Count model selection and post-estimation to evaluate composite flour technology adoption in Senegal-West Africa

Presented by

Kodjo Kondo
PhD Candidate,
UNE Business School

Supervisors

Emeritus Prof. Euan Fleming
Prof. Oscar Cacho
Associate Prof. Renato Villlano

Canberra, September 25, 2015

CORAF/WECARD

une
University of New England

Outline

1. Background
2. Research Objectives & Questions
3. Materials & Methods
4. Results & Discussion
5. Implications
6. Way-forward
7. Acknowledgements

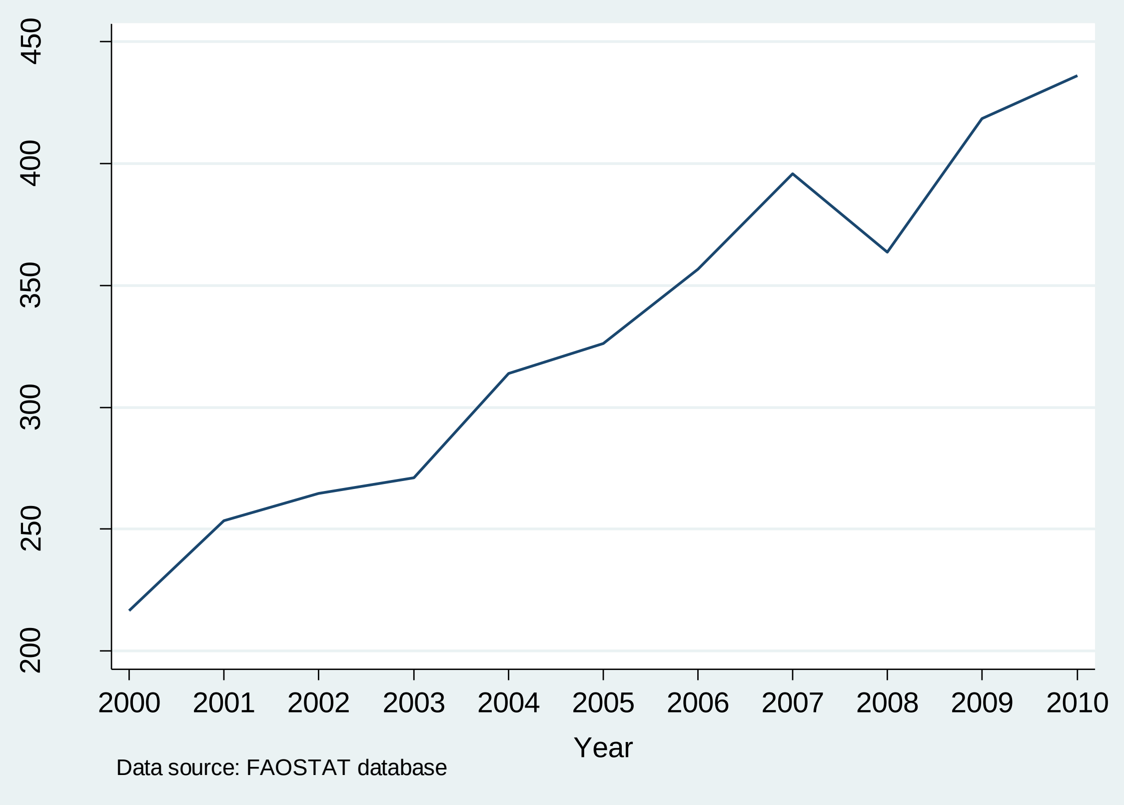
1. Background

Bread introduced in early 1900s

3 million loaves of bread produced daily

15 million population

1. Background



Data source: FAOSTAT database

Trend in value of wheat imports in Senegal (nominal prices)

+

1. Background

WAAPP / PPAAO
WEST AFRICA AGRICULTURAL PRODUCTIVITY PROGRAMME
PROGRAMME DE PRODUCTIVITE AGRICOLE EN AFRIQUE DE L'OUEST

2. Research Objective and Questions

Research Objective

Identify drivers and impediments to composite flour adoption to enable decision making towards limiting dependence on wheat imports.

Research Questions

1. What is the current rate of adoption of the composite flour technology?
2. What are the factors influencing the bakers' decisions to adopt composite flour and the quantities used?
3. Do the 80 kg mixers provided, the training programs and membership of FNBS influence both decisions positively?

