated Poisson-lognormal distribution are rarely used in applied work, although they incorporate some advantages compared with their negative binomial alternatives. I present a command that enables Stata users to estimate Poisson-lognormal hurdle models. I use adaptive Gauss–Hermite quadrature to approximate the likelihood function, and I evaluate the performance of the estimator in Monte Carlo experiments. The model is applied to the number of doctor visits in a sample of the U.S. Medical Expenditure Panel Survey.

Keywords: st0218, ztpnm, count-data analysis, hurdle models, overdispersion, Poisson-lognormal hurdle models

1 Introduction

Hurdle models, first discussed by Mullahy (1986), are very popular for modeling count data. For example, the number of doctor visits or hospitalizations may serve as proxies for demand for health care. These measures may be determined by a two-part decision process. At first, it is up to the patient whether to visit a doctor. After the first contact, though, the physician influences the intensity of treatment (Pohlmeier and Ulrich 1995). Thus the use of a single-index count-data model (such as Poisson or negative binomial models) seems to be inappropriate in many health care applications.

The hurdle model typically combines a binary model to model participation (for example, modeling the patient’s decision to visit the doctor) with a zero-truncated count-data model to model the extent of participation for those participating (for example, modeling the number of doctor visits). In contrast with a single-index model, the hurdle model permits heterogeneous effects for individuals below or above the hurdle. In many applications, the hurdle is set at zero and can therefore also solve the problem of excess zeros, that is, the presence of more zeros in the data than what was predicted by single-index count-data models.

There are many possible combinations of binary and truncated count-data models. An often-used model combines a probit or logit model with a zero-truncated negative binomial model (for example, Vesterinen et al. [2010] and Wong et al. [2010]). The zero-truncated negative binomial model is known to account for overdispersion that may be caused by unobserved heterogeneity. In this model, the heterogeneity is introduced at the level of the parent (untruncated) distribution.

Santos Silva (2003) describes an alternative method of estimating hurdle models if unobserved heterogeneity is present. He proposes using a truncated distribution and doing the mixing over this distribution only. Winkelmann's (2004) proposal of a hurdle model based on the zero-truncated Poisson-lognormal distribution follows this method. In many applications, this model seems to fit the data much better than its negative binomial alternative. The command introduced here makes it possible to estimate Poisson-lognormal hurdle models using adaptive Gauss–Hermite quadrature. It can be used with cross-sectional data, but also with panel data if one is willing to pool the data over time.

2 Model

Generally, the probability function of a hurdle model can be written as

$$f(y) = \begin{cases} g(0) & \text{if } y = 0 \\ \frac{1-g(0)}{1-h(0)}h(y) & \text{if } y \geq 1 \end{cases} \quad (1)$$

where the zeros and the positive counts are determined by the probability $g(0)$ and the truncated probability function $h(y|y > 0) = h(y)/\{1 - h(0)\}$, respectively. The numerator in (1) represents the probability of crossing the hurdle $\{1 - g(0)\}$ and is multiplied by a truncated probability function to ensure that the probabilities sum up to one. The likelihood contribution is

$$L_i^H = g(0)^{(1-d_i)} \times \left[\{1 - g(0)\} \frac{h(y)}{1 - h(0)} \right]^{d_i}$$

where d_i indicates whether individual i crosses the hurdle. Assuming that both functions are independent conditional on covariables, the maximization procedure can be divided into two separate parts. Firstly, one can maximize a binary model with d_i as a dependent variable using the full sample. Secondly, the parameters of h can be estimated separately by a truncated regression using only observations with positive counts.

Winkelmann (2004) proposes combining a probit with a zero-truncated Poisson-lognormal model in which the mixing is done over the truncated Poisson distribution. The likelihood function of this model is given by

$$L^H = \prod_{i=1}^N \{1 - \Phi(\mathbf{x}_i' \gamma)\}^{(1-d_i)} \times \{\Phi(\mathbf{x}_i' \gamma) P^+(y_i | \mathbf{x}_i, \epsilon_i)\}^{d_i}$$

where $P^+(y_i | \mathbf{x}_i, \epsilon_i)$ is the probability function of the zero-truncated part and $\Phi(\mathbf{x}_i' \gamma)$ is the cumulative distribution function of the standard normal distribution. The following discussion is limited to the estimation of the truncated part. The probability function of the zero-truncated Poisson-lognormal model is given by

$$P^+(y_i | \mathbf{x}_i, \epsilon_i) = \frac{\exp(-\lambda_i) \lambda_i^{y_i}}{\{1 - \exp(-\lambda_i)\} y_i!}$$

where λ_i is defined by $\exp(\mathbf{x}_i\beta)\zeta_i$ and $\zeta_i = \exp(\epsilon_i)$. \mathbf{x}_i is a vector of observed characteristics and ϵ_i denotes unobserved variables that might cause overdispersion.

