CORNELL

Department of Applied Economics and Management

Christopher B. Barrett
Associate Professor
Co-Director, African Food Security & Natural Resources Management Program

Cornell University
315 Warren Hall
Ithaca, NY 14853-7801

Telephone: 607-255-4489
Fax: 607-255-9984
E-mail: cbb2@cornell.edu
Web: aem.cornell.edu/faculty_sites/cbb2/

July 9, 2003

Memorandum

To: BASIS CRSP project team

Subject: Empirical analysis of welfare dynamics

Back in March, at our annual team meeting, I promised to draft up some notes on empirical methods for the analysis of welfare dynamics using the panel data we have collected under the BASIS project. It has, like usual, taken longer than I had hoped to get to this. I offer you my sincere apologies. Anyway, here goes.

Welfare Indicators

As we have discussed several times, there are multiple indicators one can use as proxies for the well-being of a household. A crude categorization can be made into flow measures such as expenditure (used in the Siaya/Vihiga, Baringo and Marsabit surveys) and income (used in all but the Siaya/Vihiga survey), and stock measures based on asset ownership: land, livestock, educated labor force, uneducated labor force, buildings, agricultural equipment, nonagricultural equipment, financial savings, etc. These can be aggregated using monetary values into a money metric measure of asset holdings or into a unitless index using factor analysis (see the attached December 2000 *World Development* paper by Sahn and Stifel for an example and detailed explanation).

For our purposes, we are interested primarily in understanding change in flow measures (income or expenditures) over time, perhaps explained by changes in stock measures (assets). But as the paper Frank Place presented in Antsirabe in March emphasized, there are lots of legitimate measures of welfare and the answers one gets to welfare dynamics questions often vary among the measures used. So we do need to be careful not to overstate findings without checking for consistency of results across multiple reasonably measures.

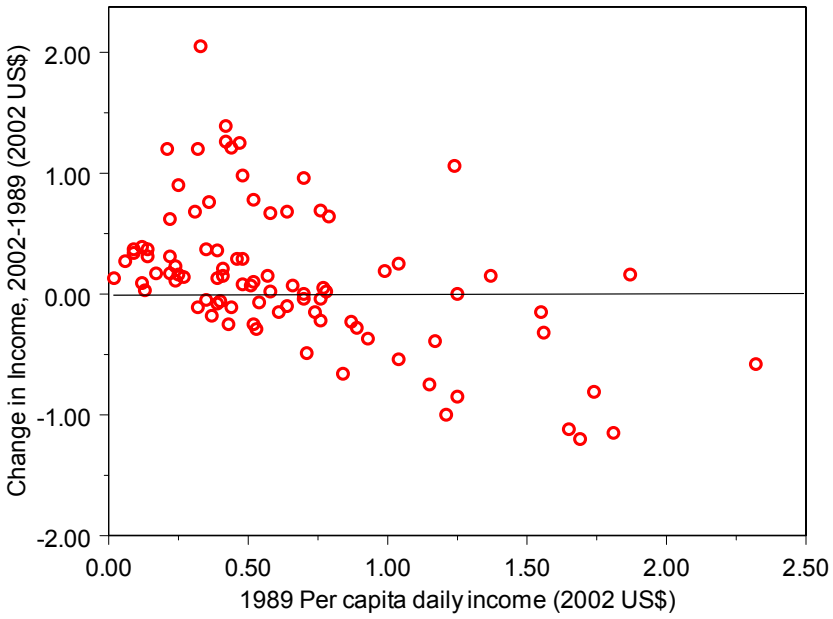
Describing welfare dynamics

We have already spoken at length about one basic way of capturing welfare dynamics, through a transition matrix that describes movement into and out of

