Welfare Dynamics Measurement: Two Definitions of a Vulnerability Line and Their Application

Hai-Anh Dang and Peter Lanjouw
DECRG, the World Bank

UNU-WIDER Conference
Inequality – Measurement, Trends, Impacts and Policies
Helsinki, Finland

September 5-6, 2014

Motivation

- In spring/summer 2012 there were intensive discussions in the WB on approaches to “boosting prosperity”
- Decision was to focus on boosting growth of the bottom 40%
  - Not explicitly inequality focused but a big foot in the door
  - This will be reported annually for all countries
  - Policy Research Report, Global Montoring Report
- Other proposals relating to prosperity had focused on vulnerability
  - People may be non-poor, but their situation still precarious
  - Idea resonates with recent FT coverage of “fragile middle”
  - Idea here is to divide population into three groups: poor, vulnerable and “prosperous (=secure)”

Motivation

- Thinking about vulnerability also enters naturally into standard poverty analysis.
  - Identifying those in need of assistance is a common objective in poverty assessments
- Exclusive focus on the poor may deflect attention from those of the non-poor facing a heightened risk of falling into poverty ("vulnerable").
  - When assessing social protection efforts one may wish to incorporate reaching the "vulnerable" in targeting schemes.
    - Not all "leakage" is the same

Vulnerability Lines

- Setting vulnerability lines is less well-established than specifying poverty lines
  - WB Poverty Assessments often estimate multiple poverty lines and designate the higher line as the "vulnerability" line
    - No attempt to explicitly address the risk of falling into poverty of those designated as vulnerable.
  - Often countries simply scale up poverty lines arbitrarily
    - In India: vulnerable population lies between 1.25-2 times the national poverty line.
    - In Vietnam: vulnerability line lies 30% higher than poverty line.

Vulnerability

- It seems desirable to link the setting of a vulnerability line to the risk of falling into poverty
  - Pritchett et al (2000) specify a vulnerability line at the level of income below which a household experiences a greater than even chance of experiencing poverty in near future.
    - “vulnerable” by this definition includes the currently poor.

