Vietnam Center for Economic and Policy Research – VEPR Seminar

Heterogeneity of the Informal Sector in Vietnam:
A Quali-Quanti Approach

LE Thi Thuy Linh

IRD-DIAL

July 25, 2014

Context and motivation

- Prevalence of the informal sector in Vietnam:
  - 2007: 50% non-farm jobs and 20% GDP (Cling et al., 2010).

- Too little attention from the State.

- Knowledge in the informal sector in Vietnam is quite limited.

Research Question

- Is the informal sector in Vietnam homogeneous or heterogeneous, and how heterogeneous is it?
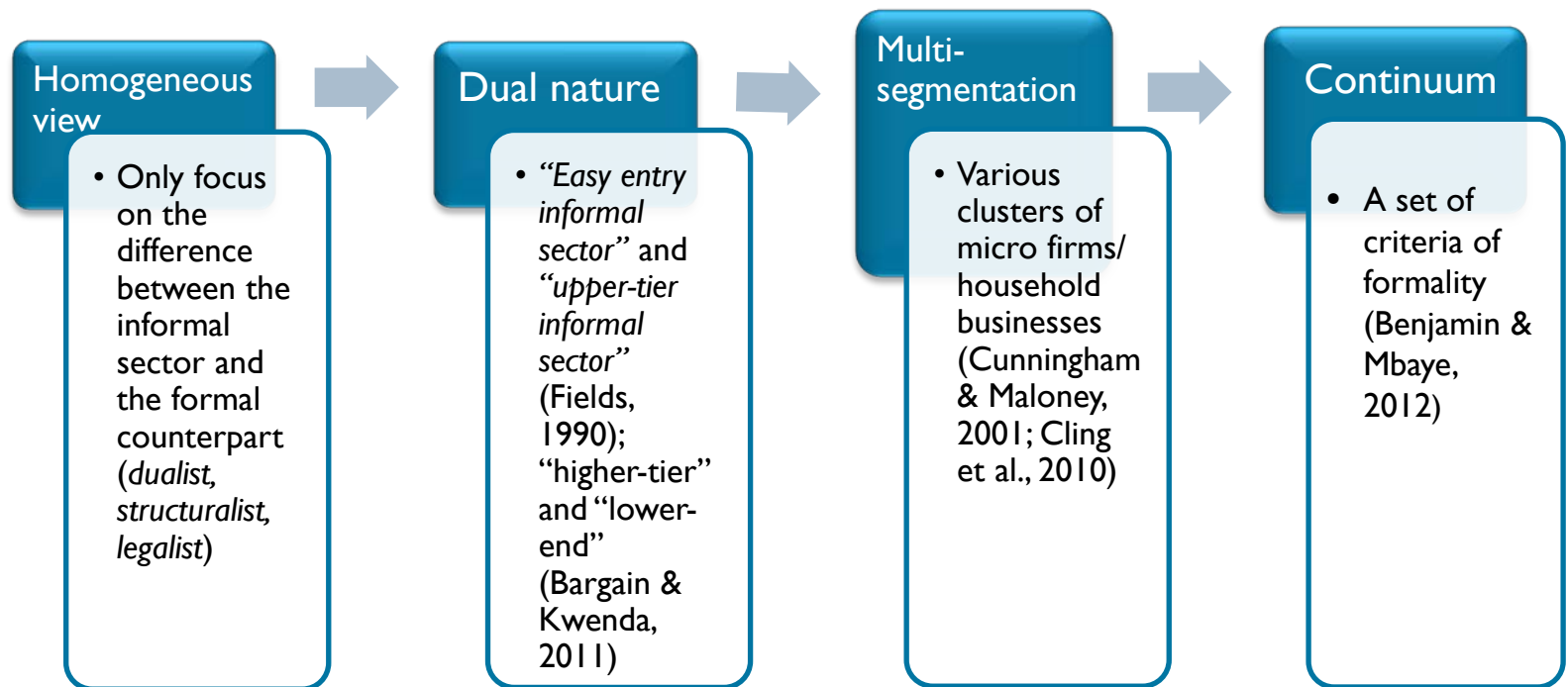
Outline

