The Vuong Test for Non-nested Models

Hossein Kavand and Macel Voia
Department of Economics, Carleton University
9 June 2017

Vuong Test

- **AIC and SBC:** useful for choosing between nested models
- **Vuong Test:** useful for choosing between non-nested models

$$\begin{cases} H_0: Model\ (1) \text{ is correct} \\ H_1: Model\ (2) \text{ is correct} \end{cases}$$

- Let: $m_i = ln\left(\frac{f_1(y_i|x_i)}{f_2(y_i|x_i)}\right)$
- $f_1(y_i|x_i)$: Probability of observing y_i using Model (1)
- $f_2(y_i|x_i)$: Probability of observing y_i using Model (2)

$$V = \frac{\sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} m_i]}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})^2}} = \frac{\sqrt{n}\bar{m}}{s_m}$$

- Limiting distribution of V is a standard normal distribution
- If $V > 1.96$, Model 1 is preferred over Model 2 ($\alpha = 5\%$)
- If $|V| < 1.96$, no decision ($\alpha = 5\%$)
- If $V < -1.96$, Model 2 is preferred over Model 1 ($\alpha = 5\%$)

Vuong test is valid even if neither of the models are correctly specified under the null hypothesis

Zero-inflated Count Models

The Number of visiting Doctors

Adverse selection: When a high-risk individual buys more insurance coverage
Moral hazard: Those who have insurance visit more doctors
(Riphahn et al (JAE 2003) :No evidence of either effect)

	(Share of total observation, %)			
	Hospital visit		Doctor Visit	
Value	Males	Females	Males	Females
0	92.21	90.18	44.05	29.51
1	6.18	7.88	13.82	13.17
2	1.09	1.28	11.63	13.42
3	0.15	0.27	8.48	11.49
04-Sep	0.21	0.25	15.29	21.83
10 and more	0.16	0.14	6.73	10.58
N	14243	13083	14243	13083

Source, German Socioeconomic Panel (1984-1995)

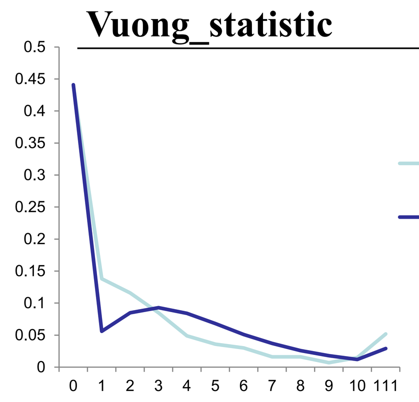
Zero-inflated count models (over-dispersion issues)
The probability of a zero outcome for the system is given by:

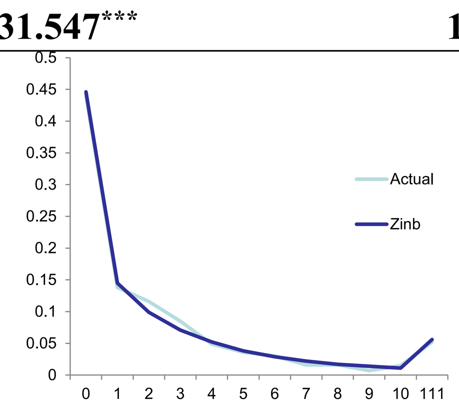
$$\Pr(y = 0) = B(0) + \{1 - B(0)\}\Pr(0)$$

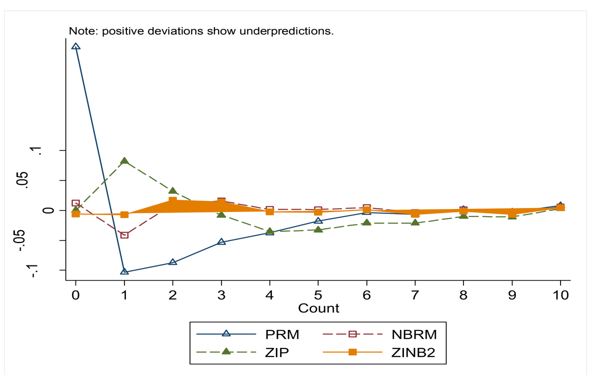
And the probability of a nonzero count is:

