

# A multilevel analysis of innovation in developing countries

Martin Srholec  
CAS Centre and TIK Centre

CAS workshop on  
on “Innovation in firms”  
Oslo, 30.10-1.11.2007

# A multilevel problem

- Multilevel modeling concerns hypothesis/problem/data with hierarchical structure
- A hierarchy refers to units clustered (grouped/related/similar, etc.) at different levels
- Acknowledged in many fields of social science or biology for a long time
- School performance is not only given by the amount of study time of a child, but also by higher-level factors such as characteristics of the class, school or national educational system
- Similarly the innovation process in firms is influenced by factors operating at different levels

# Innovation is multilevel

- Schumpeter focused on constraining factors in the environment: the forces of habit, resistance to innovation
- Survival of firms propelled by innovation, but determined by the environment in the evolutionary economics
- National, regional and sectoral innovation systems: organizations, relations and institutions that enable innovation in firms
- A need to reach beyond the dichotomy between methodological individualism vs. collectivism in empirical research on innovation
- A need for multilevel analysis of interactions between various levels of the analysis in the innovation process

# Multilevel modeling

- A multilevel model is a statistical model that relates a dependent variable to explanatory variables at more than one level
- Sociology, educational research, health science, demography, epidemiology or biology
- Not used in research on innovation (and rarely in economics)

Suppose a multilevel model has 2-level structure with firms at level-1 located in countries at level-2. A standard linear 2-level model with one explanatory variable at each level is the following:

Level-1 linear model:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$

Level-2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_j + u_{1j}$$

$$y_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10} x_{ij} + \gamma_{11}z_jx_{ij} + (u_{0j} + u_{1j}x_{ij} + e_{ij})$$

where  $y_{ij}$  is the dependent variable,  $x_{ij}$  is the level-1 predictor,  $z_j$  is the level-2 predictor,  $e_{ij}$ ,  $u_{0j}$  and  $u_{1j}$  are random effects (normally distributed residual terms for each equation),  $i$  is the firm ( $i = 1 \dots n$ ) and  $j$  is the country ( $j = 1 \dots m$ ).

# Why multilevel modeling?

- Single-level models assume that observations (and hence residuals) are independent from each other
- Single-level models may suffer from the fallacy of the wrong level
- Analyses that use exclusively micro data to account for the contextual effects often suffer from endogeneity
- A proper recognition of data hierarchies allows us to examine new lines of questions
- And at last but not least the concept of innovation systems implicitly predicts a nested structure of micro data

# Why not just dummies?

- Using dummies might be a useful quick-fix solution, if we are interested exclusively in the level-1 relationships
- A dummy is a “catch-all” variable stripped of the context for which we can only speculate what it really represents
- If there are many higher-level units, “dummy” models will have many more parameters, decreasing the degrees of freedom
- If the higher-level dummies significantly improve predictive power of the model, which indeed is often the case in the literature; a multilevel model should be given priority

# Dataset

- Micro data from the Productivity and Investment Climate Survey (PICS) organized/harmonized by the World Bank
- A sample of 15,000 manufacturing firms with more than 10 employees in 30 developing countries
- Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Egypt, El Salvador, Guatemala, Honduras, Hungary, India, Kazakhstan, Lebanon, Madagascar, Mexico, Nicaragua, Panama, Paraguay, Peru, Philippines, Poland, Romania, South Africa, Tanzania, Thailand, Turkey, Uruguay, Vietnam, Zambia
- Number of observations: 1,548 in Brazil, 1,512 in India, 1,179 in Thailand, 1,056 in Turkey, 946 in Egypt ... 157 in Tanzania and 132 in Lebanon

# Firm-level variables

1. INNPDT = “Have you developed a major new product line?”
2. INNPCS = “Have you introduced new technology that has substantially changed the way that the main product is produced?”
3. INNOV = INNPCT  $\wedge$  INNPCS
4. SIZE = Number of permanent employees (logs)
5. AGE = Number of years since the firm has started (logs)
6. FA = More than 50% of foreign ownership
7. SKILL = “Do you offer formal (beyond “on the job”) training to your permanent employees?”
8. ISO = “Have you received ISO 9000, 9002 or 14,000?”
9. WWW = “Do you regularly use a website in interactions with clients and suppliers?”