Inference is based on the density of y_i conditional on \mathbf{x}_i ,

$$P^+(y_i|\mathbf{x}_i) = \int_{-\infty}^{\infty} P^+(y_i|\mathbf{x}_i, \epsilon_i) f(\epsilon_i) d\epsilon_i \quad (2)$$

where $f(\epsilon_i)$ is the prior density of the unobserved heterogeneity. Because the mixing is done over the zero-truncated Poisson distribution, the calculation of the likelihood contributions is computationally more demanding than in the negative binomial alternative. There is no analytical solution for the integral in (2), and thus it has to be approximated using simulation or quadrature. The likelihood of the zero-truncated negative binomial model has a closed-form expression because the mixing is done prior to the truncation.

To complete the model, we need an assumption about the prior density of the unobserved heterogeneity. Because there is no analytical solution of the integral in (2), even if we assume a gamma density, we may think about using an assumption that is more theoretically motivated. If, for example, ϵ_i captures many independent variables that cannot be observed by the researcher, then normality of ϵ_i can be established by central limit theorems (Winkelmann 2008). Assuming that ϵ_i is normally distributed (that is, ζ_i is log-normally-distributed), Gauss–Hermite quadrature can be used to approximate the likelihood. Consider a change of variable $\nu_i = \epsilon_i/\sqrt{2}\sigma$:

$$P^+(y_i|\mathbf{x}_i) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} P^+(y_i|\mathbf{x}_i, \sqrt{2}\sigma\nu_i) \exp(-\nu_i^2) d\nu_i$$

Written in this form, the integral can be evaluated by Gauss–Hermite quadrature to get rid of the unobserved heterogeneity, and the likelihood function becomes

$$L = \prod_{i=1}^N \frac{1}{\sqrt{\pi}} \sum_{r=1}^R P^+(y_i|\mathbf{x}_i, \sqrt{2}\sigma\nu_r) w_r$$

where ν_r and w_r are the nodes and weights for the quadrature.

3 The ztpnm command

3.1 Stata implementation

The zero-truncated Poisson-lognormal model discussed above is implemented in Stata as `ztpnm`. The command enables users to estimate hurdle models based on the zero-truncated Poisson-lognormal distribution using standard or adaptive Gauss–Hermite quadrature. Adaptive quadrature shifts and scales the quadrature points to place them under the peak of the integrand, which most likely improves the approximation (see section 6.3.2 in Skrondal and Rabe-Hesketh [2004] for a detailed discussion). Many Stata commands such as `xtpoisson` implement an approach proposed by Liu and Pierce

(1994). They argue that the mode of the integrand and the curvature at the mode can be used as scaling and shifting factors. Instead of calculating these factors, `ztpnm` uses the corresponding values of the standard (untruncated) Poisson-lognormal model to implement adaptive quadrature. The reason for this approach is a built-in command in Stata that gives the corresponding values for the standard Poisson-lognormal model but not for the zero-truncated model. The integrand, however, is very similar in both models, which indicates that these values are good guesses for the scaling and shifting factors of the zero-truncated model. Figure 1 shows that the integrands are almost identical for higher values of the dependent variable.