3. Methods

3.1 Data

Mixed-Method Approach
April to August 2014

Qualitative

- Desk review (CORAF)
- 3 key informant interviews

Quantitative

150 bakers in 4 districts in Dakar region

Data entry -> SPSS
Data mgnt & analyses -> Stata

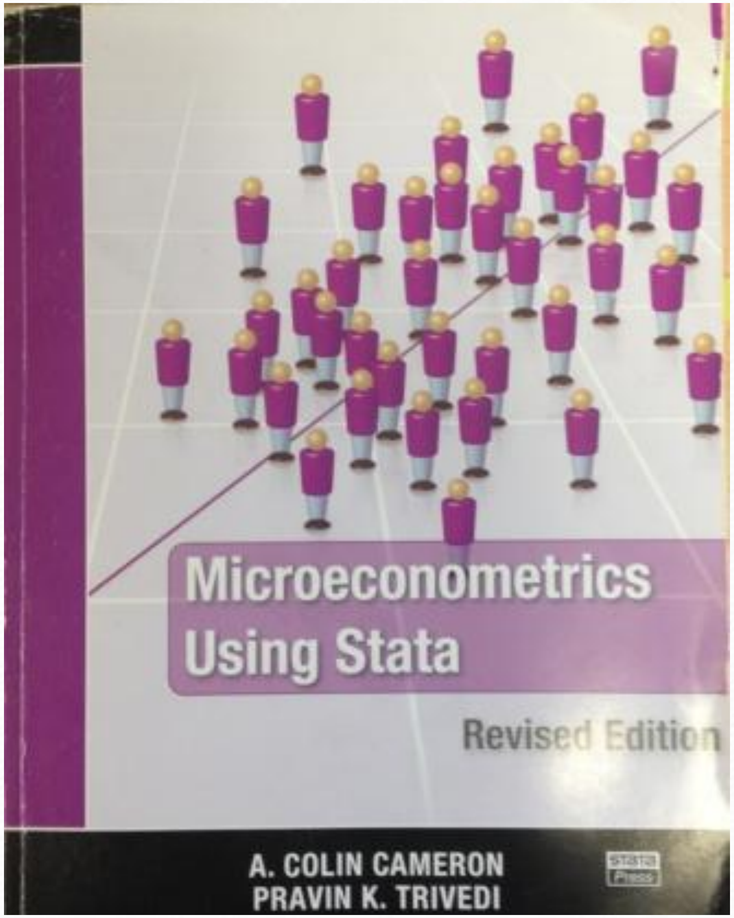
3. Methods

3.2 Analytical Framework

Decision process Assumption	Nature of the dependent variable	
	Continuous	Counts
Joint	Tobit	Poisson/NBRM
Separate	Cragg’s two-part model	Mullahy’s hurdle model
1st stage	○ logit/probit	○ logit/probit
2nd stage	○ OLS / truncated reg.	○ ZTP /ZTNB
	Heckit	ZIP /ZINB