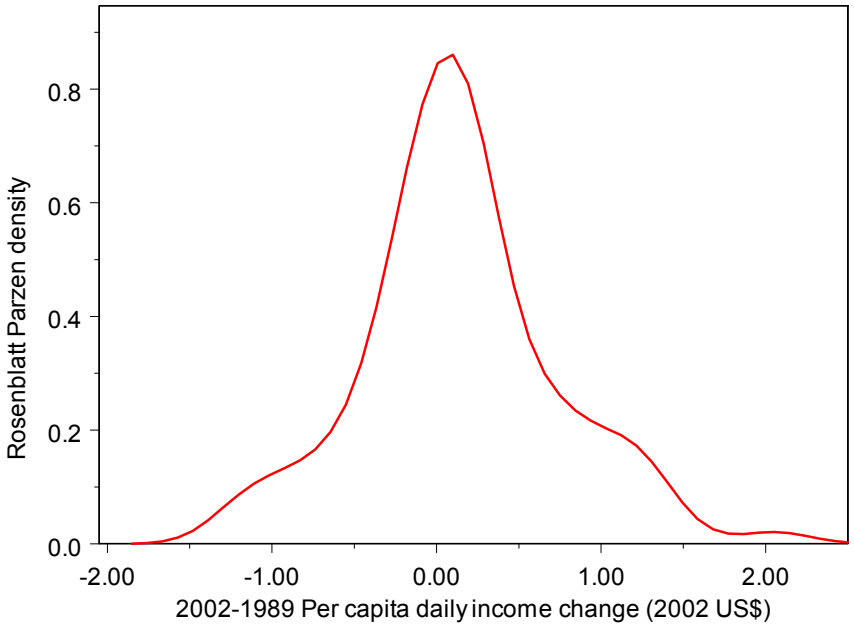
poverty across survey periods. Each site team constructed transition matrices and presented them at Antsirabe. This was extremely helpful. Thank you.

As we discussed, however, transition matrices may accurately capture improvements or deterioration that bump people across the poverty line. But they fail to capture changes within the subpopulation that is poor (or non-poor) in each survey period. If we are interested in broader patterns of welfare dynamics, not just movements around the poverty line, regardless of whether these changes are relatively big or small, then we need to use some other methods of describing what the data reveal. There are several ways of doing this. Let me try to lay them out for you quickly, with a few illustrations thrown in to help make some of the points a bit more clearly (I hope).

Method 1: Perhaps the simplest way of exploring this is to plot the distribution of income (or expenditure) changes in one or both of two ways. First, one can simply plot the observed income changes against the initial period income. Note that measurement error and the transitory component to income should generate significant regression to the mean, manifest in an inverse relation between these two variables, as depicted in the following plot of the Madzuu data (all observations are 2002 real US$ daily per capita income). This will lead to systematic overstatement of economic mobility and understatement of chronic poverty associated with poverty traps. So while it is worth looking at results derived from this method, it is very important not to stop there.



Method 2: A second way of exploring the distribution of income changes is simply to plot that distribution. One good way to do this is via a kernel smoother, generating what is commonly known as a Rosenblatt-Parzen density. Many econometric or statistical packages do this as a push button operation (the following graphic, using the same Madzuu data, was done in S-Plus in about two minutes). This sort of distribution reveals the general pattern of income growth in the community, without any association back to starting points. Here we can see that there were significantly more relatively big gains (≥$1.00/day per capita) than losses of equivalent magnitude. But most households experienced relatively little change in income. The degree of dispersion of the distribution around zero is an indicator of the degree of economic mobility in the community, keeping in mind again that measurement error will tend to generate some false dispersion. So this too will overstate true economic mobility. The sharpness of the peak here suggests negligible-to-modest economic mobility in Madzuu, especially over a 13 year span.