1) Literature review
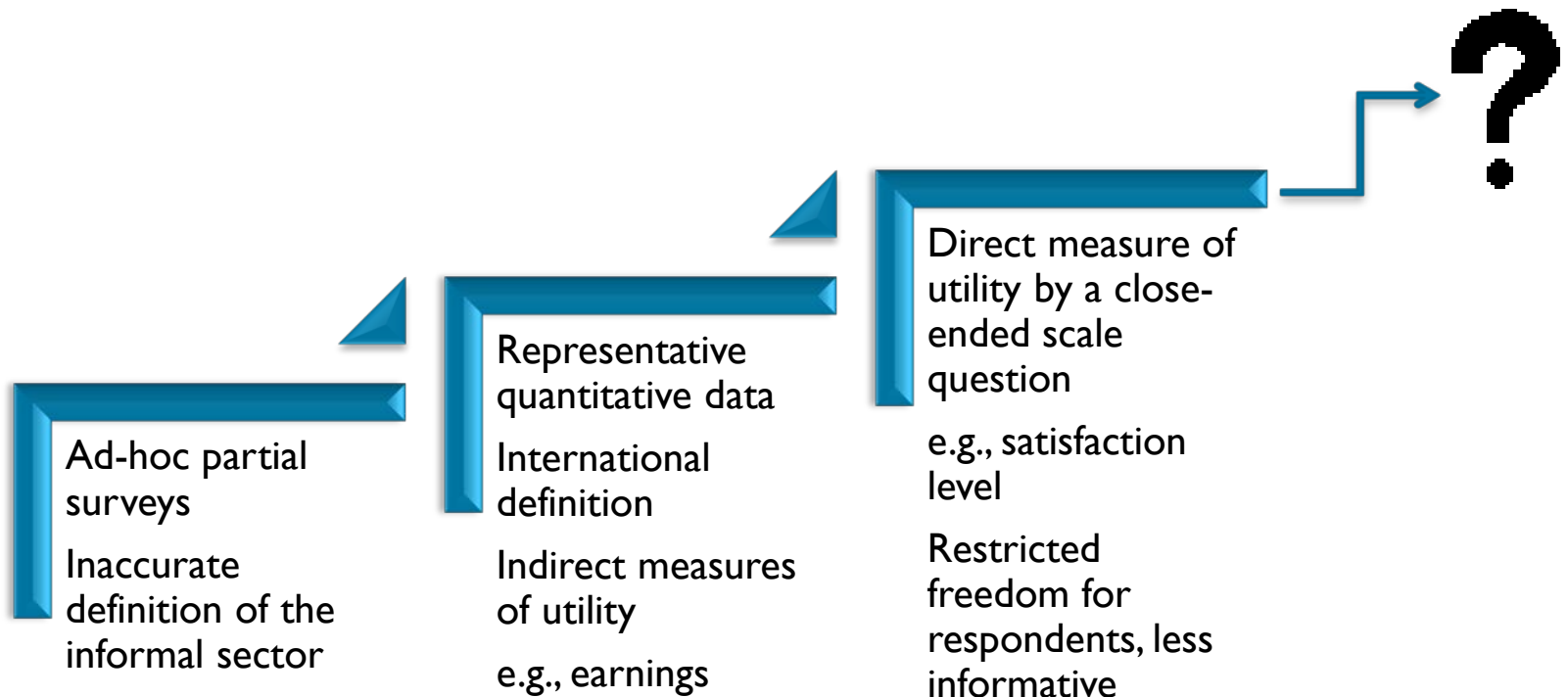2) Data
3) Methodology
4) Key findings
5) Policy implication

1. Literature review

- Heterogeneity of the informal sector is widely acknowledged, but at different levels.

Homogeneous view
- Only focus on the difference between the informal sector and the formal counterpart (*dualist, structuralist, legalist*)

Dual nature
- *"Easy entry informal sector"* and *"upper-tier informal sector"* (Fields, 1990); "higher-tier" and "lower-end" (Bargain & Kwenda, 2011)

Multi-segmentation
- Various clusters of micro firms/ household businesses (Cunningham & Maloney, 2001; Cling et al., 2010)

Continuum
- A set of criteria of formality (Benjamin & Mbaye, 2012)

Filling research gap

- What existing literature has done:

Ad-hoc partial surveys

Inaccurate definition of the informal sector

Representative quantitative data

International definition

Indirect measures of utility

e.g., earnings

Direct measure of utility by a close-ended scale question

e.g., satisfaction level

Restricted freedom for respondents, less informative

?

Filling research gap (cont.)