3. Methods

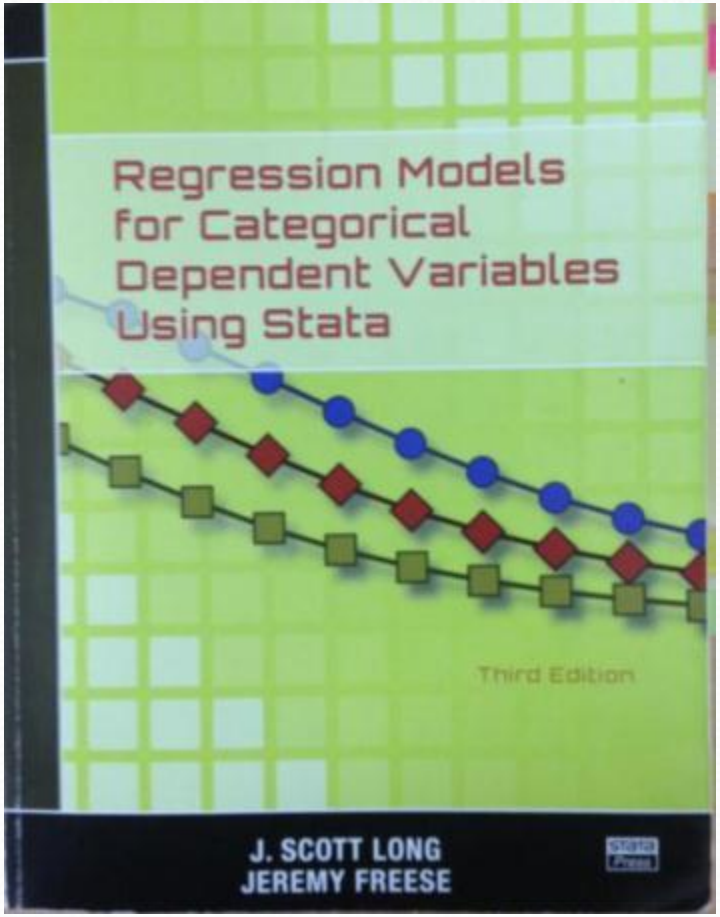
3.2 Analytical Framework



Microeconometrics Using Stata

Revised Edition

A. COLIN CAMERON
PRAVIN K. TRIVEDI



Regression Models for Categorical Dependent Variables Using Stata

Third Edition

J. SCOTT LONG
JEREMY FREESE

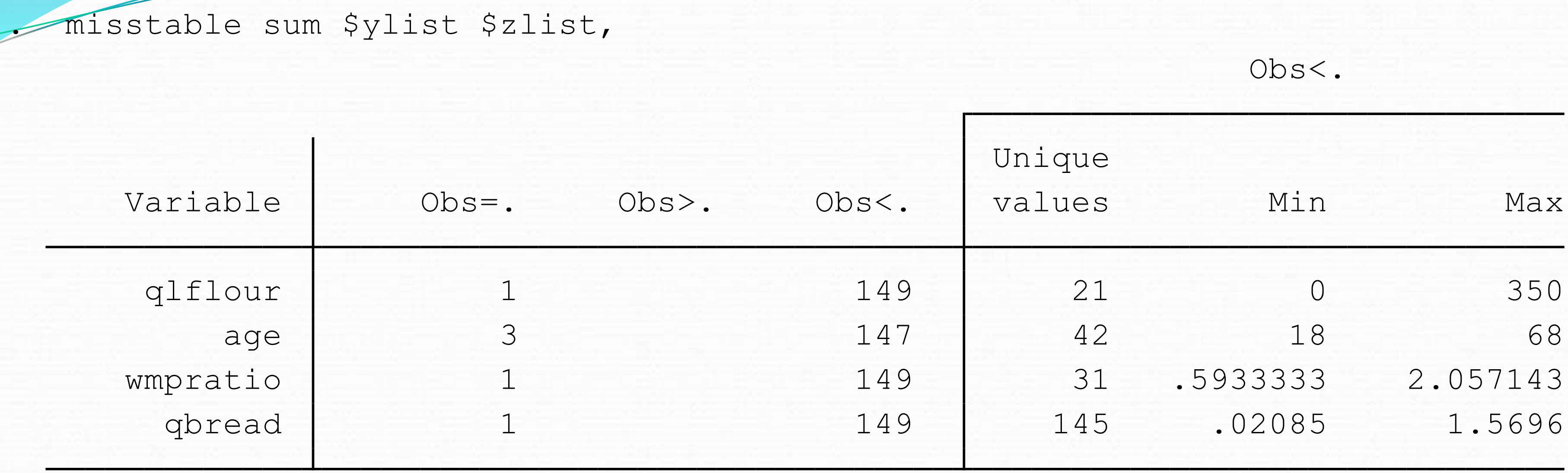
3. Methods

3.3 Choice of Estimation Variables

variable name	variable label
qlflour	Quantity of local flour used (kg/day)
owner	1 if owner-managed
age	Age of the respondent (years)
exper	Experience in c. bread baking (years)
educ	1 if received any education
bakeries	Number of bakeries owned
fnbsmb	1 if member of FNBS
cftrain	1 if trained in cf production
cbtrain	1 if trained in c. bread baking
mixer50	1 if bakery uses a 50kg mixer
mixer80	1 if bakery uses a 80kg mixer
mixer100	1 if bakery uses a 100kg mixer
wmpratio	Wheat-millet price ratio
qbread	Quantity of bread produced (tonnes)
district	1=Dakar,2=Pikine,3=Rufisque,4=Guediawaye

3. Methods

3.4 Missing values



```
. misstable sum $ylist $zlist,
                                                            Obs<.
                                               +------------------------------
                |                              | Unique
       Variable |     Obs=.     Obs>.     Obs<.| values        Min         Max
  --------------+------------------------------+------------------------------
        qlflour |         1                 149|     21          0         350
            age |         3                 147|     42         18          68
       wmpratio |         1                 149|     31   .5933333    2.057143
         qbread |         1                 149|    145     .02085      1.5696
  --------------+------------------------------+------------------------------
```

```
. mark nomiss

. markout nomiss $ylist $zlist

. drop if nomiss==0 //To drop all obse
```

Acock, (2014, pp.400-402)

3. Methods

3.5 Model Selection

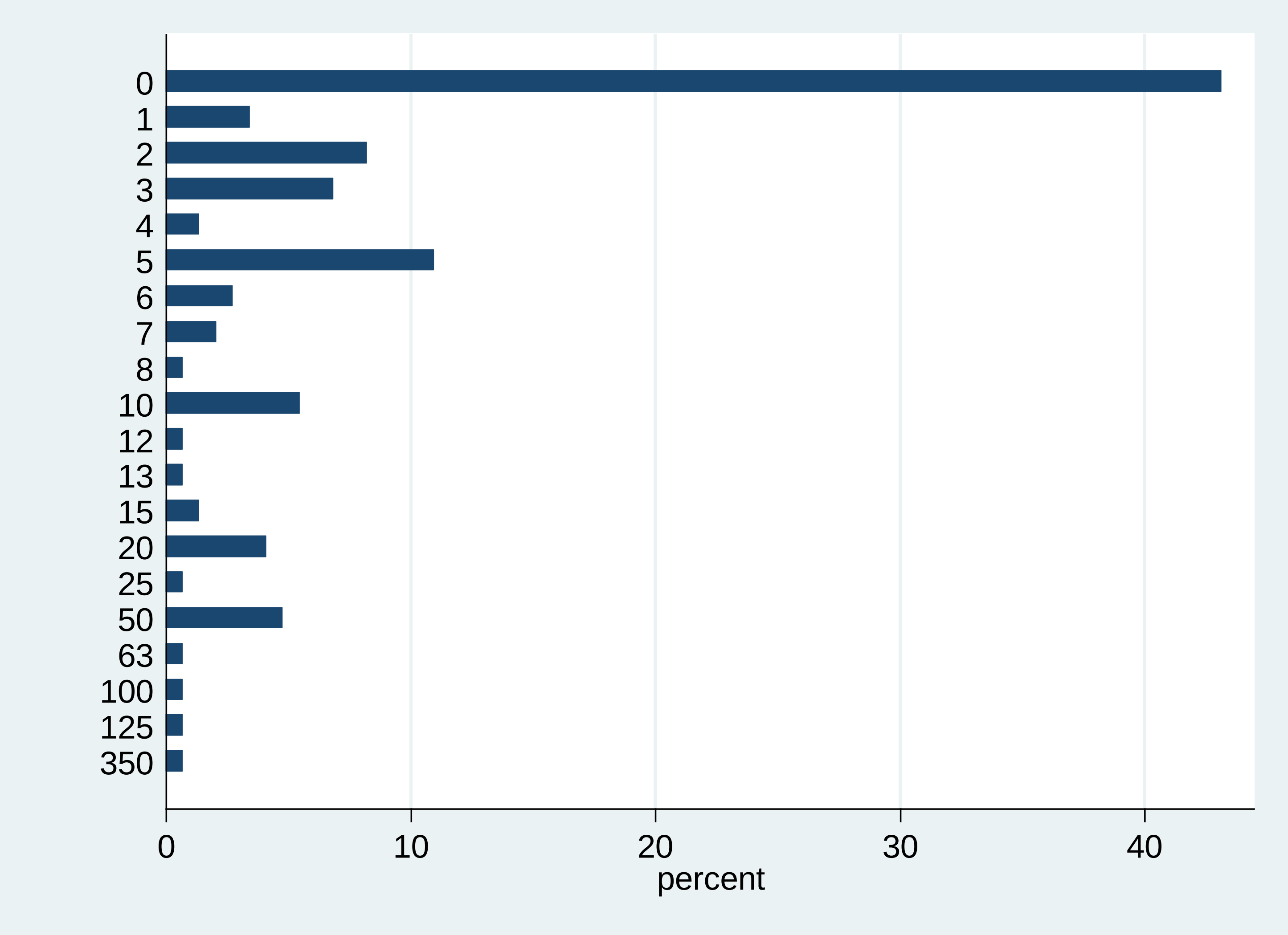
Which model to use?