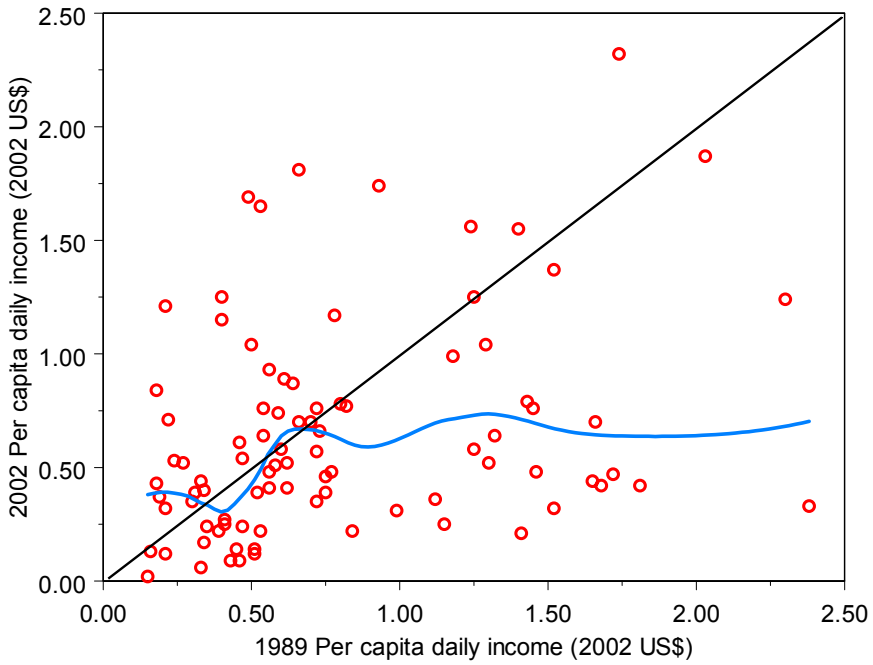


Method 3: Since measurement error creates upward bias in unfiltered measures of economic mobility, one obvious approach is to filter the data and redo the preceding two analyses. One candidate approach is to use expected income (or expenditure), based on the regression of observed income (expenditure) on a vector of assets, following the approach taken by Carter and May (1999 and 2001 *World Development* articles, previously shared with the team … let me know if you need a copy). Of course, the problem with this approach is that it will filter out

not only random measurement error but also true random changes in income, i.e., true transitory income. So this filtering approach will understate true economic mobility. I've not undertaken a demonstration of this here, not least of which because it doesn't seem worth the space to report the regression specification and results on which a plot would depend. But a combination of Methods 1-3 should bracket the true change in welfare over periods. In combination, these methods thus seem quite useful.

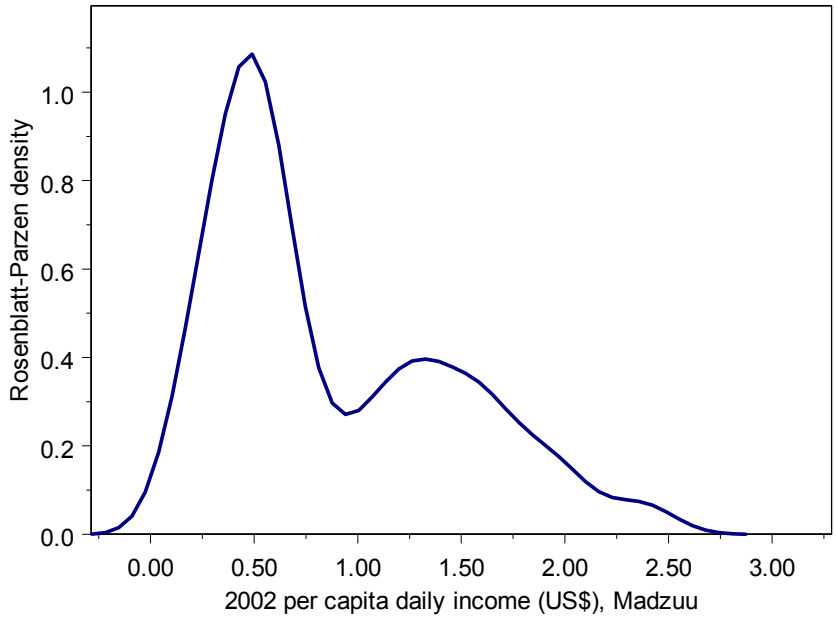
Ultimately, our project emphasizes the possibility and the consequences of structural poverty traps. This requires us to allow for patterns in the data that are often overlooked in some conventional analyses, as I discuss in the attached paper that I will present at the IAAE meetings in Durban next month. In particular, we want to allow for the possibility of multiple equilibria (if one's method assumes them away, one necessarily cannot find them!). Indications of such phenomena can be found through either of several methods.

Method 4: The first method simply plots successive periods' observations, one against the other, and then estimates the autoregression of latter period income (or expenditure) on earlier period income (expenditure). In order to allow for the sort of S-shape that characterizes poverty traps, however, one needs to either estimate a high order polynomial parametric regression via OLS (e.g., y_2 regressed on y_1, y_1^2, y_1^3, and y_1^4) or do this nonparametrically. The second degree LOESS regression for the Madzuu data is the squiggly blue line in the



graphic below. It exhibits the distinctive S-shaped income dynamics characteristic of poverty traps, with stable equilibria around $0.40 and $0.70/day – and another upward bump coming around $1.30/day (although not hitting the dynamic equilibrium reflected by the solid diagonal line) – and an unstable equilibrium at about $0.55/day per capita. Those are pretty troubling numbers.