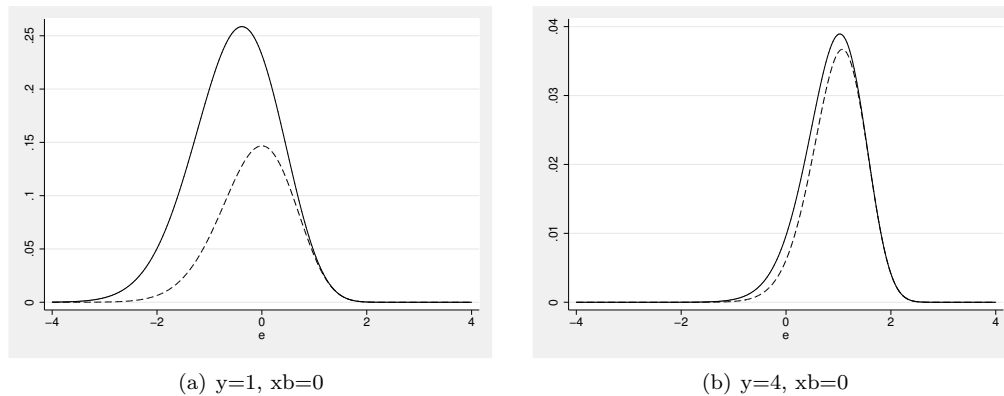


Figure 1. Integrands of zero-truncated models (solid curves) and standard models (dashed curves)

3.2 Syntax

```
ztpnm devar [indepvars] [if] [in] [, irr nonadaptive intpoints(#)
      nonconstant predict(newvar) vce(vcetype) quadcheck quadoutput vuong
      maximize_options]
```

where *devar* has to be a strictly positive outcome.

3.3 Options

`irr` reports incidence-rate ratios. `irr` is not allowed with `quadcheck`.

`nonadaptive` uses standard Gauss–Hermite quadrature; the default is adaptive quadrature.

`intpoints(#)` chooses the number of points used for the approximation. The default is `intpoints(30)`. The maximum is 195 without `quadcheck` and 146 with `quadcheck`. Generally, a higher number of points leads to a more accurate approximation, but it takes longer to converge. It is highly recommended to check the sensitivity of the results; see `quadcheck`.

`noconstant` suppresses the constant term (intercept) in the model.

`predict(newvar)` calculates the estimate of the conditional mean of n given $n > 0$; that is, $E(n|n > 0)$, which is $\exp(xb + \epsilon)/\Pr(n > 0|x)$. Gauss–Hermite quadrature is used to calculate the conditional mean.

`vce(vctype)` specifies the type of standard error reported, which includes `oim`, `robust`, `cluster clustvar`, or `opg`; see [R] *vce_option*.

`quadcheck` checks the sensitivity of the quadrature approximation by refitting the model with more and fewer quadrature points. `quadcheck` is not allowed with `irr`.

`quadoutput` shows the iteration log and the output of the refitted models.

`vuong` performs a Vuong test of `ztpnm` versus `ztnb`.

maximize_options: `difficult`, `technique(algorithm_spec)`, `iterate(#)`, `[no]log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, `nonrtolerance`, and `from(init_specs)`; see [R] `maximize`. These options are seldom used.

`difficult` is the default.

4 Examples

Santos Silva (2003) points out that in the estimation of mixture models under endogenous sampling, the distribution of the unobserved heterogeneity can be specified in both the actual population and the artificial or truncated population. Independence between the unobservables and the covariables can thus be assumed in both populations. The choice between these two alternatives is not innocuous, and this seemingly slight difference may lead to substantially different results. In hurdle models, the mechanism generating the positive counts is assumed to be different from the one generating the zeros. The population of interest for the positive counts is the truncated population because only observations in this group carry information on the mechanism that generates the positives. Thus covariables and unobservables should be independent in the truncated population.

The popular hurdle model is based on the negative binomial distribution. In this model, the unobserved heterogeneity follows a gamma distribution where covariables and unobservables are assumed to be independent in the actual population. However, independence in the actual population generally rules out independence in the truncated population. The following example illustrates this problem. Here counts are generated

by a Poisson process¹ where λ is assumed to depend on a covariable x and an unobserved variable e . Although x and e are orthogonal in the actual population, they are correlated in the truncated population. The reason is that given a certain value of x , it is more likely to get truncated if the unobserved individual effect is lower than average. As a consequence, x and e are correlated in the truncated population.

```
. set obs 3000
obs was 0, now 3000
. set seed 1234
. generate x = invnormal(runiform())
. generate e = invnormal(runiform())*0.7
. generate lambda = exp(0.5 + 0.5*x + e)
. genpoisson y, mu(lambda)
. pwcorr x e, star(0.1)
```

	x	e
x	1.0000	
e	0.0028	1.0000

```
. drop if y==0
(756 observations deleted)
. pwcorr x e, star(0.001)
```

	x	e
x	1.0000	
e	-0.1027*	1.0000

In the following Monte Carlo experiment, x and e are orthogonal in the truncated population. Therefore, the estimates of the truncated negative binomial regression applied to the simulated datasets are expected to be different from the true values. See McDowell (2003) for an example of a data-generating process that can be consistently estimated by a zero-truncated negative binomial model. Assume that McDowell's `lambda` additionally contains an appropriate heterogeneity term. The hurdle model described in section 2 assumes that the unobservables are independent of the covariables in the truncated population. This makes it possible to estimate the parameters consistently.