3. Methods

3.5 Model selection

```
. catplot qlflour, percent
```

✓ Count data models



Distribution of the dependent variable

Poisson Regression Model (PRM)

$$\Pr(y_i|x_k) = \frac{\exp(-\mu_i)\ \mu_i^{y_i}}{y_i!} \quad \text{where} \quad \mu_i = E(y_i|\,x_k) = \exp\,(x_k\beta_k) \quad \& \quad var\,(y_i|x_k) = \mu_i$$

```
. quietly poisson $ylist $xlist, nolog irr vce(robust)

. estat gof

         Deviance goodness-of-fit =  2596.768
         Prob > chi2(129)         =    0.0000

         Pearson goodness-of-fit  =  3809.871
         Prob > chi2(129)         =    0.0000
```

Negative Binomial Regression Model (NBRM)

$$Pr(y_i|\alpha,\mu) = \frac{\Gamma(y+1/\alpha)}{\Gamma(y+1)\Gamma(1/\alpha)}\left(\frac{1}{1+\alpha\mu}\right)^{1/\alpha}\left(1-\frac{1}{1+\alpha\mu}\right)^{y}$$

where

$$var\left(y_i|\mu_i,\alpha\right) = \mu_i + \alpha\mu_i^{\,p}$$

Cameron and Trivedi (2013, p. 74)

Hardin and Hilbe (2014, pp. 286-287)

```
. nbreg $ylist $xlist, nolog irr vsquish

Negative binomial regression                    Number of obs      =       146
                                                LR chi2(16)        =     56.30
Dispersion      = mean                          Prob > chi2        =    0.0000
Log likelihood = -371.68983                     Pseudo R2          =    0.0704
```

Estimates for predictors omitted

```
 .
-------------+----------------------------------------------------------------
    /lnalpha |   1.023622   .1485063                       .7325553    1.314689
-------------+----------------------------------------------------------------
       alpha |   2.783258   .4133315                        2.08039    3.723594
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 2169.19              Prob >= chibar2 = 0.000
```

Negative Binomial Regression Model (NBRM)

* Discriminating between NB1 and NB2

Hardin and Hilbe (2014)

```
. nbregp  $ylist $xlist, nolog vsquish vce(robust)

Negative binomial-P regression                   Number of obs   =        146
                                                 Wald chi2(16)   =      84.49
Log pseudolikelihood = -371.6014                 Prob > chi2     =     0.0000
```

Estimates for predictors omitted

```
 .
 _____________|_______________________________________________________________
            /P |   2.080491    .337584     6.16   0.000     1.418839    2.742144
      /lntheta |   .8388544   .8220121                     -.7722598    2.449969
 _____________|_______________________________________________________________
         theta |   2.313715   1.901902                      .4619679    11.58798
 _____________|_______________________________________________________________
 Likelihood-ratio test of P=1:     chi2 =     13.53 Prob > chi2      =     0.0002
 Likelihood-ratio test of P=2:     chi2 =      0.18 Prob > chi2      =     0.6741
```

3.5 Model selection

Single Decision process

* Discriminating between NB1 and NB2

Long and Freese (2014)

```
. quietly nbreg $ylist $xlist, nolog dispersion (constant) vce (robust)
. fitstat, using (nb2)

                          |     Current         Saved    Difference
--------------------------+--------------------------------------------
 .
IC                        |
                      AIC |     792.737       779.380        13.357
           AIC divided by N |       5.430         5.338         0.091
       BIC (df=18/18/0)   |     846.441       833.085        13.357

Note: Some measures based on pseudolikelihoods.

Difference of   13.357 in BIC provides very strong support for saved model.
```

3. Methods

3.5 Model selection

Single Decision process

Are some non-adopters constrained ?

