

December 4, 2002

Memorandum

To: BASIS CRSP project team

Subject: Thoughts on data analysis

Hello, everyone. Once again, I am late with providing guidance on next steps in our BASIS project. My sincere apologies to you.

In January 2002, I wrote and sent around to each of you a memo on data collection and analysis. I hope you still have that, as much of the discussion in that memo remains highly relevant now. (I'll attach a copy of that memo, just in case you've misplaced it or, for newer team members, in case you never got it.) We still aim to describe welfare and natural resource dynamics - individually and jointly -- in each of our sites over the period spanned by the various survey rounds, to do econometric analysis of the causal factors underlying poverty traps, and to do econometric analysis and use descriptive statistics and secondary data to calibrate and validate the CLASSES bioeconomic modeling tool for at least some of our project sites.

By now, data collection has taken place in each survey site and data entry and cleaning is at least underway, nearly done in one or two sites. Other than in Embu, where we discovered belatedly that the prior surveys on which we planned to build a panel suffered fundamental flaws, we should now have multiple observations across time on each of many households in a site (i.e., panel data). So panel data analysis can commence as soon as the data are ready.

Establishing welfare measures and associated transition matrices

The first task is to establish welfare measures for each household in each period, and then the associated transition matrices describing changes over time in those measures (see the stylized example below). These transition matrices are valuable in their own right as important descriptions of the system. We will use these in descriptive outputs from the project for each site.

Stylized Transition Matrix	Poor in first year	Non-poor in first year
Poor in last year	Poor/poor	Non-poor/poor
Non-poor in last year	Poor/non-poor	Non-Poor/Non-Poor

The transition matrices are also critical inputs into the qualitative follow-up work that is to be undertaken by the anthropologists and sociologists working with us in these sites. Recall that, in addition to focus group and key

informant interviews in each of our sites, they are going to revisit eight households in each site to try to flesh out the numbers with the human stories behind their transitions (or lack of transition)) in welfare status. To do the qualitative follow-up work correctly, the anthropologists/sociologists will need to have listing of survey households sorted into each of the four cells in the transition matrix prior to their field work. They are to draw two households from each of the four cells in the transition matrix. I am attaching the generic terms of reference we have drafted for that qualitative work. Any comments on it – refinements, corrections, concerns, other suggested questions – would be most welcomed, but we will need them by the end of December to be able to incorporate them in the site-specific plans being developed. They need to get to the field in February to start this work.

Back to the welfare measures. We must use a consistent measure across sites and periods if we are to undertake meaningful comparisons. Recall that not all of our sites had comprehensive, reliable expenditures data from the baseline survey on which BASIS builds. We therefore opted to use income as the common measure of welfare we could use for comparison across all sites. Kenya and Madagascar both have established poverty lines; unfortunately, they are not the same. So let us use the crude, international “extreme poverty line” of US\$1/person per day and also the international “poverty line” of US\$2/person per day. In each case, you will need to multiple 365 days times the annual average exchange rates for the correct year to establish the annual income per person that defines the poverty line. Please send me that figure ASAP so that we can make sure these are indeed comparable across sites (i.e., people don’t use significantly different exchange rates). I suspect that with the \$2/person per day poverty line, we will get very few entries in some of the non-poor/non-poor cells in some of our sites. Hence the need to use both metrics.

We need to follow the same procedures across all data sets for measuring income. Income is not just cash receipts. In each data set, please compute income in the

following manner (and let me know *very* soon if this is infeasible for some reason). Income equals cash income plus the cash value of home-consumed food production (including milk and meat from slaughtered animals). Cash income includes wage and salary earnings, sales of home-produced items less expenditures on inputs for those items (i.e., cash profits, the difference between cash revenues and cash expenditures), any interest or rental income on financial or physical (e.g., houses, farm land) assets, plus net transfers (i.e., money received from others – pensions, remittances, gifts, etc. – less money given to others for similar purposes). Home-consumed food production should be valued at prevailing market prices for the goods in question.¹ This is biased measure (upwardly biased in the case of net sellers, downwardly biased in the case of net buyers, i.e., it tends to magnify income differentials in so far as incomes tend to be positively related to net food sales among farm households). But we don't have farm-specific (shadow) prices readily available, so this will have to do for now.