- What this research contribute:

  ◦ Representative **<u>textual</u>** data = direct measure of utility of HBs in a non-restricted manner.

  ◦ **<u>Exploratory</u>** analytical methods (i.e., correspondence and cluster analysis) limit the pre-conceived ideas of researchers.

  ◦ The very first economic paper that solves the **<u>linguistic</u>** difficulties associated with Vietnamese texts.

2. Data

- Household Businesses and Informal Sector Survey, Ho Chi Minh City 2009: **representative** quantitative and qualitative data.

- Informal sector = all private unincorporated enterprises that:
  - produce at least some of their goods and services for sale or barter
  - are not registered (i.e., having no business license)
  - are engaged in non-agricultural activities

  NB: *"unincorporated enterprises"* = household business (HB)
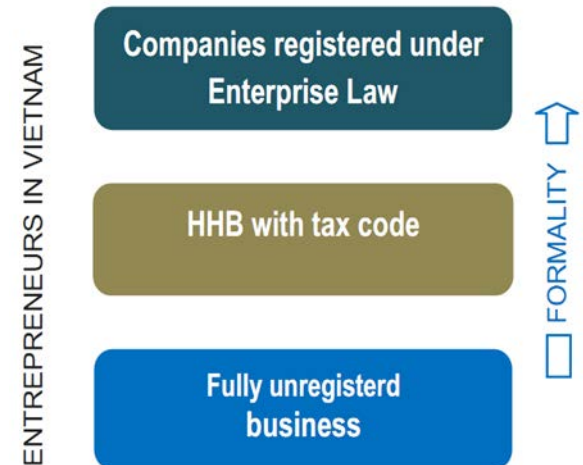
Source: Economica Vietnam (2013)

- Sample size = 1023 HBs

8

2. Data (cont.)

- **<u>Qualitative data</u>:** three open-ended questions were added into an informal sector survey for the very first time.

    - *"Please state any advantages while running a business like yours."*
    - *"Please state any disadvantages while running a business like yours."*
    - *"State any subjects related to your business that have not been mentioned or clarified before."*

3. Methodology

- Textual data = multi-dimensional and structure is unknown.

    → Exploratory analyses

Correspondence Analysis

- A multivariate statistical technique to extract the most important and uncorrelated dimensions from large data.

Cluster Analysis

- Division of data into groups of similar objects.

How to apply on texts ?

- Lexical *"contingency table"*:
  - ◦ columns = distinct **words** used
  - ◦ rows = HBs (individual observations)

- Vietnamese words – a challenge:
  - ◦ Special letters: ê, ă, â, ô, ơ, ư
  - ◦ Tone marks: *"level," "hanging," "sharp," "asking," "tumbling,"* and *"heavy."*
  - ◦ **Word boundary** : words are often formed by two or more separated syllables, which can only be recognized by context.

New coding method

- Letters:

    ă = aw; â = aa; đ = dd; ê = ee; ô = oo; ơ = ow; ư = w

- Tone marks

` (*"hanging"*)		= f
´ (*"sharp"*)		= s
̉ (*"asking"*)		= r
~ (*"tumbling"*)		= x
. (*"heavy"*)		= j

- Word boundary

"cạnh tranh"	= canhjtranh (competition)
"nguyên vật liệu"	= nguyeenvaatjlieeuj (materials)

4. Key findings

1) The informal sector is heterogeneous and can be considered as a **<u>continuum</u>** of HBs.

2) Generally low satisfaction of working in the informal sector.

4.1. Descriptive statistics

Table 2: Summary of Step Numerical Coding of Texts		
	Disadvantages	Advantages
Total number of responses	505	433
Total number of words	5188	3498
Number of distinct words	867	585
Source: HB&IS HCMC (2009); authors' calculations.		

→ Low satisfaction of working in the informal sector.

4.1. Descriptive statistics (cont.)

Table 4: Most Frequent Words Used by Informal HBs			
Disadvantages	Frequency	Advantages	Frequency
Customers	102	Home	78
Prices	79	Customers	67
Capital	56	Stable	57
Location	35	Premise	54
Competition	32	Location	50
Source: HB&IS HCMC (2009); authors' calculations.			