C:\Users\kkondo3\
uments\phd\6_Jou

Zero Inflated Negative Binomial (ZINB)

```
. zinb $ylist $xlist, inflate(i.owner age i.educ i.mixer50 i.district) nolog vuong vsquish

Zero-inflated negative binomial regression        Number of obs     =       146
                                                  Nonzero obs       =        83
                                                  Zero obs          =        63

Inflation model = logit                           LR chi2(16)       =     66.10
Log likelihood  =  -357.117                       Prob > chi2       =    0.0000
```

Estimates for predictors omitted

```
.
--------------+-----------------------------------------------------------------------
     /lnalpha |   .2362462    .2027765     1.17   0.244    -.1611884     .6336808
--------------+-----------------------------------------------------------------------
        alpha |   1.266486    .2568136                      .8511317     1.884534
--------------------------------------------------------------------------------------
Vuong test of zinb vs. standard negative binomial: z =     2.62  Pr>z = 0.0044
```

3. Methods

3.5 Model selection

Two – stage decision

Adoption decision -> Binary choice logit

$$\Pr(y_i=0|x_k) = \frac{\exp(x_k\gamma_k)}{1+\exp(x_k\gamma_k)} = \pi_i$$

```
. quietly logit $ylist $xlist, nolog vce (robust) vsquish
. est store logit
```

```
Classified + if predicted Pr(D) >= .5
True D defined as qlflour != 0
--------------------------------------------------
Sensitivity                      Pr( +| D)   79.52%
Specificity                      Pr( -|~D)   71.43%
Positive predictive value        Pr( D| +)   78.57%
Negative predictive value        Pr(~D| -)   72.58%
--------------------------------------------------
False + rate for true ~D         Pr( +|~D)   28.57%
False - rate for true D          Pr( -| D)   20.48%
False + rate for classified +    Pr(~D| +)   21.43%
False - rate for classified -    Pr( D| -)   27.42%
--------------------------------------------------
Correctly classified                         76.03%
--------------------------------------------------
```

3.5 Model selection

Two – stage decision

Quantity decisions -> ZTP / ZTNB

$$E(y_i|y_i > 0, x_k) = \frac{\Pr(y_i|x_k)}{1 - (1 + \alpha\mu_i)^{-1/\alpha}}$$

x2list = x1list c.exper##c.exper

$$E(y_i|x_k) = (1 - \pi_i) * E(y_i|y_i > 0, x_k)$$

```
. tnbreg $ylist $x2list if $ylist>0, nolog vsquish

Truncated negative binomial regression               Number of obs    =         83
Truncation point: 0                                  LR chi2(18)      =      57.23
Dispersion       = mean                              Prob > chi2      =     0.0000
Log likelihood =  -271.2018                          Pseudo R2        =     0.0954
```

Estimates for predictors omitted

```
  .
________________|_______________________________________________________________
       /lnalpha |   .1034232   .2498786                      -.3863298    .5931763
________________|_______________________________________________________________
          alpha |   1.108961   .2771055                       .6795464    1.809727
________________|_______________________________________________________________
LR test of alpha=0: chibar2(01) = 1115.52                 Prob >= chibar2 = 0.000
```

```
. est store ztnb
```

3. Methods

3.5 Model selection

Final tests

Likelihood ratio test

NBRM VS HRM (LOGIT + ZTNB)

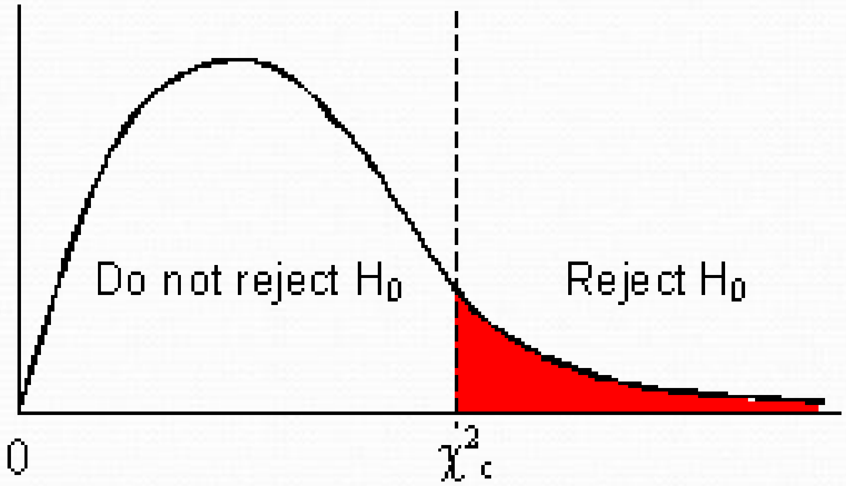
```
quietly nbreg  $ylist $xlist, nolog vce(robust)
scalar llnbreg=e(ll)
quietly logit $ylist $xlist, nolog vce(robust)
scalar lllogit=e(ll)
quietly ztnb $ylist $x2list if $ylist>0, nolog///
vce(robust)
scalar llztnb=e(ll)
scalar nbregtest= -2*( llnbreg-(lllogit+llztnb))
display "nbregtest = " nbregtest
```