There's always an issue of "equivalence scales" in comparing households that have different demographic composition, including the same households over time, as everyone ages, some die, new children are born, and people migrate in and out of the household between periods. This is a thorny technical issue to which there really is no satisfactory answer. And, unfortunately, analytical results are not always robust to reasonable changes in the way in which we normalize household income (or expenditures or assets or any other measure describing the collective household unit) to control for differences in composition across observations. So we will want to make explicit the composition of the household in each data set so that we – or others who use our data in the future – can readily employ alternative scales. For present purposes, let us keep it simple and use the straight "per capita" approach, wherein all those resident in the household count equally (i.e., we do not weight by age and/or gender according to energy intake or expenditure). Just add up the number of people in each household during a survey period and divide the income (or expenditure) measure by that number to get per capita income.

Now in many ways, the preferred welfare measure is expenditures, due to consumption smoothing in the face of volatile income, misreporting of income, etc.. So where we can compute expenditures and the associated transition matrices, we should, once again using the \$1/person per day and \$2/person per day poverty lines. But please do the income based measures first since those are the ones that are fully comparable across sites.

¹ I think it infeasible to impute rental values from housing stock, so we'll skip that. Likewise, we won't try to value leisure time.

In some of our sites (e.g., the arid and semi-arid pastoralism sites in Northern Kenya), assets (e.g., livestock holdings) could arguably be a better measure of welfare. But the concept of asset poverty remains underdeveloped in the literature and a bit hard to communicate to lay audiences. So although we will indeed work with asset holdings as an indicator of welfare, initially we want to get the transition matrices constructed for expenditures or income. We can worry about asset transition matrices later. (Constructing asset transition matrices also requires an intermediate step of building asset indices so that we can summarize the range of assets a household owns into a single, scalar-valued measure of wealth. This will take added time this spring.)

Action items:

1. Report the poverty line values for each year and site to me as soon as possible. (By December 15, 2002, please)
2. Get data entered and basic cleaning done (i.e., identify and correct or eliminate infeasible responses), by December 31, 2002, please (hopefully, it's done already in some sites, e.g., Madzu, Madagascar, Baringo, Marsabit).
3. Send the full panel data (i.e., original data as well as new data, in comparable formats), and any associated documentation on data entry or issues that emerged during the conduct of the survey, to me at Cornell for back up storage and so that we can assist with data analysis from this end, where needed. Be sure to report explicitly the code used for missing values (we do not want to confuse true "0" values with missing values).
4. Compute transition matrices following the instructions above, and send them to me, Festus, Frank and Jhon by January 31, 2003, please.
5. Likewise by January 31, 2003, please make an Excel file listing all the survey households, sorted into four groups, each one corresponding to a different cell in the stylized transition matrix. Include the income and household demographic composition data for each and the data from the shocks module. The sociologists/anthropologists will need these data for their qualitative work. They may also want the full data on the chosen households prior to their interviews so that they go into the in-depth interviews well-informed (and able to cross-check our data a bit for us!). Please cooperate with them to the maximal extent possible on this.

Responsibilities:

I am expecting that Paswel will handle the above tasks for the Madzu data, that Justine or Frank will handle these tasks for the ICRAF Siaya/Vihiga data, that Andrew Mude will handle this for the Baringo and Dirib Gumbo data, and that Jean Claude will handle this for the two sites in Madagascar. ***I would ask that you each confirm this to me via email in the coming few days. Please also confirm whether the timing outline above is feasible, given the particulars of***

your data and other obligations on your time. If you need assistance, please let me know as soon as possible. I may be able to enlist a graduate student or two here to help over our winter break. They will definitely be helping out some in spring semester as the more detailed empirical analysis proceeds.

Conclusion

My apologies for another long memo, some of which is probably pretty basic to most of you. But I felt it important to ensure we are all following the same basic methods since errors at this stage will compound rapidly once we get further into the analytical work. Please email me - and the whole group - with comments or suggestions for changes or additions.

Finally, Jhon should be getting in touch with all of us very soon with details on our March 12-14 team meeting in Antsirabe, Madagascar. Those who will be attending should plan their schedule and travel accordingly.

Thank you very much for all your hard work on this project. It was a real treat spending time with the bioeconomic modeling student cohort and I am *extremely* excited to get going on the empirical analysis of these data. Yesterday I spoke with Michael Carter, Director of the BASIS CRSP, and it appears that they might hold their first policy workshop on poverty traps and persistent poverty in rural areas in Kenya in August, in part to feature our project. If so, it will be important to have plenty of at least provisional empirical results to share with them. And August is really not terribly far off, so we must get started.

Keep up the great work!

Warmest regards from snowy, cold Ithaca,



Christopher B. Barrett
Associate Professor