

Entrepreneurship – Aggregation Issues

Aart Kraay
The World Bank

SME Learning
Entrepreneurship and Development Workshop
April 24, 2006

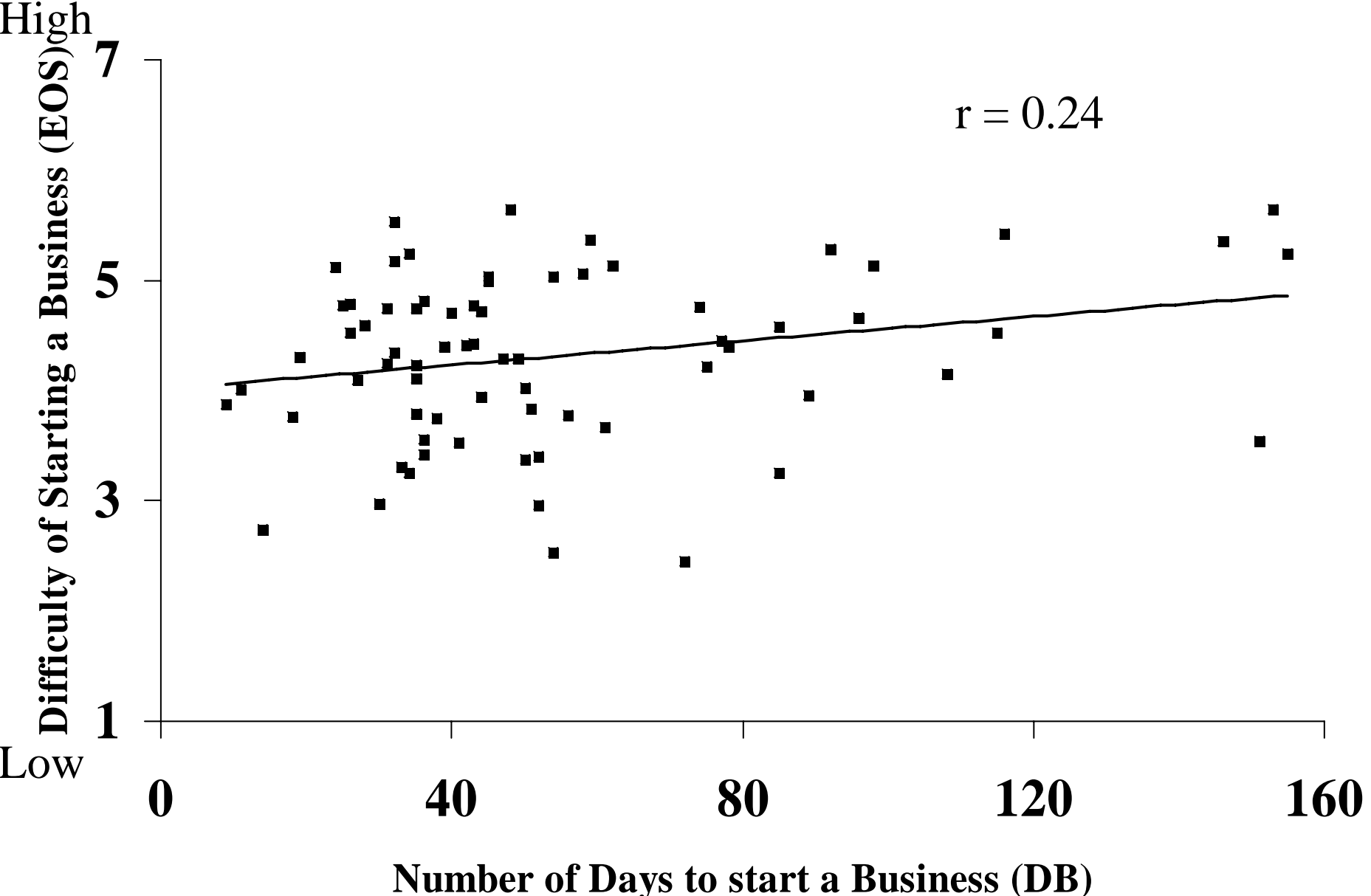
Outline

- Lots of interesting data on:
 - proxies for entrepreneurship
 - hard to observe directly, like many other things...
 - determinants of entrepreneurship
- Issues for discussion
 - other proxies for entrepreneurship?
 - aggregate index of entrepreneurship
 - how, why and how x why?