4.1. Descriptive statistics (cont.)

Table 5: The Main Difficulties Faced by Informal HBs		
The most difficulty faced	Frequency	Percentage
Competition	119	30.75
Lack of customers	57	14.73
Access to loan	55	14.21
Premise, space	49	12.66
Cash flow	25	6.46
Machine, equipment	21	5.43
Material supply	18	4.65
Crime, theft, disorder	13	3.36
Inflation, exchange rate	11	2.84
Transportation	9	2.33
Other	7	1.81
Access to land	3	0.78
Source: HB&IS HCMC (2009); authors' calculations.		

→ Question design can greatly affect data outcome.

4.2. Correspondence Analysis

Table 6: The First Five Factors of CA				
	Advantages		Disadvantages	
Factor	Eigenvalue	Percentage	Eigenvalue	Percentage
1	0.5848	**1.96**	0.5114	**1.39**
2	0.5597	**1.88**	0.4925	**1.34**
3	0.5311	1.78	0.4878	1.32
4	0.5093	1.71	0.4578	1.24
5	0.494	1.66	0.4537	1.23
Source: HB&IS HCMC (2009); authors' calculations.				

Disadvantages

Table 8: Selective Points with Extreme Contributions on the First Two Principal Axes CA on "Disadvantages" of Informal HBs					
Axis 1			**Axis 2**		
Words	**Translation**	**Coordinate**	**Words**	**Translation**	**Coordinate**
Vốn	capital	-7.32	Sớm	early	-16.33
Tài chính	finance	-7.01	Dậy	get up	-15.89
Vay	borrow	-4.97	Thức	awake	-13.14
Hỗ trợ	support	-4.88	Sang	morning	-10.72
Tay nghề	workmanship	-4.48	Vất vả	strenuous	-6.23
Mở rộng	enlarge	-4.43	Vốn	capital	-5.69
Diện tích	area, surface	-4.26	Tài chính	finance	-3.76
Ổn định	stable	-3.04	Vay	borrow	-3.06
Quy mô	scale	-3.02	Gia đình	family	-3.04
Máy móc	machines	-3.00	Sản xuất	production	-2.94
Kinh tế	economy	-2.85	Mở rộng	enlarge	-2.76
Mặt bằng	premise	-2.84	Diện tích	area, surface	-2.53
Giải tỏa	land clearance	-2.66	Già	old	-2.4
Vật giá	price of materials	-2.65	Chồng	husband	-2.04
Địa điểm	location	-2.52	Con	children	-1.82
Mất	lose, lost	2.45	Thị trường	market	2.07
Trung Quốc	China	2.51	Chất lượng	quality	2.20
Bụi	dust, dusty	2.81	Bụi	dust, dusty	2.25
Xăng	petrol	3.38	Siêu thị	supermarket	2.28
Xe ba bánh	tricycles	3.41	Nắng	sunny	2.57
Kẹt	(traffic) jams	3.74	Nước	water	2.57
Chất lượng	quality	3.85	Mưa	rain	2.68
Cấm	ban	4.36	Khách hàng	customers	2.79
Vất vả	strenuous	4.37	Dọn dẹp	tidy up	2.83
Khách hàng	customers	4.54	Buýt	bus	2.92
Đường	road, street	4.94	Kẹt	(traffic) jam	3.53
Sang	morning	8.31	Xăng	petrol	4.09
Thức	awake	9.40	Xe	vehicle	4.65
Dậy	get up	10.87	Đường	street, road	5.36
Sớm	early	11.46	Giá	prices	6.64
Source: HB&IS HCMC (2009); authors' calculations.					

entrepreneurship (left of Axis 1)

personal (left, lower)

internal (right upper)

external (right lower)

Figure 3: The Space of Informal HB Characteristics (Projection of Illustrative Variables in the First Factorial Plane of CA on Disadvantages)

external

personal

entrepreneurship

internal

turnov_q1
a4d_indep   age<40
G1_prof_know service size=1
acc_no   cust_indiv
wearner_no
nofutur_change
premi_unstab
cust-enterp
wearner_yes
premi_home
wilg_reg_yes futur_yes
trade
G1_stabler_profit size3-6 ynot_reg_others equip_q4 size=2
cust_hb
a4d_better_inc
supplier_indiv turnov_q5
G1_fam_trad
a4d_fam_trad
accounting_simple

