

Exporting CAPI data to Stata: Experience from *surveybe*

Joachim De Weerd

Research Director

EDI (Economic Development Initiatives)

Introduction

- A lot data analysed in Stata are collected through surveys.
- Movement away from paper questionnaires, in favour of electronic data collection:
 1. Web surveys
 2. Phone surveys,
 3. Computer Assisted Personal Interviewing (CAPI), etc.
- Web and phone surveys are small
 - Few dozen variables, easy to manually clean and label
 - Typically one flat file: no problems of relational integrity
- But CAPI is now handling long, complex, multi-topic questionnaires:
 - Hundreds of variables, very hard to manually clean and label
 - Dozens of files all linked together: ensure relational integrity

Why a surge in CAPI

- Market developments:
 - Tablets are becoming cheaper (need a tablet for larger screen)
 - Mobile phone networks can handle data transfer from the field and are prolific, even in rural areas in developing countries
 - Software options are becoming available (e.g. *surveybe*)
- Compelling list of benefits:
 - Errors are identified and resolved *during* interview
 - One-click export to Stata in labelled data set
 - Data available immediately
 - lower costs
 - Exponentially expands level of creativity: do stuff you cannot do on paper (call soundex algorithms, use regular expressions, access large look-up tables during interviews)

Reduce Errors:

We expect error to be reduced by:

- Automated routing (skips)
- Pre-coded drop down menus or radio-buttons
- Automatically captured GPS coordinates and interview times
- Elaborate system of validation checks
- Complex reports

Clean and labelled data set

- Variable labels: question text
- Value labels: drop-down text
- Complete relational integrity ($_m==3!!$)

Q.5 What was or is the primary use of the Tema parcel during the second cropping season?

Select...

Select...

Own Cultivated (annual crops) (1)

Own Cultivated (perennial crops) (2)

Rented out (3)

Cultivated by mailo tenant (4)

Fallow (5)

Pasture (6)

Woodlot / Forest (7)

Data Quality Matters

	(1) CAPI	(3) Pen and Paper	Significance of <i>t</i> -test that (1) = (3)
Poverty Headcount	83.0	68.3	***
Gini (95% CI)	.24 (.22-.25)	.30 (.27-.32)	***
<i>N</i>	1200	319	

From Caeyers, Chalmers and De Weerd (2010)

In a randomised experiment we see that inequality goes up when consumption expenditures are measured on paper; measurement error is being picked up as income inequality

Testing for attenuation bias

	No. of years of schooling (children aged 7-14)	Schooling expenditures on school-going children	Child slept under a treated bednet night before survey
	Age FE	Age FE	LPM Age FE
Log total consumption per aeu (δ)	0.204 (0.135)	7,775*** (1,630)	0.134*** (0.036)
log total consumption per aeu * CAPI assignment	0.292** (0.144)	6,279*** (1,722)	0.154*** (0.040)
<i>N</i>	2,683	2,137	5,148

From Caeyers et. al (2010)

We see evidence of attenuation bias when using paper questionnaires: coefficients estimated with data collected on paper are biased towards zero.

SURVEYBE DEMO

Food for Thought

- Using different scripts and different languages
- Exporting the system of skip & validation rules
- Comments