First, results are presented for simulated data, and then results are presented for a sample of the U.S. Medical Expenditure Panel Survey. The artificial count-data variable could be, for example, the number of doctor visits in a certain period. The outcome is generated using the inverse transformation method where the zeros and the positive counts come from different data-generating processes. The program `mc_ztpnm` simulates only the positive counts y . There is one explanatory variable x with parameter $\beta = 0.5$ and a constant of 0.5. The unobserved heterogeneity e is assumed to be normal with a standard deviation of $\sigma = 0.7$. The program simulates the data with `obs()` observations and estimates a regression model of y on x , which has to be specified in `command()`.

1. `genpoisson`, used below, can be obtained from <http://www.stata.com/users/rgutierrez/gendist/genpoisson.ado> or by typing `findit genpoisson` in Stata.

```

. program define mc_ztpnm, rclass
1.     syntax, COMmand(string) [obs(integer 3000) Beta(real 0.5)
> CONSTant(real 0.5) SIGMa(real 0.7) options(string)]
2.     quietly{
3.         drop _all
4.         set obs `obs'
5.         generate x = invnormal(runiform())
6.         generate e = invnormal(runiform())*`sigma'
7.         generate xb = `constant' + `beta'*x + e
8.         generate z = runiform()
9.         generate double fy_cdf=0
10.        generate y=.
11.        forvalues k=1/200 {
12.            generate double fy`k' = (exp(xb)^`k'*exp(-exp(xb)))/
> ((1-exp(-exp(xb)))*exp(lnfactorial(`k')))
13.            replace fy_cdf = fy_cdf + fy`k'
14.            replace y=`k' if fy_cdf - fy`k'<z & fy_cdf>z
15.        }
16.    }
17.    `command' y x, robust `options'
18. end

```

The `simulate` command is used to replicate `mc_ztpnm` 50 times with 3,000 cross-sectional observations for each replication. The first part of the results, below, shows the estimates from zero-truncated negbin II regressions using `ztnb`, and the second part displays the results from zero-truncated Poisson-lognormal regressions. The average estimate of β from the truncated negbin II models is 0.5834, which is noticeably different from the true value ($\beta = 0.5$). The estimates of `ztnb` are biased if the observed and unobserved variables are independent in the truncated population. The truncated Poisson-lognormal model is estimated using the `ztpnm` command. The default is adaptive quadrature with 30 nodes. The average estimate of β is now almost identical to the true value, and the average log likelihood is highest.