# Country-level variables

1. GAP = technology gap = PCT patents per capita in the United States minus the particular country
2. R&D = Number of researchers in R&D per capita
3. EDU = Literacy rate in adult population
4. FIN = Domestic credit to private sector in % of GDP
5. GINI = inequality of income = GINI coefficient
6. GOV = Factor analysis on the “Governance Matters” indicators

# A logit multilevel model

Since INNOV is binary, we need to specify a non-linear multilevel model. For this purpose, we distinguish between the sampling model (3.1), a link function (3.2) and a structural part of the multilevel model (3.3). Only the level-1 model differs from the linear case and can be written down as follows:

$$(3.1) \quad E(y_{ij} = 1 \mid \beta_j) = \varphi_{ij}$$

$$(3.2) \quad \text{Log} [\varphi_{ij} / (1 - \varphi_{ij})] = \eta_{ij}$$

$$(3.3) \quad \eta_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + \dots + \beta_{nj} x_{nij}$$

where  $\eta_{ij}$  is the log of the odds of success (to introduce an innovation, for instance). Although  $\varphi_{ij}$  is constrained to be in the interval (0,1), the logit transformation allows  $\eta_{ij}$  to take any value and therefore can be substituted to the structural model. Note that the predicted log-odds can be converted to an odds by  $\exp(\eta_{ij})$  and to the predicted probability  $\varphi_{ij}$  by  $\exp\{\eta_{ij}\}/(1+\exp\{\eta_{ij}\})$ .

# Basic model

Although there are no level-2 predictors, we allow the level-1 intercept and slopes to vary across countries by including the random effects:

Level-1 logit model:

$$E(\text{INNOV}_{ij} = 1 \mid \beta_j) = \varphi_{ij}$$

$$\text{Log} [\varphi_{ij} / (1 - \varphi_{ij})] = \beta_{0j} + \beta_{1j}\text{SIZE}_{ij} + \beta_{2j}\text{AGE}_{ij} + \beta_{3j}\text{FA}_{ij} + \beta_{4j}\text{SKILL}_{ij} + \beta_{5j}\text{ISO}_{ij} + \beta_{6j}\text{WWW}_{ij}$$

Level-2 model:

$$\beta_{0j} = \gamma_{00} + \mathbf{u}_{0j}$$

$$\beta_{1j} = \gamma_{10} + \mathbf{u}_{1j}$$

$$\beta_{2j} = \gamma_{20} + \mathbf{u}_{2j}$$

$$\beta_{3j} = \gamma_{30} + \mathbf{u}_{3j}$$

$$\beta_{4j} = \gamma_{40} + \mathbf{u}_{4j}$$

$$\beta_{5j} = \gamma_{50} + \mathbf{u}_{5j}$$

$$\beta_{6j} = \gamma_{60} + \mathbf{u}_{6j}$$

where there are seven fixed effects ( $\gamma_{00} \dots \gamma_{60}$ ) and seven random effects ( $u_{0j} \dots u_{6j}$ )