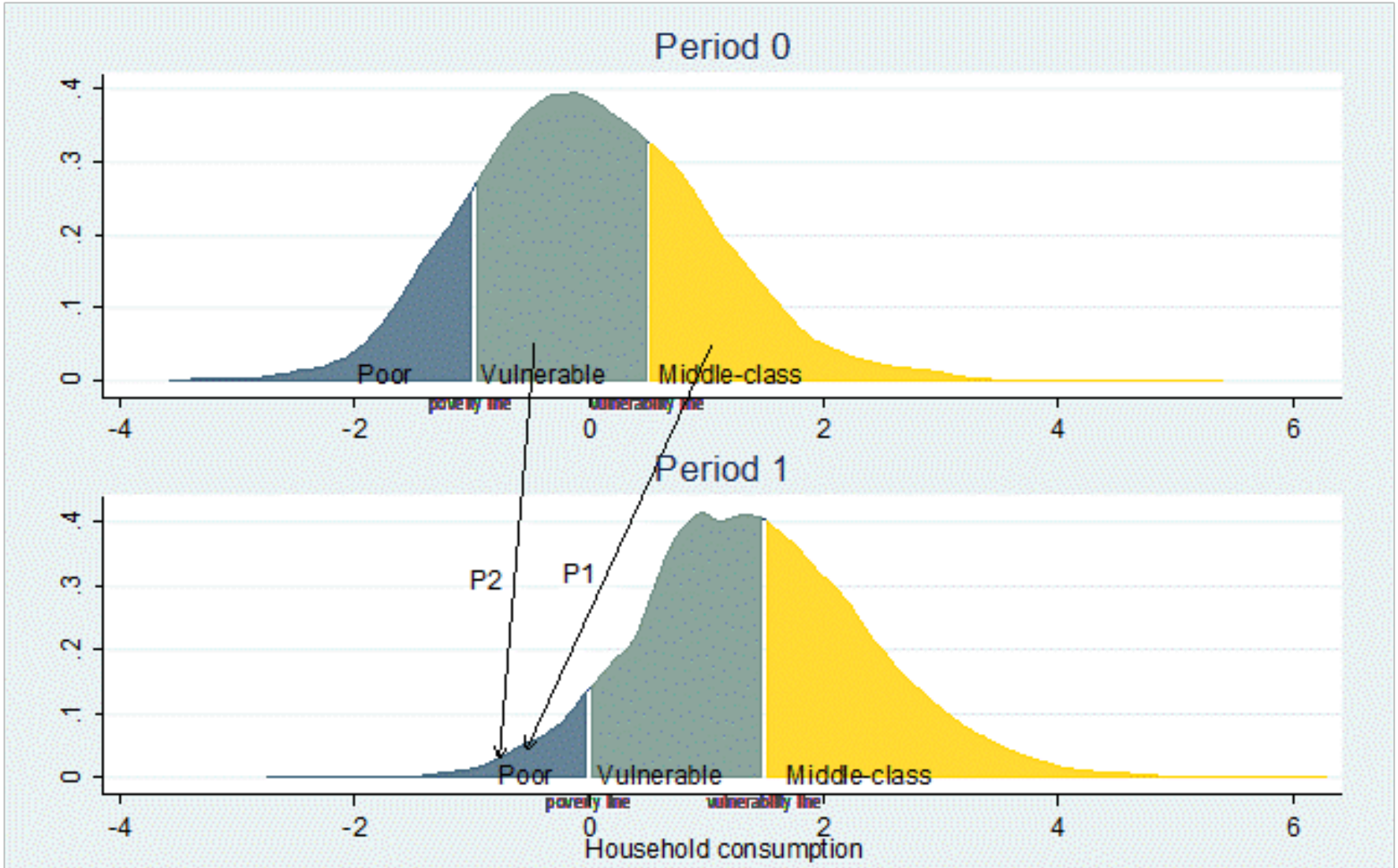
Our proposal

- Link vulnerability to the notion of susceptibility to something harmful that has not yet occurred.
  - Distinguish the vulnerable from the poor
- Set a vulnerability line that builds on the poverty line and separates the population into three groups:
  - Poor, vulnerable, secure
- Those above the vulnerability line may also be thought of as the “middle class”, “resilient”.
  - Considerable attention in literature in defining these latter groups, but no current consensus.

Two approaches

Approach 1

- Specify the highest acceptable level of risk of falling into poverty, more than that is considered excessive/unacceptable.
- Identify population whose risk of falling into poverty in next period is at this level *or lower*.
- Define vulnerability line as the lower bound income level of this population (ie, the middle class).
- Population that lies between the poverty line and this vulnerability line is designated as vulnerable.



Note
Household consumption distribution in each period is divided into 3 groups: poor, vulnerable, and middle-class, which are shaded differently. P1 and P2 respectively represents the insecurity index (percentage of the population that are in the middle class group in period 0 but in the poor group in period 1) and the vulnerability index (percentage of the population that are in the vulnerable group in period 0 but in the poor group in period 1)

Two approaches

Approach 2

- Specify the highest acceptable level of risk of falling into poverty, more than that is considered excessive/unacceptable.
- Identify population that is clearly not poor, but whose risk of falling into poverty in the next period is at this level *or higher*.
- Define the vulnerability line as the upper bound income level for this population (ie, the vulnerable).

Features

- Specifying a common cut-off risk level allows comparisons of vulnerability even across settings where income levels and poverty lines vary.
- Estimation of these vulnerability lines is simple
  - "light" demands of underlying panel data
  - Can be straightforwardly implemented also with synthetic panels.
  - Can be contrasted with other proposed approaches
    - Chaudhuri (2003) uses cross-section data but makes restrictive assumptions
    - Lopez-Calva and Ortiz-Juarez (2014) rely on panel data and make a number of parametric assumptions
- Note: The approach is intended to identify population groups rather than establish vulnerability of individual households.

Implementation

- Approach 1

  estimate: $P^1 = \frac{P(y_1 \leq Z_1 \cap y_0 > V_0)}{P(y_0 > V_0)}$

- Approach 2

  estimate: $P^2 = \frac{P(y_1 \leq Z_1 \cap Z_0 < y_0 < V_0)}{P(Z_0 < y_0 < V_0)}$

- Solve empirically for V_0

Data

- US: panel income data from PSID for 2005, 2007 and 2009 (reference to previous year)
  - 5,335 panel households
- Vietnam: panel consumption data from VHLSS for 2004, 2006, 2008
  - 1,800 panel households for 2004, 2006 and 2008
  - 3,735 panel households for 2006 and 2008
- India: cross-sectional consumption data from NSS for 2004/5, 2009/10

Approach 1 in US and Vietnam

		United States			Vietnam		
No	Vulnerability index	Vulnerability line	Increase (%)	Pop. share with consumption above V-line	Vulnerability line	Increase (%)	Pop. share with consumption above V-line
1	6	N/A	N/A	N/A	2560	0	84.2
2	5	N/A	N/A	N/A	2800	9	80.4
3	4	13305	0	89.0	3080	20	75.7
4	3	17905	35	83.5	3320	30	71.6
5	2	26505	99	73.6	3920	53	61.0
6	1	61305	361	40.8	5320	108	40.7