nbregtest = 50.211523

ZINB VS HRM (LOGIT + ZTNB)

```
quietly zinb  $ylist $xlist, inflate(i.owner age ///
  i.educ i.mixer50 i.district) vce(robust)
scalar llzinb=e(ll)
quietly logit $ylist $xlist, nolog vce(robust)
scalar lllogit=e(ll)
quietly ztnb $ylist $x2list if $ylist>0, nolog vce(robust)
scalar llztnb=e(ll)
scalar zinbtest= -2*(llzinb - (lllogit+llztnb))
display "zinbtest = " zinbtest
```

zinbtest = 21.065822



(df1, 0.05) = 3.84

HRM finally favoured

3.5 Model selection Summary

#	Test	Test type	Null Hypothesis	Test statistics	Critical values (df1, 0.05)	Decision	Model favoured
1	PRM vs NBRM	LR	Alpha (α) = 0	2169.19[a]	3.84	Reject H_0	NBRM
2	NB-1 vs NB-2	LR	p =1	13.53[b]	3.84	Reject H_0	NB-2
			p =2	0.18	3.84	Accept H_0	
3	ZTP vs ZTNB		Alpha (α) = 0	1115.52	3.84	Reject H_0	ZTNB
4	NB-2 vs HRM	LR	NB-2 better than HRM	50.21[c]	3.84	Reject H_0	HRM
5	ZINB vs NB-2	Vuong	ZINB better than NB-2	2.62[d]	6.31	Accept H_0	ZINB
6	ZINB vs HRM	LR	ZINB better than HRM	21.07	3.84	Reject H_0	HRM

4. Results

4.1 Adoption rate

```
. fre composite

composite — 1 if baker uses composite flour, 0 otherwise
─────────────────────────┬──────────────────────────────────────────
                         │      Freq.    Percent      Valid       Cum.
─────────────────────────┼──────────────────────────────────────────
Valid   0 Non_adopters   │         63      43.15      43.15      43.15
        1 Adopters       │         83      56.85      56.85     100.00
        Total            │        146     100.00     100.00
─────────────────────────┴──────────────────────────────────────────
```

Conditional & unconditional daily local flour consumption

```
. tabstat qlflour, statistics (N mean sd min max) by (composite)

Summary for variables: qlflour
     by categories of: composite ( 1 if baker uses composite flour, 0 otherwise)

   composite │         N      mean        sd       min       max
─────────────┼──────────────────────────────────────────────────
Non_adopters │        63         0         0         0         0
    Adopters │        83  17.68675  42.64829         1       350
─────────────┼──────────────────────────────────────────────────
       Total │       146  10.05479  33.25467         0       350
─────────────┴──────────────────────────────────────────────────
```

4.2 Factors Affecting Adoption and Extent of Adoption

```
. estout logit ztnb, eq(1) style(fixed) cells((b (star fmt(%9.3f)) ///
>   z(par fmt(%9.3f)))  collabels("") mlabels("logit_OR" "ztnb_IRR")  ///
>   eqlabels(,none) stats(N df_m chi2 p, ///
>   labels ("Observations" "df" "Chi2" "prob>chi2") ///
>   fmt(%9.0f %9.3f)) varlabels(lnalpha:_cons _alpha) legend nobase ///
>   starlevels(* 0.10 ** 0.05 *** 0.01) eform
```

4. Results

HRM

	logit_OR		ztnb_IRR	
1.owner ✗	1.795	(1.108)	0.310***	(-4.213)
age ✓	0.882	(-0.925)	1.200*	(1.891)
c.age#c.age	1.002	(0.946)	0.998*	(-1.686)
1.educ	6.366	(1.565)	0.261	(-1.376)
bakeries ✓	1.532**	(2.526)	1.105	(1.133)
1.fnbsmb ✓	1.124	(0.274)	1.660*	(1.786)
1.cftrain ✓	2.778*	(1.692)	1.456	(1.013)
1.cbtrain	0.504	(-1.284)	0.680	(-1.285)
1.mixer50 ✓	8.956***	(2.791)	1.111	(0.373)
1.mixer80	2.495	(1.479)	1.261	(0.677)
1.mixer100	1.301	(0.463)	1.165	(0.501)
wmpratio ✓	23.693*	(1.812)	5.378*	(1.841)
qbread ✓	2.249	(0.997)	5.343***	(3.180)
2.district	0.562	(-1.005)	0.488	(-1.503)
3.district ✗	3.509	(1.409)	0.252***	(-2.712)
4.district	2.607	(1.213)	0.688	(-1.020)
exper ✓			1.109**	(2.180)
c.exper#c.~r			0.996***	(-2.901)
_cons	0.009	(-1.399)	0.036*	(-1.822)
_alpha			1.109	(0.395)
Observations	146		83	
df	16.000		18.000	
Chi2	31.882		103.662	
prob>chi2	0.010		0.000	