$$\Pr(y = k; k > 0) = \{1 - B(0)\}\Pr(k)$$

	ZIP	ZINB2
Doctor visit equation		
Public Insurance	0.079***	0.097*
	(0.025)	(0.057)
Add-on Insurance	-0.084*	-0.039
	(0.043)	(0.096)
Observations	14243	14243
AIC	70905.853	54536.989
BIC	71238.670	54877.369
Log lik.	-35408.9	-27223.5
Vuong_statistic	**31.547*** **	**11.555*** **







Hazard Models

The Weibull Hazard Model:

$$h_i(t) = \alpha t^{\alpha-1}\exp(\beta' X_i) \quad ; \alpha > 0$$

Discrete Finite Mixture Hazard Model (hshaz code in Stata**):**

$$E[(h(t|v, X)] = \sum_{k=1}^{K}[p_k\, h(t|v_k, X)]$$

Survival After first year

Quality	Percentage
Pre-entry experienced firms	69
Other firms	31
	100

Thompson's (REStat, 2005) shipbuilding data (1825-1914)

First Stage

$$\Pr(\text{survive} > 1\text{year}|X_{it}, Z_{it}) = \Phi(\delta' X_{it} + \gamma' Z_{it})$$

Z_{it}: A vector of exclusion restrictions

The Pinto (2008) Formula:

$$\frac{1}{N(S_t)}\sum_{i \in S_t} Y_{i,t} - \frac{1}{N(S_1)}\sum_{i \in S_1} Y_{i,1} = \textit{Selection component} + \textit{Survival component}$$

Probit model

Exclusion Restrictions (Z)	Marginal Effects
Firm's entry share (number)	0.181**
Firm's entry share (tonnage)	0.133**
Pinto entry-Selection component	0.309**