Note: Vulnerability lines are in $US per household and D'000 per capita respectively for the US and Vietnam. The relative increases of the vulnerability line from the poverty line for each country is shown under the columns "Increase" (columns 4 and 7). All numbers are estimated with true panel data and weighted with population weights. Estimation sample sizes are 5,335 panel households for the US, and 3,735 panel households for Vietnam. The incremental values for

Approach 2 in US and Vietnam

		United States			Vietnam		
No	Vulnerability index	Vulnerability line	Increase (%)	Pop. share with consumption above poverty line but less than V-line	Vulnerability line	Increase (%)	Pop. share with consumption above poverty line but less than V-line
1	22	N/A	N/A	N/A	3340	30	12.9
2	21	N/A	N/A	N/A	3380	32	13.5
3	20	N/A	N/A	N/A	3500	37	15.3
4	19	21105	59	9.0	3660	43	18.2
5	18	22205	67	10.3	3700	45	19.0
6	17	22805	71	11.0	3780	48	20.6
7	16	24305	83	12.6	3920	53	23.3
8	15	25105	89	13.9	4140	62	26.7
9	14	27905	110	16.8	4260	66	28.6
10	13	29005	118	18.5	4500	76	32.2
11	12	30405	129	20.1	4860	90	37.1
12	11	32405	144	22.2	5140	101	40.9
13	10	36905	177	26.4	5480	114	45.5
14	9	45905	245	34.1	6080	138	51.6
15	8	50605	280	39.2	7080	177	59.3
16	7	59105	344	46.1	8980	251	69.6
17	6	74405	459	56.8	17500	584	81.8
18	5	104405	685	69.9	N/A	N/A	N/A
19	4	290905	2086	86.9	N/A	N/A	N/A

Note: Vulnerability lines are in $US per household and D'000 per capita respectively for the US and Vietnam. The relative increases of the vulnerability line from the poverty line for each country is shown under the columns "Increase" (columns 4 and 7). All numbers are estimated with true panel data and weighted with population weights. Estimation sample sizes are 5,335 panel households for the US, and 3,735 panel households for Vietnam. The incremental values for iteration are $US100 and D20,000 respectively for the US and Vietnam. The exchange rate is US$1 for D16,302 in 2008 (World Bank, 2013).

Vulnerability in US and Vietnam

- Approach 2 with a cut-off risk of 10%
- Estimate vulnerability line in 2006
  - Examine risk of falling into poverty between 2006 and 2008
- Inflate/deflate vulnerability line to 2008 and 2004, respectively using CPI data
- Compare percentage of population poor, vulnerable and “prosperous” in 2004 and 2008

Trends in Vulnerability and Poverty in US and Vietnam: 2004-2008

	US		Vietnam	
	2004	2008	2004	2008
Poor	8.9	10.0	20.0	14.3
Vulnerable	26.4	27.9	51.0	49.7
Middle class	64.7	62.1	29.0	36.0

Application to India

- No panel data available in India
- Dang, Lanjouw, Luoto and McKenzie (2014), and Dang and Lanjouw (2013) outline a procedure to construct synthetic panels out of cross section data.
- Based on imputation models, reliant on time-invariant correlates of consumption.
- Fairly extensive validation work suggest approach is reasonably reliable.

The proposed approach
(Dang, Lanjouw, Luoto and McKenzie, JDE 2014)

- Combines ideas of poverty-mapping with pseudo-panel ideas.
- Will set out for case of 2 rounds, can be extended easily to multiple rounds.
- Let x_{i1} be characteristics of household *i* in time period 1, which are observed in both the round 1 and round 2 surveys:
  - All time-invariant characteristics (language, religion, ethnicity)
  - Characteristics of household head if the head doesn't change across rounds (sex, place of birth, parental education, etc.)
  - Can include time-varying characteristics that can easily be recalled for round 1 in round 2
    - E.g. whether household head was employed in round 1, place of residence in round 1, whether household has a TV in round 1, etc.
  - Can also include time invariant ancillary variables (Census, GIS, etc.) at a more aggregated level

Projections

- Project round 1 consumption or income onto $x_{i1:}$

$$y_{i1} = \beta_1' x_{i1} + \varepsilon_{i1}$$

- Project round 2 consumption or income onto same set of characteristics as they appear at time of second round:

$$y_{i2} = \beta_2' x_{i2} + \varepsilon_{i2}$$

- Then we are interested in knowing quantities such as: $Pr(y_{i1} < p | y_{i2} > p)$