	Dependent variable:		
	INNOV	INNPDT	INNPCS
<u>Level-1 fixed effects:</u>			
Intercept <sub>ij</sub> ( $\gamma_{00}$ )	0.244 (0.143)*	-0.392 (0.128)***	-0.540 (0.142)***
SIZE <sub>ij</sub> slope ( $\gamma_{10}$ )	0.190 (0.041)***	0.140 (0.035)***	0.205 (0.032)***
AGE <sub>ij</sub> slope ( $\gamma_{20}$ )	-0.027 (0.032)	-0.021 (0.034)	-0.054 (0.022)**
FA <sub>ij</sub> slope ( $\gamma_{30}$ )	-0.123 (0.077)	-0.120 (0.084)	-0.178 (0.083)**
SKILL <sub>ij</sub> slope ( $\gamma_{40}$ )	0.652 (0.070)***	0.506 (0.075)***	0.589 (0.071)***
ISO <sub>ij</sub> slope ( $\gamma_{50}$ )	0.300 (0.063)***	0.314 (0.081)***	0.230 (0.048)***
WWW <sub>ij</sub> slope ( $\gamma_{60}$ )	0.497 (0.051)***	0.472 (0.060)***	0.410 (0.047)***
<u>Level-2 random effects:</u>			
Intercept <sub>ij</sub> ( $u_{0j}$ )	0.591 (574.0)***	0.465 (411.1)***	0.577 (530.0)***
SIZE <sub>ij</sub> slope ( $u_{1j}$ )	0.030 (78.1)***	0.021 (78.3)***	0.014 (53.7)***
AGE <sub>ij</sub> slope ( $u_{2j}$ )	0.016 (62.6)***	0.022 (86.6)***	0.003 (35.6)
FA <sub>ij</sub> slope ( $u_{3j}$ )	0.038 (29.6)	0.092 (40.6)*	0.091 (49.9)***
SKILL <sub>ij</sub> slope ( $u_{4j}$ )	0.083 (68.2)***	0.115 (96.4)***	0.096 (83.8)***
ISO <sub>ij</sub> slope ( $u_{5j}$ )	0.050 (41.3)*	0.120 (75.2)***	0.008 (38.1)
WWW <sub>ij</sub> slope ( $u_{6j}$ )	0.025 (38.8)	0.051 (49.5)***	0.029 (38.6)
Index of dispersion	0.9824	0.9863	0.9819
Level-1 firms	15,113	15,155	15,127
Level-2 countries	30	30	30

# Intercept-as-outcome model

incorporates the country-level variables as predictors of the level-1 intercept:

Level-1 logit model:

$$E(\text{INNOV}_{ij} = 1 \mid \beta_j) = \varphi_{ij}$$

$$\text{Log} [\varphi_{ij} / (1 - \varphi_{ij})] = \beta_{0j} + \beta_{1j}\text{SIZE}_{ij} + \beta_{2j}\text{AGE}_{ij} + \beta_{3j}\text{FA}_{ij} + \beta_{4j}\text{SKILL}_{ij} + \beta_{5j}\text{ISO}_{ij} + \beta_{6j}\text{WWW}_{ij}$$

Level-2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{GAP}_j + \gamma_{02}\text{R\&D}_j + \gamma_{03}\text{EDU}_j + \gamma_{04}\text{FIN}_j + \gamma_{05}\text{GINI}_j + \gamma_{06}\text{GOV}_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

$$\beta_{4j} = \gamma_{40} + u_{4j}$$

$$\beta_{5j} = \gamma_{50} + u_{5j}$$

$$\beta_{6j} = \gamma_{60} + u_{6j}$$

where we assume that the level-2 predictors explain the different average propensity to innovate across countries, but let the level-1 effects of SIZE, AGE, FA, SKILL, ISO and WWW remain “unconditional” at the country level.