Method 5: If there are multiple dynamic equilibria, then these should manifest themselves in cross-section as a multi-modal income distribution as households move to one or the other equilibrium, as I discuss in the IAAE paper, drawing on Danny Quah's celebrated "twin peaks" research (*Economic Journal* 1996, attached as a .pdf). In the Madzuu data, we indeed find multiple modes, as the graphic below shows. It is peculiar to me, however, that the second mode is the bump that doesn't reach the dynamic equilibrium point in the preceding graphic. This is one to puzzle over a bit, I suppose, and I'd welcome people's thoughts on why we're seeing this. Of course, it isn't terribly surprising that the two stable and one unstable equilibrium depicted in the previous graphic all blend together here, as they all like in the 0.40-0.70 interval, probably with significant movement between them given their proximity to one another, and thus significant density between the three equilibria at any given moment in time.



Method 6: A central ingredient to poverty traps is that not only does household income increase with its asset stocks, a pretty obvious and automatic point, but

the returns on assets increase as well, at least over some range, as households accumulate more assets – a far less obvious and far from automatic point. We can get at this by plotting returns on assets (income divided by asset stocks, e.g., income per hectare or per livestock or divided by the factor analytic asset index) versus the asset stock and then testing whether this relationship has an upward slope at some point. Such a point would suggest the presence of a threshold effect like that associated with an unstable equilibrium in the autoregression approach (see Method 4). Those thresholds are the key characteristics of a poverty trap. Sorry, I've not merged data files to do this for any of our sites yet, so I have no illustration to offer at this point.

Explaining Welfare Dynamics

Ultimately, we want not only to identify economic immobility by the ex ante poor, we also want to try to begin explaining these patterns. What causes changes in welfare indicators? Do all proximate factors that affect income have similar effects or do changes in one have a more pronounced impact on welfare than changes in the others do? Ultimately, we're trying to understand movement around thresholds, especially what causes people to fall beneath them.

Method 7: Oxford's Stefan Dercon (2000 Oxford working paper) offers a way to try to allocate observed changes in household income across the effects of asset accumulation, changing exogenous returns to assets (e.g., due to policy changes that cause price shifts), and shocks on income change. For several reasons, related to both data and theory, I'm not wholly convinced that this will work well for us, but let me lay out his approach nonetheless for you to consider.

Assume a risk neutral household maximizes expected profit through allocation of fixed and variable inputs to a single production process (technology) producing some composite good. The household generates income, Y, using x, a vector of n variable inputs, and k, a vector of j fixed inputs, subject to output and input prices, p and p^x, respectively, and u, a vector of m stochastic factors, such as agro-climatic conditions. Optimization, some (strong) assumptions about functional forms, algebra, taking logarithms and then subtracting the first period from the second – I'll spare you the mathematical gymnastics – yields a closed form solution for income change that is directly estimable:

$$\Delta \ln Y_i = \Delta \ln \alpha_i + \beta_1 \Delta \ln p_i + \beta_2 \ln p_{it} + \sum_n \theta_{1n} \Delta \ln p_{in}^x + \sum_n \theta_{2n} \ln p_{\text{int}}^x + \sum_j \phi_{1j} \Delta \ln k_{ij} + \sum_j \phi_{2j} \ln k_{ijt} + \sum_m \gamma_{1m} \Delta \ln u_{im} + \sum_m \gamma_{2m} \ln u_{imt}$$

where α reflects a general technology shifter (Solow's A, for those of you who know growth theory) and just becomes the constant term, with the rest of the Greek characters representing estimable parameters. If one adds an error term to the equation, an ordinary least squares regression will produce estimates of the different elasticities of income relative to initial prices, changes in prices, initial input levels, changes in fixed inputs, and the experience of shocks, controlling for heterogeneity in the form of household fixed effects. If price changes or input level changes (e.g., land cultivated) can be linked directly to a policy change, or if shocks can be linked directly to a particular event, then this approach allows for assessment of whether observed income responds to changes in economic policy or to shocks. This is certainly interesting and potentially valuable information, and probably worth trying in our data, in spite of my technical reservations about the technique.