Don't observe for the same household

Proposed method

- *Step one:* Use the sample of households observed in round 1, and regress y_{i1}^{1} on x_{i1}^{1}
  - Obtain the OLS estimator $\hat{\beta}_1$ and the residuals:

$$\hat{\varepsilon}_{i1}^{1} = y_{i1}^{1} - \hat{\beta}_1{}' x_{i1}^{1}$$

  - Superscript 1 denotes that these are observations for households observed in round 1 only.
- *Step two:* For each household observed in round 2, take a random draw with replacement from the empirical distribution of residuals, then combine with parameter estimate and known *x* to estimate round 1 income or consumption:

$$\hat{y}_{i1}^{2} = \hat{\beta}_1{}' x_{i1}^{2} + \tilde{\varepsilon}_{i1}^{2}$$

Proposed method

- *Step Three:* calculate movements into and out of poverty using $\hat{y}_{i1}^2$ in place of the unobserved round 1 variable:

$$Pr(\hat{y}_{i1}^2 < p | y_{i2}^2 > p)$$

- *Step Four:* Repeat steps 1-3 R times, and take average of the quantity of interest over the R replications.

Under what conditions will this be consistent?

- *Condition 1:* the underlying population sampled is the same in round 1 and round 2
  - Requires measure of consumption to be same from round to round,
  - Assumption implies that households in period2 that have similar characteristics to those of households in period 1 would have achieved the same consumption levels in period 1 or vice versa.

Under what conditions will this be consistent?

- *Condition 2:* ε_{i1} is independent of $y_{i2.}$ This requires ε_{i1} to be independent of ε_{i2}

  (otherwise the distribution of $\varepsilon_{i1}|y_{i2} > p$ *is not the same as the unconditional distribution of* ε_{i1}*)*
  - Won't hold if:
    - Error term contains individual fixed effect
    - If shocks to consumption or income are non-transitory.

➢ We expect in many cases this condition to be violated.
  - So long as errors positively correlated (which seems likely in most cases), this will *overstate* mobility, providing an *upper bound* on movements into and out of poverty.
  - If errors are negatively correlated then our method wouldn't provide a bound.
    - We don't expect a negative correlation *on average*
    - DLLM (2014) demonstrate this empirically with real panel data.

Lower bound method

- Instead assume the prediction error for household *i* in round 1 is the same as it is for round 2 (perfect positive autocorrelation).
- *Step One:* for sample of households surveyed in round 2, obtain OLS residuals: $\hat{\varepsilon}_{i2}^{2} = y_{i2}^{2} - \hat{\beta}_{2}{}' x_{i2}^{2}$
- *Step Two:* then estimate round 1 income or consumption as $\tilde{y}_{i1}^{2} = \hat{\beta}_{1}{}' x_{i1}^{2} + \hat{\varepsilon}_{i2}^{2}$
- *Step Three:* Use the estimated *y* from step 2 to calculate poverty dynamic of interest.

Validation Datasets in DLLM (2014)

- Choose two genuine panels from Vietnam and Indonesia:
- VLSS 1992/93 and 1997/98
  - Period over which poverty fell from 58% to 37%, more households exiting poverty than entering
  - Panel of approximately 4800 households
- Indonesian Family Life Survey 1997 and 2000 (IFLS2 and 3)
  - Static in terms of overall poverty levels, household moving into and out of poverty at similar rates
  - Panel of 7500

Validation of method

- Randomly split each genuine panel into two sub-samples, A and B.
  - Use sub-sample A from round 1 and sub-sample B from round 2 as two repeated cross-sections.
  - Then carry out our method by using sub-sample A to impute round 1 values for sub-sample B, and compare to results we would get using genuine panel for sub-sample B.

How well does the approach do at estimating overall rates of movements into and out of poverty?

Table 3: Poverty Dynamics from "Pseudo" Panel and Actual Panel Data

Indonesia	Lower Bounds		Truth		Upper Bounds	
1997, 2000 Statuses	Basic	Full	95% CI		Basic	Full
Poor, Poor	0.115	0.105	0.047	0.070	0.024	0.037
Poor, Nonpoor	0.015	0.031	0.065	0.088	0.097	0.090
Nonpoor, Poor	0.021	0.030	0.065	0.088	0.111	0.099
Nonpoor, Nonpoor	0.848	0.832	0.759	0.801	0.766	0.774

Vietnam	Lower Bounds		Truth		Upper Bounds	
1992, 1998 Statuses	Basic	Full	95% CI		Basic	Full
Poor, Poor	0.360	0.322	0.275	0.360	0.227	0.288
Poor, Nonpoor	0.241	0.274	0.261	0.324	0.331	0.308
Nonpoor, Poor	0.000	0.039	0.034	0.060	0.138	0.077
Nonpoor, Nonpoor	0.398	0.366	0.300	0.386	0.305	0.327