In addition, Vuong (1989) tests are performed to more formally select between both models (`ztnb` and `ztpnm`). Note that the models are overlapping rather than strictly nonnested. The zero-truncated Poisson-lognormal and negbin II models collapse to a zero-truncated Poisson model under the restrictions that σ and α are zero. The Vuong test can only be interpreted if these conditions are rejected. Otherwise, it is not possible to discriminate between the two models. If σ and α are not equal to zero, a test statistic larger than 1.96 indicates that the truncated Poisson-lognormal model is more appropriate. If the value of the statistic is smaller than -1.96 , the truncated negbin II model is better. In this example, σ and α are significantly different from zero by definition, and the average value of the Vuong statistic is 1.97, which indicates a better fit of the zero-truncated Poisson-lognormal model.

```

. simulate _b ll=e(l1) N=e(N), reps(50) seed(1234): mc_ztpnm, com(ztnb)
      command: mc_ztpnm, com(ztnb)
      [_eq3]ll: e(l1)
      [_eq3]N: e(N)

Simulations (50)
-----|-----|-----|-----|-----|-----|-----
      | 1 | 2 | 3 | 4 | 5 |
      |-----|-----|-----|-----|-----|-----
      ..... 50

. summarize, sep(3)

      Variable | Obs | Mean | Std. Dev. | Min | Max
      -----|-----|-----|-----|-----|-----
      y_b_x | 50 | .5834298 | .023775 | .5415038 | .6273451
      y_b_cons | 50 | .4490598 | .0419869 | .3584312 | .5225616
      lnalpha_b_s | 50 | -.1442036 | .1081242 | -.4058383 | .0702536

      _eq3_ll | 50 | -5127.739 | 71.44369 | -5341.06 | -4964.11
      _eq3_N | 50 | 3000 | 0 | 3000 | 3000

. simulate _b ll=e(l1) N=e(N) vuong=e(vuong), reps(50) seed(1234): mc_ztpnm,
> com(ztpnm) options(vuong)
      command: mc_ztpnm, com(ztpnm) options(vuong)
      [_eq3]ll: e(l1)
      [_eq3]N: e(N)
      [_eq3]vuong: e(vuong)

Simulations (50)
-----|-----|-----|-----|-----|-----
      | 1 | 2 | 3 | 4 | 5 |
      |-----|-----|-----|-----|-----|-----
      ..... 50

. summarize, sep(3)

      Variable | Obs | Mean | Std. Dev. | Min | Max
      -----|-----|-----|-----|-----|-----
      eq1_b_x | 50 | .5016896 | .0201121 | .4668793 | .542971
      eq1_b_cons | 50 | .4974964 | .0277192 | .4312099 | .5555598
      lnsigma_b_s | 50 | -.3553845 | .0313651 | -.4273075 | -.2912054

      _eq3_ll | 50 | -5117.823 | 70.90838 | -5324.094 | -4952.258
      _eq3_N | 50 | 3000 | 0 | 3000 | 3000
      _eq3_vuong | 50 | 1.967658 | .6502108 | .8654493 | 3.760262
    
```

Table 1 displays the simulation results for different values of σ . The higher σ , the further away are the estimates of `ztnb` from the true values. The correlation between x and e in the truncated sample also increases with rising σ . The estimates of `ztpnm` are consistent for all three parameterizations. There is no decision necessary if $\sigma = 0$ because both models collapse to a zero-truncated Poisson model in this case.

Table 1. Simulation results for different values of σ

σ	$\widehat{\beta}_{ztnb}$	$\widehat{\beta}_{ztpnm}$	$\text{Corr}(x, e)^+$
0.3	0.5165	0.5010	-0.0536
0.5	0.5436	0.5004	-0.0856
0.7	0.5834	0.5017	-0.1027

The real-data example analyzes a cross-section sample from the U.S. Medical Expenditure Panel Survey for 2003. The dependent variable is the annual number of doctor visits of individuals 65 years and older (see Cameron and Trivedi [2010] for more information about this dataset). Suppose we are interested in the effect that is attributable to Medicaid insurance on the number of doctor visits. The outcome is analyzed using a hurdle model because there are more zeros in the data than what was predicted by a Poisson count-data model (see figure 2).

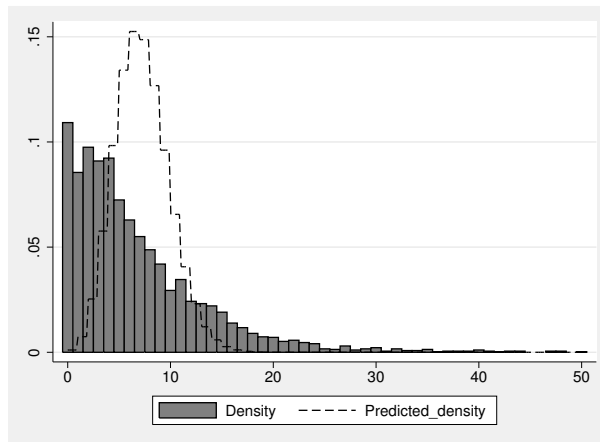


Figure 2. Observed frequencies versus probability mass function (\sim Pois[mean of docvis])

You can think about the demand for health care as a two-part decision process. At first, it is up to the patient to decide whether to visit a doctor. But after the first contact, the physician influences the intensity of treatment (Pohlmeier and Ulrich 1995). Assuming that the error terms of the binary and the truncated models are uncorrelated, the maximization process can be separated. In this case, one can first maximize a binary model with at least one doctor visit as the dependent variable, using the full sample. Second, one can estimate a zero-truncated regression separately using only observations with positive counts.