0.2
0
-0.3
-0.8
-1.3
-1.8
-0.5   -0.4   -0.3   -0.2   -0.1   0   0.1   0.2
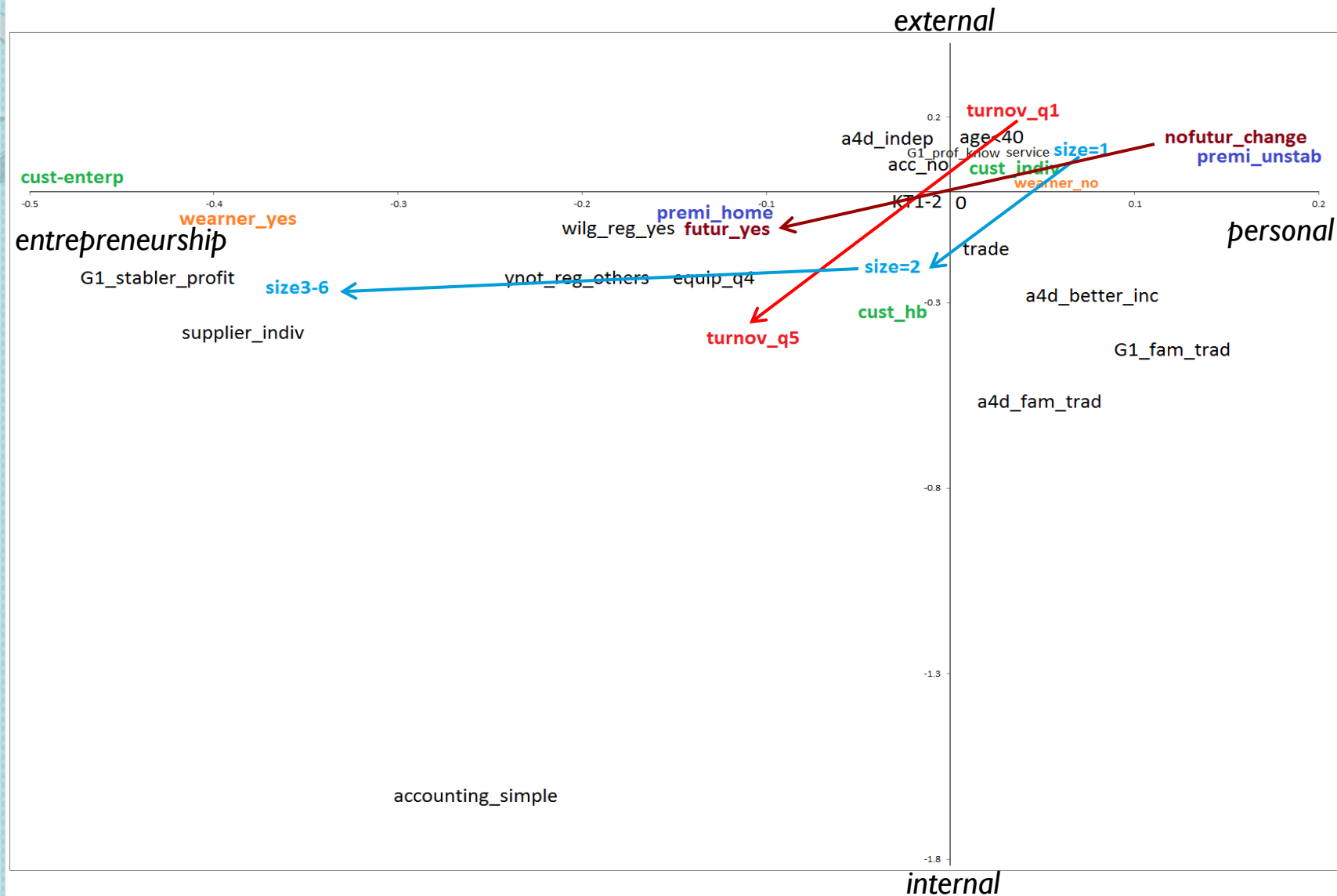
Source: HB&IS 2009, author's calculation

4.2. Correspondence Analysis
A broader view on both informal and formal HBs

personally

individual

Business and economic (supply side)

globally

Economic and market-related (demand side)