	Dependent variable:		
	INNOV	INNPDT	INNPCS
<u>Level-1 fixed effects:</u>			
Intercept <sub>ij</sub> ( $\gamma_{00}$ )	0.252 (0.107)**	-0.384 (0.101)***	-0.532 (0.096)***
SIZE <sub>ij</sub> slope ( $\gamma_{10}$ )	0.200 (0.042)***	0.146 (0.037)***	0.208 (0.033)***
AGE <sub>ij</sub> slope ( $\gamma_{20}$ )	-0.024 (0.032)	-0.023 (0.035)	-0.058 (0.022)**
FA <sub>ij</sub> slope ( $\gamma_{30}$ )	-0.101 (0.078)	-0.104 (0.088)	-0.181 (0.081)**
SKILL <sub>ij</sub> slope ( $\gamma_{40}$ )	0.636 (0.069)***	0.499 (0.075)***	0.581 (0.070)***
ISO <sub>ij</sub> slope ( $\gamma_{50}$ )	0.359 (0.062)***	0.325 (0.080)***	0.245 (0.048)***
WWW <sub>ij</sub> slope ( $\gamma_{60}$ )	0.490 (0.050)***	0.466 (0.059)***	0.414 (0.048)***
<u>Level-2 fixed effects:</u>			
GAP <sub>i</sub> ( $\gamma_{01}$ )	0.381 (0.109)***	0.360 (0.139)**	0.315 (0.101)***
R&D <sub>i</sub> ( $\gamma_{02}$ )	0.398 (0.102)***	0.335 (0.141)**	0.273 (0.136)*
EDU <sub>i</sub> ( $\gamma_{03}$ )	-0.056 (0.078)	-0.021 (0.102)	0.275 (0.095)***
FIN <sub>i</sub> ( $\gamma_{04}$ )	0.199 (0.073)**	0.208 (0.104)*	0.181 (0.072)**
GINI <sub>i</sub> ( $\gamma_{05}$ )	0.537 (0.064)***	0.544 (0.089)***	0.410 (0.074)***
GOVERN <sub>i</sub> ( $\gamma_{06}$ )	-0.255 (0.134)*	-0.154 (0.172)	-0.280 (0.144)*
<u>Level-2 random effects:</u>			
Intercept <sub>ij</sub> ( $u_{0j}$ )	0.346 (330.28)***	0.312 (320.39)***	0.280 (208.07)***
SIZE <sub>ij</sub> slope ( $u_{1j}$ )	0.033 (76.57)***	0.024 (77.41)***	0.015 (53.81)***
AGE <sub>ij</sub> slope ( $u_{2j}$ )	0.016 (62.34)***	0.024 (87.00)***	0.004 (35.47)
FA <sub>ij</sub> slope ( $u_{3j}$ )	0.028 (30.16)	0.081 (40.28)*	0.081 (50.10)***
SKILL <sub>ij</sub> slope ( $u_{4j}$ )	0.077 (67.71)***	0.111 (96.12)***	0.093 (83.36)***
ISO <sub>ij</sub> slope ( $u_{5j}$ )	0.032 (41.78)*	0.102 (75.28)***	0.006 (38.21)
WWW <sub>ij</sub> slope ( $u_{6j}$ )	0.020 (38.63)	0.047 (49.43)***	0.033 (38.76)
Index of dispersion	0.9865	0.9873	0.9836
Level-1 firms	15,113	15,155	15,127
Level-2 countries	30	30	30

# Slopes-as-outcomes model

in which we allow the country-level variables to influence also slope of the SIZE firm-level predictor:

Level-1 logit model:

$$E(\text{INNOV}_{ij} = 1 \mid \beta_j) = \varphi_{ij}$$

$$\text{Log} [\varphi_{ij} / (1 - \varphi_{ij})] = \beta_{0j} + \beta_{1j}\text{SIZE}_{ij} + \beta_{2j}\text{AGE}_{ij} + \beta_{3j}\text{FA}_{ij} + \beta_{4j}\text{SKILL}_{ij} + \beta_{5j}\text{ISO}_{ij} + \beta_{6j}\text{WWW}_{ij}$$

Level-2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{GAP}_j + \gamma_{02}\text{R\&D}_j + \gamma_{03}\text{EDU}_j + \gamma_{04}\text{FIN}_j + \gamma_{05}\text{GINI}_j + \gamma_{06}\text{GOV}_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{GAP}_j + \gamma_{12}\text{R\&D}_j + \gamma_{13}\text{EDU}_j + \gamma_{14}\text{FIN}_j + \gamma_{15}\text{GINI}_j + \gamma_{16}\text{GOV}_j + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

$$\beta_{4j} = \gamma_{40} + u_{4j}$$

$$\beta_{5j} = \gamma_{50} + u_{5j}$$

$$\beta_{6j} = \gamma_{60} + u_{6j}$$

where the new fixed parameters ( $\gamma_{11} \dots \gamma_{16}$ ) indicate the cross-level interactions