For both countries, round 1 year is predicted, round 2 is "truth"

Results seem encouraging

- Bounds not that wide:
  - Full model would lead us to estimate 3-9% of households in Indonesia and 27-31% of households in Vietnam exited poverty over 2 rounds.
  - Genuine panel would say 7-9% in Indonesia and 26-32% in Vietnam

Imposing parametric assumptions

- Assume ε_{i1} and ε_{i2} have a bivariate normal distribution
- ρ is the correlation coefficient between these two error terms (assumed positive).
- DLLM bounds assume ρ being equal to its maximum value (1) and minimum value (0)
- But the true value of ρ in all likelihood lies somewhere in between these two values.
- Parametric approach allows us to "plug" in a value of ρ

Imposing Parametric Assumptions

- DLLM (2014) explore existing panel surveys to get a sense of plausible values for ρ
    - Evidence from Indonesia, Vietnam, Chile, Nicaragua and Peru suggest ρ might range between 0.5-0.8
    - Preliminary evidence from SILC data from Europe suggest a range for ρ between 0.3-0.8
- Dang and Lanjouw (2013) propose an approach to estimate ρ directly, based on birth cohort-aggregated consumption correlations.

Validation of Synthetic Panel Approach to Measuring Vulnerability in Vietnam

- Take panel data for Vietnam for 2006-2008 and treat the two halves as though they were cross section data
- Apply synthetic panel method
- Re-estimate (approach 2) vulnerability line
  - Original, true, vulnerability line = D5,480,000
  - Synthetic panel approach = D5,500,000

Validation: Vietnam

			2008			
			Poor	**Vulnerable**	**Middle class**	**Total**
Panel A: True Panels	**2006**	**Poor**	9.9	5.4	0.4	15.7
			(0.8)	(0.5)	(0.2)	(0.9)
		Vulnerable	4.7	33.3	9.5	47.5
			(0.5)	(1.0)	(0.7)	(1.1)
		Middle class	0.2	9.4	27.1	36.8
			(0.1)	(0.7)	(0.9)	(1.0)
		Total	14.8	48.2	37.0	100
			(0.9)	(1.2)	(1.1)	
			2008			
			Poor	**Vulnerable**	**Middle class**	**Total**
Panel B: Synthetic Panels	**2006**	**Poor**	**10.8**	**5.3**	**0.1**	**16.2**
			(0.3)	(0.1)	(0.0)	(0.3)
		Vulnerable	**4.8**	31.0	7.8	43.6
			(0.1)	(0.2)	(0.1)	(0.2)
		Middle class	**0.1**	**9.3**	30.8	40.2
			(0.0)	(0.1)	(0.3)	(0.4)
		Total	**15.7**	45.7	**38.6**	100.0
			(0.4)	(0.2)	(0.4)	

Trends: US, Vietnam, India

	US		Vietnam		India	
	2004	2008	2004	2008	2004	2009
Poor	8.9	10.0	20.0	14.3	36.8	31.5
Vulnerable	26.4	27.9	51.0	49.7	46.3	48.5
Middle class	64.7	62.1	29.0	36.0	16.8	20.0

Note: cut-off risk for India is set at 20%, not 10%

Vulnerability lines:

US =2.77 * Z

Vietnam = 2.14 * Z

India = 2.05 * Z

Conclusions

- We consider an alternative 3-way breakdown of the population: poor, non-poor but vulnerable, non-poor and secure
  - We do not explicitly focus on the “rich”
- One could consider the latter group as the “prosperous”
- We focus on the definition of a “vulnerability line” that would permit such a population breakdown
  - Try to link the line explicitly to a notion of risk

Conclusion

- The two alternative approaches we propose are simple to apply with panel data
  - Make “light” use of panel data
- Of course, panel data are rare, particularly in the developing world
- However, parallel research has pointed to ways to construct “synthetic panels” with cross-section surveys
  - Growing experience suggests method works quite well for a number of applications
- Synthetic panels can readily accommodate the vulnerability line calculations proposed here

Conclusion

- Empirical application to US, Vietnam and India in the period 2004-2008/9 point to some interesting findings:
  - "prosperity" in the US declined, increased in Vietnam
  - "vulnerability" in the US increased, declined in Vietnam.
  - poverty in US rose, fell in Vietnam
- Using a different "cut-off" level (20%) the picture in India points to:
  - falling poverty, rising vulnerability, rising prosperity