Non-economic, non-market

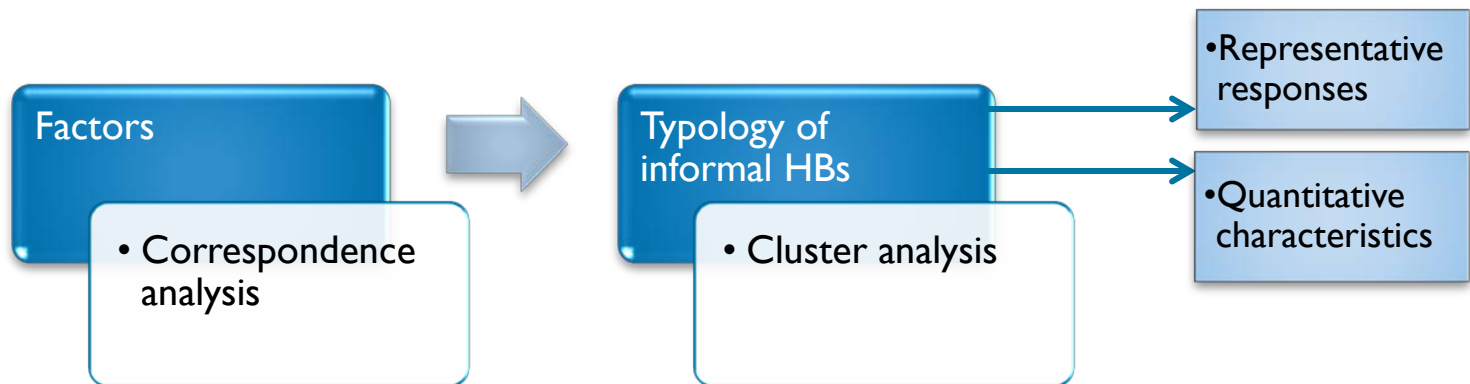| Table 9: Selective Points with Extreme Contributions on the First Two Principal Axes CA on "Disadvantages" of HBs					
Axis 1			**Axis 2**		
Words	**Translation**	**Coordinate**	**Words**	**Translation**	**Coordinate**
Sớm	early	-13.10	Giá	prices	-10.47
Dậy	get up	-12.57	Cạnh tranh	competition	-7.22
Thức	awake	-10.44	Gia công	processing	-6.5
Sang	morning	-9.74	Đối thủ	competitors	-4.06
Khách	customers	-5.99	Thị trường	markets	-3.31
Xe	vehicle	-5.69	Lợi nhuận	profit	-3.15
Đường	street, road	-4.97	Điện	electricity	-3.05
Vất vả	strenuous	-4.51	Chi phí	costs	-2.85
Cấm	forbid	-4.06	Tiêu thụ	consume	-2.69
Quản lý	manage	-3.67	Dân cư	inhabitants	-2.67
Bụi	dust	-3.42	Thuế	tax	-2.60
Xe ba bánh	tricycles	-3.23	Tay nghề	workmanship	-2.51
Xăng	petrol	-3.16	Sản phẩm	products	-2.45
Kẹt	(traffic) jam	-3.05	Doanh thu	revenue	-2.39
Lề	edge	-2.52	Chất lượng	quality	-2.29
Chật hẹp	narrow	3.14	Bụi	dust	2.96
Quy mô	scale	3.65	Xe ba bánh	tricycles	3.61
Máy móc	machines	4.09	Nước	water	4.02
Lãi suất	interest rate	4.10	Cấm	forbid	4.37
Địa điểm	location	4.32	Nắng	sunny	4.6
Sản xuất	production	4.44	Dọn dẹp	tidy up	5.26
Ngành nghề	industry	4.50	Vất vả	strenuous	5.38
Diện tích	surface area	4.7	Sáng	morning	5.67
Mặt bằng	premise	5.42	Ngập	flooded	6.42
Di dời	move	5.56	Đường	street, road	6.50
Vay	borrow	6.99	Mùa	season	7.37
Mở rộng	enlarge	7.31	Dậy	get up	8.98
Tài chính	finance	8.40	Sớm	early	9.42
Kinh doanh	business	9.03	Trời	weather	13.35
Vốn	capital	10.81	Mưa	rain	15.21
Source: HB&IS HCMC (2009); authors' calculations.					