	Dependent variable:		
	INNOV	INNPDT	INNPCS
<u>Level-1 fixed effects:</u>			
Intercept <sub>ii</sub> ( $\gamma_{00}$ )	0.246 (0.109)**	-0.392 (0.101)***	-0.541 (0.098)***
SIZE <sub>ii</sub> slope ( $\gamma_{10}$ )	0.183 (0.038)***	0.129 (0.032)***	0.196 (0.025)***
AGE <sub>ii</sub> slope ( $\gamma_{20}$ )	-0.026 (0.032)	-0.020 (0.034)	-0.057 (0.021)**
FA <sub>ii</sub> slope ( $\gamma_{30}$ )	-0.081 (0.077)	-0.097 (0.086)	-0.160 (0.086)*
SKILL <sub>ii</sub> slope ( $\gamma_{40}$ )	0.637 (0.068)***	0.503 (0.075)***	0.586 (0.069)***
ISO <sub>ii</sub> slope ( $\gamma_{50}$ )	0.318 (0.058)***	0.293 (0.081)***	0.237 (0.048)***
WWW <sub>ii</sub> slope ( $\gamma_{60}$ )	0.502 (0.049)***	0.475 (0.057)***	0.417 (0.048)***
<u>Level-2 fixed effects:</u>			
For Intercept <sub>ii</sub> ( $\beta_{0i}$ )			
GAP <sub>i</sub> ( $\gamma_{01}$ )	0.359 (0.107)***	0.375 (0.140)**	0.291 (0.118)**
R&D <sub>i</sub> ( $\gamma_{02}$ )	0.425 (0.101)***	0.313 (0.141)**	0.322 (0.135)**
EDU <sub>i</sub> ( $\gamma_{03}$ )	-0.071 (0.076)	-0.007 (0.105)	0.178 (0.091)*
FIN <sub>i</sub> ( $\gamma_{04}$ )	0.159 (0.070)**	0.194 (0.102)*	0.026 (0.093)
GINI <sub>i</sub> ( $\gamma_{05}$ )	0.608 (0.064)***	0.472 (0.091)***	0.538 (0.072)***
GOVERN <sub>i</sub> ( $\gamma_{06}$ )	-0.274 (0.149)*	-0.086 (0.166)	-0.278 (0.170)
For SIZE <sub>ii</sub> slope ( $\beta_{1i}$ )			
GAP <sub>i</sub> ( $\gamma_{11}$ )	0.065 (0.034)*	0.024 (0.026)	0.030 (0.041)
R&D <sub>i</sub> ( $\gamma_{12}$ )	0.015 (0.052)	-0.043 (0.039)	-0.023 (0.045)
EDU <sub>i</sub> ( $\gamma_{13}$ )	-0.001 (0.019)	0.043 (0.017)**	0.051 (0.024)**
FIN <sub>i</sub> ( $\gamma_{14}$ )	0.111 (0.020)***	0.012 (0.019)	0.103 (0.031)***
GINI <sub>i</sub> ( $\gamma_{15}$ )	-0.129 (0.024)***	-0.115 (0.017)***	-0.090 (0.019)***
GOVERN <sub>i</sub> ( $\gamma_{16}$ )	0.033 (0.037)	0.114 (0.019)***	0.014 (0.025)
<u>Level-2 random effects:</u>			
Intercept <sub>ii</sub> ( $u_{0i}$ )	0.359 (336.17)***	0.315 (310.75)***	0.295 (221.24)***
SIZE <sub>ii</sub> slope ( $u_{1i}$ )	0.022 (47.48)***	0.017 (56.84)***	0.008 (30.77)
AGE <sub>ii</sub> slope ( $u_{2i}$ )	0.016 (62.30)***	0.023 (86.73)***	0.003 (35.44)
FA <sub>ii</sub> slope ( $u_{3i}$ )	0.025 (30.19)	0.092 (40.03)*	0.099 (50.10)***
SKILL <sub>ii</sub> slope ( $u_{4i}$ )	0.076 (67.68)***	0.112 (96.07)***	0.090 (83.33)***
ISO <sub>ii</sub> slope ( $u_{5i}$ )	0.035 (41.50)*	0.103 (74.87)***	0.007 (38.05)
WWW <sub>ii</sub> slope ( $u_{6i}$ )	0.021 (38.66)	0.043 (49.43)**	0.036 (38.83)
Index of dispersion	0.9866	0.9879	0.9879
Level-1 firms	15,113	15,155	15,127
Level-2 countries	30	30	30