One reason why the Dercon method troubles me for our purposes is that a global regression implicitly weights the conditioning domain (the independent variables) by the frequency of observations found there. We're ultimately interested in the neighborhood of thresholds, but standard regression methods will essentially find the equilibria (the modes from Method 5) and ignore the troughs around the threshold points. This inherently washes out much of what we are after. I haven't yet thought up a fix to this, although intuition suggests quantile regression techniques might work for this (see Koenekcer and Hallock *Journal of Economic Perspectives* 2001 … sorry I can't seem to find a .pdf version to attach). Stay tuned.

Method 8: My IAAE paper proposes a simple technique for getting at the core issue of whether poverty traps exist and where the relevant threshold effects might occur. This takes the ratio in the change in income between the survey periods to the change in assets over the same time, dY/dA, as the dependent variable. This is then regressed on the initial period asset stocks using nonparametric regression methods or parametric methods using a reasonably high-order polynomial specification capable of capturing significant nonlinearities reflective of the sigmoid-shaped patterns characteristic of poverty traps. So the basic regression function looks like:

$$dY/dA_i = \alpha_i + \Sigma_j \beta_{ij} A_j + \psi$$

Of course, there's a certain spurious negative correlation between dY/dA and A_0 (since $dA = A_1 - A_0$), but that makes finding a positive relationship that much more robust a result. Measurement error can again pose problems, but so long as

measurement error in assets and income is positively correlated within a survey round – which is usually the case – then the effects of measurement error are at least attenuated relative to looking just at straight dY functions (see Method 1). No one has yet attempted this approach, to the best of my knowledge, so we'll see how it works!

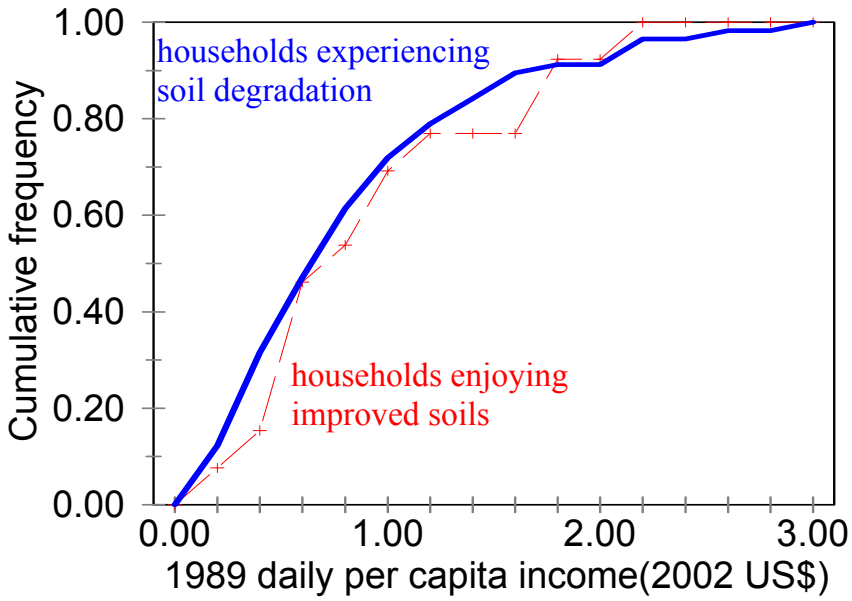
Describing the relationship between welfare and soil dynamics