Figure 5: The Space of HB Characteristics (Projection of Illustrative Variables in the First Factorial Plane of CA on Disadvantages)
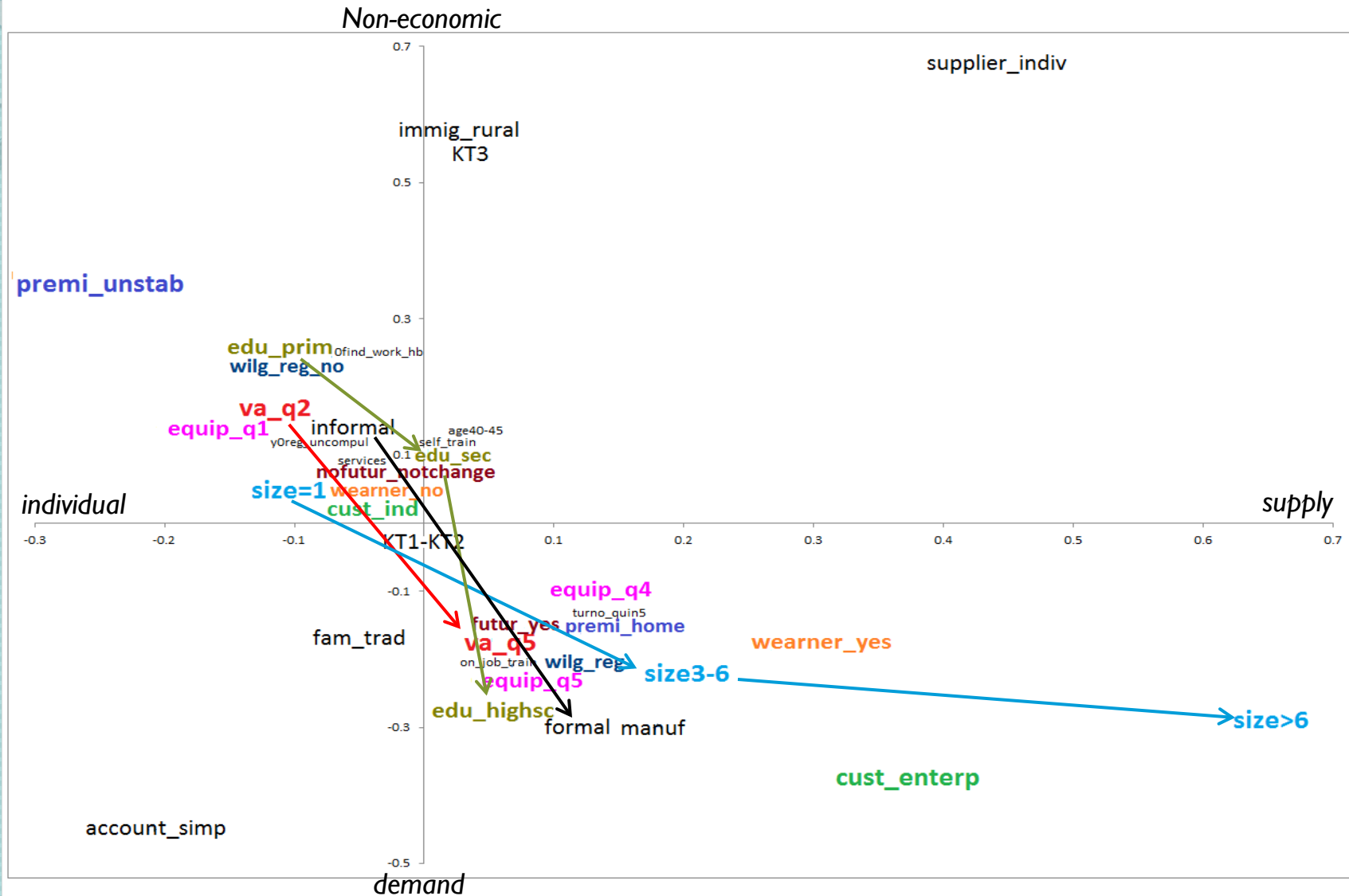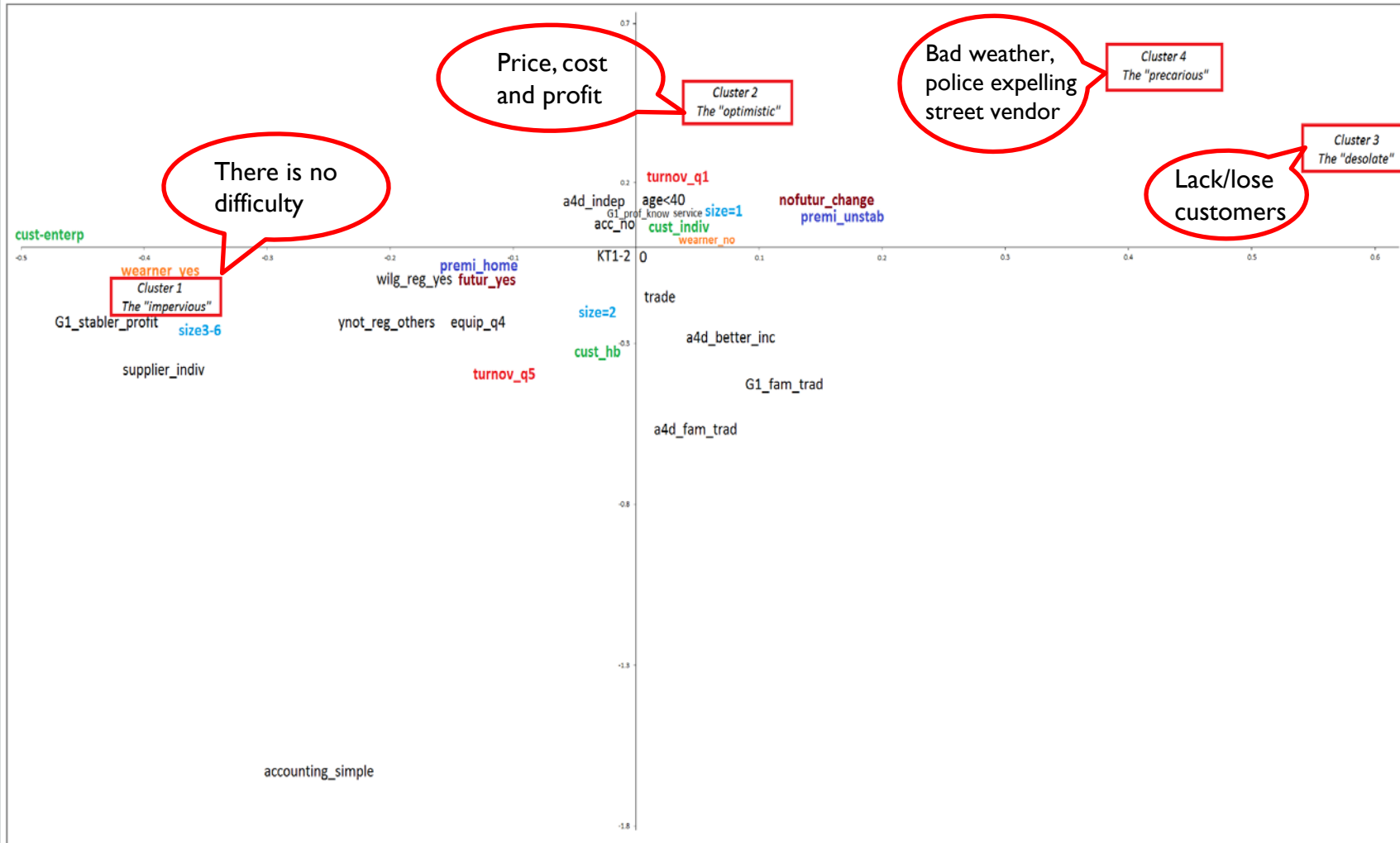
Source: HB&IS 2009, author's calculation

4.3. Cluster analysis

Factors
- Correspondence analysis

Typology of informal HBs
- Cluster analysis

- Representative responses

- Quantitative characteristics

Figure 6: Cluster View of Informal HBs on the First Factorial Plane of CA on Disadvantages

Source: HB&IS 2009, author's calculation

Policy implication

- A *"one size fits all"* scheme would not be appropriate for this highly heterogeneous informal sector.

- A policy package could include:
  - ◦ Anti-poverty policies for lower-end HBs and professional policies on business environment for higher-end HBs.
  - ◦ Premise stabilization for *"precarious HBs,"* vocational training for *"desolate HBs,"* inflation control for *"optimistic HBs."*

5. Conclusion (cont.)

- Value-added of this research work:
  - <u>New finding</u>: Continuum of HBs in Vietnam.
  - <u>New approach</u>:
    - Textual data
    - Exploratory analysis
    - Vietnamese language treatment

- Avenue for future research:
  - Sensitivity tests
  - Dynamics of the informal sector
  - Employ HBs' coordinates on the factors as continuous variables in econometric models
  - Development of statistical packages that systematically handle Vietnamese texts.

Thank you
for your attention !

Q & A