# Conclusions

- Empirical research on innovation can benefit from multilevel modeling
- Quantitative econometric methods to test predictions of the contextual theories of innovation
- Also more than 2-level and cross-classified models
- High demands on data and computation power

# Overview of the sample

County	Year	Observations	INNOV	INNPDIT	INNPCS	SIZE	AGE	FA	SKILL	ISO	WWW
Argentina	2006	507	0.84	0.77	0.69	3.96	3.24	0.09	0.53	0.28	0.77
Bolivia	2006	261	0.88	0.77	0.75	3.55	2.94	0.10	0.64	0.19	0.44
Brazil	2003	1,548	0.85	0.68	0.67	4.00	2.70	0.05	0.66	0.19	0.73
Chile	2004	544	0.58	0.48	0.28	4.23	3.22	0.13	0.73	0.37	0.74
Colombia	2006	446	0.84	0.71	0.72	3.54	2.77	0.02	0.58	0.14	0.44
Ecuador	2003	376	0.76	0.55	0.54	3.77	2.90	0.10	0.74	0.19	0.60
Egypt	2004	946	0.20	0.14	0.11	3.48	2.82	0.02	0.11	0.09	0.18
El Salvador	2003	363	0.80	0.64	0.56	3.86	2.82	0.07	0.50	0.06	0.37
Guatemala	2003	333	0.74	0.55	0.48	3.93	2.77	0.10	0.59	0.04	0.34
Honduras	2003	311	0.67	0.49	0.48	3.99	2.44	0.18	0.61	0.07	0.30
Hungary	2005	246	0.45	0.38	0.20	4.01	2.68	0.17	0.48	0.31	0.91
India	2005	1,512	0.47	0.43	0.18	3.66	2.68	0.01	0.18	0.27	0.39
Kazakhstan	2005	252	0.53	0.35	0.39	3.86	2.27	0.05	0.35	0.13	0.60
Lebanon	2006	132	0.73	0.55	0.46	3.62	3.27	0.02	0.64	0.21	0.52
Madagascar	2005	259	0.54	0.38	0.35	3.77	2.56	0.31	0.46	0.07	0.33
Mexico	2006	734	0.46	0.40	0.40	3.85	2.86	0.07	0.40	0.29	0.37
Nicaragua	2003	269	0.74	0.52	0.55	3.41	2.77	0.12	0.45	0.04	0.22
Panama	2006	166	0.75	0.62	0.60	3.55	3.15	0.08	0.49	0.13	0.42
Paraguay	2006	299	0.79	0.72	0.68	3.44	3.02	0.07	0.51	0.09	0.37
Peru	2006	274	0.90	0.79	0.78	3.91	2.82	0.11	0.65	0.23	0.49
Philippines	2003	538	0.59	0.48	0.41	4.19	2.79	0.19	0.23	0.18	0.25
Poland	2005	259	0.65	0.49	0.52	3.74	2.87	0.08	0.71	0.28	0.88
Romania	2005	292	0.62	0.38	0.52	4.01	2.61	0.12	0.36	0.22	0.72
South Africa	2003	554	0.82	0.69	0.61	4.70	2.92	0.16	0.66	0.43	0.71
Tanzania	2003	157	0.51	0.34	0.36	3.60	2.59	0.20	0.52	0.15	0.31
Thailand	2004	1,179	0.77	0.59	0.60	5.08	2.66	0.15	0.81	0.45	0.37
Turkey	2005	1,056	0.59	0.36	0.44	4.39	2.77	0.03	0.59	0.48	0.73
Uruguay	2006	271	0.80	0.69	0.65	3.49	3.14	0.10	0.41	0.14	0.41
Vietnam	2005	862	0.62	0.43	0.45	4.78	2.20	0.10	0.59	0.39	0.30
Zambia	2002	167	0.72	0.50	0.51	4.24	2.63	0.23	0.32	0.05	0.25
Total	..	15,113	0.65	0.51	0.47	4.04	2.76	0.09	0.51	0.25	0.49