* p<0.10, ** p<0.05, *** p<0.01

4. Results

Post-estimation results

Variables	Adoption model (logit)		Quantity model (ZTNB)	
	%OR 1dx	AME	% E(y) 1dx	AME
Bakers' characteristics				
owner	79.5	0.104	-69.0***	-19.075***
age	-11.8	-0.002	20.0*	0.60**
Age^2	0.2		-0.2*	
educ	536.6	0.301*	-73.9	-48.94
bakeries	53.2**	0.074***	10.5	1.78
exper			10.9**	0.54
Exper^2			-0.4***	
Institutional factors				
fnbsmb	12.4	0.021	66.0*	8.13*
cftrain	177.8*	0.178*	45.6	6.89
cbtrain	-49.6	-0.114	-32.0	-6.86
Production factors				
mixer50	795.6***	0.340***	11.1	1.88
mixer80	149.5	0.160	26.1	4.066
mixer100	30.1	0.045	16.5	2.60
wmpratio	2269.3*	0.550*	437.8*	29.88
qbread	124.9	0.141	434.3***	29.76**
Location				
Pikine vs Dakar	-43.8	-0.105	-51.2	-13.79
Rufisque vs Dakar	250.9	0.216	-74.8***	-20.15*
Guediawaye vs Dakar	160.7	0.170	-31.2	-8.40
Rufisque vs Pikine		0.321***		-6.36
Guediawaye vs Pikine		0.275***		5.39
Guediawaye vs Rufisque		-0.046		11.75
N		146		83
Predictions				
Prb (*qlflour* = 0)		0.432		
Prb (*qlflour* > 0)		0.568		
Expected *qlflour*				17.76

* p<0.10; ** p<0.05; *** p<0.01

Post-estimation results

. **mchange** bakeries wmpratio, amount (one sd range) stat (change from to se pvalue) decimals(3)

Digging into the effects of significant continuous variables on prb. adoption

```
logit: Changes in Pr(y) | Number of obs = 146

Expression: Pr(composite), predict(pr)

             |    Change       From         To    Std Err    p-value
-------------+------------------------------------------------------
bakeries     |
          +1 |     0.073      0.568      0.642      0.027      0.007
         +SD |     0.122      0.568      0.691      0.044      0.005
       Range |     0.480      0.515      0.996      0.047      0.000
wmpratio     |
          +1 |     0.378      0.568      0.946      0.086      0.000
         +SD |     0.081      0.568      0.649      0.042      0.055
       Range |     0.631      0.309      0.941      0.206      0.002

Average predictions

             |  Non_ado~s   Adopters
-------------+----------------------
   Pr(y|base)|     0.432      0.568
```

Post-estimation results

Impact of FNBS membership on the extent of adoption

```
. ttest qlflour if composite==1, by (fnbsmb)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Non_memb |      34    9.176471    2.009545    11.71756     5.08802    13.26492
 Members |      49    23.59184    7.727442     54.0921    8.054772     39.1289
---------+--------------------------------------------------------------------
combined |      83    17.68675    4.681258    42.64829    8.374235    26.99926
---------+--------------------------------------------------------------------
    diff |           -14.41537    9.442962                -33.2039    4.373167
------------------------------------------------------------------------------
    diff = mean(Non_memb) - mean(Members)                         t =  -1.5266
Ho: diff = 0                                     degrees of freedom =       81

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0654         Pr(|T| > |t|) = 0.1308          Pr(T > t) = 0.9346
```

4. Results

Impact of FNBS membership on the extent of adoption

```
. mtable, at(fnbsmb =(0 1)) atmeans ///
>   pr(1 2 3 5 10 15 20 50 60 70 80) width(7) brief

Expression: Pr(qlflour), predict(pr())
```

```
Expression: Pr(qlflour), predict(pr())

      |  fnbsmb       1       2       3       5      10      15      20      50      60      70      80
------+---------------------------------------------------------------------------------------------------
    1 |       0   0.083   0.073   0.065   0.053   0.033   0.021   0.014   0.001   0.000   0.000   0.000
    2 |       1   0.056   0.050   0.046   0.040   0.029   0.022   0.017   0.004   0.002   0.001   0.001
```