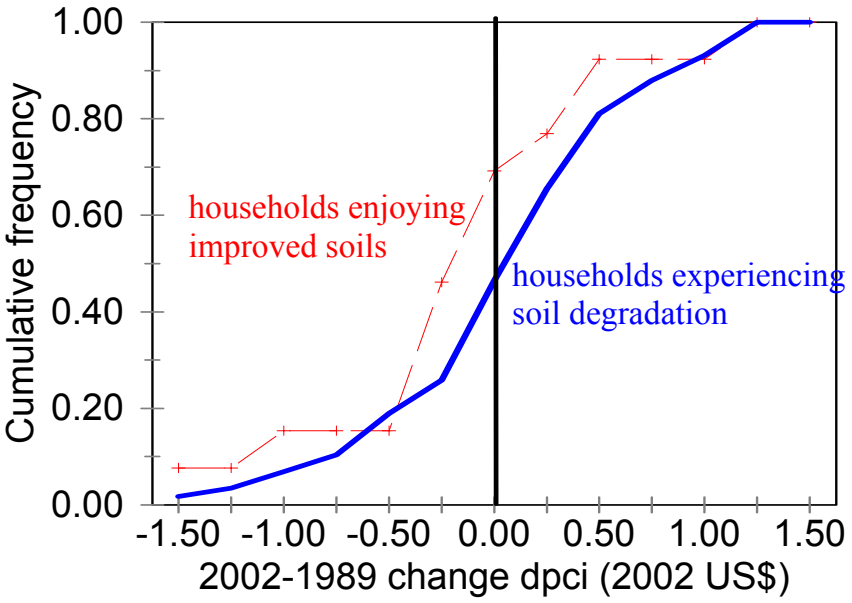
Finally, a key issue in our project relates welfare dynamics to the dynamics of natural capital, especially soils. So we want to look at this relationship explicitly. This may be tricky because of questions about the correspondence between farmers' subjective assessment of soil fertility change over time and scientific measurement of the same phenomenon (see the February 2003 (vol. 111, issue 3-4) special issue of *Geoderma* on local soil knowledge, especially the Gray and Morant paper). But at this point that is all we have to work with, so let's use the data.

We can plot the cumulative frequency of (i) initial period daily per capita income (dpci) and (ii) change in dpci for those households who indicate they believe their soils have improved versus those whose soils have degraded to see if there's any pattern. Our original hypothesis was that higher initial incomes are associated with increasing soil quality and that changes in these two variables would be positively correlated.

In Madzuu, the resulting pattern is (to my eyes, at least) counterintuitive. The 1989 income distribution for those whose soils improved subsequently was indeed superior to that for households whose soils deteriorated, as evidenced by the second order stochastic dominance manifest in the accompanying graphic. So it does indeed seem that households with improving soils indeed started with better income prospects. However, the income *change* distribution for households with degrading soils second order stochastically dominates the distribution for households with improving soils (see the graphic)! Oh well, the data don't always cooperate. I'd welcome an explanation of this observation (other than the obvious: measurement error or Chris botched the analysis).

We may want to try other approaches to exploring the relationship between soil and welfare dynamics, although the ordinal nature of the farmer soil quality change assessments sharply limits what we can do through regression analysis. I welcome people's ideas on next steps here.





Action items and responsibilities

The next step is obviously to start doing this empirical analysis. For each site other than Embu we now have the data entered and cleaned, so we should be able to do much of this reasonably quickly.

On November 14-15, 2003, the BASIS CRSP is holding a small workshop for some USAID folks with six papers (three by anthropologists and three by economists)

on "Combatting Persistent Poverty in Africa: Structure, Causes and Solutions." We will host it here at Cornell, at the Management Entity's request. I hope to be able to present a paper coauthored by much of our team that uses the above methods (or some combination across sites) to explore the within site and between site variation in asset, income and expenditure measures of welfare dynamics in order to begin to untangle the community- and household-specific causality behind persistent poverty and the possibility of poverty traps. The paper will be due in mid-to-late October, so I would ask that you get me as much of the above analysis as possible by September 30. Feel free to send results piecemeal as you generate them and please do not be bashful about asking for clarification, further information or assistance, from me or others on the team.

I am expecting that Paswel will handle the above tasks for the Madzuu data, that Justine or Frank will handle these tasks for the ICRAF Siaya/Vihiga data, that Andrew Mude and I 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 each of you confirm this to me via email in the coming few days, including confirming the feasibility of the proposed timing.***

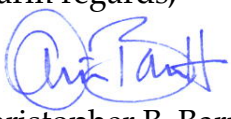
Conclusion

My apologies for another long and tardy memo, some of which may be pretty basic to many of you. But I think it is important that we try to follow very similar empirical procedures in analyzing the data from each site so that we can have greater confidence that inter-site differences reflect true differences, not just differences of method.

Please email me – and the whole group – with comments, suggestions for changes or additions, or requests for clarifications to any of the information above.

Thank you very much for all your hard work on this project. It is a pleasure and honor to work with all of you. Keep up the great work!

Warm regards,



Christopher B. Barrett
Associate Professor