Measuring Entrepreneurship

- Main proxy is new firms/total firms based on business registry data (heroic data collection effort!)
- Other proxies?
 - household survey data – “have you tried to start a business in past year?” -- potential to get at failed businesses too....
 - firm survey data (e.g. ICAs, or piggybacking on existing census data) – get distribution of firm age in a (hopefully) representative sample
 - exit data – high exit rates signal of high-risk entrepreneurial entry activity?
- How do such alternatives correlate with business registry data?
 - intrinsic value of survey-based perceptions data in getting at reality “on the ground”

Subjective and Objective Measures of Ease of Business Entry: Developing Country Sample



Indicator of Entrepreneurship: Option 1

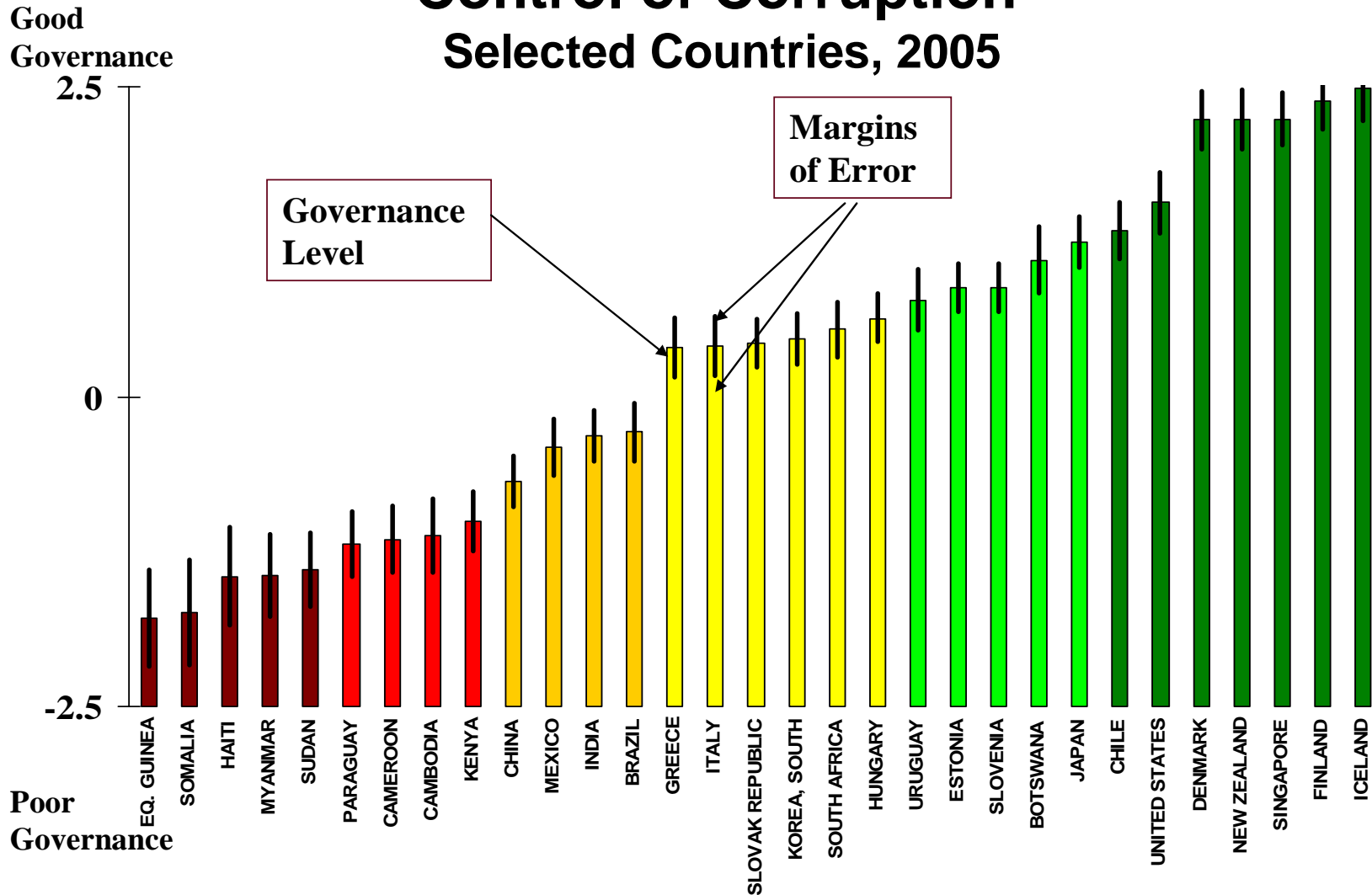
- entrepreneurship in country j , e_j , is unobservable but we can observe $k=1, \dots, K$ proxies for it:

$$E_{jk} = \alpha_k + \beta_k(e_j + u_{jk})$$

- this unobserved components model identified for $K \geq 3$ proxies (would we have this many...?)
 - provides method of rescaling (estimates of α_k and β_k)
 - provides natural estimates of e and its variance
 - estimate of $e_j = E[e_j | E_{j1}, \dots, E_{jK}]$
 - precision of estimate = $V[e_j | E_{j1}, \dots, E_{jK}]$
 - *requires identifying assumption of uncorrelated errors*

UCM Application: Worldwide Governance Indicators

Control of Corruption Selected Countries, 2005



Source for data: 'Governance Matters V: Governance Indicators for 1996-2005', D. Kaufmann, A. Kraay and M. Mastruzzi, September 2006.

Note: Colors are assigned according to the following criteria: Dark Red, below 10th percentile rank among all countries in the world; Light Red between 10th and 25th; Orange, between 25th and 50th; Yellow, between 50th and 75th; Light Green between 75th and 90th; Dark Green above 90th.

Indicator of Entrepreneurship: Option 2

- Suppose only one proxy available, E_j , but have some sense of N determinants (as per suggestive empirical work presented today):

$$E_j = \delta_1 Z_{1j} + \dots + \delta_N Z_{Nj} + \varepsilon_j$$

- Use fitted values from regression as indicator of entrepreneurship (or policy environment conducive to entrepreneurship)
 - appealing in principle.....weight determinants by how much they matter!

Indicator of Entrepreneurship: Option 2

- but nearly insuperable difficulties in credibly identifying marginal effects of determinants to use as weights
- e.g. $N=2$, $Z_1 = \text{“rule of law”}$, $Z_2 = \text{“regulation”}$
 - clearly need instruments, maybe:
 - settler mortality for rule of law
 - legal origins for entry regulation
 - subtle weak instruments problem
 - each instrument must explain just one endogenous variable
 - sloppy habit in profession of “recycling” instruments
 - do we believe legal origins as instrument for finance?
 - if so can’t “recycle” here if finance matters for entrepreneurship – invalid exclusion restriction!
 - if $Z_3 = \text{finance}$ included, then need yet another instrument!
 - imagine how things get harder with $N > 2$!

Why Do We Need an Aggregate Indicator – And What Does This Imply for Aggregation Method?

- Use 1: Awareness and Advocacy
 - premium on transparency, regular updating, consistent over time methodology
 - percentile matching and simple averaging as alternative to UCM
 - explicit margins of error important for policy dialogue
- Use 2: As Dependent Variable
 - use Option 1 based on multiple proxies, otherwise interpretation is very hard
 - less measurement error in dependent variable, easier to find significant effects
 - but don't forget key (strong?) assumption: measurement error in proxies can't be correlated with regression error term...

Why, How, and Why x How, Cont'd

- Use 3: as RHS variable in a regression

$$y_j = \theta e_j + v_j$$

- e_j unobservable, so insert average of proxies (a la Option 1)

$$y_j = \theta(w_1 E_{j1} + \dots + w_K E_{jK}) + v_j$$

- surprisingly this is a bad idea! Lubotsky and Wittenberg (ReStat 2006) show that minimum variance estimate of θ is sum of coefficients in regression, even if E_{jk} very collinear

$$y_j = \theta_1 E_{j1} + \dots + \theta_K E_{jK} + v_j$$

- point is that weights that make sense for aggregation (w_k) need not make sense for subsequent analysis

Conclusions

- Aggregate indicator of entrepreneurship potentially very useful for advocacy
- Option 1 (averaging multiple proxies) seems best bet
- Option 2 (fitted values from determinants regression) extremely hard to do credibly, could however be used as indicator of “entrepreneur-friendly” policy environment
- For some empirical applications no need to aggregate (e.g. effects of entrepreneurship on outcomes)