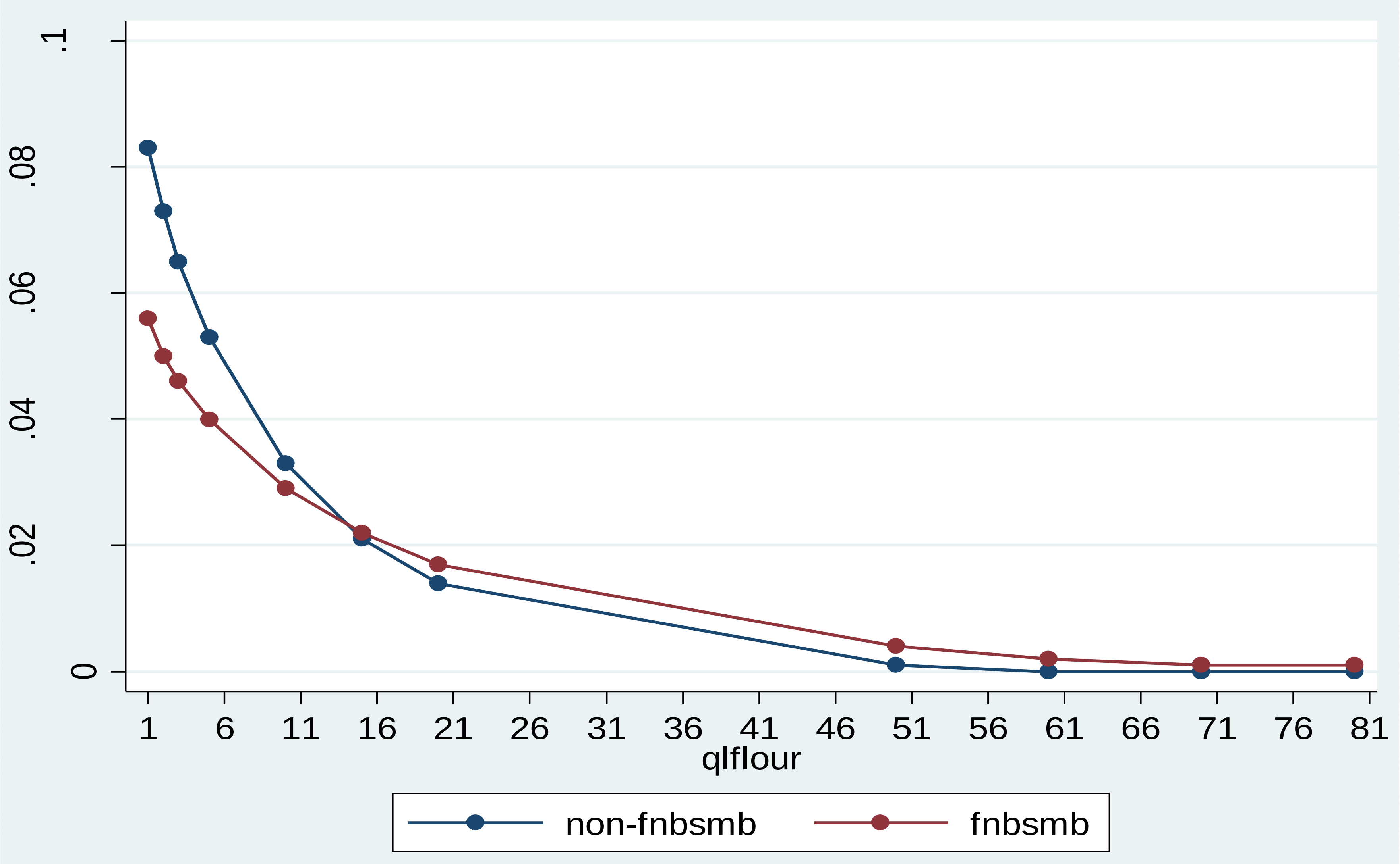
C:\Users\kkondo3\
uments\phd\6_Jou

```
. use  "…Prob counts by fnbsmb and non fnbsmb.dta", clear

. twoway (connected nonfnbsmb fnbsmb qlflour ), ytitle (Probability)  ylabel(0(.02).1, gmax) xlabel(1 (5) 80)
```

4. Results

Impact of FNBS membership on the extent of adoption



5. Implications

Strategies to enhance adoption and consumption of local flour to lessen importation of main ingredient might:

1. Promote 50 kg dough capacity mixers;
2. Intensify the professional training on composite flour production
3. Ensure the availability of local flour at lower price than wheat flour through diverse strategies
4. Institutionalise the use of the local flour in bread baking and
5. Strengthen the contractual arrangement between the value chain actors.

6. Way Forward

1. Provide evidence of technical and allocative inefficiencies among bakers
2. Identify factors affecting efficiencies including CF adoption.
3. Identify the determinants of adoption of improved varieties and good agronomic practices among millet and cassava farmers in Senegal and Ghana
4. Assess the impact of adoption on farmers’ technical and economic performances.

7. Acknowledgement

Funding

AUSTRALIA

DFAT

CORAF/WECARD

Technical and logistics support

ISRA

Bureau d'analyses macro-économiques

Institut sénégalais de recherches agricoles

Collaboration with

WAAPP / PPAAO

WEST AFRICA AGRICULTURAL PRODUCTIVITY PROGRAMME

PROGRAMME DE PRODUCTIVITE AGRICOLE EN AFRIQUE DE L'OUEST

ASPRODEB

FNBS

References

Acock, A. C. (2014), *A Gentle Introduction to Stata,* Stata Press, College Station, Tx.

Cameron, C. A. & Trivedi, P. K. (2010), *Microecometrics Using Stata,* Stata Press, College Station, Tx.

Cameron, C. A. & Trivedi, P. K. (2013), *Regression Analysis of Count Data,* Cambridge University Press, New York.

Cragg, J. G. (1971), "Some statistical models for limited dependent variables with application to the demand for durable goods", *Econometrica* Vol. 39, No., pp. 829–844.

Greene, W. H. (2012), *Econometric Analysis,* Pearson Education, Harlow, England.

Hardin, J. W. & Hilbe, J. M. (2014), "Regression models for count data based on the negative binomial (p) distribution", *Stata Journal*, Vol. 14, No., pp. 280-291.

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, No. 1, pp. 153-161.

Lambert, D. (1992), "Zero-inflated Poisson regression, with an application to defects in manufacturing", *Technometrics*, Vol. 34, No., pp. 1-14.

Long, S. J. & Freese, J. (2014), *Regression Models for Categorical Dependent Variables Using Stata,* Stata Press, College Station, Tx.

Mullahy, J. (1986), "Specification and testing of some modified count data models", *Journal of Econometrics*, Vol. 33, No., pp. 341-365.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, Vol. 26, No. 1, pp. 24-36.

Wooldridge, J. M. (2010), Econometric Analysis of Cross Section and Panel Data, The MIT Press, Cambridge, Massachusetts, London.

http://www.surveydesign.com.au/statabooks.html

jërëjëf