The following example discusses only the choice of the truncated part. Firstly, a zero-truncated negbin II model is applied. The results indicate a weakly significant increase of 10% in the expected number of doctor visits, conditional on use, which is attributable to Medicaid insurance.

```
. use mus17data
. global xs private medicaid age age2 educyr actlim totchr
. generate ytrunc = docvis
. replace ytrunc = . if ytrunc==0
(401 real changes made, 401 to missing)
```

```
. ztnb ytrunc $xs, robust
(output omitted)
Zero-truncated negative binomial regression      Number of obs   =      3276
Dispersion   = mean                            Wald chi2(7)    =      474.34
Log likelihood = -9452.899                      Prob > chi2     =      0.0000
```

ytrunc	Robust		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
private	.1095567	.0382086	2.87	0.004	.0346692	.1844442
medicaid	.0972309	.0589629	1.65	0.099	-.0183342	.212796
age	.2719032	.0671328	4.05	0.000	.1403254	.403481
age2	-.0017959	.000445	-4.04	0.000	-.002668	-.0009238
educyr	.0265974	.0050938	5.22	0.000	.0166137	.0365812
actlim	.1955384	.040658	4.81	0.000	.1158503	.2752266
totchr	.2226969	.0135761	16.40	0.000	.1960882	.2493056
_cons	-9.19017	2.517163	-3.65	0.000	-14.12372	-4.256621
/lnalpha	-.5259629	.0544868			-.632755	-.4191708
alpha	.590986	.0322009			.5311265	.6575919

Secondly, a zero-truncated Poisson lognormal model is estimated using the `ztpnm` command. Additionally, the accuracy of the likelihood approximation is checked using `ztpnm`'s `quadcheck` option, which reestimates the model with more and fewer quadrature points. The table below displays the fitted model with 30 quadrature points and, additionally, the results of the reestimated models with 20 and 40 points. The estimates and the likelihoods are almost the same in all three cases, which indicates that the likelihood approximation is good enough to rely on the results. Now the effect that is attributable to Medicaid is around 3% lower than the negbin II results and is no longer significant at the 10% level.

The Vuong test statistic is positive and larger than 1.96, which rejects the truncated negbin II model in favor of the truncated Poisson-lognormal model. Because σ is a boundary parameter, it should be tested separately instead of using the reported t -values. A likelihood-ratio test for $\sigma = 0$ gives $\bar{\chi}_{01}^2 = -2(-12998 + 9412) = 7172$. The level of significance is $\Pr(\chi_1^2 > 7172)/2 = 0.000$ (see also Gutierrez, Carter, and Drukker [2001]). The null hypothesis is rejected, which is probably the case in almost all applications.

```
. ztpnm ytrunc $xs, robust vuong quadcheck
Getting starting values from zero-truncated Poisson:
(output omitted)
Fitting zt-Poisson normal mixture model:
(output omitted)
Iteration 4: log pseudolikelihood = -9412.1934
Refitting model intpoints() = 20
Refitting model intpoints() = 40
```

```
*****
Quadrature check:
*****
```

Variable	qpoints_20	qpoints_30	qpoints_40
eq1			
private	.10347116	.10364263	.10365141
medicaid	.0721706	.07117263	.07111751
age	.26818424	.26861088	.26863663
age2	-.00176665	-.00176924	-.00176939
educyr	.02492124	.02486335	.02486017
actlim	.16483045	.16434532	.16431735
totchr	.22196453	.2220153	.22201903
_cons	-9.2211787	-9.2376583	-9.2386537
lnsigma			
_cons	-.3619317	-.36278788	-.36283427
Statistics			
ll	-9412.1877	-9412.1934	-9412.1935

```
*****
Fitted model:
*****
```