* $p < 0.10$, ** $p < 0.05$

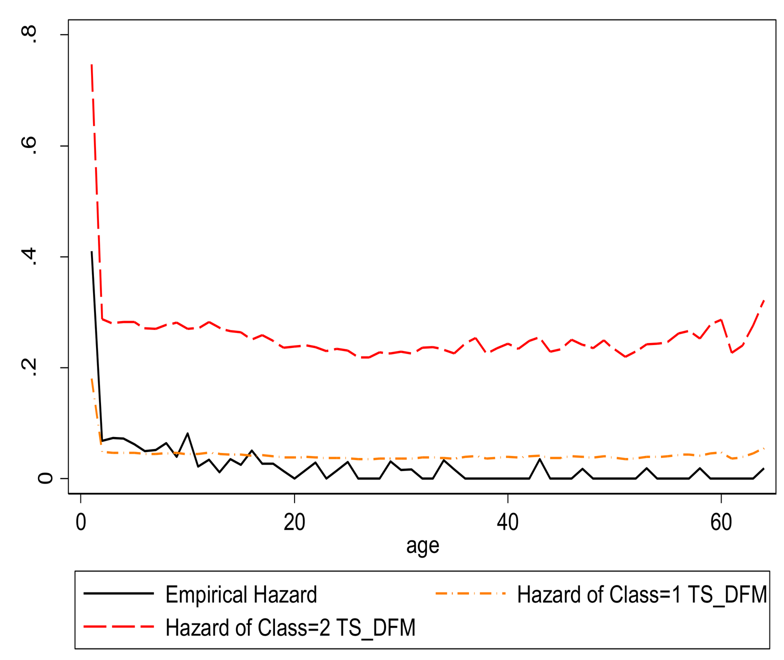
Second Stage

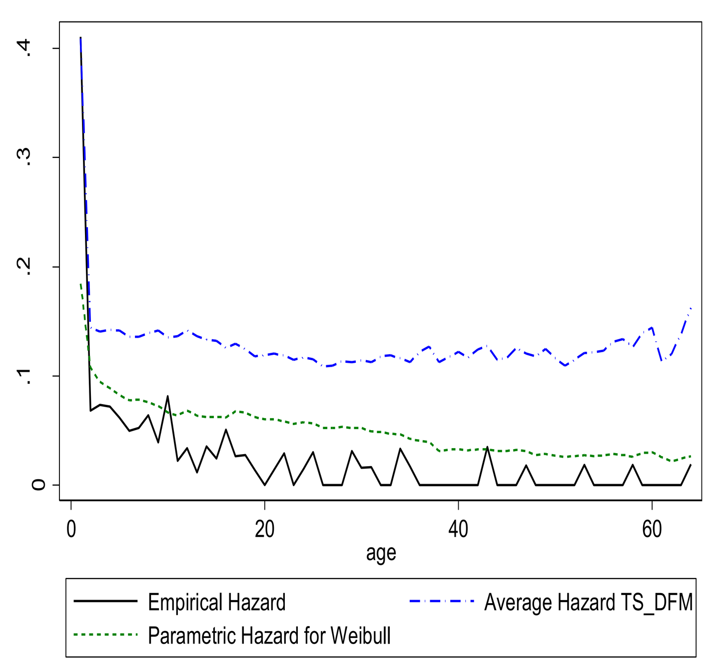
Hazard Ration: Different Model Specification

	Weibull Model	Finite Mixture
Pre-entry Experience:		
In Shipbuilding	0.630**	-
	(0.139)	
In engine building	0.524**	-
	(0.122)	
Miscellaneous	0.788	-
	(0.202)	
Selection		
Pr(survive>1 year)	-	0.020**
		(0.007)
Mixture Parameters		
Probability (class=1)	-	0.596**
		(0.099)
Probability (class=2)	-	0.404**
		(0.099)
Observations	14243	14243
Log lik.	-35408.9	-27223.5
Vuong_statistic		
Weibull vs DFM	27.985**	
Vuong_statistic (age<10)		
Weibull vs DFM	-2.124**	

* $p < 0.10$, ** $p < 0.05$

Hazard Prediction





Calculating Vuong for Hazard Models

```
***************Weibull model*************************
 streg …
.predictnl xb_weib=xb(); gen p_weib=e(aux_p); predict haz, hazard
predict sur, surv; gen density_weib1=haz*sur      if  e(sample)  & fail1==1
gen density_weib0=sur                    if  e(sample)  & fail1==0
replace      density_weib1=density_weib0 if density_weib1==.
 ************** Finite mixture model******************
hshaz …
predictnl xb_fm1_ex=   xb(); predictnl xb_fm2_ex=  e(m2) + xb()
gen hazard_fm1_ex= 1-exp(-exp(xb_fm1_ex))
gen hazard_fm2_ex= 1-exp(-exp(xb_fm2_ex))
bysort id (age):gen fm1_ex = exp(xb_fm1_ex)
bysort id (age):gen sum_fm1_ex = sum(exp(xb_fm1_ex))
bysort id (age):gen sum_fm2_ex = sum(exp(xb_fm2_ex))
bysort id (age):gen survival_fm1_ex = exp(-sum_fm1_ex[_N]) if _n==_N
bysort id (age):gen survival_fm2_ex = exp(-sum_fm2_ex[_N]) if _n==_N
gen  density_fm1_ex=e(pr1)*(hazard_fm1_ex/(1-hazard_fm1_ex))* ///
survival_fm1_ex+e(pr2)*(hazard_fm2_ex/(1-hazard_fm2_ex))*survival_fm2_ex if  fail1==1
gen  density_fm0_ex=e(pr1)*survival_fm1_ex+e(pr2)*survival_fm2_ex if  fail1==0
replace      density_fm1_ex=density_fm0_ex if density_fm1_ex==.
************** Vuong: Weibull versus Finite(H0:weibul)***********************************
gen weib_fm_density_ex=ln(density_weib1)-ln(density_fm1_ex)
quietly sum weib_fm_density_ex, d
gen sd_fmweib_ex= r(sd); gen num_fmweib_ex=r(N); gen mean_fmweib_ex=r(mean)
gen vuong_weibfm_ex= (mean_fmweib_ex*num_fmweib_ex^(0.5))/sd_fmweib_ex
```