Number of quadrature points: 30

```
Zero-truncated Poisson normal mixture model      Number of obs   =      3276
                                                    Wald chi2(7)    =      585.30
Log pseudolikelihood = -9412.1934                Prob > chi2     =      0.0000
```

ytrunc	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
private	.1036426	.0330535	3.14	0.002	.038859	.1684262
medicaid	.0711726	.0448151	1.59	0.112	-.0166633	.1590086
age	.2686109	.0577928	4.65	0.000	.1553391	.3818826
age2	-.0017692	.0003837	-4.61	0.000	-.0025214	-.0010171
educyr	.0248634	.0043052	5.78	0.000	.0164253	.0333014
actlim	.1643453	.0342507	4.80	0.000	.0972152	.2314755
totchr	.2220153	.0115839	19.17	0.000	.1993113	.2447193
_cons	-9.237658	2.165506	-4.27	0.000	-13.48197	-4.993345
/lnsigma	-.3627879	.0192127	-18.88	0.000	-.400444	-.3251318
sigma	.695734	.0133669	52.05	0.000	.6700225	.7224321

```
*****
Vuong test of ztpnm vs. ztnb: z =      2.66 Pr>z = 0.0039
```

Finally, the same model is reestimated using nonadaptive quadrature with a very high number of nodes. This, of course, takes much longer to converge. The results are very similar to the estimates from the previous regression that uses adaptive quadrature to approximate the likelihood, but they vary a little bit depending on the number of quadrature points. This indicates that the likelihood approximation is not as accurate as the adaptive approximation.

```
. ztpnm ytrunc $xs, robust intpoints(120) quadcheck nonadaptive
(output omitted)
*****
Quadrature check:
*****
```

Variable	qpoints_80	qpoint-120	qpoint-160
eq1			
private	.10250717	.10340271	.10374735
medicaid	.07526089	.07241452	.07044432
age	.26699716	.26804252	.26878583
age2	-.00175964	-.00176579	-.00177027
educyr	.02513045	.02493826	.02482422
actlim	.16622169	.16496933	.16404298
totchr	.2219509	.2219467	.22200684
_cons	-9.1749427	-9.2156541	-9.244254
lnsigma			
_cons	-.35845104	-.36152824	-.36341895
Statistics			
ll	-9411.9925	-9412.1236	-9412.2131

(output omitted)

5 Conclusion

Hurdle models based on the zero-truncated Poisson-lognormal distribution are rarely used in applied work, although they incorporate some advantages compared with their negative binomial alternatives. These models are appealing from a theoretical point of view and, additionally, perform much better in many applications. The new Stata `ztpnm` command allows accurate and quick estimation of a zero-truncated Poisson-lognormal model with adaptive quadrature. This command can be used to model the positive counts in a hurdle model.

6 Acknowledgments

I am grateful to Florian Heiss and Rainer Winkelmann for their helpful comments. Financial support by Munich Center of Health Sciences is gratefully acknowledged.

7 References

- Cameron, A. C., and P. K. Trivedi. 2010. *Microeconometrics Using Stata*. Rev. ed. College Station, TX: Stata Press.
- Gutierrez, R. G., S. Carter, and D. M. Drukker. 2001. sg160: On boundary-value likelihood-ratio tests. *Stata Technical Bulletin* 60: 15–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 269–273. College Station, TX: Stata Press.
- Liu, Q., and D. A. Pierce. 1994. A note on Gauss–Hermite quadrature. *Biometrika* 81: 624–629.
- McDowell, A. 2003. From the help desk: Hurdle models. *Stata Journal* 3: 178–184.
- Mullahy, J. 1986. Specification and testing of some modified count data models. *Journal of Econometrics* 33: 341–365.
- Pohlmeier, W., and V. Ulrich. 1995. An econometric model of the two-part decision-making process in the demand for health care. *Journal of Human Resources* 30: 339–361.
- Santos Silva, J. M. C. 2003. A note on the estimation of mixture models under endogenous sampling. *Econometrics Journal* 6: 46–52.
- Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.
- Vesterinen, J., E. Pouta, A. Huhtala, and M. Neuvonen. 2010. Impacts of changes in water quality on recreation behavior and benefits in Finland. *Journal of Environmental Management* 91: 984–994.
- Vuong, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307–333.
- Winkelmann, R. 2004. Health care reform and the number of doctor visits—An econometric analysis. *Journal of Applied Econometrics* 19: 455–472.
- . 2008. *Econometric Analysis of Count Data*. 5th ed. Berlin: Springer.
- Wong, I. O. L., M. J. Lindner, B. J. Cowling, E. H. Y. Lau, S.-V. Lo, and G. M. Leung. 2010. Measuring moral hazard and adverse selection by propensity scoring in the mixed health care economy of Hong Kong. *Health Policy* 95: 24–35.

About the author

Helmut Farbmacher is a PhD student at the University of Munich with research interest in applied health economics.